ENDPOINT SINGULARITY AND TEMPORAL HETEROGENEITY IN FLOW-MATCHING TRANSPORT

Anonymous authors
Paper under double-blind review

ABSTRACT

Flow matching learns a time-dependent velocity field transporting a source distribution toward the data distribution. We study the endpoint geometry of this transport.

A lower-dimensional target forces the endpoint flow map to lose rank. In a quadratic Hamilton–Jacobi formulation, this appears through a Riccati equation for the velocity Jacobian whose critical eigenvalue corresponds to endpoint rank loss. We further confirm this criticality empirically: the divergence of a single trained model grows sharply as $t \to 1$.

Fine-tuning modifies the learned velocity field and can be viewed as a perturbation of it. We measure the temporal profile of this perturbation and find that it is non-uniform in flow time. We use this structure to define a peak-based query distribution for dynamics-level watermark detection, in which the query budget is concentrated on flow times where the watermark signal has larger energy. On watermarked flow-matching models, the peak-based distribution is more robust to fine-tuning than uniform and fixed-mixture sampling, retaining higher detection accuracy at intermediate fine-tuning steps.

1 INTRODUCTION

Flow matching (Lipman et al., 2023; Liu et al., 2023) learns a velocity field

$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = v_\theta(x_t, t), \qquad t \in [0, 1], \tag{1}$$

whose flow transports a source distribution toward the data distribution. Let $\Phi_{s\to t}$ denote the flow map of the ODE, so that $x_t = \Phi_{s\to t}(x_s)$ for any trajectory of the velocity field. Its differential $D\Phi_{s\to t}$ is the Jacobian of the endpoint t with respect to the initial state at s. This matrix is the object that accumulates the geometry of the transport.

Our starting observation is geometric. If the target support has intrinsic dimension $k < d$, a full-dimensional source neighborhood must collapse in at least $d - k$ directions at the endpoint. Thus the endpoint flow map becomes rank deficient in the ideal collapse limit. This gives a principled reason to study the terminal regime rather than treating all flow times as equivalent.

For analysis we use a quadratic variational formulation with a terminal penalty $F(\rho_1)$ in place of the hard terminal constraint of the Benamou–Brenier problem (Benamou & Brenier, 2000). Its first-order optimality conditions give a potential flow $v = \nabla\phi$ and hence

$$\operatorname{div} v := \nabla \cdot v = \Delta\phi. \tag{2}$$

The velocity Jacobian satisfies a Riccati equation whose critical eigenvalue produces endpoint rank loss of the flow-map differential. Near criticality, $\operatorname{div} v$ is $O(1)$ for $t < 1$ and singular at $t = 1$. The transport is therefore non-uniform in time: the flow map behaves differently in the interior of the trajectory and at the endpoint. A finite penalty prevents exact rank loss and replaces the divergence pole by a finite endpoint value.

We then ask whether the same temporal structure is visible in a practical model. We measure how the learned velocity and its divergence evolve as functions of flow time, both for a single trained model and for the perturbation induced by fine-tuning. The profiles are non-uniform in flow time. On

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.

MNIST, CIFAR-10, and a toy problem with a closed-form optimum, the divergence grows sharply toward $t = 1$ and the fine-tuning perturbation concentrates at early flow times. The trained endpoint is quasi low-dimensional: the flow map is close to rank-deficient but not exactly rank-deficient, and the endpoint divergence is finite.

The temporal structure of the fine-tuning perturbation has a direct application in dynamics-level watermarking. If the watermark signal is not equally strong at every flow time, a fixed query budget should be concentrated where the signal energy is largest. We propose a peak-based query distribution that uses the measured power profile of the watermark, and we compare it against uniform and fixed-mixture sampling.

The paper is organized as follows. Section 2 derives the variational formulation and its Euler–Lagrange equations. Section 3 characterizes endpoint rank loss through the Riccati dynamics. Section 4 reports the temporal profiles of the divergence and of the fine-tuning perturbation. Section 5 describes the peak-based query distribution and its baselines. Section 6 concludes.

2 Flow Matching and the Variational Formulation

This section sets up the flow-matching and variational background used throughout the paper. The linear flow-matching construction is standard (Lipman et al., 2023; Liu et al., 2023) and so is the dynamic optimal-transport formulation with a hard terminal constraint (Benamou & Brenier, 2000). We relax that constraint to a terminal penalty $F(\rho_1)$ in equation 6 and derive the corresponding Euler–Lagrange conditions in Lemma 2.1. The derivation of the lemma follows the standard dynamic optimal-transport derivation (Villani et al., 2009) and is included for completeness.

For a pair (x_0, x_1), linear flow matching uses the interpolation

$$x_t = (1-t)x_0 + tx_1, \qquad u_t = x_1 - x_0. \tag{3}$$

Hence each conditional characteristic is a straight line,

$$\dot{x}_t = x_1 - x_0, \qquad \ddot{x}_t = 0. \tag{4}$$

The population-optimal field is

$$v^*(x,t) = \mathbb{E}[x_1 - x_0 \mid x_t = x]. \tag{5}$$

Under optimal-transport coupling, the interpolation minimizes the kinetic energy, which motivates the following dynamic variational formulation. Because in practice the target distribution is not known exactly, we replace the hard terminal constraint $\rho_1 = \rho_{\mathrm{tgt}}$ of the standard Benamou–Brenier problem by a terminal penalty $F(\rho_1)$:

$$\begin{aligned} \min_{\rho, v} \quad & \frac{1}{2}\int_0^1 \int_{\mathbb{R}^d} \rho_t(x)\, \|v_t(x)\|^2 \, \mathrm{d}x\, \mathrm{d}t + F(\rho_1) \\ \text{s.t.} \quad & \partial_t \rho_t + \nabla \cdot (\rho_t v_t) = 0, \qquad \rho_0 = \rho_{\mathrm{src}}. \end{aligned} \tag{6}$$

Here F may be zero, a hard density constraint imposed as a limit of penalties, or an information-theoretic penalty such as a KL divergence.

