BEYOND UNIFORM LIPSCHITZNESS: PATHWISE JACOBIAN STABILITY FOR FLOW MATCHING SAMPLERS

Anonymous authors
Paper under double-blind review

ABSTRACT

Flow matching samplers integrate a learned velocity field. Existing error bounds often use its largest Jacobian norm over all states, even when generated paths visit only a small part of the space. We develop a framework that weights each velocity error and solver defect by the Jacobian growth that follows it between the compared paths. An exact error identity yields signed segment and tube bounds, with classical uniform Lipschitz estimates as a special case. In an explicit conditional flow matching example with nonzero velocity error, the signed bound equals the terminal Wasserstein error, while the classical bound is more than 500 times larger. We also determine what the population loss guarantees without a uniform Jacobian bound on the learned field. For a fixed population flow and a fixed linear growth class, the sharp worst case rate in state dimension d is $\sqrt{\eta}[1+\log(1/\eta)]^{(d+4)/4}$, where η is the population velocity loss. Polynomial state weights reduce the logarithmic factor. Tests on trained samplers connect path growth to terminal sensitivity and use later growth to select paths for numerical refinement. The same flow derivative gives a weighted Poincaré inequality for the output law.

1 INTRODUCTION

Flow matching trains an ODE sampler by regressing a velocity field along a probability path from a simple reference distribution to the data distribution (Lipman et al., 2023). It is closely connected to continuous normalizing flows (Chen et al., 2018), probability flow ODEs, stochastic interpolants, rectified flows, and flow matching on general geometries (Song et al., 2021; Albergo et al., 2025; Liu et al., 2023; Chen & Lipman, 2024). At sampling time, the learned field v_θ is integrated from a latent Gaussian variable,

$$\frac{d}{dt}X_t^\theta = v_\theta(t, X_t^\theta), \qquad X_0^\theta = Z \sim \gamma := \mathcal{N}(0, \mathrm{I}_d),$$

and the terminal state $T_\theta(Z) := X_1^\theta$ is returned as a sample. The learned velocity can differ from the population velocity, and numerical integration adds solver error. Statistical analyses bound velocity error for specified estimators (Fukumizu et al., 2025; Zhou & Liu, 2025; Kunkel & Trabs, 2025). Classical perturbation formulas propagate velocity and solver error through derivatives of the subsequent flow (Jentzen et al., 2019; Benton et al., 2024). Flow matching error bounds control these derivatives through spatial regularity of the velocity field (Benton et al., 2024; Gao et al., 2024a). A common condition is

$$\|\nabla_x v_\theta(t,x)\|_{\mathrm{op}} \le L(t) \quad \text{for all } x \in \mathbb{R}^d, \qquad \int_0^1 L(t)\,dt < \infty.$$

This route is clean, and in strong regularity regimes it can imply sharp discretization bounds and functional inequalities (Stéphanovitch, 2026). The spatial supremum includes regions that generated trajectories may never visit. Let $X_t^\star$ and X_t^θ be the population and learned paths from the same start, and write $\Delta_t = X_t^\theta - X_t^\star$. Their difference satisfies

$$\dot{\Delta}_t = \big[v_\theta(t, X_t^\theta) - v_\theta(t, X_t^\star)\big] + \big[v_\theta(t, X_t^\star) - u_t(X_t^\star)\big].$$

The second term is the velocity error measured by flow matching. Standard bounds control the first by $L(t)\|\Delta_t\|$ (Albergo & Vanden-Eijnden, 2023, Proposition 3), (Benton et al., 2024, Theorem 1).

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

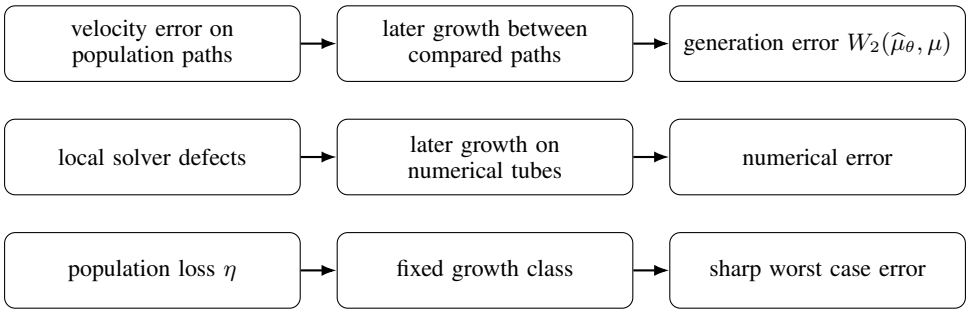

Figure 1: The first two rows bound generation and numerical error for a given sampler. The last row gives the sharp guarantee when only population loss and a fixed growth class are known.

This raises our first question: *Can we bound sampling error without a uniform Lipschitz bound over the entire state space?*

We develop a framework that keeps the velocity error at the time it occurs and measures the Jacobian growth that follows it between the compared paths. The resulting bound applies to generation error and, with local solver defects in place of velocity error, to numerical error. The distinction matters even for small networks: two conditional flow matching (CFM) fields (Lipman et al., 2023; Tong et al., 2024) with two tanh units can have the same population loss and the same absolute Jacobian at every state, yet an unbounded ratio of terminal errors across a family (Appendix I). The example in Appendix G.5 also has nonzero velocity error and finite tube growth moments although its spatial Jacobian supremum is infinite.

The bound for a given sampler uses both its velocity error and the growth that follows that error. A population loss measures the first quantity but does not record where the error meets later growth. This gives a second question: *If only the population velocity loss and fixed growth bounds are known, what terminal error can we guarantee?* We prove matching upper and lower rates over a fixed class of learned fields whose Jacobians may grow linearly with the state. In CFM, the population excess risk equals the squared velocity error along the population marginals. The comparison argument applies whenever a flow matching model supplies these population quantities.

Figure 1 summarizes the two questions and the corresponding numerical error bound. The first two rows use the growth after an error occurs. The third asks what remains when only population loss and a fixed growth class are known.

Proposition 6 gives the exact asymptotics for the two network errors as their common active region moves into the Gaussian tail.

Contributions. The paper answers the two questions above. Theorem 2 weights each velocity error by the later growth between the population and learned paths. Theorem 4 gives the corresponding result for local solver defects. Theorem 3 gives matching upper and lower rates when only population loss and a fixed growth class are known. A polynomial state weight reduces the tail contribution to that rate. The experiments measure each relaxation of the error bound on one explicit CFM sampler and test path growth and numerical refinement on trained CFM models. Section 6 gives a separate consequence for the output law. Related work appears in Appendix A.

2 Error Bounds Along Flow Matching Paths

Let $\gamma = \mathcal{N}(0, \mathrm{I}_d)$ and let the learned flow be generated by

$$\frac{d}{dt} X_t^\theta(z) = v_\theta(t, X_t^\theta(z)), \qquad X_0^\theta(z) = z.$$

Write $T_\theta = X_1^\theta$ and $\widehat{\mu}_\theta = (T_\theta)_{\#}\gamma$, where $S_{\#}\gamma$ is the law of $S(Z)$ for $Z \sim \gamma$. We assume finite second moments when using W_2. The path, tube, and integrability conditions needed for each bound appear before its theorem and are stated precisely in the appendix.

The quadratic Wasserstein distance W_2 is the square root of the least mean squared displacement over couplings of two laws. Sending the same Gaussian initial state through two flows gives one such coupling. Thus an $L^2(\gamma)$ bound on paired terminal states gives a bound between their output laws. The variational equation controls sensitivity to initialization. The mean value formula then compares two trajectories on the segment between them. We use these facts first for velocity error, then for numerical defects. Under a uniform Jacobian bound, the resulting estimates reduce to classical bounds.

2.1 Jacobian Growth on Paths and Segments

For one trajectory, we integrate the Jacobian norm along that path. For two trajectories, we also evaluate it between the paths.

Definition 1 (Pathwise Jacobian growth and stability factor)**.** *For z such that the learned flow is well defined, define the* pathwise Jacobian growth

$$\mathcal{J}_\theta(z) := \int_0^1 \|\nabla_x v_\theta(t, X_t^\theta(z))\|_{\mathrm{op}}\, dt.$$

The associated pathwise stability factor *is*

$$\mathsf{A}_\theta(z) := \exp(\mathcal{J}_\theta(z)).$$

The integral $\mathcal{J}_\theta$ measures Jacobian magnitude along the path, and A_θ bounds the resulting sensitivity. An integrable uniform Lipschitz bound makes $\mathsf{A}_\theta(Z)$ essentially bounded for $Z \sim \gamma$.

For two trajectories $X_t(z)$ and $\widehat{X}_t(z)$ with the same initialization, define their comparison segment by

$$\Gamma_t(z) := \{X_t(z) + \lambda(\widehat{X}_t(z) - X_t(z)) : \lambda \in [0,1]\}.$$

The *tube Jacobian growth* and *tube stability factor* are

$$\mathcal{J}_{\theta,\mathrm{tube}}(z) := \int_0^1 \sup_{x \in \Gamma_t(z)} \|\nabla_x v_\theta(t,x)\|_{\mathrm{op}}\, dt, \qquad \mathsf{A}_{\theta,\mathrm{tube}}(z) := \exp(\mathcal{J}_{\theta,\mathrm{tube}}(z)).$$

The comparison segment appears when the mean value formula is applied to the difference between two velocity evaluations. Generator sensitivity uses the Jacobian on one path, while trajectory comparisons use its values on this segment. Theorem 7 relates tube growth to path growth through a local Hessian bound and the distance between the paths.

The sensitivity bound assumes that the learned flow exists uniquely on $[0,1]$ for almost every initialization, is differentiable in its initial state, and satisfies the variational equation. We also require $v_\theta(t,\cdot)$ to be continuously differentiable near the path and $\mathcal{J}_\theta(z)$ to be finite. Assumption 1 states these conditions precisely.

Theorem 1 (Pathwise generator sensitivity)**.** *Under the flow regularity assumptions in Assumption 1, for γ-a.e. z,*

$$\|D_z T_\theta(z)\|_{\mathrm{op}} \le \mathsf{A}_\theta(z).$$

The proof is given in Theorem 6. The derivative with respect to the initial state solves a linear ODE with the path Jacobian as its coefficient (Hairer et al., 1993, Section I.14). The logarithmic norm $m(B) = \lambda_{\max}((B + B^\top)/2)$ also records contraction. With $\mathcal{G}_\theta(z) = \int_0^1 m(\nabla_x v_\theta(t, X_t^\theta(z)))\, dt$, the same proof gives

$$\log \|D_z T_\theta(z)\|_{\mathrm{op}} \le \mathcal{G}_\theta(z) \le \mathcal{J}_\theta(z).$$

Definition 3 gives the appendix version.

Definition 2 (Pathwise Stability Curve)**.** *The Pathwise Stability Curve is the quantile function of $\mathcal{J}_\theta(Z)$. For $\alpha \in (0,1)$, define*

$$Q_\theta(\alpha) := \inf\{r \ge 0 : \mathbb{P}(\mathcal{J}_\theta(Z) \le r) \ge \alpha\}.$$

Then $e^{Q_\theta(\alpha)}$ is the α-quantile of $\mathsf{A}_\theta(Z)$.

Theorem 1 yields, for $\delta \in (0,1)$,

$$\mathbb{P}_{Z\sim\gamma}\left(\|D_Z T_\theta(Z)\|_{\mathrm{op}} \le e^{Q_\theta(1-\delta)}\right) \ge 1 - \delta. \tag{1}$$

A linear interpolation example. Let $Z \sim \mathcal{N}(0,1)$ and $X_1 = Z + \kappa Z^3$, with $\kappa > 0$. The interpolation $X_t = Z + \kappa t Z^3$ is its population flow. Its spatial Jacobian supremum is $1/t$ for $t > 0$, so the integrated uniform bound is infinite. Along each path,

$$\mathcal{J}(Z) = \log(1 + 3\kappa Z^2), \qquad \mathsf{A}(Z) = 1 + 3\kappa Z^2.$$

The factor has every finite moment and equals the terminal derivative. The velocity derivative is nonnegative, so the spatial one-sided bound has the same divergence. Appendix G.1 gives the construction and its weighted Poincaré inequality.

2.2 From Velocity Loss to Generation Error

We next compare the learned sampler with the population flow matching dynamics. Let u_t be the population flow matching velocity and let

$$\frac{d}{dt} X_t^\star(z) = u_t(X_t^\star(z)), \qquad X_0^\star(z) = z$$

be the population flow. Let $T_\star(z) := X_1^\star(z)$ and $\mu := (T_\star)_\# \gamma$.

We separate velocity error from its later amplification. The CFM excess loss measures squared velocity error along the population marginals. The comparison theorem propagates this error to the output law (Lipman et al., 2023; Benton et al., 2024).

For a latent point z, define the pathwise velocity error along the population path by

$$\mathcal{E}_\theta(z) := \int_0^1 \|v_\theta(t, X_t^\star(z)) - u_t(X_t^\star(z))\| \, dt.$$

Application to conditional flow matching. In CFM, an interpolation state $X_t^{\mathrm{int}} \sim \mu_t$ is paired with a conditional velocity target U_t. The population velocity is $u_t(x) = \mathbb{E}[U_t \mid X_t^{\mathrm{int}} = x]$ (Lipman et al., 2023). For example, independent linear interpolation uses $X_t^{\mathrm{int}} = (1-t)Z + tX_1$ and $U_t = X_1 - Z$, with $Z \sim \gamma$ independent of the data sample X_1. Assume the population ODE has marginals μ_t and the risks below are finite. Its trajectories and the sampled interpolation share these marginals, although their individual paths generally differ. Define $\mathcal{L}_{\mathrm{CFM}}(v) = \int_0^1 \mathbb{E}\|v(t, X_t^{\mathrm{int}}) - U_t\|^2 \, dt$. Expanding the square and conditioning on X_t^{int} gives

$$\eta := \int_0^1 \mathbb{E}_{X \sim \mu_t} \|v_\theta(t, X) - u_t(X)\|^2 \, dt = \mathcal{L}_{\mathrm{CFM}}(v_\theta) - \mathcal{L}_{\mathrm{CFM}}(u), \qquad \|\mathcal{E}_\theta\|_{L^2(\gamma)} \le \sqrt{\eta}. \quad (2)$$

Conditioning on X_t^{int} removes the cross term, and Cauchy–Schwarz in time gives the inequality. Subtracting the population minimum removes the conditional variance of the velocity label. Thus the CFM objective controls the residual on population marginals, while the comparison theorem propagates it along population ODE paths. Appendix H.1 gives the proof and the population Jacobian for Gaussian interpolation.

For this comparison, use the segment $\Gamma_t^{\star,\theta}(z)$ joining $X_t^\star(z)$ and $X_t^\theta(z)$, and write

$$L_t^{\star,\theta}(z) := \sup_{x \in \Gamma_t^{\star,\theta}(z)} \|\nabla_x v_\theta(t,x)\|_{\mathrm{op}}, \qquad \mathcal{J}_{\theta,\mathrm{tube}}(z) = \int_0^1 L_t^{\star,\theta}(z) \, dt.$$

The exact difference equation uses the Jacobian averaged over this segment:

$$M_t(z) := \int_0^1 \nabla_x v_\theta\big(t, X_t^\star(z) + \lambda[X_t^\theta(z) - X_t^\star(z)]\big) \, d\lambda.$$

Both flows must exist on $[0,1]$ with the same initialization. We require the learned field to be continuously differentiable near the comparison segments, with finite tube growth and source error. Assumption 2 states these conditions precisely.

Theorem 2 (Velocity error and generation error)**.** *Under Assumption 2, for γ-a.e. z,*

$$\begin{aligned} \|T_\theta(z) - T_\star(z)\| &\le \int_0^1 \exp\left(\int_s^1 m(M_r(z)) \, dr \right) \|v_\theta(s, X_s^\star(z)) - u_s(X_s^\star(z))\| \, ds \\ &\le \int_0^1 \exp\left(\int_s^1 L_r^{\star,\theta}(z) \, dr \right) \|v_\theta(s, X_s^\star(z)) - u_s(X_s^\star(z))\| \, ds \\ &\le \mathsf{A}_{\theta,\mathrm{tube}}(z) \mathcal{E}_\theta(z). \end{aligned} \quad (3)$$

If the terminal states have finite second moments, synchronous coupling gives

$$\begin{aligned} W_2(\widehat{\mu}_\theta, \mu) &\le \left\| \int_0^1 \exp\left(\int_s^1 m(M_r(Z))\, dr \right) \|v_\theta(s, X_s^\star(Z)) - u_s(X_s^\star(Z))\|\, ds \right\|_{L^2(\gamma)} \\ &\le \left\| \int_0^1 \exp\left(\int_s^1 L_r^{\star,\theta}(Z)\, dr \right) \|v_\theta(s, X_s^\star(Z)) - u_s(X_s^\star(Z))\|\, ds \right\|_{L^2(\gamma)} \\ &\le \|\mathsf{A}_{\theta,\mathrm{tube}} \mathcal{E}_\theta\|_{L^2(\gamma)}. \end{aligned} \tag{4}$$

The proof in Theorem 8 integrates the exact difference equation and couples terminal states through their common Gaussian start. The first bound keeps the signed Jacobian averaged between the paths. The second takes the largest Jacobian norm on each comparison segment. The third replaces all later weights by one factor. Classical nonlinear variation of constants also propagates velocity error through a later flow derivative (Benton et al., 2024, Proposition 2), (Jentzen et al., 2019). Here the comparison is expressed through the two observed trajectories. In one dimension, the first bound is an equality when the residual keeps one sign along each population path. If both terminal maps are increasing, their common Gaussian coupling realizes W_2.

The segment and tube bounds use both paths and keep each residual paired with its later growth. A statistic on one generated path controls sensitivity to initialization, while a generation error bound compares two paths. Proposition 1 gives an explicit condition under which a small residual and local curvature control the separation and tube growth from the learned path. Corollary 3 gives a loss based probability form. Appendix H.1 supplies sufficient moment conditions.

Comparison with uniform bounds. If $\|D_x v_\theta(t,x)\|_{\mathrm{op}} \le L(t)$ for all states and L is integrable, then

$$\|\mathsf{A}_{\theta,\mathrm{tube}} \mathcal{E}_\theta\|_2 \le e^{\int_0^1 L(t)\, dt} \left[\int_0^1 \mathbb{E}_{X\sim\mu_t} \|v_\theta(t,X) - u_t(X)\|^2\, dt \right]^{1/2}.$$

The right side is the global Lipschitz estimate in Benton et al. (2024, Theorem 1). Keeping the time weights in Equation 4 also recovers the coefficient in Gao et al. (2024a, Proposition 6.2(ii)). Appendix C.2 gives both comparisons.

3 What Population Loss Guarantees

Suppose u and v_θ are continuous with jointly continuous spatial derivatives. For finite $K, L \ge 0$ and every $t \in [0,1]$ and $x \in \mathbb{R}^d$, assume

$$\begin{aligned} \|u_t(0)\| + \|D_x u_t(x)\|_{\mathrm{op}} &\le L, \\ \|v_\theta(t,x)\| + \|D_x v_\theta(t,x)\|_{\mathrm{op}} &\le K(1 + \|x\|). \end{aligned} \tag{5}$$

Start both flows from the standard Gaussian law and let μ_t be the population flow marginals. We use the integrated squared velocity error η from Equation 2; for CFM it equals the population excess loss.

Theorem 3 (Sharp bounds from population loss)**.** *Under Equation 5, there is a finite $C_{d,K,L}$ such that, for $0 < \eta \le 1$,*

$$W_2(\widehat{\mu}_\theta, \mu) \le C_{d,K,L} \sqrt{\eta} [1 + \log(1/\eta)]^{(d+4)/4}. \tag{6}$$

At zero loss, the terminal laws agree. For each fixed population field satisfying the stated bound, there is a finite $K_* = K_*(d,L)$ such that the rate is sharp for every fixed $K \ge K_*$. More precisely, for every sufficiently small $\varepsilon > 0$, some learned field satisfying the same growth bound has $\eta \le \varepsilon$ and

$$W_2(\widehat{\mu}_\theta, \mu) \ge c_{d,K,L} \sqrt{\varepsilon} [1 + \log(1/\varepsilon)]^{(d+4)/4},$$

where $c_{d,K,L} > 0$ is independent of ε.

Theorem 15 and Corollary 14 give the proof. The population Jacobian is bounded, while the learned Jacobian may grow linearly with the state under the same constant K at every loss budget. The upper bound combines a local energy estimate in population flow coordinates with a Gaussian tail bound. For the lower bound, a cube of side R^{-1} at distance R has Gaussian mass of order $\phi(R)R^{-d}$. The

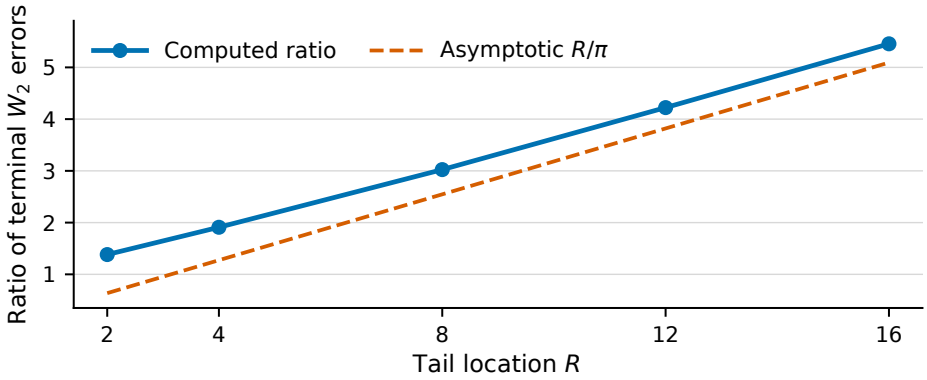

Figure 2: Two CFM fields differ only in the sign of their output weights. At each R, they have the same population loss and absolute Jacobian. Their terminal error ratio grows with R and approaches R/π asymptotically. Proposition 6 gives the formulas. Numerical checks of all three asymptotics appear in Appendix I.

Table 1: Three guarantees with different available information. The classical row assumes a global learned Jacobian bound (Benton et al., 2024, Theorem 1). The pathwise row uses residuals and their subsequent growth on comparison paths. The final row gives the sharp worst case rate from a loss budget and the fixed class in Equation 5.

Available information	Learned Jacobian condition	Terminal W_2 bound
Population loss η	$\|D_x v_\theta(t,x)\|_{\mathrm{op}} \le L(t)$ for all x	$\le e^{\int_0^1 L(t)dt}\sqrt{\eta}$
Residual and comparison paths	Finite growth on the compared segments	$\le \|\mathsf{A}_{\theta,\mathrm{tube}}\mathcal{E}_\theta\|_2$
Loss budget $\eta \le \varepsilon$	$\|v_\theta\|+\|D_x v_\theta\|_{\mathrm{op}} \le K(1+\|x\|)$	Worst case $\asymp \sqrt{\varepsilon}[1+\log(1/\varepsilon)]^{(d+4)/4}$

construction moves part of this mass a distance of order R. Its loss is of order $\phi(R)R^{-2d-2}$ and its squared transport error is of order $\phi(R)R^{2-d}$. Their ratio is of order R^{d+4}, while $\log(1/\eta)$ is of order R^2. This gives the exponent $(d+4)/4$ in the terminal error rate. The constants depend on the fixed d, K, and L. Theorem 2 can be much sharper for a given sampler because it retains the actual residual and the growth on its comparison paths.

The two networks in Figure 2 also belong to a fixed growth class: their population field is zero and both learned fields satisfy Equation 38 with $K = 1$. Proposition 6 gives $W_2^2(\mu_R^+, \gamma)/\eta_R \sim R/3$, while $\log(1/\eta_R) \sim R^2/2$. Thus no constant depending only on this fixed class can give $W_2 \le C\sqrt{\eta}$ for every field in it. This example requires a logarithmic factor, whereas the lower bound in Theorem 3 determines the largest possible exponent.

Theorem 2 bounds one sampler using its residual and compared paths. Theorem 3 uses only a loss budget and a fixed growth class. Table 1 displays the different inputs.

How state weights improve the guarantee. For $a \ge 0$, define the weighted population loss

$$\eta_a = \int_0^1 \mathbb{E}_{X\sim\mu_t}(1+\|X\|)^a \|v_\theta(t,X) - u_t(X)\|^2\, dt.$$

Corollary 15 gives, for $0 < \eta_a \le 1$,

$$W_2(\widehat{\mu}_\theta, \mu) \lesssim_{d,K,L,a} \sqrt{\eta_a}[1+\log(1/\eta_a)]^{(d+4-a)_+/4},$$

where $r_+ = \max\{r, 0\}$. For sufficiently large fixed K, the corresponding worst case over the same growth class has a matching lower bound at every sufficiently small loss budget. The logarithmic factor disappears at $a = d+4$: the weight makes a velocity error in the Gaussian tail more costly. In CFM, conditioning shows that the weighted regression objective has the same unrestricted population minimizer u; the exact identity is proved in Appendix H.1.

4 NUMERICAL ERROR FROM LOCAL DEFECTS

Numerical integration adds a local defect at each step. Standard global error analysis propagates each defect through the later exact flow (Hairer et al., 1993). We keep the Jacobian growth encountered by the corresponding numerical comparison tube, so defects at different times receive different weights.

Let $0 = t_0 < t_1 < \cdots < t_N = 1$ be a time grid and $h_k = t_{k+1} - t_k$. Let $\Phi^\theta_{s,t}$ be the exact learned subflow from time t to time s. Consider a generic one-step method

$$\widetilde{X}^\theta_0(z) = z, \qquad \widetilde{X}^\theta_{k+1}(z) = \Psi^\theta_k(\widetilde{X}^\theta_k(z)).$$

For example, explicit Euler has $\Psi^\theta_k(x) = x + h_k v_\theta(t_k, x)$. Define the one-step local defect

$$\tau^\theta_k(z) := \|\Psi^\theta_k(\widetilde{X}^\theta_k(z)) - \Phi^\theta_{t_{k+1},t_k}(\widetilde{X}^\theta_k(z))\|.$$

For each step, the numerical comparison tube consists of exact learned subflows started from the segment joining the exact and numerical states at t_k. We assume that these subflows and the numerical states exist, that $v_\theta(t,\cdot)$ is continuously differentiable near each tube, and that the local tube integrals and defects are finite. Assumption 3 defines the local integrals ℓ^θ_k and the total factor $\mathsf{A}_{\theta,h,\mathrm{tube}} = \exp(\sum_k \ell^\theta_k)$.