2.1 Euler–Lagrange Equations

We now derive the first-order optimality conditions for equation 6, which will provide the potential-flow structure used throughout the paper.

Lemma 2.1 (Euler–Lagrange equations)**.** *Let ϕ_t be the Lagrange multiplier for the continuity constraint in equation 6. The first-order necessary conditions for optimality are*

$$v_t = \nabla \phi_t, \tag{7}$$

and

$$\partial_t \phi_t + \frac{1}{2}\|\nabla \phi_t\|^2 = 0. \tag{8}$$

The terminal functional determines the boundary condition at $t = 1$ through

$$\phi_1(x) = -\frac{\delta F}{\delta \rho_1}(x), \tag{9}$$

and the initial density is fixed by $\rho_0 = \rho_{\mathrm{src}}$.

Proof. Introduce ϕ_t for the continuity constraint:

$$\mathcal{A} = \int_0^1 \int \rho_t \frac{\|v_t\|^2}{2} \,\mathrm{d}x\,\mathrm{d}t + F(\rho_1) + \int_0^1 \int \phi_t \big(\partial_t \rho_t + \nabla \cdot (\rho_t v_t)\big) \,\mathrm{d}x\,\mathrm{d}t. \tag{10}$$

Integrating by parts,

$$\mathcal{A} = \int_0^1 \int \rho_t \Big[\frac{\|v_t\|^2}{2} - \partial_t \phi_t - v_t \cdot \nabla \phi_t\Big] \,\mathrm{d}x\,\mathrm{d}t + F(\rho_1) + \int \phi_1 \rho_1 \,\mathrm{d}x - \int \phi_0 \rho_0 \,\mathrm{d}x. \tag{11}$$

Varying in v_t gives $\rho_t v_t - \rho_t \nabla \phi_t = 0$, hence $v_t = \nabla \phi_t$. Varying in ρ_t for $0 < t < 1$ gives $\frac{1}{2}\|v_t\|^2 - \partial_t \phi_t - v_t \cdot \nabla \phi_t = 0$, which after substituting $v_t = \nabla \phi_t$ becomes the Hamilton–Jacobi equation. Because ρ_1 appears only in the endpoint terms $F(\rho_1)$ and $\int \phi_1 \rho_1$, stationarity with respect to ρ_1 gives

$$\frac{\delta F}{\delta \rho_1}(x) + \phi_1(x) = 0 \quad \Longrightarrow \quad \phi_1(x) = -\frac{\delta F}{\delta \rho_1}(x). \tag{12}$$

The initial density is constrained, so $\delta \rho_0 = 0$ and no initial condition on ϕ is produced. □

Remark 2.2 (Role of the terminal functional)**.** *The terminal functional F enters only through the boundary condition equation 9. The interior Euler–Lagrange equations equation 7 and equation 8 do not depend on F. Consequently, the potential-flow structure, the Hamilton–Jacobi equation, and the Riccati dynamics of Section 3 hold for any choice of F.*

In this potential-flow formulation,

$$J := \nabla v = \nabla^2 \phi, \qquad \operatorname{div} v = \Delta \phi. \tag{13}$$

The divergence is therefore the Laplacian of the same potential whose gradient is the velocity.

3 Dimensional Collapse and Endpoint Singularity

A natural way to characterize the geometry of a flow is through the deformation of its flow map $D\Phi_{s\to t}$. Finite-time deformation diagnostics such as the finite-time Lyapunov exponent (FTLE) characterize this deformation through the singular spectrum of $D\Phi_{s\to t}$, with particular emphasis on the largest singular value and finite-time stretching (Shadden et al., 2005; Haller, 2015). Here the relevant phenomenon is at the opposite end of the spectrum: dimensional collapse corresponds to singular values approaching zero, indicating the loss of locally distinguishable directions under the transport.

This distinction matters for flow matching. When a full-dimensional source distribution is transported toward a target concentrated on a lower-dimensional set, the terminal flow map cannot remain locally invertible in all directions. Some singular directions must collapse as $t \to 1$. We first make this statement precise at the level of the flow map, then show how the same collapse appears as a singularity in the instantaneous velocity divergence.

Let $\Phi_{0\to 1}$ denote the terminal flow map. Suppose a full-dimensional neighborhood of the source is transported into a k-dimensional C^1 target manifold, with $k < d$. Then

$$\operatorname{rank} D\Phi_{0\to 1} \le k. \tag{14}$$

Hence at least $d - k$ singular directions collapse at $t = 1$.

To relate this to the velocity field, we use Liouville's identity for the flow-map differential (Arnold et al., 1989),

$$\frac{\mathrm{d}}{\mathrm{d}t} \log |\det D\Phi_{s\to t}(x_s)| = \operatorname{div} v(x_t, t). \tag{15}$$

Thus, dimensional collapse is a statement about the singular values of $D\Phi$, while the divergence is its instantaneous volume-rate diagnostic.

We can now make the collapse explicit in the potential-flow model. Taking the gradient of equation 8 and then restricting to a characteristic gives the Riccati equation

$$\dot{J} = -J^2, \qquad J := \nabla v = \nabla^2 \phi. \tag{16}$$

Solving equation 16 gives the flow-map differential

$$J_t = J_s[I + (t-s)J_s]^{-1}, \qquad D\Phi_{s\to t} = I + (t-s)J_s. \tag{17}$$

The derivations are given in Appendix A.

Proposition 3.1 (Critical eigenvalue and endpoint rank loss)**.** *Let J_s be the velocity Jacobian at reference time $s \in [0,1)$, and let μ_s be an eigenvalue of J_s with eigenvector u_s. The flow-map differential acts on u_s by the scalar factor*

$$m(t) = 1 + (t-s)\mu_s, \tag{18}$$

so that

$$D\Phi_{s\to t}(x_s)\, u_s = m(t)\, u_s. \tag{19}$$

Consequently, $D\Phi_{s\to 1}$ is rank-deficient in the direction u_s if and only if $\mu_s = -1/(1-s)$.

Proof. By equation 17, $D\Phi_{s\to t}(x_s) = I + (t-s)J_s$. Applying both sides to u_s and using $J_s u_s = \mu_s u_s$ gives $D\Phi_{s\to t}(x_s)u_s = [1 + (t-s)\mu_s]u_s$. At $t = 1$ the factor is $1 + (1-s)\mu_s$, which vanishes if and only if $\mu_s = -1/(1-s)$. In that case u_s lies in the kernel of $D\Phi_{s\to 1}$ and the differential cannot have full rank. □

Corollary 3.2 (Divergence singularity at the critical eigenvalue)**.** *Under the same hypothesis, the eigenvalue of J_t corresponding to μ_s evolves as*

$$\mu(t) = \frac{\mu_s}{1 + (t-s)\mu_s}, \tag{20}$$

and

$$\operatorname{div} v(x_t, t) = \operatorname{tr} J_t = \sum_i \frac{\mu_s^{(i)}}{1 + (t-s)\mu_s^{(i)}}. \tag{21}$$

If any eigenvalue $\mu_s^{(i)}$ equals $-1/(1-s)$, the corresponding term has a pole at $t = 1$.

Proof. The eigenvalue evolution equation 20 follows from $J_t = J_s[I + (t-s)J_s]^{-1}$ applied to the eigenvector u_s. The divergence is the trace, which is the sum of the eigenvalues equation 21. The i-th term has a pole at $t = 1$ exactly when $1 + (1-s)\mu_s^{(i)} = 0$, i.e. at the critical value $-1/(1-s)$. □

Remark 3.3 (Endpoint regularization)**.** *Corollary 3.2 describes the hard-constraint limit, in which the endpoint is forced onto a lower-dimensional target and the divergence has a pole at $t = 1$. In the penalized problem equation 6, the terminal constraint is relaxed to a finite penalty $F(\rho_1)$, and the endpoint is pulled toward the target without being forced onto it. The critical eigenvalue is then approached but not attained, and the pole of Corollary 3.2 is replaced by a finite maximum. A stiff penalty recovers the hard-constrained singularity, while a finite penalty produces a finite endpoint divergence.*

4 Fine-Tuning and Temporal Heterogeneity

Section 3 established that near-critical endpoint geometry makes the flow map non-uniform in time. We now measure four diagnostics on trained models, all as functions of flow time. Two concern a single trained model and measure its endpoint behavior: the divergence $\operatorname{div} v(t)$, which is the instantaneous volume-rate diagnostic, and the single-model SNR against the variational reference. Two concern the perturbation induced by fine-tuning: the matched-state perturbation Δv_t between a pre-trained and a fine-tuned model, and the change in divergence $\Delta(\operatorname{div} v)(t)$ that it induces.

We begin with the fine-tuning perturbation. Let v_{pre} and v_{FT} denote the pre-trained and fine-tuned velocity fields, evaluated at the same states along a pre-trained trajectory x_t^{pre}:

$$\Delta v_t := v_{\mathrm{FT}}(x_t^{\mathrm{pre}}, t) - v_{\mathrm{pre}}(x_t^{\mathrm{pre}}, t). \tag{22}$$

If the flow were temporally uniform, the magnitude of Δv_t would be roughly constant in t. Its temporal profile is measured by

$$A_v(t) := \mathbb{E}_{x_0}\left[\|\Delta v_t\|^2\right]^{1/2}, \tag{23}$$

normalized as

$$\widetilde{A}_v(t) := \frac{A_v(t)}{\int_0^1 A_v(s)\,\mathrm{d}s}. \tag{24}$$

A flat profile would indicate approximately uniform perturbation over flow time; a non-uniform profile establishes temporal heterogeneity.

The geometric counterpart of the perturbation is

$$\Delta(\operatorname{div} v)(x,t) := \nabla \cdot \Delta v(x,t) = \operatorname{div} v_{\mathrm{FT}}(x,t) - \operatorname{div} v_{\mathrm{pre}}(x,t). \tag{25}$$

When both fields belong to the potential-flow class,

$$\Delta v = \nabla \Delta\phi, \qquad \Delta(\operatorname{div} v) = \Delta(\Delta\phi)\,, \tag{26}$$

so a temporal change in $\Delta(\operatorname{div} v)$ measures a temporal change in the Laplacian of the fine-tuning-induced potential perturbation.

We turn next to the single-model diagnostics. The divergence $\operatorname{div} v(x_t, t)$ measures the instantaneous volume change of the flow map through Liouville's identity equation 15. The straight-line variational reference provides a second signal: along a characteristic the reference velocity $v(x_0, 0)$ is constant, and the deviation of the learned velocity from this constant is

$$\Delta v(x_t, t) := v(x_t, t) - v(x_0, 0), \tag{27}$$

with corresponding single-model signal-to-noise ratio

$$\mathrm{SNR}(t) := \frac{\mathbb{E}_{x_0} \left\| v(x_0, 0) \right\|^2}{\mathbb{E}_{x_0} \left\| \Delta v(x_t, t) \right\|^2}. \tag{28}$$

The protocol is described in Appendix C.1. We report the four profiles in the order in which they were introduced.

For the single-model divergence, the measured $\operatorname{div} v(t)$ is close to -1 for $t \lesssim 0.9$ and decreases sharply thereafter, reaching values of order -10^2 to -10^3 at $t = 1$. The profile is reported in Appendix C.2 and matches the divergence singularity of Corollary 3.2.