Theorem 4 (Numerical defects and discretization error)**.** *Under Assumption 3, for γ-a.e. z,*

$$\|\widetilde{X}^\theta_N(z) - X^\theta_1(z)\| \le \sum_{k=0}^{N-1} \exp\left(\sum_{j=k+1}^{N-1} \ell^\theta_j(z)\right) \tau^\theta_k(z) \le \mathsf{A}_{\theta,h,\mathrm{tube}}(z) \sum_{k=0}^{N-1} \tau^\theta_k(z). \tag{7}$$

If the terminal states have finite second moments, then, with $\widetilde{\mu}_{\theta,h} := (\widetilde{X}^\theta_N)_\#\gamma$,

$$W_2(\widetilde{\mu}_{\theta,h}, \widehat{\mu}_\theta) \le \left\| \sum_{k=0}^{N-1} \exp\left(\sum_{j=k+1}^{N-1} \ell^\theta_j\right) \tau^\theta_k \right\|_2 \le \left\| \mathsf{A}_{\theta,h,\mathrm{tube}} \sum_{k=0}^{N-1} \tau^\theta_k \right\|_2 .$$

The proof in Theorem 9 propagates each local defect through the remaining steps. This is the discrete counterpart of Equation 3. A defect matters when it is large and followed by strong growth. Keeping the individual weights gives the time allocation rule in Appendix H.3 and motivates the numerical test in Section 5.

Corollary 1 (Total sampling error and uniform bounds)**.** *Under the hypotheses of Theorems 2 and 4, with finite second moments of all terminal states,*

$$W_2(\widetilde{\mu}_{\theta,h}, \mu) \le \left\| \mathsf{A}_{\theta,h,\mathrm{tube}} \sum_{k=0}^{N-1} \tau^\theta_k \right\|_2 + \|\mathsf{A}_{\theta,\mathrm{tube}} \mathcal{E}_\theta\|_2.$$

If $\|D_x v_\theta(t,x)\|_{\mathrm{op}} \le L(t)$ for all states and L is integrable, both tube factors are at most $e^{\int_0^1 L(t)\,dt}$, so

$$W_2(\widetilde{\mu}_{\theta,h}, \mu) \le e^{\int_0^1 L(t)\,dt} \left(\|\mathcal{E}_\theta\|_2 + \left\| \sum_{k=0}^{N-1} \tau^\theta_k \right\|_2 \right).$$

The triangle inequality combines generation and numerical error. A uniform Jacobian bound recovers the classical Grönwall estimate. For $\dot{x}_t = a_t x_t$ with $a_t \ge 0$, a perturbation grows exactly by $e^{\int_0^1 a_t dt}$, so exponential amplification already occurs in a scalar linear ODE. Our bounds keep the growth associated with each source error. Corollaries 8–9 give the numerical terms for order-p solvers and Euler.

Euler step count. For a fixed learned field and terminal time, the defect bound gives an $O(N^{-1})$ numerical term under a finite tube and material derivative moment. Appendix H.3 gives a finite step count and a time allocation rule.

5 EXPERIMENTS

We test three consequences of the theory. An explicit CFM sampler shows the cost of each relaxation in Theorem 2. Trained samplers show how path growth relates to terminal sensitivity in Theorem 1. They also test whether later growth helps locate the numerical error described by Theorem 4.

5.1 COMPARING ERROR BOUNDS FOR ONE CFM SAMPLER

To compare the bounds on one sampler, we use independent linear CFM from a standard Gaussian to $\frac{1}{2}\mathcal{N}(-2, 0.5^2) + \frac{1}{2}\mathcal{N}(2, 0.5^2)$. Its population velocity u_t is explicit. We set $v_\delta(t, x) = u_t(x) + \delta(1 + \tanh x)$ at the fixed amplitudes $\delta \in \{0.01, 0.03, 0.10\}$. The residual is nonzero and the largest absolute Jacobian on each comparison segment has a closed form. Both flows preserve order in one dimension, so the common Gaussian coupling gives terminal W_2 exactly. Appendix H.4.1 gives the numerical quadrature checks.

The positive residual and increasing flow maps make the signed segment bound equal to terminal W_2 in this example. The successive relaxations in Figure 3 each widen the bound. Across the three perturbations, the tube bound is 6.7 to 35.8 times the exact error. Taking a spatial maximum widens it by another factor of 9.3 to 27.4, and replacing the time resolved residual by population loss widens it by about 2.8. The classical bound is 515 to 937 times W_2. Appendix H.4.1 reports all values and numerical checks.

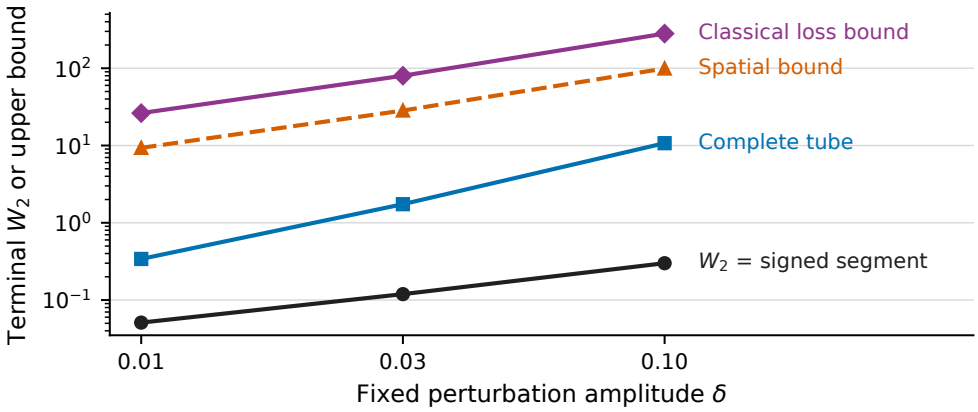

Figure 3: Error and upper bounds for one explicit CFM sampler at three fixed nonzero perturbations. The signed segment bound equals W_2 under the positive residual and monotone flow maps of this example. The tube bound uses Jacobians on the comparison segments. The spatial bound uses their maximum over all states, and the classical bound also replaces the residual by population loss. Numerical checks appear in Appendix H.4.1.

5.2 MEASURING JACOBIAN GROWTH ALONG GENERATED PATHS

We first compare growth along generated paths with two spatial references. In an exact population CFM flow, the time integral of the spatial Jacobian supremum up to $T = 0.999$ is 198.03, while the 99th percentile of path growth is 3.97. Across ten trained two dimensional CFM models, the 99th percentile of path growth is 4.18 ± 0.35. A finite spatial grid over the scanned states gives an empirical reference of 11.74 ± 1.74, and a global layer norm bound gives 99.32 ± 18.57. Appendix H.4 gives both protocols.

Path growth and terminal sensitivity. Across the ten trained models, the mean Spearman correlations with terminal log sensitivity are 0.8553 for $\mathcal{J}_\theta$ and 0.9971 for signed growth $\mathcal{G}_\theta$. Within each checkpoint, the gap $\mathcal{G}_\theta - \log \|D_Z T_\theta\|_{\mathrm{op}}$ has a median of 0.068 and a 99th percentile of 0.265, averaged over checkpoints. These correspond to factors of 1.07 and 1.30. The signed quantity retains

contraction, while $\mathcal{J}_\theta$ counts Jacobian magnitude. Theorem 1 bounds the sensitivity measured here; the generation error bound additionally needs growth between two paths.

Locating numerical error. On five baseline CFM models, we compare a local Heun defect score with the same defects weighted by later Jacobian growth on a 16 step pilot path. The top 20 percent selected by the weighted score contain $82.3\% \pm 6.2\%$ of the pilot Euler squared endpoint error, compared with $28.5\% \pm 4.9\%$ for defects alone. Refining the selected paths to 32 Euler steps retains 19.2 steps per path on average. Its RMSE is $21.48\% \pm 4.31\%$ lower than uniform 20 step Euler and $28.15\% \pm 3.41\%$ lower than defect only selection. The growth score also requires 16 Jacobian norm evaluations per path. These comparisons use matched retained Euler steps, with scoring work reported separately. Appendix H.4 gives the endpoint reference checks and evaluation counts.

6 Weighted Poincaré Inequalities from Terminal Derivatives

A Poincaré inequality bounds variance by an average squared gradient. Applying the Gaussian inequality to $f \circ T_\star$ gives a weight determined by the derivative of the population terminal map. Larger terminal stretch produces a larger gradient weight. This result concerns the regularity of the output law and uses the same flow derivative that controls path sensitivity. The appendix states the chain rule and integrability assumptions and gives the learned output law version.

Theorem 5 (Weighted Poincaré inequality from a flow)**.** *Suppose the population map $T_\star$ satisfies Assumption 4, the population flow satisfies the variational equation, and the displayed integrals are finite. For every smooth compactly supported f,*

$$\begin{aligned}\mathrm{Var}_\mu(f) &\le \mathbb{E}_{Z\sim\gamma}\left\|DT_\star(Z)^\top \nabla f(T_\star(Z))\right\|^2 \\ &\le \mathbb{E}_{Z\sim\gamma}\left[e^{2\int_0^1 m(D_x u_t(X_t^\star(Z)))dt}\|\nabla f(T_\star(Z))\|^2\right].\end{aligned}$$

The first inequality applies the Gaussian Poincaré inequality to $f \circ T_\star$ (Bakry et al., 2014). The second uses the variational equation, as in Theorem 1. Appendix F gives the matrix weight and the learned map version. The cubic flow in Appendix G.1 has an explicit weight although its spatial Jacobian supremum has a divergent time integral.

7 Conclusion

Flow matching loss measures velocity error on population marginals. The error at the sampler output also depends on how the learned flow amplifies it after it occurs. Our comparison identity keeps this timing and the Jacobian growth between the compared paths. It gives generation and solver error bounds and recovers uniform Lipschitz estimates as a special case. In the explicit CFM example, the signed bound equals the terminal Wasserstein error, while the classical bound is more than 500 times larger. When only the population loss and a fixed growth class are known, matching upper and lower bounds identify the sharp worst case logarithmic factor. Weighting the loss by $(1 + \|x\|)^{d+4}$ removes that factor. On trained samplers, path growth tracks terminal sensitivity and helps identify paths that benefit from numerical refinement. The population flow derivative also gives a weighted Poincaré inequality for the output law. These results distinguish the guarantee from population loss alone from the sharper bound available when the sampler paths are known.

AI use statement

We used Codex to assist with numerical experiments, reference checking, and manuscript editing. We also used GPT-6 to help identify and correct minor errors in the mathematical proofs. The authors reviewed all content and take full responsibility for the final version of this paper.

References

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

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

Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348. Springer, 2014.

Joe Benton, George Deligiannidis, and Arnaud Doucet. Error Bounds for Flow Matching Methods. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856.

Sergey G Bobkov and Michel Ledoux. Weighted Poincaré-Type Inequalities for Cauchy and Other Convex Measures. *The Annals of Probability*, pp. 403–427, 2009.

Yang Cao and Linda Petzold. A posteriori error estimation and global error control for ordinary differential equations by the adjoint method. *SIAM Journal on Scientific Computing*, 26(2):359–374, 2004.

Patrick Cattiaux, Ivan Gentil, and Arnaud Guillin. Weak logarithmic Sobolev inequalities and entropic convergence. *Probability theory and related fields*, 139(3):563–603, 2007.

Patrick Cattiaux, Nathael Gozlan, Arnaud Guillin, and Cyril Roberto. Functional inequalities for heavy tailed distributions and application to isoperimetry. *Electronic Journal of Probability*, 15 (13):346–385, 2010.

Patrick Cattiaux, Arnaud Guillin, and Li-Ming Wu. Some remarks on weighted logarithmic Sobolev inequality. *Indiana University Mathematics Journal*, pp. 1885–1904, 2011.

Ricky T. Q. Chen and Yaron Lipman. Flow Matching on General Geometries. In *The Twelfth International Conference on Learning Representations*, 2024.

Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.

Sitan Chen, Sinho Chewi, Holden Lee, Yuanzhi Li, Jianfeng Lu, and Adil Salim. The probability flow ODE is provably fast. *Advances in Neural Information Processing Systems*, 36:68552–68575, 2023.

Aryeh Dvoretzky, Jack Kiefer, and Jacob Wolfowitz. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *The Annals of Mathematical Statistics*, pp. 642–669, 1956.

Chris Finlay, Jörn-Henrik Jacobsen, Levon Nurbekyan, and Adam Oberman. How to train your neural ode: the world of jacobian and kinetic regularization. In *International conference on machine learning*, pp. 3154–3164. PMLR, 2020.

Kenji Fukumizu, Taiji Suzuki, Noboru Isobe, Kazusato Oko, and Masanori Koyama. Flow matching achieves almost minimax optimal convergence. In *The Thirteenth International Conference on Learning Representations*, 2025.

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

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

Mudit Gaur, Prashant Trivedi, Shuchin Aeron, Amrit Singh Bedi, George K Atia, and Vaneet Aggarwal. Generative modeling with continuous flows: Sample complexity of flow matching. *arXiv preprint arXiv:2512.01286*, 2025.

Aditi Gupta, Soon Hoe Lim, Annan Yu, and N Benjamin Erichson. Sharpen Your Flow: Sharpness-Aware Sampling for Flow Matching. *arXiv preprint arXiv:2605.11547*, 2026.

Ernst Hairer, Syvert P. Nørsett, and Gerhard Wanner. *Solving Ordinary Differential Equations I: Nonstiff Problems*, volume 8 of *Springer Series in Computational Mathematics*. Springer, 2 edition, 1993.

Yihan He, Qishuo Yin, Yuan Cao, Jianqing Fan, and Han Liu. A theory on flow matching with neural networks. *arXiv preprint arXiv:2606.10089*, 2026.

Juha Heinonen. *Lectures on Analysis on Metric Spaces*. Universitext. Springer, New York, 2001.

Michael F Hutchinson. A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines. *Communications in Statistics-Simulation and Computation*, 18(3):1059–1076, 1989.

Arieh Iserles. *A First Course in the Numerical Analysis of Differential Equations*. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2 edition, 2008.

Arnulf Jentzen, Felix Lindner, and Primož Pušnik. On the Alekseev-Gröbner formula in Banach spaces. *Discrete and Continuous Dynamical Systems - B*, 24(8):4475–4511, 2019. ISSN 1531-3492.

Robert L Jerrard and Cy Maor. Vanishing geodesic distance for right-invariant sobolev metrics on diffeomorphism groups. *Annals of Global Analysis and Geometry*, 55(4):631–656, 2019.

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.

Sangyun Lee, Beomsu Kim, and Jong Chul Ye. Minimizing trajectory curvature of ode-based generative models. In *International Conference on Machine Learning*, pp. 18957–18973. PMLR, 2023.

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.

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.

Pascal Massart. The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *The annals of Probability*, pp. 1269–1283, 1990.

Peter W Michor and David Mumford. Vanishing geodesic distance on spaces of submanifolds and diffeomorphisms. *Documenta Mathematica*, 10:217–245, 2005.

Michael Röckner and Feng-Yu Wang. Weak Poincaré inequalities and L2-convergence rates of Markov semigroups. *Journal of Functional Analysis*, 185(2):564–603, 2001.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-Based Generative Modeling through Stochastic Differential Equations. In *International Conference on Learning Representations*, 2021.

Arthur Stéphanovitch. Lipschitz regularity in Flow Matching and Diffusion Models: sharp sampling rates and functional inequalities. *arXiv preprint arXiv:2604.06065*, 2026.

Chenxi Tao and Seung-Kyum Choi. On the role of strain and vorticity in numerical integration error for flow matching. *arXiv preprint arXiv:2605.06680*, 2026.

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

Zhengyu Zhou and Weiwei Liu. An Error Analysis of Flow Matching for Deep Generative Modeling. In *Forty-second International Conference on Machine Learning*, 2025.

APPENDIX CONTENTS

A Prior Work on Flow Matching Error Bounds **13**

B Flow Regularity and Generator Sensitivity **13**

C Proofs for Generation Error **14**

C.1 Local Tube Control for Generation Error 15

C.2 Comparison with Global Lipschitz Bounds 18

D Proofs for ODE Discretization **19**

E Estimating Path Growth **21**

E.1 Estimating the Pathwise Stability Curve 21

E.2 Estimating the Squared Frobenius Norm 22

E.3 Tail Bounds for Terminal Error 22

F Functional Inequalities and Correction Proofs **23**

F.1 Transport-induced Weighted Poincaré Inequality 23

F.2 Converse Weighted Transfer to Weak Poincaré 24

F.3 Weak Poincaré Convergence 25

F.4 Warm-start χ^2, KL, and TV correction 26

F.5 Weighted Logarithmic Sobolev Inequality 27

G Examples of Non-Uniform Pathwise Stability **27**

G.1 A linear interpolation with finite pathwise growth 27

G.2 Dissipative Residual Flow 28

G.3 Linear Growth Conditions for Residual Fields 29

G.4 Gaussian-Path Residual Flow 29

G.5 A Bounded Perturbation with an Unbounded Jacobian 30

H CFM Loss, Sampling Error, and Experiments **30**

H.1 CFM Loss and Velocity Error 30

H.2 Generation Error from Population Loss 31

H.3 Euler Step Counts and Time Allocation 43

H.4 Numerical Experiments 45

I Two Small Networks with Different Sampling Errors **48**

A Prior Work on Flow Matching Error Bounds

Training loss and statistical error. Flow matching fits a velocity field for a prescribed probability path (Lipman et al., 2023). Conditional paths provide regression targets whose conditional expectation is the marginal velocity (Lipman et al., 2023; Tong et al., 2024). Work on finite sample learning bounds and minimax rates controls how closely a trained field approaches that velocity (Fukumizu et al., 2025; Gao et al., 2024b; Zhou & Liu, 2025; Kunkel & Trabs, 2025; Gaur et al., 2025; He et al., 2026). Our generation error bounds take the velocity mismatch as input and study how the ODE propagates it.

From velocity error to terminal error. Classical ODE perturbation formulas propagate a velocity mismatch through the subsequent flow (Jentzen et al., 2019). For stochastic interpolants, Albergo & Vanden-Eijnden (2023, Proposition 3) bound terminal W_2 under a uniform spatial Lipschitz condition on the learned field. Related flow matching bounds use the same type of global stability input (Benton et al., 2024, Theorem 1), while the Gaussian interpolation analysis of Gao et al. (2024a, Proposition 6.2(ii)) retains a time integrated global coefficient. Our comparison theorem retains the Jacobian encountered between the compared paths. The sharp population loss result asks a separate question: what one can guarantee without observing those individual paths.

Sensitivity, solvers, and output laws. The variational equation controls derivatives of a flow map (Hairer et al., 1993, Section I.14). Local solver defects propagate through the later dynamics (Hairer et al., 1993; Iserles, 2008; Cao & Petzold, 2004); studies of straighter paths, Jacobian penalties, and time grids address complementary numerical choices (Lee et al., 2023; Finlay et al., 2020; Tao & Choi, 2026; Gupta et al., 2026). Gaussian Poincaré and the chain rule give the direct weighted inequality after transport (Bakry et al., 2014). Further weak inequalities require their own reciprocal weight and semigroup assumptions (Röckner & Wang, 2001; Bobkov & Ledoux, 2009; Cattiaux et al., 2010).

B Flow Regularity and Generator Sensitivity

This appendix states the classical ODE regularity assumptions and proves the sensitivity bounds used in Section 2. The logarithmic norm gives the refinement that retains contraction.

Assumption 1 (Flow regularity). *Let $v_\theta : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ be a time-dependent vector field. There exists a Borel set $\mathcal{Z}_\theta \subseteq \mathbb{R}^d$ with $\gamma(\mathcal{Z}_\theta) = 1$ such that for every $z \in \mathcal{Z}_\theta$:*

1. *the ODE*
$$\frac{d}{dt} X_t^\theta(z) = v_\theta(t, X_t^\theta(z)), \qquad X_0^\theta(z) = z,$$
admits a unique solution on $[0,1]$;
2. *for every $t \in [0,1]$, the map $x \mapsto v_\theta(t,x)$ is continuously differentiable on an open neighborhood of the path $\{X_t^\theta(z) : t \in [0,1]\}$;*
3. *the flow map $z \mapsto X_t^\theta(z)$ is differentiable along the path and its Jacobian satisfies*
$$\frac{d}{dt} D_z X_t^\theta(z) = \nabla_x v_\theta(t, X_t^\theta(z)) D_z X_t^\theta(z), \qquad D_z X_0^\theta(z) = \mathrm{I}_d;$$
4. *the pathwise Jacobian growth is finite:*
$$\mathcal{J}_\theta(z) = \int_0^1 \|\nabla_x v_\theta(t, X_t^\theta(z))\|_{\mathrm{op}}\, dt < \infty.$$

Definition 3 (Logarithmic-norm refinement). *For a matrix $A \in \mathbb{R}^{d\times d}$, its Euclidean logarithmic norm measures instantaneous norm growth in the linear ODE $\dot{\xi} = A\xi$ (Hairer et al., 1993, Section I.10). It is defined by*
$$m(A) := \lambda_{\max}\left(\frac{A + A^\top}{2}\right).$$

For a learned flow, define

$$\mathcal{G}_\theta(z) := \int_0^1 m\big(\nabla_x v_\theta(t, X_t^\theta(z))\big)\, dt.$$

Since $m(A) \le \|A\|_{\mathrm{op}}$ for every matrix A, one has $\mathcal{G}_\theta(z) \le \mathcal{J}_\theta(z)$.

Theorem 6 (Pathwise generator sensitivity)**.** *Under Assumption 1, for every $z \in \mathcal{Z}_\theta$ and every $t \in [0,1]$,*

$$\begin{aligned}\|D_z X_t^\theta(z)\|_{\mathrm{op}} &\le \exp\left(\int_0^t m(\nabla_x v_\theta(s, X_s^\theta(z)))\, ds\right) \\ &\le \exp\left(\int_0^t \|\nabla_x v_\theta(s, X_s^\theta(z))\|_{\mathrm{op}}\, ds\right).\end{aligned}$$

In particular, $\|D_z T_\theta(z)\|_{\mathrm{op}} \le e^{\mathcal{G}_\theta(z)} \le e^{\mathcal{J}_\theta(z)} = \mathsf{A}_\theta(z)$.

Proof. Fix a unit vector $h \in \mathbb{R}^d$ and set $\xi_t = D_z X_t^\theta(z) h$. The variational equation gives

$$\dot{\xi}_t = \nabla_x v_\theta(t, X_t^\theta(z))\xi_t, \qquad \xi_0 = h.$$

The coefficient has integrable operator norm by Assumption 1. Uniqueness for this linear equation implies that ξ_t never vanishes. Thus, almost everywhere,

$$\begin{aligned}\frac{d}{dt}\log\|\xi_t\| &= \frac{\langle \xi_t, \nabla_x v_\theta(t, X_t^\theta(z))\xi_t\rangle}{\|\xi_t\|^2} \\ &\le m(\nabla_x v_\theta(t, X_t^\theta(z))).\end{aligned}$$

Integrating and taking the supremum over unit vectors h proves the first bound. The second follows from $m(B) \le \|B\|_{\mathrm{op}}$. □

Corollary 2 (Moment and quantile control of terminal stretch)**.** *Under Assumption 1, for every $q \ge 1$,*

$$\mathbb{E}_\gamma \|D_z T_\theta(Z)\|_{\mathrm{op}}^q \le \mathbb{E}_\gamma[e^{q\mathcal{G}_\theta(Z)}] \le \mathbb{E}_\gamma[e^{q\mathcal{J}_\theta(Z)}],$$

whenever the right-hand side is finite. Moreover, if $\mathbb{P}(\mathcal{J}_\theta(Z) \le R) \ge 1-\delta$, then

$$\mathbb{P}(\|D_z T_\theta(Z)\|_{\mathrm{op}} \le e^R) \ge 1-\delta.$$

Proof. Both statements follow immediately from Theorem 6. □

Remark 1 (Uniform Lipschitzness as a special case)**.** *If there exists $L \in L^1([0,1])$ such that*

$$\|\nabla_x v_\theta(t,x)\|_{\mathrm{op}} \le L(t), \qquad \forall (t,x) \in [0,1]\times\mathbb{R}^d,$$

then

$$\mathcal{J}_\theta(Z) \le \int_0^1 L(t)\, dt, \qquad \mathsf{A}_\theta(Z) \le \exp\left(\int_0^1 L(t)\, dt\right).$$

Thus an integrable uniform Lipschitz bound gives a deterministic upper bound on the pathwise stability factor.

C Proofs for Generation Error

This appendix proves the results used in Section 2.2.

Assumption 2 (Exact–learned comparison)**.** *There exists a Borel set $\mathcal{Z}_{\star,\theta} \subseteq \mathbb{R}^d$ with $\gamma(\mathcal{Z}_{\star,\theta}) = 1$ such that, for every $z \in \mathcal{Z}_{\star,\theta}$:*

1. *the population and learned ODEs*

$$\dot{X}_t^\star(z) = u_t(X_t^\star(z)), \qquad \dot{X}_t^\theta(z) = v_\theta(t, X_t^\theta(z)),$$

are well defined on $[0,1]$ with $X_0^\star(z) = X_0^\theta(z) = z$;

2. *the learned vector field $x \mapsto v_\theta(t,x)$ is continuously differentiable on a neighborhood of the segment*
$$\Gamma_t^{\star,\theta}(z) := \{X_t^\star(z) + \lambda(X_t^\theta(z) - X_t^\star(z)) : \lambda \in [0,1]\};$$

3. *the tube Jacobian growth is finite:*
$$\mathcal{J}_{\theta,\mathrm{tube}}(z) := \int_0^1 \sup_{x \in \Gamma_t^{\star,\theta}(z)} \|\nabla_x v_\theta(t,x)\|_{\mathrm{op}}\, dt < \infty;$$

4. *the pathwise velocity error is finite:*
$$\mathcal{E}_\theta(z) := \int_0^1 \|v_\theta(t, X_t^\star(z)) - u_t(X_t^\star(z))\|\, dt < \infty.$$

C.1 Local Tube Control for Generation Error

Classical ODE error bounds multiply each velocity error or local numerical defect by the growth that follows it (Hairer et al., 1993; Iserles, 2008). Here this growth depends on the random trajectories being compared.

For a vector-valued map $v : \mathbb{R}^d \to \mathbb{R}^d$, write
$$\|\nabla_x^2 v(x)\|_{\mathrm{op}\to\mathrm{op}} := \sup_{\|u\|=1} \|D_x(\nabla_x v)(x)[u]\|_{\mathrm{op}}.$$
This is the operator norm of the derivative of the Jacobian map $x \mapsto \nabla_x v(x)$.