For the fine-tuning perturbation, the measured $\widetilde{A}_v(t)$ is non-uniform in t: it peaks around $t \approx 0.3$ and decays toward the endpoint. The coefficient of variation is 0.79 on MNIST and 0.43 on CIFAR-10. The profiles are reported in Appendix C.3.

For the single-model SNR, the measured $\mathrm{SNR}(t)$ is of order 10 throughout the interior of the flow and drops to order 1 at $t = 1$. The drop is consistent with the endpoint region being harder to fit, since the velocity must compress strongly in the collapsing directions there. The profile is reported in Appendix C.4.

For the change in divergence induced by fine-tuning, the measured $\Delta(\operatorname{div} v)(t)$ is $O(1)$ throughout the interior and grows sharply at $t = 1$. The change at several fine-tuning steps is reported in Appendix C.3.

5 Watermark Detection by Temporal Peak Sampling

A watermark embeds a message into a channel that the detector observes with noise. In the classical setting the channel is stationary: the signal-to-noise ratio does not depend on the observation index, and the optimal detector queries all indices equally. When the channel is non-stationary, the budget should instead be allocated toward indices where the signal has energy, as in matched filtering.

Dynamics-level watermarks embed a message into the velocity field of a generative model, with flow time $t \in [0, 1]$ as the channel index. Prior work (Wang, 2026) treats this channel as stationary and queries t uniformly. Sections 3 and 4 show that it is not stationary: the flow-map differential is regular in the interior and singular as $t \to 1$, and the fine-tuning perturbation itself is non-uniform in t. Uniform querying is therefore suboptimal.

We use the standard additive dynamics-level watermark. With codeword $c_m \in \mathbb{R}^r$ and projection $P \in \mathbb{R}^{d \times r}$ onto a fixed r-dimensional subspace, the watermarked velocity field is

$$v_{\mathrm{WM}}(x,t) = v_{\mathrm{pre}}(x,t) + \epsilon\, a(t)\, P c_m, \qquad a(t) = \sin(2\pi t), \tag{29}$$

with embedding strength $\epsilon > 0$ and temporal envelope $a(t)$. The envelope controls the temporal profile of the injection; both the low-rank form and the envelope are standard. The detector projects the received field onto the secret directions:

$$Y(x,t) = a(t)\left\langle P^\top v(x,t), c_m \right\rangle. \tag{30}$$

Under the null, Y has zero mean; under the alternative, its mean is $\epsilon\, a(t)^2 \left\|c_m\right\|^2$.

5.1 Power profile

The watermark signal is not equally strong at every flow time. Its time-resolved energy is captured by the power profile

$$P(t) = \mathbb{E}_x\left[Y(x,t)^2\right]. \tag{31}$$

The profile is estimated on a fixed grid of flow times by Monte Carlo sampling of $x \sim \mathcal{N}(0, I)$, normalized by its maximum, and smoothed with a Savitzky–Golay filter. Peaks are extracted with a standard peak-finding routine using a minimum height and a minimum separation between peaks. We use the top $K = 2$ peaks $\{t_1^*, \ldots, t_K^*\}$ with heights $\{h_1, \ldots, h_K\}$.

5.2 Peak-based query distribution

Given a query budget B, we allocate samples to the top-K peaks with weights $w_i = h_i / \sum_j h_j$ and sample

$$t \sim \sum_{i=1}^{K} w_i\, \mathcal{N}(t_i^*, \sigma^2), \tag{32}$$

with kernel width $\sigma = 0.03$ and clamping to $[0, 1]$. If no peak exceeds the minimum height, the sampler falls back to uniform. The mixture is fixed once from the calibration model and reused at detection time.

5.3 Detector and baselines

For each query the detector evaluates $Y(x_i, t_i)$ and forms the signature

$$\hat{s} = \frac{1}{B}\sum_{i=1}^{B} Y(x_i, t_i), \tag{33}$$

with decoded message

$$\hat{m} = \arg\max_{m'} \left\langle \hat{s}, c_{m'} \right\rangle. \tag{34}$$

The signature is a plain average of the detector readouts over the sampled queries.

We compare three distributions for the query times, which share the same embedding equation 29, the same detector statistic equation 30, and the same codebook, and differ only in where the budget is spent: uniform, $t \sim \mathcal{U}(0,1)$; a fixed gaussian mixture, $\frac{1}{2}\mathcal{N}(0.25, \sigma^2) + \frac{1}{2}\mathcal{N}(0.75, \sigma^2)$; and the peak-based mixture equation 32. Algorithm 1 summarizes the procedure.

The power-peak detector spends the budget where the measured signal energy is largest and forms a plain average of the detector readouts. The peak-based sampler retains higher detection accuracy than uniform and fixed-mixture sampling over a wider range of fine-tuning steps. On MNIST it holds full accuracy through 200 fine-tuning steps while both baselines fall to chance by 200 steps. On CIFAR-10 the three methods track each other through 200 steps, the peak-based sampler separates at 300 steps (0.90 against 0.55 for the gaussian mixture and 0.35 for uniform), and it retains its advantage through 400 steps. Full accuracy curves are reported in Appendix C.5.