Theorem 7 (Local tube control)**.** *Let $X_t(z)$ and $\widehat{X}_t(z)$ be two paths with the same initialization, and let $\Gamma_t(z)$ be the segment between them. Suppose $v_\theta(t,\cdot)$ is twice differentiable near this segment and set*
$$H_t(z) := \sup_{x \in \Gamma_t(z)} \|\nabla_x^2 v_\theta(t,x)\|_{\mathrm{op}\to\mathrm{op}}.$$
Then
$$\mathcal{J}_{\theta,\mathrm{tube}}(z) \le \int_0^1 \left[\|\nabla_x v_\theta(t, X_t(z))\|_{\mathrm{op}} + H_t(z)\|\widehat{X}_t(z) - X_t(z)\|\right] dt.$$
Exponentiating gives the corresponding bound on $\mathsf{A}_{\theta,\mathrm{tube}}(z)$.

Proof. For $x \in \Gamma_t(z)$, the mean value inequality gives
$$\begin{aligned}\|\nabla_x v_\theta(t,x)\|_{\mathrm{op}} &\le \|\nabla_x v_\theta(t, X_t(z))\|_{\mathrm{op}} + H_t(z)\|x - X_t(z)\| \\ &\le \|\nabla_x v_\theta(t, X_t(z))\|_{\mathrm{op}} + H_t(z)\|\widehat{X}_t(z) - X_t(z)\|.\end{aligned}$$
Take the supremum over $x \in \Gamma_t(z)$ and integrate in time. □

The next condition combines the classical variational estimate (Hairer et al., 1993, Section I.14) with Taylor's formula and a first exit argument. It bounds the separation using growth on the learned path and curvature in a specified neighborhood.

Proposition 1 (A local condition for tube control)**.** *Fix an initialization z. Suppose the population and learned trajectories are absolutely continuous solutions on $[0,1]$, start at z, and the learned ODE has unique solutions. The ODE identities hold almost everywhere in time. Set*
$$E := \mathcal{E}_\theta(z), \qquad A := \exp\left(\int_0^1 \|\nabla_x v_\theta(t, X_t^\theta(z))\|_{\mathrm{op}}\, dt\right).$$
Assume $E < \infty$ and $A < \infty$. If $E = 0$, the two paths coincide. For $E > 0$, let $\rho = 2AE$ and suppose that $v_\theta(t,\cdot)$ is twice continuously differentiable on a neighborhood of $\{x : \|x - X_t^\theta(z)\| \le \rho\}$ for every t. Assume that the derivatives are jointly measurable and that
$$H := \int_0^1 \sup_{\|x - X_t^\theta(z)\| \le \rho} \|\nabla_x^2 v_\theta(t,x)\|_{\mathrm{op}\to\mathrm{op}}\, dt < \infty.$$

If $2A^2EH < 1$*, then*

$$\sup_{0\le t\le 1} \|X_t^\star(z) - X_t^\theta(z)\| < 2AE, \tag{8}$$

$$\mathcal{J}_{\theta,\mathrm{tube}}(z) \le \mathcal{J}_\theta(z) + 2AEH. \tag{9}$$

Consequently,

$$\|T_\theta(z) - T_\star(z)\| \le A\exp(2AEH)E.$$

Proof. Write $d_t = X_t^\star(z) - X_t^\theta(z)$ and $r_t = u_t(X_t^\star(z)) - v_\theta(t, X_t^\star(z))$. Then $\int_0^1 \|r_t\|\,dt = E$. As long as $\|d_t\| \le \rho$, Taylor's formula gives

$$\dot{d}_t = \nabla_x v_\theta(t, X_t^\theta(z))d_t + r_t + R_t, \qquad \|R_t\| \le \tfrac{1}{2}H_t\|d_t\|^2,$$

where H_t is the supremum in the definition of H. The homogeneous linear equation has amplification at most A on every subinterval of $[0,1]$, by the variational bound (Hairer et al., 1993).

If the paths first reach distance ρ at time t_0, variation of constants gives

$$\rho = \|d_{t_0}\| \le AE + \frac{A}{2}H\rho^2 = AE(1 + 2A^2EH) < 2AE = \rho,$$

a contradiction. Continuity proves equation 8, so every comparison segment lies in the stated neighborhood. The mean value formula for the Jacobian gives

$$\sup_{x\in\Gamma_t^{\star,\theta}(z)} \|\nabla_x v_\theta(t,x)\|_{\mathrm{op}} \le \|\nabla_x v_\theta(t, X_t^\theta(z))\|_{\mathrm{op}} + H_t\|d_t\|.$$

Integration proves equation 9. The generation bound gives the terminal estimate. If $E = 0$, the population path also solves the learned ODE, so uniqueness makes the two paths equal. □

Corollary 3 (A loss based local error bound)**.** *Fix $\varepsilon > 0$. Suppose the population and learned paths are absolutely continuous solutions from the same initial state for almost every z, the learned ODE has unique solutions, and $\mathsf{A}_\theta(z) < \infty$. Assume $v_\theta(t,\cdot)$ is twice continuously differentiable on a neighborhood of each ball $\{x : \|x - X_t^\theta(z)\| \le 2\varepsilon\mathsf{A}_\theta(z)\}$, with jointly measurable derivatives. Suppose $\mathbb{E}_\gamma\mathcal{E}_\theta(Z)^2 \le \eta < \infty$. Define*

$$H_\varepsilon(z) := \int_0^1 \sup_{\|x - X_t^\theta(z)\| \le 2\varepsilon\mathsf{A}_\theta(z)} \|\nabla_x^2 v_\theta(t,x)\|_{\mathrm{op}\to\mathrm{op}}\,dt$$

and let $C_\varepsilon = \{z : H_\varepsilon(z) < \infty,\ 2\varepsilon\mathsf{A}_\theta(z)^2 H_\varepsilon(z) < 1\}$. Then

$$\mathbb{P}_\gamma\big(Z \in C_\varepsilon,\ \|T_\theta(Z) - T_\star(Z)\| \le \varepsilon\mathsf{A}_\theta(Z)\exp\big(2\varepsilon\mathsf{A}_\theta(Z)H_\varepsilon(Z)\big)\big) \ge \mathbb{P}_\gamma(Z \in C_\varepsilon) - \frac{\eta}{\varepsilon^2}.$$

For CFM, η may be taken as the population excess loss in Equation 2.

Proof. Markov's inequality gives $\mathbb{P}_\gamma(\mathcal{E}_\theta(Z) > \varepsilon) \le \eta/\varepsilon^2$. On $C_\varepsilon \cap \{\mathcal{E}_\theta \le \varepsilon\}$, the radius $2\mathsf{A}_\theta\mathcal{E}_\theta$ in Proposition 1 is at most $2\varepsilon\mathsf{A}_\theta$. Its local curvature integral is therefore at most H_ε, and its smallness condition holds. The proposition gives the displayed terminal bound on this event. □

Theorem 8 (Pathwise error propagation)**.** *Under Assumption 2, for every $z \in \mathcal{Z}_{\star,\theta}$ and every $t \in [0,1]$,*

$$\begin{aligned}
\|X_t^\theta(z) - X_t^\star(z)\| &\le \int_0^t \exp\left(\int_s^t m(M_r(z))\,dr\right)\|v_\theta(s, X_s^\star(z)) - u_s(X_s^\star(z))\|\,ds \\
&\le \int_0^t \exp\left(\int_s^t L_r^{\star,\theta}(z)\,dr\right)\|v_\theta(s, X_s^\star(z)) - u_s(X_s^\star(z))\|\,ds \\
&\le \exp\left(\int_0^t L_r^{\star,\theta}(z)\,dr\right)\int_0^t \|v_\theta(s, X_s^\star(z)) - u_s(X_s^\star(z))\|\,ds.
\end{aligned}$$

Here M_r and $L_r^{\star,\theta}$ are defined before Theorem 2. At $t = 1$, the last term is $\mathsf{A}_{\theta,\mathrm{tube}}(z)\mathcal{E}_\theta(z)$.

Proof. Fix $z \in \mathcal{Z}_{\star,\theta}$ and set $\Delta_t := X_t^\theta(z) - X_t^\star(z)$. Then

$$\dot{\Delta}_t = v_\theta(t, X_t^\theta) - u_t(X_t^\star) = \{v_\theta(t, X_t^\theta) - v_\theta(t, X_t^\star)\} + \{v_\theta(t, X_t^\star) - u_t(X_t^\star)\}.$$

By the fundamental theorem of calculus along $\Gamma_t^{\star,\theta}(z)$,

$$\dot{\Delta}_t = M_t(z)\Delta_t + v_\theta(t, X_t^\star) - u_t(X_t^\star),$$

where

$$M_t(z) = \int_0^1 \nabla_x v_\theta(t, X_t^\star(z) + \lambda\Delta_t)\, d\lambda, \qquad L_t^{\star,\theta}(z) = \sup_{x \in \Gamma_t^{\star,\theta}(z)} \|\nabla_x v_\theta(t,x)\|_{\mathrm{op}}.$$

Therefore, in the comparison sense,

$$\frac{d}{dt}\|\Delta_t\| \le m(M_t(z))\|\Delta_t\| + \|v_\theta(t, X_t^\star) - u_t(X_t^\star)\|.$$

Since $\Delta_0 = 0$, Grönwall's inequality gives

$$\|\Delta_t\| \le \int_0^t \exp\left(\int_s^t m(M_r(z))\, dr\right) \|v_\theta(s, X_s^\star) - u_s(X_s^\star)\|\, ds.$$

The logarithmic norm is convex, so $m(M_t(z)) \le \int_0^1 m(\nabla_x v_\theta(t, X_t^\star + \lambda\Delta_t))\, d\lambda \le L_t^{\star,\theta}(z)$ (Hairer et al., 1993, Section I.10). This gives the tube bound. In one dimension, variation of constants gives equality before taking absolute values. If the residual has one sign in time, taking its absolute value preserves equality. When both terminal maps are increasing, monotone coupling identifies the $L^2(\gamma)$ endpoint displacement with W_2. □

Remark 2 (One-sided tube stability)**.** *Using the logarithmic norm m from Definition 3, define the one-sided tube coefficient*

$$\ell_t^{\star,\theta}(z) := \sup_{x \in \Gamma_t^{\star,\theta}(z)} m(\nabla_x v_\theta(t,x))\,.$$

Then the same comparison argument gives the sharper time localized estimate

$$\|T_\theta(z) - T_\star(z)\| \le \int_0^1 \exp\left(\int_s^1 \ell_r^{\star,\theta}(z)\, dr\right) \|v_\theta(s, X_s^\star(z)) - u_s(X_s^\star(z))\|\, ds.$$

Consequently,

$$\|T_\theta(z) - T_\star(z)\| \le \exp\left(\int_0^1 \left(\ell_r^{\star,\theta}(z)\right)_+ dr\right) \mathcal{E}_\theta(z).$$

This can be sharper than the operator-norm tube factor when the learned dynamics is one-sided contractive along the compared tube.

Indeed, for $\Delta_t = X_t^\theta - X_t^\star$,

$$\frac{d}{dt}\|\Delta_t\| \le \ell_t^{\star,\theta}(z)\|\Delta_t\| + \|v_\theta(t, X_t^\star) - u_t(X_t^\star)\|$$

in the comparison sense; variation of constants gives the displayed bound.

Remark 3 (Amplification after each source error)**.** *The proof gives the first inequality in Equation 3:*

$$\|T_\theta(z) - T_\star(z)\| \le \int_0^1 \exp\left(\int_s^1 L_r^{\star,\theta}(z)\, dr\right) \|v_\theta(s, X_s^\star(z)) - u_s(X_s^\star(z))\|\, ds.$$

Each residual is weighted by the growth that follows its time of occurrence. Bounding this factor by $\mathsf{A}_{\theta,\mathrm{tube}}(z)$ gives the product form.

Corollary 4 (W_2 generation error)**.** *Assume $T_\star(Z), T_\theta(Z) \in L^2(\gamma)$. Under Assumption 2,*

$$W_2(\widehat{\mu}_\theta, \mu) \le \|\mathsf{A}_{\theta,\mathrm{tube}}\mathcal{E}_\theta\|_{L^2(\gamma)}.$$

More generally, for $q \in (1, \infty]$ and Hölder conjugate q',

$$W_2(\widehat{\mu}_\theta, \mu) \le \|\mathsf{A}_{\theta,\mathrm{tube}}\|_{L^{2q}(\gamma)} \|\mathcal{E}_\theta\|_{L^{2q'}(\gamma)}.$$

Proof. The synchronous coupling $Z \mapsto (T_\theta(Z), T_\star(Z))$ has marginals $\widehat{\mu}_\theta$ and μ. Since the terminal variables are square-integrable,

$$W_2(\widehat{\mu}_\theta, \mu)^2 \le \mathbb{E}\|T_\theta(Z) - T_\star(Z)\|^2.$$

Theorem 8 gives

$$W_2(\widehat{\mu}_\theta, \mu) \le \|\mathsf{A}_{\theta,\mathrm{tube}}\mathcal{E}_\theta\|_{L^2(\gamma)}.$$

The Hölder form follows from $\|AB\|_{L^2} \le \|A\|_{L^{2q}}\|B\|_{L^{2q'}}$. □

Corollary 5 (Population velocity-risk control)**.** *Assume $X_t^\star(Z) \sim \mu_t$ for every $t \in [0,1]$. Then*

$$\|\mathcal{E}_\theta\|_{L^2(\gamma)} \le \left(\int_0^1 \mathbb{E}_{X\sim\mu_t}\|v_\theta(t,X) - u_t(X)\|^2\,dt\right)^{1/2}.$$

Consequently, under the hypotheses of Corollary 4, if $\mathsf{A}_{\theta,\mathrm{tube}} \in L^\infty(\gamma)$, then

$$W_2(\widehat{\mu}_\theta, \mu) \le \|\mathsf{A}_{\theta,\mathrm{tube}}\|_{L^\infty(\gamma)} \left(\int_0^1 \mathbb{E}_{X\sim\mu_t}\|v_\theta(t,X) - u_t(X)\|^2\,dt\right)^{1/2}.$$

Proof. By Cauchy–Schwarz in time,

$$\mathcal{E}_\theta(Z)^2 = \left(\int_0^1 \|v_\theta(t, X_t^\star(Z)) - u_t(X_t^\star(Z))\|\,dt\right)^2 \le \int_0^1 \|v_\theta(t, X_t^\star(Z)) - u_t(X_t^\star(Z))\|^2\,dt.$$

Taking expectation over $Z \sim \gamma$ and using $X_t^\star(Z) \sim \mu_t$ gives the first claim. The second follows from Corollary 4. □

Corollary 6 (Tail bound for terminal error)**.** *Under Assumption 2, for any $R > 0$ and $\varepsilon > 0$,*

$$\mathbb{P}(\|T_\theta(Z) - T_\star(Z)\| > \varepsilon) \le \mathbb{P}(\mathcal{J}_{\theta,\mathrm{tube}}(Z) > R) + \mathbb{P}(\mathcal{E}_\theta(Z) > \varepsilon e^{-R}).$$

Proof. On the event $\{\mathcal{J}_{\theta,\mathrm{tube}}(Z) \le R\} \cap \{\mathcal{E}_\theta(Z) \le \varepsilon e^{-R}\}$, Theorem 8 gives $\|T_\theta(Z) - T_\star(Z)\| \le e^R \varepsilon e^{-R} = \varepsilon$. Taking complements proves the claim. □

C.2 Comparison with Global Lipschitz Bounds

Suppose the hypotheses of Theorem 2 hold and $\|D_x v_\theta(t,x)\|_{\mathrm{op}} \le L(t)$ for all states, with $L \ge 0$ integrable. Then $L_t^{\star,\theta}(z) \le L(t)$ and

$$\begin{aligned} W_2(\widehat{\mu}_\theta, \mu) &\le \|\mathsf{A}_{\theta,\mathrm{tube}}\mathcal{E}_\theta\|_2 \\ &\le e^{\int_0^1 L(t)\,dt}\|\mathcal{E}_\theta\|_2 \\ &\le e^{\int_0^1 L(t)\,dt}\left[\int_0^1 \mathbb{E}_{X\sim\mu_t}\|v_\theta(t,X) - u_t(X)\|^2\,dt\right]^{1/2}. \end{aligned}$$

The last expression recovers Benton et al. (2024, Theorem 1) under the common flow assumptions. The joint bound can be smaller because it keeps the growth on each comparison tube together with its source error.

For a more precise time coefficient, retain the first bound in Equation 4. Cauchy–Schwarz in time and the population marginals give

$$\begin{aligned} W_2^2(\widehat{\mu}_\theta, \mu) &\le \mathbb{E}_\gamma\left[\left(\int_0^1 e^{\int_s^1 L(r)\,dr}\|v_\theta(s, X_s^\star) - u_s(X_s^\star)\|\,ds\right)^2\right] \\ &\le \left(\int_0^1 e^{2\int_s^1 L(r)\,dr}\,ds\right)\int_0^1 \mathbb{E}_{X\sim\mu_t}\|v_\theta(t,X) - u_t(X)\|^2\,dt. \end{aligned}$$

For $L(t) = C_3$, the time coefficient is $(e^{2C_3} - 1)/(2C_3)$, with value one at $C_3 = 0$. This is the coefficient in Gao et al. (2024a, Proposition 6.2(ii), Equation (6.10)). Their Equation (6.9) measures velocity error along the learned law and uses population regularity. The comparison here uses population velocity loss and learned tube growth. It refines the velocity propagation bound for the same fields. Smoothing, stopping, and statistical estimation contribute additional terms in analyses that optimize those choices.

D PROOFS FOR ODE DISCRETIZATION

Let $0 = t_0 < t_1 < \cdots < t_N = 1$, $h_k = t_{k+1} - t_k$, and $h = \max_k h_k$. Let $\Phi^\theta_{s,t}$ be the exact learned subflow from time t to time s.

Assumption 3 (Numerical tube regularity)**.** *There exists a Borel set $\mathcal{Z}_{\theta,h} \subseteq \mathbb{R}^d$ with $\gamma(\mathcal{Z}_{\theta,h}) = 1$ such that, for every $z \in \mathcal{Z}_{\theta,h}$:*

1. *the exact learned flow and numerical trajectory are well defined. For each k, the subflows $\Phi^\theta_{s,t_k}(x)$ exist for all $s \in [t_k, t_{k+1}]$ and all x on the segment joining $X^\theta_{t_k}(z)$ and $\widetilde{X}^\theta_k(z)$. They are continuously differentiable in x near this segment, and their derivatives satisfy the variational equation;*

2. *the learned vector field is continuously differentiable on a neighborhood of the numerical error tube*
$$\Gamma^{\theta,h}_{k,s}(z) := \left\{ \Phi^\theta_{s,t_k}\Big(X^\theta_{t_k}(z) + \lambda(\widetilde{X}^\theta_k(z) - X^\theta_{t_k}(z)) \Big) : \lambda \in [0,1] \right\};$$

3. *the local tube Jacobian growths*
$$\ell^\theta_k(z) := \int_{t_k}^{t_{k+1}} \sup_{x \in \Gamma^{\theta,h}_{k,s}(z)} \|\nabla_x v_\theta(s,x)\|_{\mathrm{op}}\, ds$$
are finite, and
$$\mathcal{J}_{\theta,h,\mathrm{tube}}(z) := \sum_{k=0}^{N-1} \ell^\theta_k(z), \qquad \mathsf{A}_{\theta,h,\mathrm{tube}}(z) := e^{\mathcal{J}_{\theta,h,\mathrm{tube}}(z)};$$

4. *the one-step local defects*
$$\tau^\theta_k(z) := \|\Psi^\theta_k(\widetilde{X}^\theta_k(z)) - \Phi^\theta_{t_{k+1},t_k}(\widetilde{X}^\theta_k(z))\|$$
are finite.

Theorem 9 (Propagation of individual local defects)**.** *Under Assumption 3, for every $z \in \mathcal{Z}_{\theta,h}$,*
$$\|\widetilde{X}^\theta_N(z) - X^\theta_1(z)\| \le \sum_{k=0}^{N-1} \exp\left(\sum_{j=k+1}^{N-1} \ell^\theta_j(z) \right) \tau^\theta_k(z).$$
Consequently,
$$\|\widetilde{X}^\theta_N(z) - X^\theta_1(z)\| \le \mathsf{A}_{\theta,h,\mathrm{tube}}(z) \sum_{k=0}^{N-1} \tau^\theta_k(z).$$

Proof. Let $D_k := \widetilde{X}^\theta_k - X^\theta_{t_k}$. Since both trajectories start from z, $D_0 = 0$. For one step,
$$D_{k+1} = \Psi^\theta_k(\widetilde{X}^\theta_k) - \Phi^\theta_{t_{k+1},t_k}(X^\theta_{t_k}).$$
Adding and subtracting $\Phi^\theta_{t_{k+1},t_k}(\widetilde{X}^\theta_k)$ gives
$$\|D_{k+1}\| \le \|\Phi^\theta_{t_{k+1},t_k}(\widetilde{X}^\theta_k) - \Phi^\theta_{t_{k+1},t_k}(X^\theta_{t_k})\| + \tau^\theta_k.$$
By the variational bound for the learned subflow restricted to the numerical tube,
$$\|\Phi^\theta_{t_{k+1},t_k}(\widetilde{X}^\theta_k) - \Phi^\theta_{t_{k+1},t_k}(X^\theta_{t_k})\| \le e^{\ell^\theta_k(z)} \|D_k\|.$$
Thus
$$\|D_{k+1}\| \le e^{\ell^\theta_k(z)} \|D_k\| + \tau^\theta_k(z).$$
Unrolling this recursion and using $D_0 = 0$ gives
$$\|D_N\| \le \sum_{k=0}^{N-1} \exp\left(\sum_{j=k+1}^{N-1} \ell^\theta_j(z) \right) \tau^\theta_k(z).$$
The simplified bound follows because $\sum_{j=k+1}^{N-1} \ell^\theta_j(z) \le \sum_{j=0}^{N-1} \ell^\theta_j(z) = \mathcal{J}_{\theta,h,\mathrm{tube}}(z)$. □

Corollary 7 (W_2 discretization error)**.** *Assume $X_1^\theta(Z), \widetilde{X}_N^\theta(Z) \in L^2(\gamma)$. Under Assumption 3,*

$$W_2(\widetilde{\mu}_{\theta,h}, \widehat{\mu}_\theta) \le \left\| \mathsf{A}_{\theta,h,\mathrm{tube}} \sum_{k=0}^{N-1} \tau_k^\theta \right\|_{L^2(\gamma)}.$$

More generally, for $q \in (1, \infty]$ and Hölder conjugate q',

$$W_2(\widetilde{\mu}_{\theta,h}, \widehat{\mu}_\theta) \le \|\mathsf{A}_{\theta,h,\mathrm{tube}}\|_{L^{2q}(\gamma)} \left\| \sum_{k=0}^{N-1} \tau_k^\theta \right\|_{L^{2q'}(\gamma)}.$$

Proof. The coupling $Z \mapsto (\widetilde{X}_N^\theta(Z), X_1^\theta(Z))$ has marginals $\widetilde{\mu}_{\theta,h}$ and $\widehat{\mu}_\theta$. Thus

$$W_2(\widetilde{\mu}_{\theta,h}, \widehat{\mu}_\theta)^2 \le \mathbb{E}\|\widetilde{X}_N^\theta(Z) - X_1^\theta(Z)\|^2.$$

The result follows from Theorem 9 and Hölder's inequality. □

Corollary 8 (Error bound for an order-p solver)**.** *Under the hypotheses of Corollary 7, suppose that, for some $p > 0$,*

$$\sum_{k=0}^{N-1} \tau_k^\theta(Z) \le C_{\theta,h}(Z) h^p.$$

Then

$$W_2(\widetilde{\mu}_{\theta,h}, \widehat{\mu}_\theta) \le \|\mathsf{A}_{\theta,h,\mathrm{tube}} C_{\theta,h}\|_{L^2(\gamma)} h^p.$$

If the displayed norm is uniformly bounded over a family of grids with $h \to 0$, the numerical error is $O(h^p)$ along that family.

Proof. Substitute the defect bound into Corollary 7. □

Corollary 9 (Explicit Euler local defect)**.** *Under the hypotheses of Corollary 7, assume additionally that v_θ is continuously differentiable in t and x on the relevant numerical tubes. For explicit Euler, $\Psi_k^\theta(x) = x + h_k v_\theta(t_k, x)$, define*

$$\mathscr{D}_t v_\theta(t,x) := \partial_t v_\theta(t,x) + \nabla_x v_\theta(t,x) v_\theta(t,x).$$

Then

$$\tau_k^\theta(z) \le \frac{h_k^2}{2} \sup_{s \in [t_k, t_{k+1}]} \left\| \mathscr{D}_t v_\theta\Big(s, \Phi_{s,t_k}^\theta(\widetilde{X}_k^\theta(z))\Big) \right\|.$$

Consequently, if

$$\mathcal{M}_{\theta,h}(z) := \sum_{k=0}^{N-1} h_k \sup_{s \in [t_k, t_{k+1}]} \left\| \mathscr{D}_t v_\theta\left(s, \Phi_{s,t_k}^\theta(\widetilde{X}_k^\theta(z))\right) \right\|,$$

then

$$\sum_{k=0}^{N-1} \tau_k^\theta(z) \le \frac{h}{2} \mathcal{M}_{\theta,h}(z),$$

and

$$W_2(\widetilde{\mu}_{\theta,h}, \widehat{\mu}_\theta) \le \frac{h}{2} \|\mathsf{A}_{\theta,h,\mathrm{tube}} \mathcal{M}_{\theta,h}\|_{L^2(\gamma)}.$$

Proof. Let $Y_s := \Phi_{s,t_k}^\theta(\widetilde{X}_k^\theta(z))$ for $s \in [t_k, t_{k+1}]$. The exact subflow satisfies

$$\Phi_{t_{k+1},t_k}^\theta(\widetilde{X}_k^\theta) = \widetilde{X}_k^\theta + \int_{t_k}^{t_{k+1}} v_\theta(s, Y_s)\, ds,$$

whereas the Euler step is

$$\Psi_k^\theta(\widetilde{X}_k^\theta) = \widetilde{X}_k^\theta + h_k v_\theta(t_k, \widetilde{X}_k^\theta).$$

Therefore

$$\tau_k^\theta \le \int_{t_k}^{t_{k+1}} \|v_\theta(s, Y_s) - v_\theta(t_k, \widetilde{X}_k^\theta)\|\, ds.$$

Since

$$\frac{d}{ds} v_\theta(s, Y_s) = \partial_t v_\theta(s, Y_s) + \nabla_x v_\theta(s, Y_s) v_\theta(s, Y_s) = \mathscr{D}_t v_\theta(s, Y_s),$$

we have

$$\|v_\theta(s, Y_s) - v_\theta(t_k, \widetilde{X}_k^\theta)\| \le \int_{t_k}^{s} \|\mathscr{D}_t v_\theta(r, Y_r)\|\, dr.$$

Integrating over $s \in [t_k, t_{k+1}]$ gives the defect bound. The remaining claims follow by summation and Corollary 7. □

Corollary 10 (Total sampling error)**.** *Assume the hypotheses of Corollaries 4 and 7. Then*

$$W_2(\widetilde{\mu}_{\theta,h}, \mu) \le \left\| \mathsf{A}_{\theta,h,\mathrm{tube}} \sum_{k=0}^{N-1} \tau_k^\theta \right\|_{L^2(\gamma)} + \|\mathsf{A}_{\theta,\mathrm{tube}} \mathcal{E}_\theta\|_{L^2(\gamma)}.$$

Proof. Use the triangle inequality

$$W_2(\widetilde{\mu}_{\theta,h}, \mu) \le W_2(\widetilde{\mu}_{\theta,h}, \widehat{\mu}_\theta) + W_2(\widehat{\mu}_\theta, \mu),$$

then apply Corollaries 7 and 4. □

Uniform bound in Corollary 1. For every comparison tube,

$$\int_0^1 \sup_{x \in \Gamma_t(z)} \|\nabla_x v_\theta(t, x)\|_{\mathrm{op}}\, dt \le \int_0^1 L(t)\, dt.$$

Exponentiating bounds $\mathsf{A}_{\theta,\mathrm{tube}}$. The same argument bounds $\mathsf{A}_{\theta,h,\mathrm{tube}}$. Substitute these bounds into Corollary 10. □

E Estimating Path Growth

This appendix gives finite sample bounds for the quantiles of pathwise Jacobian growth and explains how Jacobian vector products estimate the squared Frobenius norm.

E.1 Estimating the Pathwise Stability Curve

For independent samples $Z_1, \dots, Z_n \sim \gamma$, estimate the path integral by

$$\widehat{\mathcal{J}}_{\theta,h}(Z_i) := \sum_{k=0}^{N-1} h_k \|\nabla_x v_\theta(t_k, \widetilde{X}_k^\theta(Z_i))\|_{\mathrm{op}}.$$

The empirical quantiles of these estimates describe their variation across initial states. The next bound applies to exact values $\mathcal{J}_\theta(Z_i)$. Numerical integration and norm estimation introduce additional errors.

Proposition 2 (A quantile bound from sampled paths)**.** *Let $Z_1, \dots, Z_n$ be independent with law γ, and let $\widehat{Q}_n(\alpha)$ be the empirical quantile of $\mathcal{J}_\theta(Z_1), \dots, \mathcal{J}_\theta(Z_n)$. For $\beta \in (0, 1)$, set*

$$\Delta_{n,\beta} := \sqrt{\frac{\log(2/\beta)}{2n}}.$$

Then with probability at least $1 - \beta$, for every $\alpha \in (0, 1 - \Delta_{n,\beta})$,

$$Q_\theta(\alpha) \le \widehat{Q}_n(\alpha + \Delta_{n,\beta}).$$

Proof. Let F and $\widehat{F}_n$ be the population and empirical distribution functions of $\mathcal{J}_\theta(Z)$. The sharp Dvoretzky–Kiefer–Wolfowitz inequality gives (Dvoretzky et al., 1956; Massart, 1990)

$$\mathbb{P}\left(\sup_r |\widehat{F}_n(r) - F(r)| > \Delta_{n,\beta}\right) \le \beta.$$

On the event $\sup_r |\widehat{F}_n(r) - F(r)| \le \Delta_{n,\beta}$, set $r = \widehat{Q}_n(\alpha + \Delta_{n,\beta})$. Then $\widehat{F}_n(r) \ge \alpha + \Delta_{n,\beta}$, and hence $F(r) \ge \alpha$. By definition of the population quantile, $Q_\theta(\alpha) \le r$. □

The moment conditions for generation and numerical error are stated in Appendices C and D. They use the L^q norms of the corresponding amplification factors together with the source error norms.

E.2 Estimating the Squared Frobenius Norm

The Frobenius norm gives a computable upper bound on the operator norm. If ξ has independent standard Gaussian or Rademacher entries, then $\mathbb{E}\xi\xi^\top = I_d$ and

$$\mathbb{E}_\xi \|\nabla_x v_\theta(t,x)\xi\|^2 = \|\nabla_x v_\theta(t,x)\|_{\mathrm{F}}^2.$$

Averaging independent Jacobian vector products therefore estimates the squared Frobenius norm without forming the full Jacobian (Hutchinson, 1989).

Proposition 3 (A second moment bound for pathwise growth)**.** *For every trajectory on which the path integral is defined,*

$$\mathcal{J}_\theta(Z)^2 \le \int_0^1 \|\nabla_x v_\theta(t, X_t^\theta(Z))\|_{\mathrm{F}}^2 \, dt.$$

The same inequality holds after taking expectations.

Proof. Cauchy–Schwarz on the unit time interval and $\|B\|_{\mathrm{op}} \le \|B\|_{\mathrm{F}}$ give

$$\begin{aligned}
\mathcal{J}_\theta(Z)^2 &= \left(\int_0^1 \|\nabla_x v_\theta(t, X_t^\theta(Z))\|_{\mathrm{op}} \, dt\right)^2 \\
&\le \int_0^1 \|\nabla_x v_\theta(t, X_t^\theta(Z))\|_{\mathrm{op}}^2 \, dt \\
&\le \int_0^1 \|\nabla_x v_\theta(t, X_t^\theta(Z))\|_{\mathrm{F}}^2 \, dt.
\end{aligned}$$

□

Remark 4. *This estimate controls $\mathbb{E}\mathcal{J}_\theta(Z)^2$. Bounds on moments of $\mathsf{A}_\theta = e^{\mathcal{J}_\theta}$ require exponential moments of $\mathcal{J}_\theta$. Generation and numerical error additionally involve the product of tube amplification and source error.*

E.3 Tail Bounds for Terminal Error

The high-probability bounds can be summarized as follows. Under Assumption 2,

$$\mathbb{P}(\|T_\theta(Z) - T_\star(Z)\| > \varepsilon) \le \mathbb{P}(\mathcal{J}_{\theta,\mathrm{tube}}(Z) > R) + \mathbb{P}(\mathcal{E}_\theta(Z) > \varepsilon e^{-R}).$$

Under Assumption 3,

$$\mathbb{P}(\|\widetilde{X}_N^\theta(Z) - X_1^\theta(Z)\| > \varepsilon) \le \mathbb{P}(\mathcal{J}_{\theta,h,\mathrm{tube}}(Z) > R) + \mathbb{P}\left(\sum_{k=0}^{N-1} \tau_k^\theta(Z) > \varepsilon e^{-R}\right).$$

These bounds separate the tail of the source error from the tail of tube amplification. Controlling both probabilities gives a high probability guarantee for the terminal error.

F FUNCTIONAL INEQUALITIES AND CORRECTION PROOFS

We first transfer the Gaussian Poincaré inequality through a terminal map. The later results assume a converse weighted Poincaré inequality to obtain a weak inequality for the ordinary gradient energy. Variance decay then uses a reversible Markov semigroup, and correction bounds additionally require a bounded initial density relative to the target.

F.1 TRANSPORT-INDUCED WEIGHTED POINCARÉ INEQUALITY

Let $T_{\mathrm{tar}} : \mathbb{R}^d \to \mathbb{R}^d$ satisfy

$$\mu = (T_{\mathrm{tar}})_{\#}\gamma, \qquad \gamma = \mathcal{N}(0, \mathrm{I}_d).$$

Assumption 4 (Transport regularity)**.** *Assume:*

1. *T_{tar} is differentiable γ-almost surely;*
2. *for every $f \in C_c^\infty(\mathbb{R}^d)$, $f \circ T_{\mathrm{tar}} \in H^1(\gamma)$ and*

$$\nabla(f \circ T_{\mathrm{tar}})(z) = DT_{\mathrm{tar}}(z)^\top \nabla f(T_{\mathrm{tar}}(z))$$

   γ-almost surely;

3. *for every finite R, $\mathbb{E}[\|DT_{\mathrm{tar}}(Z)\|_{\mathrm{op}}^2 \mathbf{1}_{\{\|T_{\mathrm{tar}}(Z)\| \le R\}}] < \infty$, and all integrals below are finite for the test functions considered.*

Define

$$G_{\mathrm{tar}}(y) := \mathbb{E}\big[DT_{\mathrm{tar}}(Z)DT_{\mathrm{tar}}(Z)^\top \mid T_{\mathrm{tar}}(Z) = y\big], \qquad a_{\mathrm{tar}}(y) := 1 \vee \lambda_{\max}(G_{\mathrm{tar}}(y)).$$

Theorem 10 (Matrix-weighted Poincaré)**.** *Under Assumption 4,*

$$\mathrm{Var}_\mu(f) \le \int \langle G_{\mathrm{tar}}(y)\nabla f(y), \nabla f(y)\rangle \, \mu(dy) \le \int a_{\mathrm{tar}}(y)\|\nabla f(y)\|^2 \, \mu(dy).$$

Proof. Since $\mu = (T_{\mathrm{tar}})_{\#}\gamma$, the Gaussian Poincaré inequality (Bakry et al., 2014) and the chain rule in Assumption 4 give

$$\begin{aligned}\mathrm{Var}_\mu(f) &= \mathrm{Var}_\gamma(f \circ T_{\mathrm{tar}}) \\ &\le \mathbb{E}_\gamma\big\|DT_{\mathrm{tar}}(Z)^\top \nabla f(T_{\mathrm{tar}}(Z))\big\|^2.\end{aligned}$$

Conditioning on $T_{\mathrm{tar}}(Z)$ yields

$$\mathrm{Var}_\mu(f) \le \int \langle G_{\mathrm{tar}}(y)\nabla f(y), \nabla f(y)\rangle \, \mu(dy).$$

The scalar bound follows from

$$\langle G_{\mathrm{tar}}\nabla f, \nabla f\rangle \le a_{\mathrm{tar}}\|\nabla f\|^2.$$

□

Corollary 11 (Growth of the target transport controls stretch moments)**.** *Suppose $\|DT_{\mathrm{tar}}(Z)\|_{\mathrm{op}} \le e^{J_{\mathrm{tar}}(Z)}$ almost surely. Then*

$$a_{\mathrm{tar}}(T_{\mathrm{tar}}(Z)) \le 1 \vee \mathbb{E}[e^{2J_{\mathrm{tar}}(Z)} \mid T_{\mathrm{tar}}(Z)].$$

For every $p \ge 1$,

$$\int a_{\mathrm{tar}}^p \, d\mu \le 1 + \mathbb{E}_\gamma e^{2pJ_{\mathrm{tar}}(Z)}.$$

Proof. Convexity of the largest eigenvalue gives

$$\begin{aligned}\lambda_{\max}(G_{\mathrm{tar}}(T_{\mathrm{tar}}(Z))) &\le \mathbb{E}[\|DT_{\mathrm{tar}}(Z)\|_{\mathrm{op}}^2 \mid T_{\mathrm{tar}}(Z)] \\ &\le \mathbb{E}[e^{2J_{\mathrm{tar}}(Z)} \mid T_{\mathrm{tar}}(Z)].\end{aligned}$$

Using the definition of a_{tar} and conditional Jensen's inequality,

$$\int a_{\mathrm{tar}}^p \, d\mu \le 1 + \mathbb{E}_\gamma \left(\mathbb{E}[e^{2J_{\mathrm{tar}}(Z)} \mid T_{\mathrm{tar}}(Z)] \right)^p \le 1 + \mathbb{E}_\gamma e^{2pJ_{\mathrm{tar}}(Z)}.$$

□

The weighted Poincaré inequality derived above has the direct form

$$\mathrm{Var}_\mu(f) \le \int a_{\mathrm{tar}} \|\nabla f\|^2 \, d\mu.$$

To obtain a weak Poincaré inequality for the standard Dirichlet form, we use the converse weighted condition below. The transfer follows the truncation argument of Cattiaux et al. (2010, Theorem 5.1); the reciprocal weight condition and its use here are stated below.

F.2 Converse Weighted Transfer to Weak Poincaré

Theorem 11 (Converse weighted Poincaré implies weak Poincaré)**.** *Let μ be a probability measure and let $\omega : \mathbb{R}^d \to (0, 1]$ be measurable. Suppose there exists $C_{\mathrm{cw}} < \infty$ such that*

$$\inf_{c \in \mathbb{R}} \int (f - c)^2 \omega \, d\mu \le C_{\mathrm{cw}} \int \|\nabla f\|^2 \, d\mu \tag{10}$$

for every $f \in C_c^\infty(\mathbb{R}^d)$. Write $\mathrm{osc}_\mu(f) = \operatorname{ess\,sup}_\mu f - \operatorname{ess\,inf}_\mu f$. Then, for every $u \in (0, 1]$,

$$\mathrm{Var}_\mu(f) \le \frac{C_{\mathrm{cw}}}{u} \int \|\nabla f\|^2 \, d\mu + \mu(\omega < u) \, \mathrm{osc}_\mu(f)^2. \tag{11}$$

Consequently, if $\mu(\omega < u) \le K u^p$ for some $K \ge 1$ and $p > 0$, then μ satisfies a weak Poincaré inequality with profile

$$\beta(r) \le C_{\mathrm{cw}} K^{1/p} r^{-1/p}, \qquad r \in (0, 1/2].$$

Proof. Fix $u \in (0, 1]$ and $f \in C_c^\infty(\mathbb{R}^d)$. Let

$$m_\omega = \frac{\int f\omega \, d\mu}{\int \omega \, d\mu},$$

which is well defined because $0 < \int \omega \, d\mu \le 1$. The weighted mean m_ω lies between the essential infimum and essential supremum of f, and hence

$$|f - m_\omega| \le \mathrm{osc}_\mu(f) \qquad \mu\text{-a.s.}$$

Using the variational characterization of variance,

$$\mathrm{Var}_\mu(f) \le \int (f - m_\omega)^2 \, d\mu.$$

Split the integral into $\{\omega \ge u\}$ and $\{\omega < u\}$. On $\{\omega \ge u\}$,

$$(f - m_\omega)^2 \le u^{-1} (f - m_\omega)^2 \omega.$$

On $\{\omega < u\}$,

$$(f - m_\omega)^2 \le \mathrm{osc}_\mu(f)^2.$$

Therefore,

$$\mathrm{Var}_\mu(f) \le u^{-1} \int (f - m_\omega)^2 \omega \, d\mu + \mu(\omega < u) \, \mathrm{osc}_\mu(f)^2.$$

By equation 10,

$$\int (f - m_\omega)^2 \omega \, d\mu = \inf_{c \in \mathbb{R}} \int (f - c)^2 \omega \, d\mu \le C_{\mathrm{cw}} \int \|\nabla f\|^2 \, d\mu.$$

This proves equation 11. If $\mu(\omega < u) \le K u^p$, choose $u = (r/K)^{1/p}$. Since $K \ge 1$ and $r \le 1/2$, this choice lies in $(0, 1]$. Substituting it into equation 11 gives the stated weak Poincaré profile. □

Corollary 12 (Stretch moments and weak Poincaré inequalities)**.** *Assume the converse weighted inequality*

$$\inf_{c\in\mathbb{R}} \int \frac{(f-c)^2}{a_{\mathrm{tar}}}\, d\mu \le C_{\mathrm{cw}} \int \|\nabla f\|^2\, d\mu.$$

If $a_{\mathrm{tar}} \in L^p(\mu)$ for some $p \ge 1$, then the weak Poincaré profile satisfies

$$\beta(r) \le C_{\mathrm{cw}} \|a_{\mathrm{tar}}\|_{L^p(\mu)} r^{-1/p}, \qquad r \in (0, 1/2].$$

In particular, if $\|DT_{\mathrm{tar}}(Z)\|_{\mathrm{op}} \le e^{J_{\mathrm{tar}}(Z)}$ and $\mathbb{E}e^{\vartheta J_{\mathrm{tar}}(Z)} \le M_\vartheta$ for some $\vartheta > 2$ and $M_\vartheta \ge 1$, then

$$\beta(r) \le C_\vartheta C_{\mathrm{cw}} M_\vartheta^{2/\vartheta} r^{-2/\vartheta}.$$

Proof. Since $a_{\mathrm{tar}} \ge 1$, Markov's inequality gives $\mu(1/a_{\mathrm{tar}} < u) \le u^p \int a_{\mathrm{tar}}^p\, d\mu$. Apply Theorem 11 with $\omega = 1/a_{\mathrm{tar}}$. For the second claim, take $p = \vartheta/2$ and use Corollary 11 to obtain $\int a_{\mathrm{tar}}^p\, d\mu \le 1 + M_\vartheta \le 2M_\vartheta$. □

F.3 Weak Poincaré Convergence

Theorem 12 (Weak Poincaré to variance decay)**.** *Let $(P_t)_{t\ge 0}$ be a conservative reversible Markov semigroup that is strongly continuous on $L^2(\mu)$, with invariant measure μ and Dirichlet form $\mathcal{E}$. Suppose that, for some $r_0 \in (0, 1]$ and some nonincreasing function $\beta : (0, r_0] \to (0, \infty)$,*

$$\mathrm{Var}_\mu(f) \le \beta(r)\mathcal{E}(f, f) + r\ \mathrm{osc}_\mu(f)^2, \qquad r \in (0, r_0],$$

for all bounded f in the form domain. Assume further that

$$\beta(r) \le C_\beta r^{-\alpha}, \qquad r \in (0, r_0],$$

for some $C_\beta > 0$ and $\alpha > 0$. Then there exists $C_{\alpha,r_0} > 0$ such that

$$\mathrm{Var}_\mu(P_t f) \le C_{\alpha,r_0} \left(1 + \frac{t}{C_\beta}\right)^{-1/\alpha} \mathrm{osc}_\mu(f)^2.$$

Proof. Let $V(t) := \mathrm{Var}_\mu(P_t f)$ and $O := \mathrm{osc}_\mu(f)^2$. If $O = 0$, the claim is immediate. The energy identity for a reversible semigroup gives (Bakry et al., 2014, Sections 1.6–1.7)

$$V'(t) = -2\mathcal{E}(P_t f, P_t f).$$

Since $\mathrm{osc}_\mu(P_t f) \le \mathrm{osc}_\mu(f)$, applying weak Poincaré to $P_t f$ gives

$$V(t) \le \beta(r)\mathcal{E}(P_t f, P_t f) + rO.$$

Let t_* be the first time at which $V(t) \le 2r_0 O$, with $t_* = 0$ if this already holds at time zero. Since V is nonincreasing and $V(0) \le O/4$, the time t_* is finite. Before t_* choose $r = r_0$. Then $r_0 O < V/2$, so the weak Poincaré inequality and $\beta(r_0) \le C_\beta r_0^{-\alpha}$ give $V' \le -V/\beta(r_0) \le -r_0^\alpha V/C_\beta$. Hence $t_* \le C_\beta r_0^{-\alpha}[\log(1/(8r_0))]_+$.

For $t \ge t_*$ with $V(t) > 0$, choose $r = V(t)/(2O) \le r_0$. The weak inequality gives $\mathcal{E}(P_t f, P_t f) \ge V(t)/(2\beta(r))$, and therefore

$$V'(t) \le -\frac{V(t)^{1+\alpha}}{2^\alpha C_\beta O^\alpha}.$$

Integration yields

$$V(t) \le \left[V(t_*)^{-\alpha} + \frac{\alpha(t - t_*)}{2^\alpha C_\beta O^\alpha}\right]^{-1/\alpha}.$$

If V reaches zero, the conclusion continues to hold by monotonicity. The bound on t_* and $V(t_*) \le O/4$ give $V(t) \le C_{\alpha,r_0} O(1 + t/C_\beta)^{-1/\alpha}$, including the initial interval $t < t_*$. □

F.4 WARM-START χ^2, KL, AND TV CORRECTION

Theorem 13 (Warm-start correction). *Let $(P_t)_{t\geq 0}$ be a conservative reversible Markov semigroup with invariant measure μ. Suppose*

$$\mathrm{Var}_\mu(P_t f) \leq \rho(t)\,\mathrm{osc}_\mu(f)^2$$

for every bounded function f in the class to which the variance-decay estimate applies. Let $\nu_0 \ll \mu$, set $g_0 = d\nu_0/d\mu$, and assume

$$0 \leq g_0 \leq B_{\mathrm{ws}} \qquad \mu\text{-a.s.}$$

Assume moreover that the variance-decay estimate applies to g_0, for instance by bounded form-domain approximation. Then

$$\chi^2(\nu_0 P_t\|\mu) \leq B_{\mathrm{ws}}^2\rho(t),$$
$$\mathrm{KL}(\nu_0 P_t\|\mu) \leq B_{\mathrm{ws}}^2\rho(t),$$

and

$$\|\nu_0 P_t - \mu\|_{\mathrm{TV}} \leq \frac{B_{\mathrm{ws}}}{2}\rho(t)^{1/2}.$$

Proof. By reversibility,

$$\frac{d(\nu_0 P_t)}{d\mu} = P_t g_0.$$

Since $\mu(g_0) = 1$ and μ is invariant, $\mu(P_t g_0) = 1$. Hence

$$\chi^2(\nu_0 P_t\|\mu) = \int (P_t g_0 - 1)^2\,d\mu = \mathrm{Var}_\mu(P_t g_0).$$

Using the variance decay assumption,

$$\chi^2(\nu_0 P_t\|\mu) \leq \rho(t)\,\mathrm{osc}_\mu(g_0)^2 \leq B_{\mathrm{ws}}^2\rho(t).$$

For KL, let $g = d\nu/d\mu$. Since $\log g \leq g - 1$,

$$\mathrm{KL}(\nu\|\mu) = \int g\log g\,d\mu \leq \int g(g-1)\,d\mu = \int g^2\,d\mu - 1 = \chi^2(\nu\|\mu).$$

For TV, Cauchy–Schwarz gives

$$\|\nu - \mu\|_{\mathrm{TV}} = \frac{1}{2}\int |g-1|\,d\mu \leq \frac{1}{2}\left(\int (g-1)^2\,d\mu\right)^{1/2} = \frac{1}{2}\chi^2(\nu\|\mu)^{1/2}.$$

Applying these inequalities to $\nu = \nu_0 P_t$ gives the claim. □

Corollary 13 (Target transport growth moment to warm-start correction). *Assume the hypotheses of Corollary 12 in its exponential-moment case, with $\vartheta > 2$ and $\mathbb{E}_\gamma e^{\vartheta J_{\mathrm{tar}}(Z)} \leq M_\vartheta$ for $M_\vartheta \geq 1$. Let $(P_t)_{t\geq 0}$ be a conservative reversible Markov semigroup that is strongly continuous on $L^2(\mu)$, with invariant measure μ. Assume that the weak Poincaré inequality extends to bounded functions in its form domain. Its Dirichlet form is*

$$\mathcal{E}(f,f) = \int \|\nabla f\|^2\,d\mu.$$

Let $\nu_{\mathrm{prop}} \ll \mu$ be a flow proposal, for instance $\nu_{\mathrm{prop}} = (T_\theta)_\#\gamma$, and set

$$g_0 := \frac{d\nu_{\mathrm{prop}}}{d\mu}.$$

Assume

$$0 \leq g_0 \leq B_{\mathrm{ws}} \qquad \mu\text{-a.s.}$$

and assume that the variance-decay estimate applies to g_0 by the approximation condition in Theorem 13. Then

$$\chi^2(\nu_{\mathrm{prop}} P_t\|\mu) \leq C_\vartheta B_{\mathrm{ws}}^2\left(1 + \frac{t}{C_{\mathrm{cw}} M_\vartheta^{2/\vartheta}}\right)^{-\vartheta/2},$$

$$\mathrm{KL}(\nu_{\mathrm{prop}}P_t\|\mu) \le C_\vartheta B_{\mathrm{ws}}^2 \left(1 + \frac{t}{C_{\mathrm{cw}}M_\vartheta^{2/\vartheta}}\right)^{-\vartheta/2},$$

and

$$\|\nu_{\mathrm{prop}}P_t - \mu\|_{\mathrm{TV}} \le C_\vartheta B_{\mathrm{ws}} \left(1 + \frac{t}{C_{\mathrm{cw}}M_\vartheta^{2/\vartheta}}\right)^{-\vartheta/4}.$$

Proof. Corollary 12 gives

$$\beta(r) \le C_\vartheta C_{\mathrm{cw}} M_\vartheta^{2/\vartheta} r^{-2/\vartheta}.$$

Apply Theorem 12 with

$$\alpha = \frac{2}{\vartheta}, \qquad C_\beta = C_\vartheta C_{\mathrm{cw}} M_\vartheta^{2/\vartheta}.$$

Then apply Theorem 13 with $\nu_0 = \nu_{\mathrm{prop}}$. □

F.5 Weighted Logarithmic Sobolev Inequality

The same change of variables gives a weighted logarithmic Sobolev inequality, which bounds entropy by a weighted gradient energy (Bakry et al., 2014, Chapter 5).

Theorem 14 (Transport-induced weighted log-Sobolev inequality)**.** *Under Assumption 4, for every smooth compactly supported f,*

$$\mathrm{Ent}_\mu(f^2) \le 2\int \langle G_{\mathrm{tar}}(y)\nabla f(y), \nabla f(y)\rangle\, \mu(dy) \le 2\int a_{\mathrm{tar}}(y)\|\nabla f(y)\|^2\, \mu(dy).$$

Proof. Since $\mu = (T_{\mathrm{tar}})_{\#}\gamma$,

$$\mathrm{Ent}_\mu(f^2) = \mathrm{Ent}_\gamma((f \circ T_{\mathrm{tar}})^2).$$

The Gaussian log-Sobolev inequality (Bakry et al., 2014, Chapter 5) gives

$$\mathrm{Ent}_\gamma((f \circ T_{\mathrm{tar}})^2) \le 2\mathbb{E}_\gamma\|\nabla(f \circ T_{\mathrm{tar}})(Z)\|^2.$$

Using the chain rule and conditioning as in Theorem 10 gives the matrix-weighted bound. The scalar bound follows from

$$\langle G_{\mathrm{tar}}\nabla f, \nabla f\rangle \le a_{\mathrm{tar}}\|\nabla f\|^2.$$

□

Remark 5 (Entropy decay)**.** *An additional transfer to a weak logarithmic Sobolev inequality can give entropy decay estimates (Cattiaux et al., 2007; 2011). The correction bounds above use variance decay together with* $\mathrm{KL} \le \chi^2$ *and* $\mathrm{TV} \le \frac{1}{2}\sqrt{\chi^2}$.