6 Discussion and Conclusion

We studied endpoint behavior in flow matching in three linked steps. First, lower-dimensional target support requires endpoint rank collapse of the flow map; in the Hamilton–Jacobi potential-flow model this appears through a Riccati criticality, and the divergence has a pole at $t = 1$. We refer to this as

Algorithm 1 Watermark detection by temporal peak sampling

Require: model v; secret projection P and codeword c_m; time grid $\{t_1, \dots, t_K\}$; peak width σ; query budget B; calibration samples n.
Ensure: decoded message $\hat{m}$.
Calibration (offline).
sample $\{x^{(j)}\}_{j=1}^{n} \sim \mathcal{N}(0, I)$.
for $k = 1, \dots, K$ **do**
  set $Y_k^{(j)} \leftarrow a(t_k) \left\langle P^\top v(x^{(j)}, t_k), c_m \right\rangle$ for each j.
  set $P(t_k) \leftarrow \frac{1}{n} \sum_j (Y_k^{(j)})^2$.
end for
normalize P by its maximum; smooth with Savitzky–Golay.
find peaks $\{t_i^*, h_i\}$ with height threshold and distance.
keep the top K peaks; set $w_i \leftarrow h_i / \sum_j h_j$.
if no peaks found **then**
  fall back to uniform sampling.
end if
Detection (online).
for $i = 1, \dots, B$ **do**
  sample $t_i \sim \sum_{k=1}^{K} w_k \mathcal{N}(t_k^*, \sigma^2)$.
  clamp t_i to $[0, 1]$.
  sample $x_i \sim \mathcal{N}(0, I)$.
  set $Y_i \leftarrow a(t_i) \left\langle P^\top v(x_i, t_i), c_m \right\rangle$.
end for
set $\hat{s} \leftarrow \frac{1}{B} \sum_{i=1}^{B} Y_i$.
decode $\hat{m} \leftarrow \arg\max_{m'} \langle \hat{s}, c_{m'} \rangle$.

the *endpoint singularity*. Second, we reported the temporal profiles of the fine-tuning perturbation and the single-model SNR and found that both are non-uniform in flow time. We refer to this as *temporal heterogeneity*. Third, we used the temporal heterogeneity to motivate a peak-based query distribution for dynamics-level watermarking.

The endpoint singularity is established as the analytic counterpart of the low-dimensional manifold hypothesis. Proposition 3.1 states that $D\Phi_{s\to 1}$ is rank-deficient in the direction u_s if and only if the velocity Jacobian has the critical eigenvalue $\mu_s = -1/(1-s)$, and Corollary 3.2 states that the corresponding term in the divergence has a pole at $t = 1$. Both statements hold for the hard-constrained problem, in which $\rho_1 = \rho_{\mathrm{tgt}}$ is imposed exactly.

A trained model does not attain this limit, and two measurements show where the departure lies. First, the divergence. In the toy example (Appendix B), three of four singular values collapse toward 2.5×10^{-3} at $t = 1 - 10^{-3}$ while the fourth remains at 1; the divergence reaches 390.68 against an analytic prediction of 3×10^3, according to Corollary 3.2. The same pattern appears on MNIST and CIFAR-10 (Appendix C.2): the divergence grows sharply toward $t = 1$ but stays finite at the last grid point. Second, the single-model SNR along a characteristic of the straight-line reference the velocity is constant, so the analytic deviation vanishes and the analytic SNR is unbounded at every flow time. The measured SNR is finite and drops near $t = 1$ (Appendix C.4). The finite endpoint divergence and the SNR drop are two measurements of the same departure from the analytic limit.

The departure from the singularity is a regularization effect. In the standard flow-matching objective the terminal constraint is not imposed exactly: the model is trained by a regression loss, and the terminal distribution approximates the target rather than matching it. The analysis models this by replacing the hard constraint with a terminal penalty $F(\rho_1)$ in equation 6, so that F represents the residual endpoint mismatch left by training. A finite penalty replaces the pole with a finite maximum (Remark 3.3) and moves the optimum away from the straight-line reference. The finite capacity of the network acts as a further penalization of the same kind: it cannot represent a singular velocity field, so the trained field is smoothed relative to any singular target. The trained endpoint is therefore quasi low-dimensional rather than exactly low-dimensional.

The temporal heterogeneity motivates the peak-based query distribution of Section 5. The method uses the measured power profile of the watermark to concentrate the query budget on flow times where the signal energy is largest, and retains higher detection accuracy under fine-tuning than uniform and fixed-mixture sampling on MNIST and CIFAR-10 (Appendix C.5).

REFERENCES

Vladimir Igorevich Arnold, Karen Vogtmann, and Alan Weinstein. *Mathematical methods of classical mechanics*, volume 60. Springer, 1989.

Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

Lawrence C Evans. *Partial differential equations*, volume 19. American mathematical society, 2022.

George Haller. Lagrangian coherent structures. *Annual review of fluid mechanics*, 47(1):137–162, 2015.

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.

Yaron Lipman, Ricky TQ 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, et al. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh International Conference on Learning Representations*, 2023.

Shawn C Shadden, Francois Lekien, and Jerrold E Marsden. Definition and properties of lagrangian coherent structures from finite-time lyapunov exponents in two-dimensional aperiodic flows. *Physica D: Nonlinear Phenomena*, 212(3-4):271–304, 2005.

Cédric Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2009.

Shuchan Wang. Dynamics-level watermarking of flow matching models with random codes, 2026. URL https://arxiv.org/abs/2605.16239.

A Derivation of the Riccati Equation and the Flow-Map Differential

We derive equation 16 from the Hamilton–Jacobi equation equation 8, then solve it to obtain equation 17. For background information, see, e.g., (Evans, 2022).

A.1 Derivation of the Riccati Equation

The velocity is a gradient, $v_t = \nabla\phi_t$, and the potential satisfies

$$\partial_t\phi_t + \frac{1}{2}\left\|\nabla\phi_t\right\|^2 = 0. \tag{35}$$

Applying ∇ to equation 35 gives

$$\nabla\big(\partial_t\phi_t\big) + \frac{1}{2}\nabla\big(\left\|\nabla\phi_t\right\|^2\big) = 0. \tag{36}$$