G Examples of Non-Uniform Pathwise Stability

The explicit examples have finite path or tube amplification moments despite a nonintegrable uniform Jacobian bound. They illustrate expansion, dissipation, and bounded velocity perturbations. We also give sufficient growth conditions for residual fields and Gaussian interpolation models.

G.1 A linear interpolation with finite pathwise growth

Proposition 4. *Let $Z \sim \mathcal{N}(0,1)$, fix $\kappa > 0$, and set*

$$X_1 = Z + \kappa Z^3, \qquad I_t = (1-t)Z + tX_1, \qquad U = X_1 - Z.$$

Write $F_t(z) = z + \kappa t z^3$. The population velocity $u_t(x) = \mathbb{E}[U \mid I_t = x]$ is

$$u_t(x) = \kappa(F_t^{-1}(x))^3,$$

and its flow is $X_t(z) = F_t(z)$. For every $t > 0$,

$$\sup_x |\partial_x u_t(x)| = \sup_x \partial_x u_t(x) = \frac{1}{t}.$$

Thus neither spatial supremum has a finite time integral on $[0,1]$*. Along each path,*

$$\mathcal{J}(z) = \log(1 + 3\kappa z^2), \qquad \mathsf{A}(z) = 1 + 3\kappa z^2 = F_1'(z). \tag{12}$$

In particular, $\mathbb{E}\mathsf{A}(Z)^p < \infty$ *for every finite* $p > 0$*, and*

$$\mathbb{E}\mathsf{A}(Z)^2 = 1 + 6\kappa + 27\kappa^2.$$

The material derivative $\partial_t u_t + u_t \partial_x u_t$ *vanishes. Explicit Euler therefore reproduces this population flow on every time grid.*

Proof. For each $t \in [0,1]$, $F_t'(z) = 1 + 3\kappa t z^2 > 0$ and $F_t(z)$ tends to $\pm\infty$ as z does. Hence F_t is a smooth increasing bijection. Conditioning on $I_t = x$ fixes $Z = F_t^{-1}(x)$, which gives the velocity formula. Since

$$\partial_t F_t(z) = \kappa z^3 = u_t(F_t(z)),$$

F_t solves the population ODE. Since $|F_t^{-1}(x)| \le |x|$, we have $0 \le \partial_x u_t(x) \le 3\kappa x^2$ on $[0,1]$. The velocity is therefore locally Lipschitz in space uniformly on compact sets, and the solution is unique. Differentiating gives

$$\partial_x u_t(F_t(z)) = \frac{3\kappa z^2}{1 + 3\kappa t z^2}.$$

For $t > 0$ this expression is nonnegative and tends to $1/t$ as $|z| \to \infty$. Integrating it for fixed z proves equation 12. Gaussian moments give the moment statements, using $\mathbb{E}Z^2 = 1$ and $\mathbb{E}Z^4 = 3$. Finally, $u_t(F_t(z)) = \kappa z^3$ is constant in time, so its material derivative is zero. If an Euler step starts at $F_{t_k}(z)$, it returns

$$F_{t_k}(z) + (t_{k+1} - t_k)\kappa z^3 = F_{t_{k+1}}(z).$$

Induction proves the claim for every grid. □

This example uses a dependent coupling between the Gaussian source and the target. As $\kappa \downarrow 0$, its pathwise amplification converges to one in every finite L^p, while the time integral of the spatial supremum remains infinite for every $\kappa > 0$.

It also makes the weighted Poincaré inequality explicit. With $T_{\mathrm{tar}} = F_1$ and $\mu = (F_1)_{\#}\gamma$, invertibility gives

$$G_{\mathrm{tar}}(y) = [1 + 3\kappa(F_1^{-1}(y))^2]^2.$$

Theorem 5 therefore yields

$$\mathrm{Var}_\mu(f) \le \int [1 + 3\kappa(F_1^{-1}(y))^2]^2 |f'(y)|^2 \, \mu(dy).$$

This weight is finite at each state, although the flow has no integrable uniform Jacobian bound.

G.2 Dissipative Residual Flow

Let $d = 1$, $Z \sim \mathcal{N}(0,1)$, and consider

$$\dot{X}_t = a(t)X_t - b(t)X_t^3,$$

where $a, b : [0,1] \to [0,\infty)$ are measurable. Assume

$$A_1 := \int_0^1 a(t)\,dt < \infty, \qquad C := \int_0^1 b(s)e^{2A_s}\,ds < \infty, \qquad A_s := \int_0^s a(r)\,dr.$$

The dissipative cubic term prevents finite time blowup on $[0,1]$. When $b \equiv 0$, the equation reduces to a linear ODE. If $b(t) > 0$ on a set of positive measure, the integrated absolute spatial Jacobian supremum is infinite, since

$$\partial_x v_t(x) = a(t) - 3b(t)x^2.$$

The pathwise Jacobian growth has finite exponential moments of all orders. Indeed,

$$\frac{d}{dt}X_t^2 = 2a(t)X_t^2 - 2b(t)X_t^4.$$

Solving this scalar logistic equation gives

$$X_t^2 = \frac{e^{2A_t} Z^2}{1 + 2Z^2 \int_0^t b(s) e^{2A_s}\, ds}.$$

Hence

$$\int_0^1 b(t) X_t^2\, dt = \frac{1}{2} \log(1 + 2CZ^2).$$

Therefore

$$\mathcal{J}(Z) = \int_0^1 |a(t) - 3b(t) X_t^2|\, dt \le A_1 + \frac{3}{2} \log(1 + 2CZ^2).$$

Consequently, for every $\vartheta > 0$,

$$\mathbb{E} e^{\vartheta \mathcal{J}(Z)} \le e^{\vartheta A_1} \mathbb{E}(1 + 2CZ^2)^{3\vartheta/2} < \infty.$$

When $b(t) > 0$ on a set of positive measure, the velocity field has an infinite spatial Jacobian supremum on that set. Its pathwise stability factor has finite moments of every order under Gaussian initialization.

Remark 6. *The cubic term limits the time spent at large states. The global one-sided Jacobian bound is also finite, since* $\sup_x \partial_x v_t(x) = a(t)$. *This example shows how contraction and pathwise operator norms avoid the infinite absolute spatial bound.*

G.3 Linear Growth Conditions for Residual Fields

Let $d \ge 1$, $Z \sim \mathcal{N}(0, I_d)$, and consider

$$\dot{X}_t = A_t X_t + b_t + r_\theta(t, X_t),$$

where $A_t \in \mathbb{R}^{d \times d}$ and $b_t \in \mathbb{R}^d$ are measurable in time. Assume r_θ is measurable in time and continuously differentiable in the state. The first two terms are affine, and the residual is nonlinear.

Assume

$$\|\nabla_x r_\theta(t, x)\|_{\mathrm{op}} \le L_0(t) + L_1(t)\|x\|$$

and

$$\|A_t x + b_t + r_\theta(t, x)\| \le K(t)(1 + \|x\|).$$

Assume K, L_0, L_1 and $t \mapsto \|A_t\|_{\mathrm{op}}$ are nonnegative integrable bounds. Grönwall's inequality gives

$$\|X_t\| \le \exp\left(\int_0^1 K(s)\, ds \right) (1 + \|Z\|).$$

Therefore,

$$\mathcal{J}(Z) \le \int_0^1 (\|A_t\|_{\mathrm{op}} + L_0(t))\, dt + \exp\left(\int_0^1 K(s)\, ds \right) (1 + \|Z\|) \int_0^1 L_1(t)\, dt.$$

Gaussian initialization gives $\mathbb{E} e^{\vartheta \mathcal{J}(Z)} < \infty$ for every $\vartheta > 0$. Thus finite pathwise exponential moments allow Jacobians that grow linearly in the state.

G.4 Gaussian-Path Residual Flow

The field $v_\theta(t, x) = \dot{m}_t + \lambda(t)(x - m_t) + r_\theta(t, x)$ is a special case with $A_t = \lambda(t) I_d$ and $b_t = \dot{m}_t - \lambda(t) m_t$. If m is absolutely continuous, $|\lambda|, \rho_0, \rho_1, L_0, L_1$ are integrable, and $\|r_\theta(t, x)\| \le \rho_0(t) + \rho_1(t)\|x\|$, $\|D_x r_\theta(t, x)\|_{\mathrm{op}} \le L_0(t) + L_1(t)\|x\|$, the preceding argument gives every finite amplification moment under Gaussian initialization.

G.5 A Bounded Perturbation with an Unbounded Jacobian

Let $Z, G \sim \mathcal{N}(0,1)$ be independent and use the conditional path $X_t^{\mathrm{int}} = \cos(\pi t/2)Z + \sin(\pi t/2)G$. Its marginals are standard Gaussian. The conditional velocity $U_t = (\pi/2)[-\sin(\pi t/2)Z + \cos(\pi t/2)G]$ is independent of X_t^{int} and has zero mean, so the population velocity is zero. For $0 < \varepsilon \le 1$, consider

$$v_\varepsilon(t,x) = \varepsilon \sin(x^2).$$

The bound $|v_\varepsilon| + |\partial_x v_\varepsilon| \le 2(1+|x|)$ holds with the same constant for every ε. Since $\partial_x v_\varepsilon = 2\varepsilon x \cos(x^2)$, both its absolute spatial supremum and its one-sided spatial supremum are infinite. Bounded velocity and local smoothness give a unique global flow with $|X_t^\varepsilon(z) - z| \le \varepsilon t$. Every point between the population state z and the learned state has absolute value at most $|z| + \varepsilon t$. Hence

$$\mathcal{J}_{\varepsilon,\mathrm{tube}}(z) \le 2\varepsilon|z| + \varepsilon^2, \qquad \mathcal{E}_\varepsilon(z) = \varepsilon|\sin(z^2)|.$$

The source error is nonzero for almost every Gaussian initial state. The estimate gives $\mathbb{E}\exp(p\mathcal{J}_{\varepsilon,\mathrm{tube}}) < \infty$ for every finite $p > 0$, uniformly over $0 < \varepsilon \le 1$. Since $\mathcal{E}_\varepsilon \le \varepsilon$, the generation error bound is finite. Thus one fixed linear growth class contains CFM perturbations with nonzero residuals and infinite global absolute and one-sided Jacobian bounds.

H CFM Loss, Sampling Error, and Experiments

H.1 CFM Loss and Velocity Error

Assume the conditional variables in Section 2.2 have finite second moments, the displayed risks are finite, and the population ODE is well posed with $X_t^\star(Z) \sim \mu_t$. Since $u_t(X_t^{\mathrm{int}}) = \mathbb{E}[U_t \mid X_t^{\mathrm{int}}]$, the cross term in

$$v_\theta(t, X_t^{\mathrm{int}}) - U_t = [v_\theta(t, X_t^{\mathrm{int}}) - u_t(X_t^{\mathrm{int}})] + [u_t(X_t^{\mathrm{int}}) - U_t]$$

has zero expectation. Expanding the square gives the conditional and marginal loss identity used in CFM (Lipman et al., 2023). For $T \in (0,1]$, let $\Delta_{\mathrm{CFM},T}$ be the excess loss integrated over $[0,T]$, without normalization by T. Then

$$\Delta_{\mathrm{CFM},T} = \int_0^T \mathbb{E}_{X\sim\mu_t}\|v_\theta(t,X) - u_t(X)\|^2\,dt, \qquad \|\mathcal{E}_{\theta,T}\|_2^2 \le T\Delta_{\mathrm{CFM},T}. \tag{13}$$

Here $\mathcal{E}_{\theta,T}$ is the integral defining $\mathcal{E}_\theta$ restricted to $[0,T]$. The inequality follows from Cauchy–Schwarz in time, as in Corollary 5. Norms of latent random variables in this appendix are under γ. Equation 13 concerns population expectations. The conditional velocity variance is common to all models, so it cancels when their expected CFM losses are compared. Finite validation samples also require control of sampling error. An upper bound on raw loss also bounds excess loss, but retains the conditional velocity variance.

Remark 7 (Moments of the product)**.** *Finite moments of the amplification factor alone do not combine with an L^2 source bound to control their product. For example, let $Z \sim \mathcal{N}(0,1)$ and define*

$$A = e^{|Z|}, \qquad E = e^{Z^2/4 - |Z|}.$$

Then $\mathbb{E}A^p < \infty$ for every finite $p > 0$, and

$$\mathbb{E}E^2 = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-2|z|}\,dz < \infty.$$

But

$$\mathbb{E}(AE)^2 = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} 1\,dz = \infty.$$

This illustrates the product integrability required by the bound. Matched higher moments are one sufficient way to ensure it.

Synchronous coupling and Hölder's inequality yield, for $1 \le r < 2$,

$$W_r(\widehat{\mu}_\theta, \mu) \le \|\mathsf{A}_{\theta,\mathrm{tube}}\|_{2r/(2-r)}\sqrt{\eta},$$

whenever the displayed moment is finite. Here $1/r = 1/2 + (2-r)/(2r)$. For $r = 2$, bounded amplification gives the same formula with its L^∞ norm. More generally, the joint L^2 bound and matched moments in Theorem 2 apply. Corollary 6 and Markov's inequality also give

$$\mathbb{P}(\|T_\theta(Z) - T_\star(Z)\| > \varepsilon) \le \mathbb{P}(\mathcal{J}_{\theta,\mathrm{tube}}(Z) > R) + e^{2R}\eta/\varepsilon^2.$$

Population Jacobian for Gaussian interpolation. Let $X_t^{\mathrm{int}} = \alpha_t X_1 + \sigma_t Z$ with independent $Z \sim \gamma$ and X_1, and $U_t = \dot{\alpha}_t X_1 + \dot{\sigma}_t Z$. Assume differentiable coefficients, $\sigma_t > 0$, finite conditional second moments, and valid differentiation under the Gaussian integral. Conditional mean differentiation expresses the Jacobian through a conditional covariance (Benton et al., 2024, Lemma 8, Appendix B.1). Applying this identity gives

$$u_t(x) = \frac{\dot{\sigma}_t}{\sigma_t} x + \left(\dot{\alpha}_t - \frac{\alpha_t \dot{\sigma}_t}{\sigma_t}\right) \mathbb{E}[X_1 \mid X_t^{\mathrm{int}} = x],$$

$$\nabla_x u_t(x) = \frac{\dot{\sigma}_t}{\sigma_t} I_d + \frac{\alpha_t}{\sigma_t^2} \left(\dot{\alpha}_t - \frac{\alpha_t \dot{\sigma}_t}{\sigma_t}\right) \mathrm{Cov}(X_1 \mid X_t^{\mathrm{int}} = x).$$

The derivative of the conditional mean is $\alpha_t \,\mathrm{Cov}(X_1 \mid X_t^{\mathrm{int}} = x)/\sigma_t^2$, obtained by differentiating its Gaussian density. This proves the second identity. For $\alpha_t = t$ and $\sigma_t = 1 - t$, the Jacobian is $-I_d/(1-t) + t\,\mathrm{Cov}(X_1 \mid X_t^{\mathrm{int}} = x)/(1-t)^3$, for $t < 1$. A learned Jacobian bound additionally requires control of $\nabla_x(v_\theta - u)$.

Sufficient ODE conditions. The variational equation controls the growth of a trajectory perturbation (Hairer et al., 1993, Section I.14). Integrating a pathwise estimate over the initial distribution then gives a distributional bound. Uniform Lipschitz assumptions make the growth factor independent of the initial state. Our assumptions retain the growth along each path or comparison tube. For example, suppose both fields are continuous, locally Lipschitz in space uniformly on compact time and state sets, and bounded by $C(1 + \|x\|)$. If the learned field has a jointly continuous spatial derivative bounded by $C(1+\|x\|)$, the flows are well posed and differentiable as required. Grönwall bounds both paths and their comparison segment by $C_0 + C_1\|z\|$. Hence $\mathcal{J}_{\theta,\mathrm{tube}}(z) \le C_2 + C_3\|z\|$, and Gaussian initialization gives every finite amplification moment. The source moments are paired with these moments as above. Appendix G gives further examples with unbounded spatial Jacobians.

H.2 Generation Error from Population Loss

Classical ODE comparisons can measure the velocity error along either of the two flows. These choices use different stability bounds. In Gao et al. (2024a, Proposition 6.2, Equations (6.9)–(6.10)), population regularity controls an error measured along the learned flow, whereas a global learned Jacobian bound controls the population loss. The next theorem uses the population loss and allows the learned Jacobian to grow linearly with the state.

Theorem 15 (Generation error from population loss)**.** *Fix $T \in (0, 1]$ and finite $K, L \ge 0$. Suppose u is continuous with a jointly continuous spatial derivative and*

$$\|u_t(0)\| + \|D_x u_t(x)\|_{\mathrm{op}} \le L. \tag{14}$$

Suppose v_θ is continuous with a jointly continuous spatial derivative and

$$\|v_\theta(t,x)\| + \|D_x v_\theta(t,x)\|_{\mathrm{op}} \le K(1 + \|x\|). \tag{15}$$

Both inequalities hold for every $t \in [0,T]$ and $x \in \mathbb{R}^d$. Start both flows from $Z \sim \gamma$ and suppose the population flow has marginals μ_t. Write $\widehat{\mu}_{\theta,T}$ for the law of the learned flow at time T. Set

$$\eta = T \int_0^T \mathbb{E}_{X \sim \mu_t} \|v_\theta(t,X) - u_t(X)\|^2 \, dt. \tag{16}$$

There is a finite $C_{d,K,L,T}$ such that, for $0 < \eta \le 1$,

$$W_2(\widehat{\mu}_{\theta,T}, \mu_T) \le C_{d,K,L,T} \sqrt{\eta} [1 + \log(1/\eta)]^{(d+4)/4}. \tag{17}$$

If $\eta = 0$, the terminal laws agree. For each fixed dimension, Proposition 5 gives a Gaussian CFM path and a fixed linear growth constant for which the exponent $(d+4)/4$ is sharp.

The regularity assumption is on the population field. The learned Jacobian may have an infinite spatial supremum. For CFM, Equation 13 identifies η with T times the excess loss integrated over $[0,T]$. If the loss is averaged over a uniform time in this interval, the factor is T^2.

The upper bound rests on a local energy estimate: moving a trajectory requires velocity energy in a neighborhood of its initial state. We first prove this estimate for a stationary Gaussian reference and then apply the argument in population flow coordinates. The lower bound collects mass in the Gaussian tail and moves it farther into the tail at the same loss scale. Retaining a polynomial weight in these two arguments gives the weighted loss rates below.

Lemma 1 (Gaussian velocity energy and terminal displacement). *Let $b : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ be continuous and continuously differentiable in space, with jointly continuous spatial derivative. Suppose that finite $K_0, K_1 \geq 0$ satisfy*

$$|b(t,y)| \leq K_0(1+|y|), \qquad \|D_y b(t,y)\|_{\mathrm{op}} \leq K_1(1+|y|).$$

Let $Y_t(z)$ solve $\dot{Y}_t = b(t, Y_t)$ with $Y_0(z) = z$, and let γ_d be the standard Gaussian law. Define

$$\mathcal{E}_b = \int_0^1 \int |b(t,y)|^2 \, \gamma_d(dy)\, dt, \qquad B = \max\{1, e^{K_0} - 1\},$$

$$A = 2^{3d+16} 3^d e^{9/2} B^2 (1+2K_1)^d, \qquad H = 2B^2\, 2^{d/2}(1+2d).$$

For every $R \geq 0$,

$$W_2^2((Y_1)_{\#}\gamma_d, \gamma_d) \leq \mathbb{E}|Y_1(Z) - Z|^2 \leq A(1+R)^{d+4}\mathcal{E}_b + He^{-R^2/4}, \qquad Z \sim \gamma_d.$$

If $\mathcal{E}_b > 0$, set $R_b = 2\sqrt{[\log(H/\mathcal{E}_b)]_+}$, where $[a]_+ = \max\{a, 0\}$. Then

$$\mathbb{E}|Y_1(Z) - Z|^2 \leq \mathcal{E}_b\big[A(1+R_b)^{d+4} + 1\big].$$

If $\mathcal{E}_b = 0$, then $Y_1(Z) = Z$ almost surely.

Proof. Write φ_d for the standard Gaussian density and ω_d for the volume of the unit Euclidean ball. Set

$$D(z) = |Y_1(z) - z|, \qquad s(z) = 1 + |z|, \qquad r(z) = \min\{D(z), s(z)^{-1}\}.$$

Linear growth gives global existence and $D(z) \leq (e^{K_0} - 1)s(z) \leq Bs(z)$. Consequently,

$$D(z) \leq Bs(z)^2 r(z).$$

Fix a point with $D(z) > 0$ and abbreviate $s = s(z)$ and $r = r(z)$. Let τ be the first time $|Y_\tau(z) - z| = r/2$, which exists because $D(z) \geq r$. Put $m(t) = |b(t, Y_t(z))|$ and $\mathcal{H} = \{t \in [0,\tau] : m(t) \geq r/4\}$. Path length bounds displacement, so

$$\int_0^\tau m(t)\, dt \geq r/2, \qquad \int_{\mathcal{H}} m(t)\, dt \geq r/4, \qquad \int_{\mathcal{H}} m(t)^2\, dt \geq r^2/16.$$

The second inequality uses $\tau \leq 1$, and the third is Cauchy–Schwarz. On $B(z,r)$, a spatial Lipschitz constant is $L_z = (1+2K_1)s \geq 1$. For $t \in \mathcal{H}$, the ball centered at $Y_t(z)$ with radius

$$\min\{r/2, m(t)/(2L_z)\} \geq r/(8L_z)$$

lies in $B(z,r)$, up to its boundary, and $|b(t,y)| \geq m(t)/2$ there. For every $y \in B(z,r)$, the choice $r \leq s^{-1}$ gives

$$\log \frac{\varphi_d(y)}{\varphi_d(z)} = -z \cdot (y-z) - |y-z|^2/2 \geq -3/2.$$

It follows that the energy on this fixed spatial ball satisfies

$$\mathcal{E}_z := \int_0^1 \int_{B(z,r)} |b(t,y)|^2 \, \gamma_d(dy)\, dt \geq c_d \frac{\varphi_d(z) r^{d+2}}{s^d}, \qquad c_d = \frac{\omega_d e^{-3/2}}{2^{3d+6}(1+2K_1)^d}.$$

We sum these local costs using a disjoint ball selection, following the covering argument in Heinonen (2001, Chapter 1). Consider centers $\mathcal{A}_R = \{z : |z| \leq R,\ D(z) > 0\}$ and their balls $B(z, r(z))$. Select a remaining ball whose radius is more than half the supremum of the remaining radii. Assign to it all remaining centers whose balls intersect it, remove those centers, and repeat. The selected balls $B_j = B(z_j, r_j)$ are pairwise disjoint. Every center is assigned: an unassigned ball of radius $r_0 > 0$ would force infinitely many disjoint selected balls of radius greater than $r_0/2$ inside $B(0, R+1)$, which is impossible. The assigned sets E_j are Borel because r is continuous, and form a finite or countable partition of $\mathcal{A}_R$. For $z \in E_j$,

$$r(z) \leq 2r_j, \qquad |z - z_j| < r(z) + r_j \leq 3r_j.$$

Writing $s_j = 1 + |z_j|$, these relations imply

$$s(z) \leq 4s_j, \qquad \varphi_d(z) \leq e^3 \varphi_d(z_j), \qquad D(z)^2 \leq 1024 B^2 s_j^4 r_j^2.$$

Since $E_j \subset B(z_j, 3r_j)$, the local energy bound gives

$$\int_{E_j} D(z)^2 \, \gamma_d(dz) \le 1024 B^2 \omega_d 3^d e^3 \varphi_d(z_j) s_j^4 r_j^{d+2}$$
$$\le A s_j^{d+4} \mathcal{E}_{z_j}.$$

All selected centers satisfy $s_j \le 1 + R$. The original balls B_j are disjoint, so $\sum_j \mathcal{E}_{z_j} \le \mathcal{E}_b$. Therefore

$$\mathbb{E}[D(Z)^2 \mathbf{1}_{\{|Z| \le R\}}] \le A(1+R)^{d+4} \mathcal{E}_b.$$

For the tail, the displacement bound and the Gaussian exponential moment give

$$\mathbb{E}[D(Z)^2 \mathbf{1}_{\{|Z| > R\}}] \le 2B^2 e^{-R^2/4} \mathbb{E}[(1+|Z|^2) e^{|Z|^2/4}]$$
$$= 2B^2 \, 2^{d/2} (1+2d) e^{-R^2/4}.$$

Here $\mathbb{E} e^{|Z|^2/4} = 2^{d/2}$ and $\mathbb{E}[|Z|^2 e^{|Z|^2/4}] = 2d \, 2^{d/2}$ follow by Gaussian integration. Adding the compact and tail estimates proves the claim for every R. The Wasserstein bound uses the coupling $(Y_1(Z), Z)$. For positive energy, choose R_b as stated. If $\mathcal{E}_b \ge H$, then $R_b = 0$ and $H \le \mathcal{E}_b$. For zero energy, let $R \to \infty$. □

Proof of Theorem 15. Let Φ_t and $\widehat{\Phi}_t$ be the population and learned flows. Both flows are complete under the stated growth conditions. Set $A = e^{LT}$. The first variational equation and the linear growth estimate (Hairer et al., 1993, Section I.14) give

$$1 + |\Phi_t(z)| \le A(1+|z|), \qquad \|D\Phi_t(z)\|_{\mathrm{op}}, \ \|(D\Phi_t(z))^{-1}\|_{\mathrm{op}} \le A.$$

The population flow is a global C^1 diffeomorphism. Applying the same growth estimate to the backward population ODE gives $1 + |\Phi_t^{-1}(x)| \le A(1+|x|)$. For $s \in [0,1]$, define

$$\begin{aligned} Y_s(z) &= \Phi_{Ts}^{-1}(\widehat{\Phi}_{Ts}(z)), \\ q(s,y) &= T[v_\theta(Ts, \Phi_{Ts}(y)) - u_{Ts}(\Phi_{Ts}(y))], \\ b(s,y) &= (D\Phi_{Ts}(y))^{-1} q(s,y). \end{aligned} \tag{18}$$

Then $Y_0(z) = z$ and $\partial_s Y_s = b(s, Y_s)$. The map Y_s is continuous in its initial state. Uniqueness follows by mapping any solution back through Φ_{Ts} to the learned ODE. We will estimate the energy of q, without differentiating b. Put $K_q = TA(KA + L)$. The assumptions imply

$$\|D_y q(s,y)\|_{\mathrm{op}} \le K_q(1+|y|), \qquad |q(s,y)| \ge A^{-1} |b(s,y)|.$$

The energy is exactly

$$\int_0^1 \mathbb{E}_{Z \sim \gamma_d} |q(s,Z)|^2 \, ds = \eta.$$

We adapt the local energy estimate in the proof of Lemma 1. Write $D(z) = |Y_1(z) - z|$, $\rho(z) = 1 + |z|$, and $r(z) = \min\{D(z), \rho(z)^{-1}\}$. The learned flow satisfies $1 + |\widehat{\Phi}_t(z)| \le e^{Kt} \rho(z)$. Together with $\|(D\Phi_t)^{-1}\|_{\mathrm{op}} \le e^{Lt}$, this gives

$$|\partial_s Y_s(z)| \le T(K+L) e^{(K+L)Ts} \rho(z).$$

Integrating over $s \in [0,1]$ yields $D(z) \le [e^{(K+L)T} - 1]\rho(z)$. Hence, with $B = \max\{1, e^{(K+L)T} - 1\}$, we have $D(z) \le B\rho(z)$ and $D(z) \le B\rho(z)^2 r(z)$. Fix $D(z) > 0$ and abbreviate $\rho = \rho(z)$ and $r = r(z)$. Let τ be the first time $|Y_\tau(z) - z| = r/2$. For $m(s) = |b(s, Y_s(z))|$ and $\mathcal{H} = \{s \in [0,\tau] : m(s) \ge r/4\}$, path length and Cauchy–Schwarz give

$$\int_{\mathcal{H}} m(s) \, ds \ge r/4, \qquad \int_{\mathcal{H}} m(s)^2 \, ds \ge r^2/16.$$

On $B(z,r)$, $q(s,\cdot)$ has Lipschitz constant at most $L_z = (1 + 2K_q)\rho$. For $s \in \mathcal{H}$, the ball centered at $Y_s(z)$ with radius

$$\min\{r/2, m(s)/(2AL_z)\} \ge r/(8AL_z)$$

lies in $B(z,r)$ and satisfies $|q(s,y)| \geq m(s)/(2A)$. On this ball the Gaussian density obeys $\varphi_d(y) \geq e^{-3/2}\varphi_d(z)$. Thus, with ω_d the unit ball volume,

$$\int_0^1 \int_{B(z,r)} |q(s,y)|^2 \, \gamma_d(dy)\, ds \geq \frac{\omega_d e^{-3/2}}{2^{3d+6} A^{d+2}(1+2K_q)^d} \frac{\varphi_d(z) r^{d+2}}{\rho^d}.$$

Apply the same disjoint ball selection as in the proof of Lemma 1 to centers with $|z| \leq R$ and $D(z) > 0$. The radii are continuous because Y_1 is continuous. For a selected ball $B(z_j, r_j)$, its assigned centers lie within $3r_j$, have radii at most $2r_j$, and satisfy $D(z)^2 \leq 1024B^2(1+|z_j|)^4 r_j^2$. Using the local estimate above and summing the energies of the disjoint balls gives

$$\mathbb{E}[D(Z)^2 \mathbf{1}_{\{|Z|\leq R\}}] \leq C_0(1+R)^{d+4}\eta, \quad C_0 = 2^{3d+16} 3^d e^{9/2} B^2 A^{d+2}(1+2K_q)^d.$$

The same Gaussian tail calculation gives

$$\mathbb{E}[D(Z)^2 \mathbf{1}_{\{|Z|>R\}}] \leq H_0 e^{-R^2/4}, \qquad H_0 = 2B^2\, 2^{d/2}(1+2d).$$

Since $\widehat{\Phi}_T(Z) = \Phi_T(Y_1(Z))$, synchronous coupling yields

$$W_2^2(\widehat{\mu}_{\theta,T}, \mu_T) \leq A^2\{C_0(1+R)^{d+4}\eta + H_0 e^{-R^2/4}\}.$$

For $\eta > 0$, choose $R = 2\sqrt{[\log(H_0/\eta)]_+}$. The tail term is at most η, which proves Equation 17 for $0 < \eta \leq 1$. For $\eta = 0$, let $R \to \infty$. □

Gaussian mixture targets in linear CFM. Let $Z, G \sim N(0, I_d)$ and a finite valued random vector M be mutually independent. Suppose $\|M\| \leq R_0$ and let $X_1 = M + \tau G$, where $\tau > 0$. This includes the mixture targets in Appendix H.4. For $X_t^{\mathrm{int}} = (1-t)Z + tX_1$, define

$$q_t = (1-t)^2 + \tau^2 t^2, \qquad a_t = \frac{\tau^2 t - (1-t)}{q_t}, \qquad b_t = \frac{1-t}{q_t}, \qquad m_t(x) = \mathbb{E}[M \mid X_t^{\mathrm{int}} = x].$$

Conditioning first on M gives $u_t(x) = a_t x + b_t m_t(x)$. The posterior weight of a center m_j is proportional to $\mathbb{P}(M = m_j)\exp\{tx\cdot m_j/q_t - t^2\|m_j\|^2/(2q_t)\}$. Differentiating this finite sum gives

$$\begin{aligned} Du_t(x) &= a_t I_d + \frac{t(1-t)}{q_t^2}\,\mathrm{Cov}(M \mid X_t^{\mathrm{int}} = x), \\ D^2u_t(x)[h,k] &= \frac{t^2(1-t)}{q_t^3}\mathbb{E}[(M - m_t(x))\langle M - m_t(x), h\rangle\langle M - m_t(x), k\rangle \mid X_t^{\mathrm{int}} = x]. \end{aligned} \tag{19}$$

The derivative formula is consistent with Gao et al. (2024a, Proposition 4.1(d)), and the resulting flow regularity is consistent with their Corollary 6.3. The bounds on $u_t(0)$ and Du_t suffice for Theorem 15. The conditional covariance has norm at most R_0^2. The third centered moment in the second line has norm at most $2R_0^3$, by $\|M - m_t(x)\| \leq 2R_0$ and Cauchy–Schwarz for the two scalar factors. Thus

$$\|u_t(0)\| \leq b_t R_0, \qquad \|Du_t(x)\|_{\mathrm{op}} \leq |a_t| + \frac{t(1-t)R_0^2}{q_t^2}, \qquad \|D^2u_t(x)\|_{\mathrm{op}} \leq \frac{2t^2(1-t)R_0^3}{q_t^3}.$$

Since $q_t \geq \tau^2/(1+\tau^2) > 0$, these bounds are finite on the full interval $[0,1]$. At the endpoint, $u_1(x) = x$, $Du_1(x) = I_d$, and $D^2u_1(x) = 0$. The regression velocity satisfies the continuity equation of the interpolation. Its global spatial Lipschitz bound gives a unique flow with these marginals. Theorem 15 therefore applies at $T = 1$ whenever the learned field satisfies Equation 15. Its input is the population excess loss integrated over all of $[0,1]$. There is no stopping error for this application. Numerical error remains separate.

A sharp rate for Gaussian CFM. The dimension dependent logarithmic power is attained on a fixed Gaussian CFM path. The construction first collects a fixed fraction of a small cube into a smaller ball, then moves that ball through the Gaussian tail. Compression and transport with small velocity energy have classical precedents (Michor & Mumford, 2005; Jerrard & Maor, 2019). We track both the Gaussian mass and the spatial derivative bound throughout the construction.

Lemma 2 (Local gathering)**.** *Fix an integer $d \geq 1$ and let $Q = (-1/2, 1/2)^d$. There are positive constants L_d, m_d, c_d, C_d such that, for every real $L \geq L_d$, a field $B_L \in C_c^\infty((0,1/2) \times Q; \mathbb{R}^d)$ and a Borel set $S_L \subset Q$ satisfy*

$$|S_L| = m_d, \qquad \sup|B_L| \leq C_d, \qquad \sup\|D_x B_L\|_{\mathrm{op}} \leq C_d L,$$

$$c_d L^{-d} \leq \int_0^{1/2} \int_Q |B_L(t,x)|^2 \, dx \, dt \leq C_d L^{-d}. \tag{20}$$

The flow of B_L maps S_L into $B(0, C_d/L)$ by time $1/2$. All constants depend only on d and fixed smooth functions.

Proof. Choose a radial $\chi \in C_c^\infty(B(0,1))$ with $0 \leq \chi \leq 1$ and $\chi = 1$ on $B(0,1/2)$. Choose $\psi \in C_c^\infty((0,1))$, nonnegative, with integral one. Set

$$P = \|\psi\|_\infty, \qquad \Theta(t) = \int_0^t \psi(s)\, ds.$$

These functions and constants remain fixed. Define

$$a = \frac{2}{d+2}, \quad A = 10\sqrt{d} + 1, \quad D = 128A, \quad c = \frac{1 - 2^{-a}}{8}, \quad \Lambda = \log(100A), \quad \kappa = P\Lambda.$$

For $L \geq 2$, put

$$n = \left\lceil \frac{d+2}{2} \log_2 L \right\rceil, \qquad \delta = 2^{-n}, \qquad \ell_j = 2^j \delta \quad (1 \leq j \leq n),$$

$$\tau_j = c\ell_j^a + D/L, \qquad w_j = \frac{\ell_j}{L\tau_j}. \tag{21}$$

Then $\ell_n = 1$, $w_j \leq \ell_j/D$, and w_j increases with j.

Partition Q into cubes of side δ. In each cube retain the ball of radius $\rho\delta$ about its center, where

$$\rho = \frac{1}{100(D + c2^a)}.$$

Let S_L be their union. These balls are disjoint and lie inside their cubes, so $|S_L| = \omega_d \rho^d =: m_d > 0$. Here ω_d is the unit ball volume. Since $L\delta^a \leq 1$,

$$w_1 = \frac{2\delta}{cL(2\delta)^a + D} \geq \frac{2\delta}{c2^a + D}.$$

Each retained ball therefore has radius at most $w_1/200$.

We operate on dyadic parent cubes in increasing order of size. Consider a parent of side $\ell = \ell_j$ and center q. Its child centers are $q + (\ell/4)\varepsilon$, where $\varepsilon \in \{-1, 1\}^d$. Assume the retained points in each child lie within radius $w_j/100$ of its center. For the first level this was just proved. For later levels it follows from $w_{j-1} \leq w_j$.

Over a time interval of length τ_j, use the same scalar function for every child:

$$r(t) = \frac{\ell}{4} - \left(\frac{\ell}{4} - 10w_j\right) \Theta(t/\tau_j), \qquad q_\varepsilon(t) = q + r(t)\varepsilon.$$

The translation field in this parent is

$$B_{q,j}^{\mathrm{tr}}(t,x) = \sum_{\varepsilon \in \{-1,1\}^d} \dot{q}_\varepsilon(t) \chi\left(\frac{x - q_\varepsilon(t)}{w_j}\right).$$

Since $D \geq 80$, $10w_j \leq \ell/8$. Distinct centers stay at distance at least $2r(t) \geq 20w_j$. Their radius-w_j supports are disjoint and remain inside the parent. Every retained point follows $x(t) = q_\varepsilon(t) + x(0) - q_\varepsilon(0)$ inside a region where its cutoff is one. Thus each retained cluster is translated rigidly. At the end all retained points lie in $B(q, Aw_j)$.

Next use an interval of length κ/L and set

$$\lambda(t) = \frac{\Lambda L}{\kappa}\psi(Lt/\kappa), \qquad B^{\mathrm{ct}}_{q,j}(t,x) = -\lambda(t)(x-q)\chi\left(\frac{x-q}{2Aw_j}\right).$$

Here $0 \le \lambda \le L$ and $\int \lambda = \Lambda$. The retained points follow the exact contraction

$$x(t) - q = \exp\left(-\int_0^t \lambda(s)\,ds\right)(x(0)-q).$$

They stay where the cutoff is one and finish within radius $Aw_j e^{-\Lambda} = w_j/100$ of q. The support radius $2Aw_j \le \ell/64$ lies inside the parent. The contraction factor is positive.

At each level perform these operations in all parents simultaneously. Parent supports are disjoint. The construction gives the induction needed at the next level. At the last level all retained points lie in $B(0, w_n/100)$.

We record the bounds including the cost of contraction. During translation,

$$|\dot{q}_\varepsilon| \le \frac{\sqrt{d}\,P\ell}{4\tau_j}, \qquad \frac{\ell}{\tau_j} \le c^{-1}.$$

Disjoint supports and $\ell/(\tau_j w_j) = L$ give speed at most C_d and derivative at most $C_d L$. The exact translation action in one parent is

$$\iint |B^{\mathrm{tr}}_{q,j}|^2 = 2^d d\left(\int \chi^2\right)\left(\int_0^1 \psi^2\right) w_j^d \frac{(\ell/4 - 10w_j)^2}{\tau_j} \le C_d \frac{\ell^{d+2}}{L^d \tau_j^{d+1}}. \tag{22}$$

During contraction, the speed is at most $2ALw_j \le C_d$ and the derivative is at most $(1 + \|D\chi\|_\infty)L$. Since $\lambda^2 \le L\lambda$,

$$\iint |B^{\mathrm{ct}}_{q,j}|^2 \le L\Lambda \int_{B(0,2Aw_j)} |x|^2 dx \le C_d L w_j^{d+2}.$$

Relative to the upper scale in equation 22, this last bound has factor at most $C_d/(L\tau_j)$. Since $L\tau_j \ge D$ and there are ℓ_j^{-d} parents, the action of the entire level is at most

$$C_d \frac{\ell_j^2}{L^d \tau_j^{d+1}} \le C_d L^{-d} \ell_j^{2-a(d+1)} = C_d L^{-d} \ell_j^a.$$

The sum of ℓ_j^a is at most $(1 - 2^{-a})^{-1}$. This proves the upper bound in equation 20.

The total active time is

$$\sum_{j=1}^n (\tau_j + \kappa/L) \le \frac{1}{8} + \frac{n(D+\kappa)}{L} \le \frac{1}{4}$$

for sufficiently large L. Start after an idle interval of length $1/16$ and extend the field by zero. The smooth time profiles vanish near every stage endpoint. The resulting field belongs to $C_c^\infty((0,1/2) \times Q)$. For each L, there are finitely many stages and cells. All spatial extensions by zero are smooth. The resulting flow is a smooth diffeomorphism.

At the top level,

$$\tau_n = c + D/L, \qquad w_n = (cL + D)^{-1}.$$

For sufficiently large L, $c \le \tau_n \le 2c$, $w_n \ge (2cL)^{-1}$, and $1/4 - 10w_n \ge 1/8$. The exact identity in equation 22 for this single parent gives action at least $c_d L^{-d}$. Finally, $w_n/100 \le (100cL)^{-1}$. This proves all claims. □

Let $\gamma_d = N(0, I_d)$ and write φ_d for its density. Write ϕ for the one-dimensional standard Gaussian density. For independent $Z_0, Z_1 \sim \gamma_d$, keep the conditional path $(1-t)Z_0 + tZ_1$ on $[0,1]$. Its population field, marginals, and flow are

$$q_t = (1-t)^2 + t^2, \qquad \sigma_t = \sqrt{q_t}, \qquad \alpha_t = \frac{2t-1}{q_t},$$

$$u(t,x) = \alpha_t x, \qquad \mu_t = N(0, q_t I_d), \qquad X_t^\star(z) = \sigma_t z.$$

In particular, the population terminal map is the identity.

Proposition 5 (Gaussian CFM lower bound)**.** *For every fixed integer $d \geq 1$, there are constants $K_d \geq 1$, R_d, and smooth fields v_R, for every real $R \geq R_d$, such that*

$$|v_R(t,x)| + \|D_x v_R(t,x)\|_{\mathrm{op}} \leq K_d(1+|x|).$$

Let T_R be their terminal flow maps and define

$$\eta_R = \int_0^1 \int |v_R(t,x) - u(t,x)|^2 \, \mu_t(dx)\, dt.$$

With positive comparison constants depending only on d,

$$\eta_R \asymp_d \frac{\phi(R)}{R^{2d+2}}, \qquad W_2^2((T_R)_{\#}\gamma_d, \gamma_d) \asymp_d \phi(R) R^{2-d}. \tag{23}$$

Consequently,

$$\frac{W_2((T_R)_{\#}\gamma_d, \gamma_d)}{\sqrt{\eta_R}} \asymp_d [1 + \log(1/\eta_R)]^{(d+4)/4}.$$

Proof. We first construct a field b_R in the coordinates $X_t = \sigma_t Y_t$ and then set

$$v_R(t,x) = \alpha_t x + \sigma_t b_R(t, x/\sigma_t). \tag{24}$$

The learned equation becomes $\dot{Y}_t = b_R(t, Y_t)$. Conditional expectation for CFM gives

$$\eta_R = \mathcal{L}_{\mathrm{CFM}}(v_R) - \mathcal{L}_{\mathrm{CFM}}(u) = \int_0^1 q_t \, \mathbb{E}_{\gamma_d} |b_R(t,Z)|^2 dt. \tag{25}$$

The raw optimum is $\mathcal{L}_{\mathrm{CFM}}(u) = d\pi/2$.

Use Lemma 2 with $L = R$. On the first quarter of the interval define

$$b_R^{\mathrm{g}}(t,y) = \frac{2}{R} B_R(2t, R(y - Re_1)).$$

It is supported in $Q_R = Re_1 + R^{-1}Q$. Its flow carries a set of initial Gaussian mass at least $c_d \phi(R) R^{-d}$ into $B(Re_1, M_d/R^2)$. Indeed, for $\xi \in Q$,

$$\varphi_d(Re_1 + \xi/R) = (2\pi)^{-(d-1)/2} \phi(R) e^{-\xi_1 - |\xi|^2/(2R^2)}.$$

The last exponential has fixed positive upper and lower bounds. Using equation 20, including its lower bound, gives

$$\int_0^{1/4} \mathbb{E}_{\gamma_d} |b_R^{\mathrm{g}}(t,Z)|^2 dt \asymp_d \phi(R) R^{-2d-2}. \tag{26}$$

The factors are $2R^{-2}$ from velocity and time scaling, R^{-d} from spatial volume, and R^{-d} from the local action.

We next transport this ball. Choose a radial $f \in C_c^\infty(B(0,1))$ with $0 \leq f \leq 1$ and $f = 1$ on $B(0,1/2)$. Put $p = 2d+4$ and choose a fixed $k \geq \max\{16p, 4M_d\}$. Choose a smooth nondecreasing step θ equal to zero for arguments at most zero and one for arguments at least one. Set

$$\Gamma(c) = (1 - \theta(c-1))c/p + \theta(c-1), \qquad \zeta(r) = 1 - \theta(2r-1).$$

In an auxiliary variable $s \in [0,1]$, solve

$$c' = Rc\Gamma(c)\zeta(c/R), \quad c(0) = k/R, \qquad S' = c, \quad S(0) = R. \tag{27}$$

For sufficiently large R, the solution is smooth, $k/R \leq c < R$, and $0 \leq c' \leq Rc$. Until $c = 1$,

$$c' = Rc^2/p, \qquad s_1 = p/k - p/R, \qquad S - R = \frac{p}{R}\log(c/(k/R)).$$

The next two stages, from 1 to 2 and from 2 to $R/2$, take at most $p/(2R)$ and exactly $\log(R/4)/R$. Thus c reaches $R/2$ before time $1/2$ and $S(1) - R \geq R/4$. The left edge $\ell = S - c/R$ is nondecreasing and satisfies

$$\ell(0) = R - k/R^2 \geq R/2, \qquad \ell(s_1) = R + \frac{p\log(R/k) - 1}{R} > R. \tag{28}$$

Define the base transport field

$$a_R(s,y) = c(s)e_1 f\left(\frac{R(y - S(s)e_1)}{c(s)}\right).$$

Every point of $B(Re_1, M_d/R^2)$ remains in the region where $f = 1$ and is translated with the center $S(s)e_1$. Its offset from that center is constant and c increases. Choose a fixed smooth increasing $\vartheta : [1/4, 1] \to [0, 1]$ with $\vartheta(1/4) = 0$, $\vartheta(1) = 1$, flat at both endpoints, and define

$$b_R(t,y) = \begin{cases} b_R^{\mathrm{g}}(t,y), & 0 \le t \le 1/4, \\ \vartheta'(t) a_R(\vartheta(t), y), & 1/4 \le t \le 1. \end{cases}$$

The fields agree to every order at the join. This schedule leaves the population CFM path and its clock fixed. Consequently,

$$\int_{1/4}^1 \mathbb{E}|b_R(t,Z)|^2 dt \le \|\vartheta'\|_\infty \int_0^1 \mathbb{E}|a_R(s,Z)|^2 ds.$$

We bound the full transport action. For $s \le s_1$, change variables $y = S(s)e_1 + c(s)\xi/R$ to obtain

$$\mathbb{E}|a_R(s,Z)|^2 = \frac{c^{d+2}}{R^d} \int f(\xi)^2 \varphi_d(Se_1 + c\xi/R)\, d\xi.$$

Here $c \le 1$ and $S/R \le 2$ for sufficiently large R. On the unit support,

$$\varphi_d(Se_1 + c\xi/R) \le C_d \phi(R)(c/(k/R))^{-p}.$$

Using $ds = p\, dc/(Rc^2)$ gives

$$\int_0^{s_1} \mathbb{E}|a_R(s,Z)|^2 ds \le C_d \phi(R) R^{-d-1} (k/R)^p \int_{k/R}^1 c^{d-p} dc \le C_d \phi(R) R^{-2d-2}.$$

For all later times, including the widened support,

$$|a_R(s,y)|^2 \le R^2 \mathbf{1}_{\{y_1 \ge \ell(s_1)\}}, \qquad \phi(\ell(s_1)) \le e(k/R)^p \phi(R).$$

The Gaussian tail bound $\mathbb{P}(G > b) \le \phi(b)/b$ for $b > 0$ then yields

$$\int_{s_1}^1 \mathbb{E}|a_R(s,Z)|^2 ds \le R\phi(\ell(s_1)) \le C_d \phi(R) R^{1-p} = C_d \phi(R) R^{-2d-3}.$$

Together with equation 26 and $1/2 \le q_t \le 1$, this proves both risk bounds in equation 23.

On the gathering support, $|b_R| \le C_d/R$ and $\|Db_R\|_{\mathrm{op}} \le C_d R$. On the transport support, both quantities are at most $C_d R$. Every such support lies in $|y| \ge R/2$. In equation 24, the spatial derivative of $\sigma_t b_R(t, x/\sigma_t)$ is exactly $Db_R(t, x/\sigma_t)$. The bounds $1/\sqrt{2} \le \sigma_t \le 1$ and $|\alpha_t| \le 1$ therefore give the common growth constant K_d. For every fixed R, the smooth field has a finite global spatial Lipschitz constant and a complete flow.

Let $\nu_R = (T_R)_\# \gamma_d$. The gathered initial mass $m_R \ge c_d \phi(R) R^{-d}$ ends with first coordinate at least $R + R/4 - M_d/R^2 \ge 6R/5$. On the other hand, $\gamma_d\{z_1 > 9R/8\} = o(\phi(R)R^{-d}) \le m_R/2$. Every coupling of ν_R and γ_d must therefore pair at least $m_R/2$ mass across a distance of at least $3R/40$. Consequently,

$$W_2^2(\nu_R, \gamma_d) \ge \frac{m_R}{2}(3R/40)^2 \ge c_d \phi(R) R^{2-d}. \tag{29}$$

For the reverse bound, the fixed population field satisfies equation 14 with $L = 1$. Theorem 15 applies with the fixed K_d. The risk bounds already proved give

$$\log(1/\eta_R) = R^2/2 + (2d+2)\log R + O_d(1).$$

Substitution into that theorem yields $W_2^2(\nu_R, \gamma_d) \le C_d \phi(R) R^{2-d}$. This completes equation 23 and the ratio claim. □

Sharpness for each fixed population flow.

Corollary 14 (Sharp error rate for a fixed population flow)**.** *Fix $d \geq 1$, $T \in (0, 1]$, and a population field u satisfying the regularity assumptions of Theorem 15 with constant L. Let Φ_t be its flow and $\mu_t = (\Phi_t)_{\#}\gamma_d$. There is a finite $K_* = K_*(d, L, T)$ with the following property. For any fixed $K \geq K_*$, let $\mathcal{V}_K$ consist of all continuous fields with jointly continuous spatial derivative satisfying*

$$|v(t,x)| + \|D_x v(t,x)\|_{\mathrm{op}} \leq K(1 + |x|) \qquad (t \in [0,T],\ x \in \mathbb{R}^d).$$

Write $\widehat{\mu}_{v,T}$ for the terminal law from initial law γ_d and define

$$\eta_u(v) = T \int_0^T \int |v(t,x) - u_t(x)|^2 \, \mu_t(dx)\, dt.$$

There are positive c, C, ε_0, depending only on d, K, L, T, such that, for every $0 < \varepsilon \leq \varepsilon_0$,

$$\begin{aligned} c\sqrt{\varepsilon}[1 + \log(1/\varepsilon)]^{(d+4)/4} &\leq \sup_{\substack{v \in \mathcal{V}_K \\ \eta_u(v) \leq \varepsilon}} W_2(\widehat{\mu}_{v,T}, \mu_T) \\ &\leq C\sqrt{\varepsilon}[1 + \log(1/\varepsilon)]^{(d+4)/4}. \end{aligned} \tag{30}$$

At zero loss the supremum is zero.

Proof. The upper bound follows from Theorem 15 and monotonicity of its bound for small positive loss. For the lower bound, we implement the collection and transport in Proposition 5 directly in physical coordinates. This requires only the first derivative of u.

Set $U(s,x) = Tu(Ts, x)$, $\Psi_s = \Phi_{Ts}$, $\ell = TL$, and $A = e^{\ell}$. The population assumptions give $|U(s,0)| + \|D_x U(s,x)\|_{\mathrm{op}} \leq \ell$. The first variational equation and linear growth estimates imply

$$\|D\Psi_s\|_{\mathrm{op}},\ \|D\Psi_s^{-1}\|_{\mathrm{op}} \leq A, \qquad 1 + |\Psi_s^{\pm 1}(x)| \leq A(1 + |x|).$$

The absolute Jacobian determinants lie between A^{-d} and A^d. We construct a field V_R on $[0,1]$ and set $v_R(t,x) = T^{-1}V_R(t/T, x)$. Then

$$\eta_u(v_R) = \int_0^1 \mathbb{E}_{Z \sim \gamma_d} |V_R(s, \Psi_s(Z)) - U(s, \Psi_s(Z))|^2 \, ds.$$

All constants below depend only on d, ℓ and fixed smooth functions.