The first term is $\partial_t v_t$, since ∂_t and ∇ commute. For the second, the identity $\nabla\left\|v\right\|^2 = 2(\nabla v)^\top v$ gives

$$\frac{1}{2}\nabla\left\|v_t\right\|^2 = (\nabla v_t)^\top v_t = J_t^\top v_t, \tag{37}$$

where $J_t := \nabla v_t$. Since $v_t = \nabla\phi_t$, the Jacobian $J_t = \nabla^2\phi_t$ is symmetric, so $J_t^\top = J_t$. Hence

$$\partial_t v_t + J_t\, v_t = 0. \tag{38}$$

Applying ∇ to equation 38 gives

$$\nabla(\partial_t v_t) + \nabla(J_t\, v_t) = 0. \tag{39}$$

The first term is $\partial_t J_t$. For the second, the product rule for a matrix–vector product gives

$$\nabla(J_t\, v_t) = (\nabla J_t)\, v_t + J_t\,(\nabla v_t) = (\nabla J_t)\, v_t + J_t^2. \tag{40}$$

The tensor $(\nabla J_t)\, v_t$ is the directional derivative $(v_t\cdot\nabla)J_t$, so

$$\partial_t J_t + (v_t\cdot\nabla)J_t + J_t^2 = 0. \tag{41}$$

Let x_t be a characteristic satisfying $\dot{x}_t = v_t(x_t)$. The material derivative of J_t along the characteristic is

$$\frac{\mathrm{d}}{\mathrm{d}t}J_t(x_t) = \partial_t J_t(x_t) + (v_t(x_t)\cdot\nabla)J_t(x_t), \tag{42}$$

and comparing with equation 41 gives

$$\frac{\mathrm{d}}{\mathrm{d}t}J_t = -J_t^2. \tag{43}$$

A.2 Solution of the Riccati Equation

We solve equation 43 with initial condition J_s at time s. Define

$$A(t) := I + (t-s)J_s, \tag{44}$$

so that $A(s) = I$ and $\dot{A}(t) = J_s$, and set

$$J_t = J_s\, A(t)^{-1}. \tag{45}$$

Differentiating equation 45 with respect to t and using the identity $\frac{\mathrm{d}}{\mathrm{d}t}A(t)^{-1} = -A(t)^{-1}\dot{A}(t)\,A(t)^{-1}$ with $\dot{A}(t) = J_s$,

$$\frac{\mathrm{d}}{\mathrm{d}t}J_t = -J_s\, A(t)^{-1}\, J_s\, A(t)^{-1}. \tag{46}$$

From equation 45, $J_t = J_s\, A(t)^{-1}$, hence

$$J_t^2 = J_s\, A(t)^{-1}\, J_s\, A(t)^{-1}, \tag{47}$$

which is exactly the right-hand side of equation 46. Therefore $\frac{\mathrm{d}}{\mathrm{d}t}J_t = -J_t^2$. The initial condition at $t = s$ is $J_s\, A(s)^{-1} = J_s$, so equation 45 solves the initial-value problem, valid as long as $A(t)$ is invertible, i.e. as long as $1 + (t-s)\mu_s \neq 0$ for every eigenvalue μ_s of J_s.

A.3 Derivation of the Flow-Map Differential

The flow-map differential $M_{t,s} := D\Phi_{s\to t}(x_s)$ satisfies the variational equation

$$\frac{\mathrm{d}}{\mathrm{d}t} M_{t,s} = J_t\, M_{t,s}, \qquad M_{s,s} = I. \tag{48}$$

We claim

$$M_{t,s} = I + (t-s)J_s. \tag{49}$$

Differentiating the right-hand side with respect to t gives J_s. On the other hand, using equation 45 for J_t,

$$J_t\, M_{t,s} = J_s\, A(t)^{-1}\, A(t) = J_s, \tag{50}$$

where the first equality substitutes $M_{t,s} = A(t)$ from equation 44. The two sides agree, and the initial condition is $M_{s,s} = I$. Hence equation 49 solves equation 48. Combining equation 45 and equation 49,

$$J_t = J_s\,[I + (t-s)J_s]^{-1}, \qquad D\Phi_{s\to t}(x_s) = I + (t-s)J_s, \tag{51}$$

which is equation 17 in the main text.

B Toy example

This section tests Proposition 3.1 and Corollary 3.2 on a controlled low-dimensional example where the optimal-transport solution is known in closed form. This example provides a ground truth against which the behavior of a trained flow-matching model can be checked directly.

B.1 Setup

Let $d = 4$ and $k = 1$. The source is $x_0 \sim \mathcal{N}(0, I_4)$. The target is supported on the e_1 axis,

$$x_1 = (z, 0, 0, 0), \qquad z \sim \mathcal{N}(0, 1).$$

The optimal-transport map is $T(x_0) = (x_0^{(1)}, 0, 0, 0)$, and the displacement interpolation is

$$x_t = (x_0^{(1)}, (1-t)x_0^{(2)}, (1-t)x_0^{(3)}, (1-t)x_0^{(4)}).$$

The velocity along the interpolation is

$$v(x_t, t) = \left(0,\ -\frac{x_t^{(2)}}{1-t},\ -\frac{x_t^{(3)}}{1-t},\ -\frac{x_t^{(4)}}{1-t}\right),$$

with velocity Jacobian

$$J_t = \mathrm{diag}\left(0,\ -\frac{1}{1-t},\ -\frac{1}{1-t},\ -\frac{1}{1-t}\right).$$

The eigenvalues at $s = 0$ are $\mu_0 = \{0, -1, -1, -1\}$. The singular values of the flow-map differential are

$$\sigma_i(D\Phi_{0\to t}) = (1,\ 1-t,\ 1-t,\ 1-t),$$

so three singular values vanish at $t = 1$ and one remains at 1. The divergence of the velocity is

$$\mathrm{div}\, v(x_t, t) = \mathrm{tr}\, J_t = -\frac{3}{1-t},$$

whose magnitude grows without bound as $t \to 1$.