Choose a radial $\chi \in C_c^\infty(B(0,1))$ with $0 \leq \chi \leq 1$ and $\chi = 1$ on $B(0, 1/2)$. For a prescribed center path $q(s)$, let $C(s) = \Psi_s(q(s))$. Given a radius $r(s) > 0$ and a contraction rate $\lambda(s) \geq 0$, define the local residual

$$W(s,x) = g(s)\chi\left(\frac{x - C(s)}{r(s)}\right)\big[C'(s) - U(s,x) - \lambda(s)(x - C(s))\big].$$

Here g is a smooth function in $[0,1]$, equal to one on the supports of q' and λ, and zero near the stage endpoints. The time profiles below have compact support inside each stage, so such a function exists. Since $C' = U(s,C) + D\Psi_s(q)q'$, a trajectory in the inner ball has offset $e = x - C$ satisfying

$$e' = (1 - g)\big[U(s, C + e) - U(s, C)\big] - \lambda e, \qquad |e(s)| \leq |e(s_0)| e^{\ell(s - s_0) - \int_{s_0}^s \lambda}.$$

Moreover,

$$|W| \leq A|q'| + (\ell + \lambda)r, \qquad \|D_x W\|_{\mathrm{op}} \leq C_\chi\left(A|q'|/r + \ell + \lambda\right), \qquad \Psi_s^{-1}(B(C,r)) \subset B(q, Ar).$$

These derivatives use $D_x U$, but no second derivative of the population flow. The function g also makes W and $D_x W$ vanish near each stage endpoint.

We first collect mass from $Q_R = Re_1 + R^{-1}Q$, where $Q = (-1/2, 1/2)^d$. Use the dyadic construction in the proof of Lemma 2, with the following constants. Choose a nonnegative $\psi \in C_c^\infty((0,1))$ of integral one and write $P = \|\psi\|_\infty$ and $\Theta(t) = \int_0^t \psi(s)\, ds$. Put

$$\begin{aligned} &a = \frac{2}{d+2}, \qquad c_0 = \frac{1 - 2^{-a}}{8}, \qquad H_0 = 20\sqrt{d} + 2, \\ &D \geq 512 H_0 A^3, \qquad \Lambda = \log(100 H_0 A^4), \qquad \kappa = P\Lambda. \end{aligned}$$

For each sufficiently large real R, set

$$n = \left\lceil \frac{d+2}{2} \log_2 R \right\rceil, \quad \delta = 2^{-n}, \quad h_j = 2^j \delta, \quad \tau_j = c_0 h_j^a + D/R, \quad w_j = \frac{h_j}{R\tau_j}.$$

The widths w_j are increasing, $w_j \le h_j/D$, and

$$\sum_{j=1}^{n} (\tau_j + \kappa/R) \le \frac{1}{8} + \frac{n(D+\kappa)}{R} \le \frac{1}{4}.$$

Start at time $1/16$. Thus collection ends before time $1/2$. Outside the stated stages, use the population field U.

In each side-δ cube of Q, retain its central ball of radius $\rho\delta$, where $\rho = [1000A^3(D + c_0 2^a)]^{-1}$. After scaling and translating into Q_R, their union S_R satisfies

$$\gamma_d(S_R) \ge c\phi(R)R^{-d}, \qquad \varphi_d(Re_1 + \xi/R) = (2\pi)^{-(d-1)/2}\phi(R)e^{-\xi_1 - |\xi|^2/(2R^2)}.$$

Here ϕ is the one dimensional standard Gaussian density and $\xi \in Q$. Since $w_1 \ge 2\delta/(D + c_0 2^a)$, at the first stage each retained group lies within $w_1/(100A^2R)$ of its physical center.

At level j, consider a parent cube in Q of side h_j and center p. During a stage of length τ_j, move its child centers along

$$q_\epsilon(s) = Re_1 + R^{-1}(p + b(s)\epsilon), \quad b(s) = h_j/4 - (h_j/4 - 10w_j)\Theta((s - s_0)/\tau_j), \quad \epsilon \in \{-1, 1\}^d.$$

Use the local residual above with $\lambda = 0$ and radius $r_{\mathrm{tr}} = w_j/(AR)$. Distinct physical centers stay at least $20r_{\mathrm{tr}}$ apart. Each support has its inverse image inside the corresponding scaled parent cube in Q_R, of side h_j/R. Starting with offsets at most $w_j/(100A^2R)$, the offset estimate keeps them below $w_j/(100AR) = r_{\mathrm{tr}}/100$ throughout this stage. At its end, all retained points are within $H_0 A w_j/R$ of the parent center $\Psi_s(Re_1 + p/R)$.

Next keep $q = Re_1 + p/R$ fixed for a stage of length κ/R and take

$$\lambda(s) = \frac{\Lambda R}{\kappa}\psi(R(s - s_0)/\kappa), \qquad r_{\mathrm{ct}} = 4H_0 A^2 w_j/R.$$

Then $\lambda \le R$ and $\int \lambda = \Lambda$. All retained points stay within $H_0 A^2 w_j/R < r_{\mathrm{ct}}/2$. Their final offsets are at most $H_0 A^2 e^{-\Lambda} w_j/R = w_j/(100A^2R)$. This proves the induction because $w_{j+1} \ge w_j$. The support inverse image has radius at most $4H_0 A^3 w_j/R \le h_j/(128R)$ and stays in that scaled parent cube. Thus simultaneous supports are disjoint at every stage. These bounds also justify the inner ball estimates by first exit. After the last level and any idle time, the retained mass lies in $B(\Psi_{1/2}(Re_1), M/R^2)$ for a fixed M.

We record the cost of these operations, including the cancellation of U. On every collection support, the population density in physical coordinates is bounded above by $C\phi(R)$. For one parent cube, translation has cost at most

$$C\phi(R)R^{-d-2}\left(h_j^2 w_j^d/\tau_j + \ell^2 \tau_j w_j^{d+2}\right),$$

whereas contraction has cost at most

$$C\phi(R)R^{-d-2}\left(Rw_j^{d+2} + \ell^2 R^{-1} w_j^{d+2}\right).$$

For the second estimate, use $\int \lambda^2 \le R\Lambda$. The cancellation term in translation is at most ℓ^2/R^2 times its first term. The first contraction term divided by the first translation term is $1/(R\tau_j) \le 1/D$. There are h_j^{-d} parents at level j, and

$$h_j^{-d} h_j^2 w_j^d/\tau_j = R^{-d} h_j^2/\tau_j^{d+1} \le CR^{-d}h_j^a, \qquad \sum_{j=1}^{n} h_j^a \le (1 - 2^{-a})^{-1}.$$

Hence the total collection loss is at most $C\phi(R)R^{-2d-2}$. Its residual speed is at most C/R and its spatial derivative is at most CR.

We now transport the collected ball. Use the scalar schedule equation 27, with $p = 2d + 4$ and a fixed $k \geq \max\{16p, 4A^2M\}$. Denote its auxiliary time by σ. Its speed $c(\sigma)$ and center $S(\sigma)$ satisfy

$$k/R \leq c < R, \quad 0 \leq c' \leq Rc, \quad S' = c, \quad S(0) = R, \quad S(1) - R \geq R/4.$$

Before the first time σ_1 with $c = 1$,

$$c' = Rc^2/p, \qquad S - R = (p/R)\log(c/(k/R)).$$

The left boundary $S - c/R$ is nondecreasing and at σ_1 equals $R + [p\log(R/k) - 1]/R$. Choose a smooth nondecreasing $\vartheta : [1/2, 1] \to [0, 1]$ that is constant near its endpoints, with endpoint values zero and one. In the local residual take

$$q(s) = S(\vartheta(s))e_1, \qquad r(s) = c(\vartheta(s))/(AR), \qquad \lambda = 0.$$

Choose $g = 1$ on the support of ϑ' and zero near the endpoints. The retained offsets stay below $AM/R^2 < r(s)/2$. The support inverse image lies in $B(Se_1, c/R)$, with exactly the radius used in the Gaussian schedule.

The squared residual is at most $C[(\vartheta')^2c^2 + \ell^2c^2/R^2]$. For $c \leq 1$, the Gaussian mass of its support is at most

$$C\phi(R)(c/R)^d(c/(k/R))^{-p}.$$

Using $(\vartheta')^2 ds \leq \|\vartheta'\|_\infty d\sigma$ and $d\sigma = p\,dc/(Rc^2)$, the first term has integrated cost at most

$$C\phi(R)R^{-d-1}(k/R)^p \int_{k/R}^1 c^{d-p}\,dc \leq C\phi(R)R^{-2d-2}.$$

The cancellation term has pointwise cost at most

$$C\phi(R)R^{-d-2}(k/R)^p c^{d+2-p} \leq C\phi(R)R^{-2d-4}.$$

This estimate also covers intervals where ϑ is constant. After σ_1, the support inverse image is contained in $\{z_1 \geq R + [p\log(R/k) - 1]/R\}$, whose Gaussian mass is at most $C\phi(R)R^{-p-1}$. The squared residual is at most CR^2, so the remaining cost is at most $C\phi(R)R^{1-p} = C\phi(R)R^{-2d-3}$. We have proved, for every sufficiently large real R,

$$\eta_u(v_R) \leq B\phi(R)R^{-2d-2}$$

with a fixed B.

During transport, both the residual speed and its spatial derivative are at most CR. Throughout collection and transport, its support inverse image satisfies $|z| \geq R/2$. Therefore $1 + |x| \geq (1 + |z|)/A \geq R/(2A)$ on these supports. Together with the bounds on U, this gives one constant $K_*(d, L, T) > 2L$ such that every v_R belongs to $\mathcal{V}_{K_*}$. Each field is continuous with jointly continuous spatial derivative. For each fixed R, its residual has bounded support and a finite global spatial Lipschitz constant, so its flow is complete.

Let $\widehat{\mu}_R$ be its terminal law and put $\nu_R = (\Psi_1^{-1})_\#\widehat{\mu}_R$. The retained mass ends within AM/R^2 of $\Psi_1(S(1)e_1)$. Its inverse image is therefore within A^2M/R^2 of $S(1)e_1$. For large R,

$$\nu_R\{z_1 \geq 6R/5\} \geq m_R := \gamma_d(S_R) \geq c\phi(R)R^{-d}, \qquad \gamma_d\{z_1 > 9R/8\} \leq m_R/2.$$

Every coupling of ν_R and γ_d moves at least $m_R/2$ mass a distance at least $3R/40$. Applying the A Lipschitz map Ψ_1^{-1} to an arbitrary terminal coupling gives

$$W_2^2(\widehat{\mu}_R, \mu_T) \geq A^{-2}W_2^2(\nu_R, \gamma_d) \geq c\phi(R)R^{2-d}.$$

Finally, $B\phi(R)R^{-2d-2}$ is continuous, strictly decreasing, and tends to zero. For every sufficiently small $\varepsilon > 0$, choose its unique solution $B\phi(R)R^{-2d-2} = \varepsilon$. Then $\eta_u(v_R) \leq \varepsilon$ and the squared terminal error is at least $c\varepsilon R^{d+4}$, while

$$\log(1/\varepsilon) = R^2/2 + (2d + 2)\log R + O(1).$$

This proves the lower bound for every sufficiently small budget and every fixed $K \geq K_*$. The population field belongs to $\mathcal{V}_K$. The zero loss statement follows from Theorem 15. □

The Gaussian mixture calculation above verifies the population assumptions on $[0, 1]$. The corollary therefore applies to each of those fixed CFM paths, with a sufficiently large fixed K.

Weighted population loss. The logarithmic factor in the unweighted loss bound comes from errors at large distances from the origin. We now quantify how a polynomial state weight changes this rate.

Corollary 15 (Error rates for weighted population loss)**.** *Fix $d \geq 1$, $T \in (0, 1]$, and a continuous population field u with jointly continuous spatial derivative such that*

$$|u_t(0)| + \|D_x u_t(x)\|_{\mathrm{op}} \leq L \qquad (t \in [0, T],\ x \in \mathbb{R}^d).$$

Let Φ_t be its flow and $\mu_t = (\Phi_t)_{\#}\gamma_d$. For a fixed $a \geq 0$, define

$$\eta_{u,a}(v) := T \int_0^T \int (1 + |x|)^a |v(t, x) - u_t(x)|^2 \, \mu_t(dx)\, dt. \tag{31}$$

There is a finite $K_* = K_*(d, L, T)$ such that the following holds for each fixed $K \geq K_*$. Let $\mathcal{V}_K$ be the class of continuous fields with jointly continuous spatial derivative satisfying

$$|v(t, x)| + \|D_x v(t, x)\|_{\mathrm{op}} \leq K(1 + |x|).$$

Write $\widehat{\mu}_{v,T}$ for the terminal law from γ_d. There are positive c, C, ε_0, depending only on d, K, L, T, a, such that, for every $0 < \varepsilon \leq \varepsilon_0$,

$$\begin{aligned} c\sqrt{\varepsilon}[1 + \log(1/\varepsilon)]^{(d+4-a)_+/4} \leq \sup_{\substack{v \in \mathcal{V}_K \\ \eta_{u,a}(v) \leq \varepsilon}} W_2(\widehat{\mu}_{v,T}, \mu_T) \\ \leq C\sqrt{\varepsilon}[1 + \log(1/\varepsilon)]^{(d+4-a)_+/4}. \end{aligned} \tag{32}$$

Here $r_+ = \max\{r, 0\}$, and the supremum is zero at zero loss. For $a \geq d + 4$, the bound $W_2^2(\widehat{\mu}_{v,T}, \mu_T) \leq C\eta_{u,a}(v)$ holds at every loss level. For $a < d + 4$, no such bound holds uniformly over this class.

Proof. Fix $v \in \mathcal{V}_K$ and let $\widehat{\Phi}_t$ be its flow. We retain the local weights in the proof of Theorem 15. Set $A = e^{LT}$ and, for $s \in [0, 1]$, write

$$Y_s(z) = \Phi_{Ts}^{-1}(\widehat{\Phi}_{Ts}(z)), \qquad q(s, z) = T[v(Ts, \Phi_{Ts}(z)) - u_{Ts}(\Phi_{Ts}(z))].$$

Put $D(z) = |Y_1(z) - z|$, $\rho(z) = 1 + |z|$, and $r(z) = \min\{D(z), \rho(z)^{-1}\}$. For centers with $|z| \leq R$ and $D(z) > 0$, the disjoint ball selection in that proof gives balls $B_j = B(z_j, r_j)$ and assigned sets E_j covering those centers. With $\rho_j = 1 + |z_j|$, the local energy and displacement estimates give

$$\int_{E_j} D^2 \, d\gamma_d \leq C\rho_j^{d+4} \int_0^1 \int_{B_j} |q(s, y)|^2 \, \gamma_d(dy)\, ds.$$

The constant is independent of R and j. For $y \in B_j$, the bound $r_j \leq 1$ implies $\rho_j \leq 2(1 + |y|)$. Thus, for $0 \leq a \leq d + 4$,

$$\mathbb{E}[D(Z)^2 \mathbf{1}_{\{|Z| \leq R\}}] \leq C_a(1 + R)^{d+4-a} \int_0^1 \int (1 + |z|)^a |q(s, z)|^2 \, \gamma_d(dz)\, ds.$$

The inverse population flow satisfies $1 + |\Phi_t^{-1}(x)| \leq A(1 + |x|)$. Changing variables gives

$$\begin{aligned} \int_0^1 \int (1 + |z|)^a |q(s, z)|^2 \, \gamma_d(dz)\, ds = T \int_0^T \int (1 + |\Phi_t^{-1}(x)|)^a |v(t, x) - u_t(x)|^2 \, \mu_t(dx)\, dt \\ \leq A^a \eta_{u,a}(v). \end{aligned}$$

The same Gaussian tail estimate and terminal coupling as in Theorem 15 therefore give

$$W_2^2(\widehat{\mu}_{v,T}, \mu_T) \leq C_a\{(1 + R)^{d+4-a}\eta_{u,a}(v) + e^{-R^2/4}\}.$$

For $a < d + 4$ and $0 < \eta_{u,a}(v) \leq 1$, take $R = 2\sqrt{\log(1/\eta_{u,a}(v))}$. This proves the upper rate. Its right side is increasing at sufficiently small loss, which gives the stated budget bound. For $a = d+4$, let $R \to \infty$ in the inner ball estimate. Monotone convergence gives $W_2^2 \leq C\eta_{u,d+4}(v)$. For $a > d + 4$, use $\eta_{u,d+4}(v) \leq \eta_{u,a}(v)$. Zero weighted loss implies zero unweighted loss.

For the lower bound, use the fields v_R constructed in the proof of Corollary 14. Their growth constant is fixed, and, for every sufficiently large real R,

$$\eta_{u,0}(v_R) \le B\phi(R)R^{-2d-2}, \qquad W_2^2(\widehat{\mu}_{v_R,T}, \mu_T) \ge c\phi(R)R^{2-d}.$$

Here ϕ is the one dimensional standard Gaussian density. The collection supports have inverse images in the cube centered at Re_1 with side length R^{-1}. During transport, the inverse support lies in $B(S(\sigma)e_1, c(\sigma)/R)$, with $R \le S(\sigma) \le 2R$ and $c(\sigma) < R$. These supports include the time gates and the intervals when the center is stationary. Hence $1 + |x| \le 4AR$ wherever $v_R - u$ is nonzero, for sufficiently large R. Consequently,

$$\eta_{u,a}(v_R) \le B_a\phi(R)R^{a-2d-2}.$$

The displayed upper bound is continuous and strictly decreasing for sufficiently large R. For every sufficiently small $\varepsilon > 0$, choose R so that this bound equals ε. Then

$$W_2^2(\widehat{\mu}_{v_R,T}, \mu_T) \ge c_a\varepsilon R^{d+4-a}, \qquad \log(1/\varepsilon) = R^2/2 + (2d + 2 - a)\log R + O_a(1).$$

This proves the lower rate for $0 \le a \le d + 4$ and rules out a uniform bound $W_2^2 \le C\eta_{u,a}$ when $a < d + 4$.

For $a > d + 4$, choose $\tau \in (0, T)$ with $e^{L\tau} \le 3/2$ and a nonnegative $g \in C_c^\infty((T - \tau, T))$ with integral one. For a unit vector e and $0 < \delta \le 1$, set $v_\delta(t, x) = u_t(x) + \delta g(t)e$. These fields have a common growth bound depending only on L, T. Let $d_z(t)$ be the learned trajectory minus the population trajectory from the same initial state. Gronwall's inequality gives

$$|d_z(t)| \le \delta \int_{T-\tau}^{t} e^{L(t-s)}g(s)\,ds \quad (t \ge T - \tau).$$

Integrating the projection onto e then gives

$$e^\top d_z(T) \ge \delta - L\int_{T-\tau}^{T} |d_z(t)|\,dt \ge \delta \int_{T-\tau}^{T} g(s)\{2 - e^{L(T-s)}\}\,ds \ge \delta/2.$$

The terminal mean difference implies $W_2 \ge \delta/2$. Meanwhile,

$$\eta_{u,a}(v_\delta) \le C_a'\delta^2, \qquad C_a' = T\|g\|_{L^2(0,T)}^2 e^{aLT}\mathbb{E}_{Z\sim\gamma_d}(1 + |Z|)^a < \infty.$$

Taking $\delta = \sqrt{\varepsilon/C_a'}$ proves the remaining lower bound for small budgets. Enlarge K_* once to include both families. This choice does not depend on a or ε. □

Connection to the CFM objective. Let $(X_t^{\mathrm{int}}, U_t)$ be a conditional path and its velocity label, with $\mathrm{Law}(X_t^{\mathrm{int}}) = \mu_t$ and $u_t(x) = \mathbb{E}[U_t \mid X_t^{\mathrm{int}} = x]$. Define the unnormalized time integral

$$\mathcal{L}_a(v) = \int_0^T \mathbb{E}\big[(1 + |X_t^{\mathrm{int}}|)^a |v(t, X_t^{\mathrm{int}}) - U_t|^2\big]\,dt.$$

Assume $\int_0^T \mathbb{E}[(1 + |X_t^{\mathrm{int}}|)^a |U_t|^2]dt < \infty$. Expanding the square and conditioning on X_t^{int} makes the cross term zero, as in the standard CFM loss identity (Lipman et al., 2023). Hence

$$T[\mathcal{L}_a(v) - \mathcal{L}_a(u)] = \eta_{u,a}(v). \tag{33}$$

The positive state weight leaves the unrestricted population minimizer u unchanged and is available from each training state. At $a = d + 4$, the corollary gives a square root bound in weighted CFM excess loss, with constants depending on dimension and the growth conditions.

H.3 Euler Step Counts and Time Allocation

Fix v_θ and $T \in (0, 1]$. Assume the comparison and numerical tube conditions on $[0, T]$, the time and space differentiability in Corollary 9, and finite second moments of the terminal states. For an integer $N_0 \ge 1$ and the uniform grid $h = T/N$, define

$$m_k := \sup_{s\in[t_k,t_{k+1}]} \|\mathscr{D}_t v_\theta(s, \Phi_{s,t_k}^\theta(\widetilde{X}_k^\theta))\|,$$

$$F_T := \|\mathsf{A}_{\theta,T,\mathrm{tube}}\mathcal{E}_{\theta,T}\|_2 + W_2(\mu_T, \mu),$$

$$C_{\mathrm{E},T} := \sup_{N\ge N_0} \left\| \mathsf{A}_{\theta,N,T,\mathrm{tube}} \sum_{k=0}^{N-1} h\, m_k \right\|_2 < \infty.$$

$$W_2(\widetilde{\mu}_{\theta,N,T},\mu) \le F_T + \frac{TC_{\mathrm{E},T}}{2N}, \qquad N \ge N_0. \tag{34}$$

For $\varepsilon > F_T$, it suffices to take $N \ge \max\{N_0, \lceil TC_{\mathrm{E},T}/[2(\varepsilon - F_T)]\rceil\}$. Euler uses one velocity evaluation per step. The supremum is over integers N. The first amplification factor compares population and learned paths up to T, and the second uses the numerical tube for the chosen grid. The Euler defect bound gives $\sum_k \tau_k \le (T/2N)\sum_k h\, m_k$. The triangle inequality yields Equation 34. The term F_T bounds generation and stopping error.

One sufficient condition for the uniform Euler constant is

$$\|v_\theta(t,x)\| \le C(1+\|x\|), \quad \|\nabla_x v_\theta(t,x)\|_{\mathrm{op}} \le C(1+\|x\|), \quad \|\mathscr{D}_t v_\theta(t,x)\| \le C(1+\|x\|)^m$$

for a field continuously differentiable in time and space, some $m \ge 0$, and all $(t,x) \in [0,T]\times\mathbb{R}^d$. Continuous and discrete Grönwall bounds place the numerical states and exact subflows in a ball of radius $C'(1+\|Z\|)$, uniformly in N. The amplification grows at most exponentially in $\|Z\|$, and the defect sum at most polynomially. Their product has a finite Gaussian L^2 norm.

For independent linear interpolation, coupling X_T^{int} with X_1 gives $W_2(\mu_T,\mu) \le (1-T)(\mathbb{E}\|Z - X_1\|^2)^{1/2}$. If $\mathcal{J}_{\theta,T,\mathrm{tube}} \le R_T$ almost surely, then

$$W_2(\widetilde{\mu}_{\theta,N,T},\mu) \le e^{R_T}\sqrt{T\Delta_{\mathrm{CFM},T}} + (1-T)(\mathbb{E}\|Z-X_1\|^2)^{1/2} + TC_{\mathrm{E},T}/(2N).$$

For positive error budgets $\varepsilon_{\mathrm{learn}}+\varepsilon_{\mathrm{stop}}+\varepsilon_{\mathrm{num}} \le \varepsilon$, it suffices to require $\Delta_{\mathrm{CFM},T} \le \varepsilon_{\mathrm{learn}}^2/(Te^{2R_T})$, $(1-T)(\mathbb{E}\|Z-X_1\|^2)^{1/2} \le \varepsilon_{\mathrm{stop}}$, and $N \ge \max\{N_0, \lceil TC_{\mathrm{E},T}/(2\varepsilon_{\mathrm{num}})\rceil\}$.

Allocating steps across time intervals. Assume the same numerical tube and Euler regularity conditions for every grid considered below. On a nonuniform grid, let ℓ_i^θ be the local numerical tube growth in Assumption 3, and define m_k as above for each step. Theorem 9, the Euler defect bound, and Minkowski's inequality give

$$W_2(\widetilde{\mu}_{\theta,h,T},\widehat{\mu}_{\theta,T}) \le \frac{1}{2}\sum_{k=0}^{N-1} h_k^2 \left\| \exp\left(\sum_{i=k+1}^{N-1}\ell_i^\theta\right) m_k \right\|_2 .$$

Partition $[0,T]$ into B intervals of lengths r_j. Consider grids that use equal steps within each interval, with at least one step per interval. Suppose $q_j > 0$ bounds each norm on the right for every step in interval j on every such grid. If interval j receives $n_j \ge 1$ steps and $\sum_j n_j = N$, then

$$W_2(\widetilde{\mu}_{\theta,h,T},\widehat{\mu}_{\theta,T}) \le \frac{1}{2}\sum_j q_j r_j^2/n_j.$$

For the continuous relaxation $n_j > 0$, Lagrange multipliers give

$$n_j = \frac{N r_j\sqrt{q_j}}{\sum_i r_i\sqrt{q_i}}, \qquad \min_{\substack{n_j>0\\ \sum_j n_j = N}} \frac{1}{2}\sum_j \frac{q_j r_j^2}{n_j} = \frac{(\sum_j r_j\sqrt{q_j})^2}{2N}.$$

By Cauchy–Schwarz, this is at most $T\sum_j r_j q_j/(2N)$, the relaxed bound for allocation proportional to interval length. For integers and $N \ge B$, start with one step per interval and repeatedly add a step where $q_j r_j^2/[n_j(n_j+1)]$ is largest. The decrease becomes smaller after each addition, so this rule minimizes the displayed integer bound. The bounds q_j must remain valid after changing the grid.

Acceleration profiles also motivate fixed budget Euler grids (Gupta et al., 2026). Here the local defect is weighted by later tube growth. Replacing the bounds q_j with an empirical profile requires controlling its approximation error on the changed grid.

Sampling resources. Table 1 in the main text compares information available for error bounds. Table 2 below records guarantees with different numerical and statistical resources. Increasing N in the Euler bound reduces only its numerical term. The quantity F_T contains generation and stopping error.

Our Euler bound fixes the learned field and terminal time T and assumes the uniform moment condition defining $C_{\mathrm{E},T}$. The geometric grid result assumes weak log concavity, bounds on the

Table 2: Resource guarantees under their respective assumptions. Each row records its resource, metric, and stopping rule. Comparing complexities requires fixing these choices and the regularity assumptions. Here L denotes the score Lipschitz constant in the final row.

Resource varied	Metric	Guarantee	Setting
N Euler evaluations	W_2	$F_T + TC_{\mathrm{E},T}/(2N)$	Fixed field and T, finite tube and defect moment (this work)
N geometric grid steps	W_2	$C\sqrt{d}\log^2(N)/N$ plus integrated drift error	Regularity and an N dependent stopping time (Stéphanovitch, 2026, Corollary 11)
n training samples	W_2	$\widetilde{O}(n^{-1/[3(d+5)]})$	Specified estimator, step size, and stopping time (Zhou & Liu, 2025, Theorem 1.7)
Predictor and corrector iterations	TV	$\widetilde{O}(L^2\sqrt{d}/\varepsilon)$ iterations for error ε	Probability flow ODE with an underdamped corrector (Chen et al., 2023, Theorem 3)

target center and integrated drift error, the stated schedule conditions, and the learned drift regularity in Definition 3 and Corollary 11 of Stéphanovitch (2026). Its stopping time is $\tau_N = 1 - (\log^2(N)/N)^{1/q}$, where q is the minimum of one and the noise schedule exponent. The additional error is $C\int_0^{\tau_N} \varepsilon_{\mathrm{drift}}(t)\,dt$.

The statistical W_2 rate holds with probability at least $1 - 1/n$ under support $[0,1]^d$, the potential and Hessian conditions in Assumptions 1.1–1.2, and the specified network, step size, and stopping time (Zhou & Liu, 2025, Theorem 1.7). The TV guarantee uses a probability flow ODE predictor and an underdamped Langevin corrector (Chen et al., 2023, Theorem 3). It assumes a finite target second moment, Lipschitz true and estimated scores, and prescribed score error, step sizes, and stopping time. Its iteration count includes both parts of the sampler.

H.4 Numerical Experiments

All $\pm$ values below are sample standard deviations across the stated models or paired runs. Relative changes are calculated within each pair before averaging. The two population experiments use deterministic integration for their primary metrics.

H.4.1 Complete Error Bounds for an Explicit CFM Perturbation

Table 3: All columns use the same population flow, perturbation, and Gaussian quadrature nodes. The segment and tube columns are the first two bounds in Equation 4. Spatial replaces the segment supremum by a supremum over all states. Classical also replaces the time resolved residual by its population loss.

δ	W_2 (segment)	Tube	Spatial	Classical	Classical / W_2
0.01	0.05106	0.3403	9.325	26.28	515
0.03	0.1193	1.740	28.36	80.01	671
0.10	0.2998	10.72	99.22	281.0	937

Let $Z \sim \mathcal{N}(0,1)$ and $X_1 \sim \frac{1}{2}\mathcal{N}(-2, 0.5^2) + \frac{1}{2}\mathcal{N}(2, 0.5^2)$ be independent. Use $X_t^{\mathrm{int}} = (1-t)Z + tX_1$ on $[0,1]$. Its marginal is the equal mixture with means $\pm 2t$ and variance $q_t = (1-t)^2 + t^2/4$. Conditional expectation gives

$$u_t(x) = \frac{t/4 - (1-t)}{q_t}x + \frac{2(1-t)}{q_t}\tanh\left(\frac{2tx}{q_t}\right).$$

We prescribe $v_\delta(t,x) = u_t(x) + \delta(1 + \tanh x)$ for $\delta \in \{0.01, 0.03, 0.10\}$. The model, amplitudes, integration interval, and evaluation rules were fixed before computing the results. We report all three amplitudes. The residual on the population trajectory is $\delta[1 + \tanh X_t^\star(z)]$, which is positive for

every finite state. The learned derivative is

$$\partial_x v_\delta(t,x) = \frac{t/4-(1-t)}{q_t} + \frac{4t(1-t)}{q_t^2}\operatorname{sech}^2\left(\frac{2tx}{q_t}\right) + \delta\operatorname{sech}^2 x.$$

It is even and nonincreasing in $|x|$. Its absolute maximum on a segment is therefore attained at an endpoint or at zero if zero lies in the segment. We evaluate the tube supremum by this formula. No spatial grid is used to approximate it. The exact global coefficient is

$$L_\delta(t) = \max\left\{\left|\frac{t/4-(1-t)}{q_t}\right|, \left|\frac{t/4-(1-t)}{q_t} + \frac{4t(1-t)}{q_t^2} + \delta\right|\right\}.$$

Both ODE flows preserve order in one dimension. Their common Gaussian coupling is therefore optimal, so the endpoint RMS equals $W_2(\widehat{\mu}_\delta, \mu)$. We call the first two bounds in Equation 4 the segment and tube bounds. The third is the product bound $\|\mathsf{A}_{\theta,\mathrm{tube}}\mathcal{E}_\theta\|_2$. In addition to the classical bound $e^{\int_0^1 L_\delta(t)dt}\sqrt{\eta}$, we evaluate the stronger global quantities

$$\left\|\int_0^1 e^{\int_s^1 L_\delta(r)dr}\delta[1+\tanh X_s^\star(Z)]\,ds\right\|_2, \qquad \left[\eta\int_0^1 e^{2\int_s^1 L_\delta(r)dr}\,ds\right]^{1/2}.$$

The first retains the source on each path and its time of occurrence. The second applies Cauchy–Schwarz in time and depends only on η. Their values are $(9.325, 28.36, 99.22)$ and $(13.84, 42.12, 147.6)$ at the three amplitudes. Table 3 reports the other quantities. The tube bound is below both global quantities at every amplitude.

Integration and checks. We integrate both states together with the source integral, the tube Jacobian integral, and the full bound. The latter is obtained by solving the scalar linear equation whose forcing is the source residual and whose coefficient is the tube supremum. The same source nodes evaluate every quantity for all amplitudes. Composite Gauss–Legendre rules use symmetric panel boundaries 0, ±0.0025, ±0.01, ±0.05, ±0.1, ±0.25, ±0.5, ±1, ±2, ±3, ±4, ±6, ±8, and ±12. The central panels are fixed from the analytic population derivative $T_\star'(0) = 0.5e^8$. We use 24, 48, 96, and 192 nodes per panel. The last level has 4992 source nodes and is not renormalized after truncation. The maximum relative change in the reported metrics from order 96 to 192 is below 4.83×10^{-4}.

The primary calculation uses the DOP853 Runge–Kutta method (Hairer et al., 1993) with relative tolerance 2×10^{-11} and absolute tolerance 2×10^{-13}. The tube supremum is evaluated analytically at the computed endpoints and at zero when the segment contains it. The quadrature comparison covers all seven metrics. Changing the ODE tolerances to $(2\times10^{-9}, 2\times10^{-11})$ changes the seven aggregate metrics by at most 6.27×10^{-7} in absolute value. All differences satisfy the prescribed threshold $10^{-8} + 5\times10^{-4}|\mathrm{reference}|$.

We also compare fixed step RK4 at 4096 and 8192 steps with DOP853 at tolerances $(2\times10^{-12}, 2\times10^{-14})$ on 15 fixed source nodes. At 8192 steps, the largest coordinatewise difference divided by $10^{-7} + 5\times10^{-4}|\mathrm{reference}|$ is 4.03×10^{-5}, below the acceptance threshold of one. The checks include the two trajectories and all five accumulated quantities. The trajectory calculations use floating point arithmetic, and the analytic estimate below controls the omitted Gaussian tail.

An independent audit uses the same 2048 Gaussian initial states for all amplitudes. In increasing amplitude order, its empirical endpoint RMS values are $(0.07096, 0.1481, 0.3128)$ and its empirical full tube values are $(0.4691, 2.243, 11.43)$. All pathwise and aggregate bound orderings hold at the stated numerical tolerance. These are averages over sampled initial states. Table 3 uses deterministic population quadrature.

Since $\int_0^1 L_\delta(t)dt \le \log(2.5) + 8 + \delta$ and the residual is at most 2δ, every pointwise error or bound in this comparison is at most $2\delta e^{\log(2.5)+8+\delta}$. The omitted Gaussian tail contributes at most

$$2\delta e^{\log(2.5)+8+\delta}\sqrt{\mathbb{P}(|Z|>12)} < 10^{-13}$$

on the L^2 scale for all tested amplitudes. All three fields satisfy a common linear growth bound, for example $K = 41$. The three prescribed perturbations compare complete error bounds at fixed amplitudes.

H.4.2 Path Diagnostics and Numerical Refinement

Exact population flow. Use independent $X_0 \sim \mathcal{N}(0,1)$ and $X_1 \sim \frac{1}{2}\mathcal{N}(-2, 0.1^2) + \frac{1}{2}\mathcal{N}(2, 0.1^2)$ with $X_t^{\mathrm{int}} = (1-t)X_0 + tX_1$. Write $m = 2$, $s = 0.1$, $a_t = (1-t)^2 + s^2t^2$, and $b_t = ((1+s^2)t - 1)/a_t$. The population velocity and derivative are

$$u_t(x) = b_t x + \frac{m(1-t)}{a_t} \tanh\left(\frac{mtx}{a_t}\right), \qquad \partial_x u_t(x) = b_t + \frac{m^2 t(1-t)}{a_t^2} \operatorname{sech}^2\left(\frac{mtx}{a_t}\right).$$

The spatial supremum is $\max\{|b_t|, |b_t + m^2 t(1-t)/a_t^2|\}$. With 20,000 source samples and 4,000 RK4 steps to $T = 0.999$, its time integral is 198.0311. The path median and Q99 are 2.0689 and 3.9686.

Trained models and sensitivity. We measure growth and terminal sensitivity on ten two dimensional CFM models. The source is $\mathcal{N}(0, I_2)$, and the target is an equally weighted mixture of eight Gaussians whose means lie on a circle of radius 3 and whose standard deviations are 0.18. Independent source and target samples define $X_t^{\mathrm{int}} = (1-t)Z + tX_1$ and $U_t = X_1 - Z$. Each tanh velocity network has two hidden layers of width 48. We train for 10,000 Adam steps with batch size 384, learning rate 0.002, and times uniform on $[0.001, 0.999]$, clipping gradients at Euclidean norm 20. In each of five paired runs, the baseline and the model with a 0.03 squared Frobenius Jacobian penalty at interpolation states share initialization and CFM minibatches.

For each model, we integrate 1,500 source paths and the variational equation with 192 RK4 steps. Write $\ell_\theta = \log \|D_Z T_\theta\|_{\mathrm{op}}$, and let $\mathcal{G}_\theta$ be the one-sided growth integral in Definition 3. The Spearman correlation with ℓ_θ is 0.9971 ± 0.0021 for $\mathcal{G}_\theta$ and 0.8553 ± 0.0503 for $\mathcal{J}_\theta$. Within each model, we take the median and 99th percentile of $\mathcal{G}_\theta(Z_i) - \log \|D_{Z_i} T_\theta(Z_i)\|_{\mathrm{op}}$ across the 1,500 paths, then average each quantile over the ten models. The results are 0.068 and 0.265.

The 99th percentile of $\mathcal{J}_\theta$ is 4.1781 ± 0.3481 across models. A spatial scan of $[-24, 24]^2$ with spacing 0.3 at 81 times gives 11.7360 ± 1.7394; this is a reference for the scanned states. The product of the three layer spectral norms gives a global upper bound of 99.3225 ± 18.5684, since the tanh derivative is at most one.

Selecting paths for refinement. For the five baseline models, use the same 1,500 source samples. Along Euler with 16 steps, let d_k be the norm of the difference between a Heun step and an Euler step from the same state. With $h = 1/16$, compare scores

$$q_{\mathrm{local}} = \sum_{k=0}^{15} d_k, \qquad q_{\mathrm{growth}} = \sum_{k=0}^{15} d_k \exp\left(h \sum_{j=k+1}^{15} \|D_x v_\theta(t_j, X_j)\|_{\mathrm{op}}\right).$$

Each rule replaces the endpoints of its top 20 percent of paths with endpoints from Euler with 32 steps. For either score, capture is the fraction of the total squared 16 step Euler endpoint error carried by the selected 20 percent of paths. We compute the fraction within each of the five baseline models and report its mean and sample standard deviation: $82.3\% \pm 6.2\%$ for growth weighting and $28.5\% \pm 4.9\%$ for local defects. For each model, RMSE and Q99 are computed from Euclidean endpoint distances to RK4 with 192 steps, using the same initial states. Table 4 reports arithmetic means of these model values. These path scores use measured growth along the pilot trajectories. The time allocation bound above uses bounds valid on numerical comparison tubes for every admissible grid. Relative to Euler with 20 steps, growth selection reduces RMSE by $21.48\% \pm 4.31\%$ and Q99 error by $35.85\% \pm 4.87\%$. Relative to local defect selection, the reductions are $28.15\% \pm 3.41\%$ and $45.17\% \pm 3.71\%$. Both errors decrease on all five models. The mean RMSE between the RK4 references with 192 and 384 steps is 1.04×10^{-8}. The retained paths average $0.8 \times 16 + 0.2 \times 32 = 19.2$ steps. A selective implementation uses $2 \times 16 + 0.2 \times 32 = 38.4$ velocity evaluations, or $17 + 0.2 \times 32 = 23.4$ if each Euler endpoint evaluation is reused at the next step. Growth selection also uses 16 Jacobian norm evaluations. These are algorithmic counts. The comparison script computes all candidate endpoints and reference trajectories offline.

Norm estimation in 64 dimensions. The source is standard Gaussian and the target has eight components with means $\{\pm 3.5 e_j : j = 1, \ldots, 4\}$ and covariance $0.3^2 I_{64}$. Three independently initialized tanh networks use two hidden layers of width 96 and 6,000 Adam steps with batch size

Table 4: Mean endpoint errors over five baseline models. Retained steps exclude path selection cost. Random selection is averaged over 100 independent draws within each model.

Rule	Retained steps	RMSE	Q99 error
Uniform Euler with 16 steps	16.0	0.10389	0.47093
Uniform Euler with 20 steps	20.0	0.08474	0.37827
Uniform Euler with 24 steps	24.0	0.07150	0.33072
Random 20% selection	19.2	0.09617	0.43955
Local defect selection	19.2	0.09239	0.43534
Growth selection	19.2	0.06610	0.23834

384 and learning rate 0.002. Record 16 states at times $0, 1/16, \ldots, 15/16$ along each of 32 paths computed with RK4 and 96 steps. Dense Jacobians provide reference norms. Eight Rademacher probes estimate $\|J\|_F^2 = \mathbb{E}\|J\xi\|^2$ (Hutchinson, 1989), with 20 repetitions. After averaging the 512 states, aggregate relative error is $0.0993\% \pm 0.0076\%$. Median error across individual states is $1.7926\% \pm 0.0394\%$. Eight power iterations using Jacobian and transposed Jacobian products estimate the operator norm with median relative error $0.0980\% \pm 0.0215\%$.

I Two Small Networks with Different Sampling Errors

The same population loss and global Jacobian bounds can lead to different sampling errors, even for two small networks. We give an explicit example with a fixed CFM path and a common growth bound. Let $\gamma = \mathcal{N}(0,1)$ and let ϕ denote its density. For independent $Z_0, Z_1 \sim \gamma$, set

$$X_t^{\mathrm{int}} = \cos(\pi t/2)Z_0 + \sin(\pi t/2)Z_1, \qquad U_t = \frac{d}{dt}X_t^{\mathrm{int}}. \tag{35}$$

Every marginal is γ. The variables X_t^{int} and U_t are jointly Gaussian with zero covariance, so the population velocity is $u_t(x) = \mathbb{E}[U_t \mid X_t^{\mathrm{int}} = x] = 0$. This uses the conditional velocity formulation of CFM (Lipman et al., 2023). For time uniform on $[0,1]$, expanding the square and conditioning on X_t^{int} gives

$$\mathcal{L}_{\mathrm{CFM}}(v) = \frac{\pi^2}{4} + \int_0^1 \mathbb{E}_{Z\sim\gamma}|v(t,Z)|^2\,dt. \tag{36}$$

Proposition 6 (Equal global bounds and different terminal errors)**.** *For $R \geq 2$, define*

$$b_R(x) = \frac{1}{2}\left[\tanh\left(\frac{R}{2}x - \frac{R^2}{2}\right) - \tanh\left(\frac{R}{2}x - \frac{R(R+2)}{2}\right)\right], \qquad v_R^{\pm}(t,x) = \pm b_R(x). \tag{37}$$

Each field is a network with one hidden layer of two tanh units and an affine output. The time input has zero weight. Both fields satisfy

$$|v_R^{\pm}(t,x)| + |\partial_x v_R^{\pm}(t,x)| \leq 1 + |x|. \tag{38}$$

They have the same population CFM loss, the same layer operator norms, and identical $|\partial_x^j v_R^{\pm}(t,x)|$ at every state for every integer $j \geq 0$. Their global absolute and one-sided Jacobian bounds also agree:

$$\sup_x |\partial_x v_R^+| = \sup_x |\partial_x v_R^-| = \sup_x \partial_x v_R^+ = \sup_x \partial_x v_R^-. \tag{39}$$

Let $T_R^{\pm}$ be their time one flow maps, let $\mu_R^{\pm} = (T_R^{\pm})_{\#}\gamma$, and write $\eta_R = \mathbb{E}b_R(Z)^2$ for their common excess loss. As $R \to \infty$,

$$\eta_R \sim \frac{\phi(R)}{R}, \qquad W_2^2(\mu_R^+, \gamma) \sim \frac{\phi(R)}{3}, \qquad W_2^2(\mu_R^-, \gamma) \sim \frac{\pi^2\phi(R)}{3R^2}. \tag{40}$$

In particular, $W_2(\mu_R^+, \gamma)/W_2(\mu_R^-, \gamma) \sim R/\pi$.

Proof. Write $p(s) = (1+e^{-s})^{-1}$. Then $b_R(x) = p(R(x-R)) - p(R(x-R-2))$, so $0 < b_R < 1$ and $|b_R'| \leq R/4$. For $x \geq R/2$, these bounds give $b_R + |b_R'| \leq 1 + R/4 \leq 1 + |x|$. For $x < R/2$, both logistic terms are at most $e^{-R^2/2}$, and hence

$$b_R(x) + |b_R'(x)| \leq (1+2R)e^{-R^2/2} \leq 1.$$

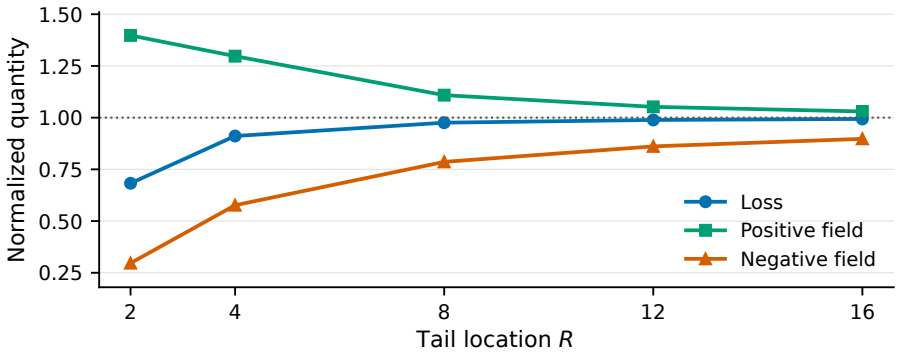

Figure 4: Finite R checks for Proposition 6. The curves show $R\eta_R/\phi(R)$, $3W_2^2(\mu_R^+,\gamma)/\phi(R)$, and $3R^2W_2^2(\mu_R^-,\gamma)/(\pi^2\phi(R))$. Each tends to one by Equation 40.

This proves equation 38. Reversing the output weights preserves all layer operator norms and the absolute spatial derivatives. The function b_R is symmetric about $R+1$, so b_R' is antisymmetric about that point. Its positive and negative extrema have equal magnitude, which proves equation 39. Equation 36 gives the common excess loss.

To compute the terminal errors, set $a = e^{-2R}$ and $y = R(x-R)$. Direct integration gives

$$H_R(x) = \frac{(1+a)y - e^{-y} + ae^{y}}{R(1-a)}, \qquad H_R'(x) = \frac{1}{b_R(x)}. \tag{41}$$

The function H_R is strictly increasing onto $\mathbb{R}$. Thus

$$T_R^{\pm}(z) = H_R^{-1}(H_R(z) \pm 1). \tag{42}$$

The flows exist for all times because their velocities are bounded. Their time one maps are increasing, and $|T_R^{\pm}(z) - z| < 1$. Increasing maps give the optimal coupling for squared distance on the line. To see this directly, define the convex function $\Psi(x) = \int_0^x T_R^{\pm}(s)\,ds$. Fenchel's inequality gives $xy \le \Psi(x) + \Psi^*(y)$, with equality when $y = T_R^{\pm}(x)$. Therefore this coupling maximizes $\mathbb{E}[XY]$ among couplings of the same marginals and minimizes $\mathbb{E}|X-Y|^2$. The expectations are finite since $T_R^{\pm}(x) = x + O(1)$. Consequently,

$$W_2^2(\mu_R^{\pm},\gamma) = \int_{\mathbb{R}} |T_R^{\pm}(z) - z|^2 \phi(z)\,dz. \tag{43}$$

For the loss, substitute $y = R(z-R)$ to obtain

$$\frac{R\eta_R}{\phi(R)} = \int_{\mathbb{R}} e^{-y-y^2/(2R^2)}[p(y) - p(y-2R)]^2\,dy \longrightarrow \int_{\mathbb{R}} e^{-y}p(y)^2\,dy = 1. \tag{44}$$

Dominated convergence applies because $0 \le p(y) - p(y-2R) \le p(y)$ and $e^{-y}p(y)^2 = p'(y)$.

For the terminal errors, let $z_R(w) = R + (w - \log R)/R$ and define

$$d_R(w) = T_R^{+}(z_R(w)) - z_R(w), \qquad D_R(w) = R\big[z_R(w) - T_R^{-}(z_R(w))\big].$$

Equations 41 and 42 give

$$(1+a)d_R + e^{-w}(1 - e^{-Rd_R}) + \frac{e^{w-2R}}{R^2}(e^{Rd_R} - 1) = 1 - a, \tag{45}$$

$$\frac{(1+a)D_R}{R} + e^{-w}(e^{D_R} - 1) + \frac{e^{w-2R}}{R^2}(1 - e^{-D_R}) = 1 - a. \tag{46}$$

Here $0 < d_R < 1$ and $0 < D_R < R$. All terms on the left are nonnegative, so

$$D_R(w) \le \log(1 + e^{w}), \qquad d_R(w) \le -\frac{1}{R}\log(1 - e^{w}) \quad (w < 0). \tag{47}$$

For fixed $w > 0$, the last term in equation 45 is at most e^{w-R}/R^2. The same equation implies $\liminf_R d_R(w) \geq 1 - e^{-w} > 0$, and then $d_R(w) \to 1 - e^{-w}$. For $w < 0$, equation 47 gives $d_R(w) \to 0$. The negative root equation and its bound give $D_R(w) \to \log(1 + e^w)$ for every fixed w.

Changing variables in equation 43 yields

$$\frac{W_2^2(\mu_R^+, \gamma)}{\phi(R)} = \int_{\mathbb{R}} e^{-w-(w-\log R)^2/(2R^2)} d_R(w)^2 \, dw, \tag{48}$$

$$\frac{R^2 W_2^2(\mu_R^-, \gamma)}{\phi(R)} = \int_{\mathbb{R}} e^{-w-(w-\log R)^2/(2R^2)} D_R(w)^2 \, dw. \tag{49}$$

The positive integrand is bounded by e^{-w} on $w > 0$ and by $\frac{1}{4} e^{-w} \log^2(1 - e^w)$ on $w < 0$. The negative integrand is bounded by $e^{-w} \log^2(1 + e^w)$. These bounds are integrable. Indeed,

$$\int_{\mathbb{R}} e^{-w} \log^2(1 + e^w) \, dw = 2 \int_0^1 \frac{-\log(1-t)}{t} \, dt = \frac{\pi^2}{3}.$$

The integral of $e^{-w} \log^2(1 - e^w)$ on $w < 0$ has the same value, by the substitution $u = e^w/(1 - e^w)$. Dominated convergence in the two terminal integrals now gives $\int_0^\infty e^{-w}(1 - e^{-w})^2 \, dw = 1/3$ and $\pi^2/3$, respectively. Together with equation 44, these limits prove equation 40. □

Why the errors differ. Near the left edge of the region where b_R is large, the positive field moves points toward larger velocity magnitudes. The negative field moves them toward smaller velocity magnitudes. The two flows therefore visit different states despite having the same velocity magnitudes at every fixed state. The derivative bounds over the entire state space lose this distinction. So do population averages of absolute derivatives, including the squared Hessian. Even the material derivatives agree at each state, since $v_R^+(v_R^+)' = v_R^-(v_R^-)' = b_R b_R'$. The visited paths retain the direction of motion.

The role of path growth is exact in this example. Let $\Phi_t^\pm(z)$ denote the flow at time t. Differentiating its defining equation gives

$$\partial_z \Phi_t^\pm(z) = \frac{b_R(\Phi_t^\pm(z))}{b_R(z)} = \exp\left(\int_0^t (v_R^\pm)'(\Phi_s^\pm(z)) \, ds \right). \tag{50}$$

Integrating the velocity along the path then yields

$$|T_R^\pm(z) - z| = b_R(z) \int_0^1 \exp\left(\int_0^t (v_R^\pm)'(\Phi_s^\pm(z)) \, ds \right) dt. \tag{51}$$

The source magnitude $b_R(z)$ is the same for both fields. The signed Jacobian growth along their different paths determines the displacement. This identity uses the logarithmic norm in one dimension and retains contraction as well as expansion.

What the example establishes. The two networks have explicit parameters on a Gaussian rotation CFM path with Gaussian source and target. Both terminal errors tend to zero, and their ratio diverges. The common growth constant in equation 38 is fixed, while the global Lipschitz constants grow with R. The construction shows that equal population losses and equal global derivative bounds leave the relative terminal errors undetermined. Signed Jacobian growth along the generated paths distinguishes the two fields.