B.2 Training a flow-matching model

We train a small MLP velocity field $v_\theta : \mathbb{R}^4 \times [0,1] \to \mathbb{R}^4$ with three hidden layers of 128 units, SiLU activations, and t concatenated to the input. The training data use the exact optimal-transport pairing: each x_0 is paired with $T(x_0)$. The flow-matching loss is

$$\mathcal{L}(\theta) = \mathbb{E}_{x_0,t}\, \|v_\theta((1-t)x_0 + t\,T(x_0),\, t) - (T(x_0) - x_0)\|^2\,.$$

Training uses Adam with learning rate 10^{-3}, cosine decay over 20,000 steps, batch size 256, and gradient clipping at 1.0. The final loss is of order 10^{-3}.

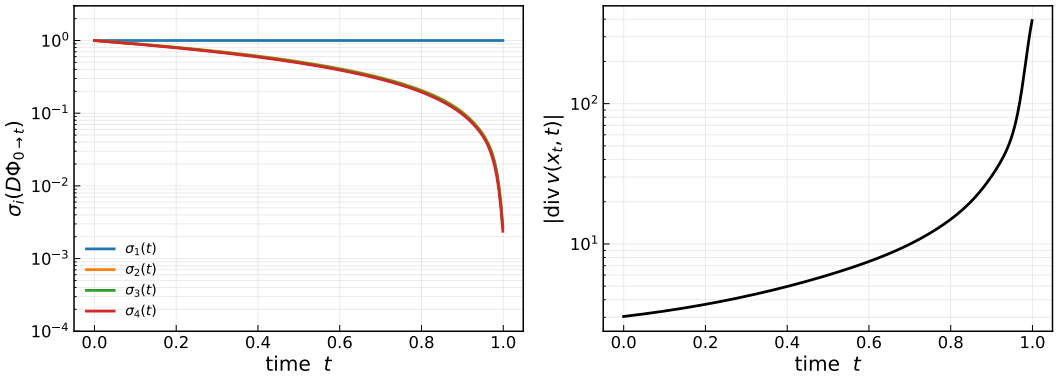

Figure 1: Toy example, $d = 4$, $k = 1$. Solid curves: measurements from the trained model. Dashed curves: analytic optimal-transport predictions. Left: singular values $\sigma_i(D\Phi_{0\to t})$ on a log scale. Three singular values decay toward zero as $t \to 1$ and reach $(2.76 \times 10^{-3},\ 2.67 \times 10^{-3},\ 2.37 \times 10^{-3})$ at $t = 1 - 10^{-3}$, while $\sigma_1(t)$ remains at 1.000. This verifies the rank loss of Proposition 3.1. Right: divergence $|\operatorname{div} v(x_t, t)|$ on a log scale. The learned divergence grows from 3.04 at $t = 0$ to 390.68 at the last plotted time. The analytic prediction at the same time is 3×10^3. The differences between the learned values and the analytic prediction are due to the network's approximation error, which is amplified near the critical value.

Table 1: Toy example, endpoints. Values at $t = 0$ and at $t = 1 - 10^{-3}$ for the trained model and the analytic optimal-transport prediction. All quantities are means over 12 trajectories.

	$t = 0$		$t = 1 - 10^{-3}$	
Quantity	Trained	Analytic	Trained	Analytic
$\sigma_1(D\Phi_{0\to t})$	1.000	1	1.000	1
$\sigma_2(D\Phi_{0\to t})$	1.000	1	2.8×10^{-3}	10^{-3}
$\sigma_3(D\Phi_{0\to t})$	1.000	1	2.7×10^{-3}	10^{-3}
$\sigma_4(D\Phi_{0\to t})$	1.000	1	2.4×10^{-3}	10^{-3}
$\lvert\operatorname{div} v(x_t, t)\rvert$	3.04	3	390	3×10^3

B.3 Measurement

We integrate the learned velocity on a grid of 201 times in $[0, 1 - 10^{-3}]$ and compute two quantities along the learned trajectory:

- the flow-map differential $M_t = D\Phi_{0\to t}$, obtained by integrating $\dot{M}_t = J(x_t, t)M_t$ with $M_0 = I$ and $J(x_t, t) = \nabla_x v_\theta(x_t, t)$ computed by autograd;
- the divergence $\operatorname{div} v(x_t, t) = \operatorname{tr} J(x_t, t)$.

Both quantities are averaged over 12 initial conditions. They are compared against the analytic predictions of Proposition 3.1 and Corollary 3.2.

B.4 Results

Figure 1 reports the two measurements alongside the analytic predictions.

Table 1 reports the values at the endpoints.

The trained model is consistent with both theoretical claims. Three of the four singular values of $D\Phi_{0\to t}$ collapse toward zero as $t \to 1$ while the fourth remains at 1, matching the rank loss of Proposition 3.1. The divergence grows monotonically toward the endpoint, matching the analytic prediction $|\operatorname{div} v| = 3/(1 - t)$ of Corollary 3.2. The quantitative differences at the last plotted time,

the collapsing singular values reaching 2.5×10^{-3} and the divergence reaching 390.68, rather than the analytic 10^{-3} and 3×10^{3} at the same time, are due to the approximation error of the trained network.

C Real-Data Experiments

This section reports the experiments on trained MNIST and CIFAR-10 models. We first describe the configuration, then report each experiment.

C.1 Configuration

C.1.1 Models and training

The clean model is a SimpleUNet with base width 64, InstanceNorm, and SiLU activations, trained by flow matching with independent coupling. Training uses Adam with learning rate 10^{-3}, batch size 4096 for MNIST and 256 for CIFAR-10, and cosine decay over $15{,}000$ and $20{,}000$ steps respectively. The same trained checkpoint is used for all experiments.

C.1.2 Fine-tuning

Fine-tuning uses LoRA adapters on all 3×3 convolutions with rank $r = 4$ and scale $\alpha = 1$. The optimizer is AdamW with learning rate 5×10^{-4}, batch size 256, and gradient clipping at 1.0. The fine-tuned class is class 5.

C.1.3 Trajectories

For each class c, we sample 20 real data points from the training set of class c, treat each as an endpoint x_1, and integrate the reverse ODE backward to obtain x_t^{pre} for $t \in [0, 1]$. Reverse integration is used rather than forward integration from noise because it gives direct control over which class the trajectory approaches at $t = 1$. The reverse ODE is integrated with `scipy.integrate.solve_ivp` using the RK45 method with `rtol`$= 10^{-3}$ and `atol`$= 10^{-5}$.

C.1.4 Matched states

Both v_{pre} and v_{FT} are evaluated at the same states x_t^{pre} along the pre-trained trajectory. This isolates the effect of fine-tuning from the effect of trajectory drift.

C.1.5 Divergence estimation

The divergence is estimated by the Hutchinson identity (Hutchinson, 1989) with 5 Gaussian directions and a centered finite-difference step of $\epsilon = 10^{-3}$ scaled by the local state magnitude. The directions are normalized to unit norm. Under this normalization, the expected value of the estimator is $\mathrm{tr}\, J/d$, so the computed quantity is the trace scaled by a constant factor $1/d$. The factor is the same for v_{pre} and v_{FT}, so $\Delta(\mathrm{div}\, v)$ is scaled by the same factor and the shape of the temporal profile is preserved. Throughout the paper we report the scaled values.

C.1.6 Time grid

The divergence is evaluated on 51 uniform times in $[0, 1]$, and $A_v(t)$ and SNR on 40 uniform times. The resolution is chosen so that the growth of the divergence near $t = 1$ is visible: the penultimate point of the divergence grid is at $t = 0.98$, and the last point is at $t = 1$.

C.1.7 Watermark

The watermark uses a 4-bit codebook with 16 codewords, projection rank 32, embedding strength 10^{-2}, envelope $a(t) = \sin(2\pi t)$, query budget 4096, peak width 0.03, and top $K = 2$ peaks. Twenty detection trials are run per checkpoint.

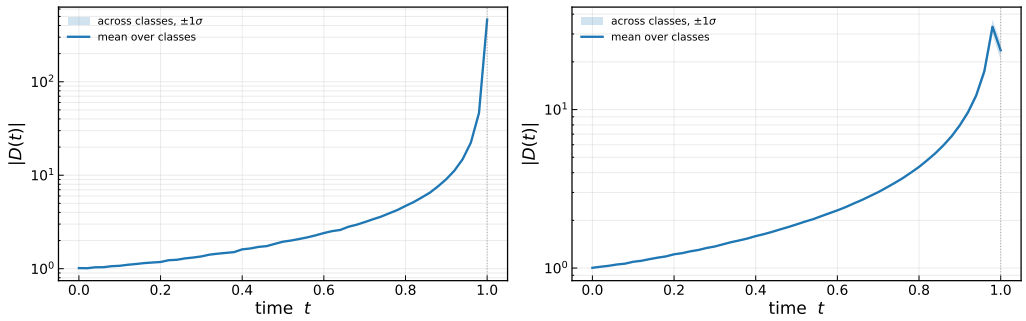

Figure 2: Single-model divergence $D(t) = \operatorname{div} v(x_t, t)$ along the pre-trained trajectory, on a log scale for $|D(t)|$. Left: MNIST. Right: CIFAR-10. Solid curve: mean across the ten classes. Shaded band: ± 1 standard deviation across classes. The divergence is approximately -1 at $t = 0$ and grows sharply toward the endpoint, consistent with the singularity of Corollary 3.2.

Table 2: Fine-tuning perturbation, scalar summaries at 500 fine-tuning steps. Values are mean $\pm$ standard deviation across the ten classes. The coefficient of variation is computed from the normalized profile; the mass split is the integral of $\widetilde{A}_v(t)$ over $[0, 0.5]$ and $[0.5, 1]$.

Dataset	mean A_v	CV	mass first half	mass second half
MNIST	3.045 ± 0.109	0.787 ± 0.016	0.829 ± 0.004	0.171 ± 0.004
CIFAR-10	2.744 ± 0.077	0.431 ± 0.009	0.678 ± 0.004	0.322 ± 0.004

C.2 Divergence profile

Figure 2 reports the single-model divergence $D(t) = \operatorname{div} v(x_t, t)$ along the pre-trained trajectory for MNIST and CIFAR-10, on a logarithmic scale for $|D(t)|$. Values are scaled by $1/d$ from the Hutchinson estimator of Appendix C.1; the factor is constant and preserves the temporal shape.

The profile is common to all classes in the interior: the across-class standard deviation is below 0.3 for $t \leq 0.8$ on both datasets, and grows toward the endpoint. MNIST decreases monotonically through $t = 1$ with no turnover; CIFAR-10 decreases monotonically through $t = 0.98$ and turns over at the final grid point.

The measured endpoint divergence is finite and below the analytic pole in every setting. In the toy example of Appendix B, where the analytic solution is known, the trained divergence reaches 390.68 at $t = 1 - 10^{-3}$ against an analytic prediction of 3×10^3. The same pattern appears on MNIST and CIFAR-10. Because it reproduces across datasets and across the controlled toy problem, it is not a property of the finite-difference estimator. It reflects the finite capacity of the trained network: the velocity field compresses strongly toward the target manifold but does not represent the exact rank-deficient limit. By Proposition 3.1, exact rank loss requires the critical eigenvalue $\mu_s = -1/(1 - s)$; the trained model approaches but does not attain it, and the gap is approximation error rather than a numerical artifact.

C.3 Fine-tuning perturbation

This appendix reports the fine-tuning perturbation $A_v(t)$ and the divergence change $\Delta(\operatorname{div} v)(t)$ defined in Section 4, for MNIST and CIFAR-10, averaged across the ten classes and shown at 50, 100, 200, and 500 fine-tuning steps on class 5. The protocol is the one described in Appendix C.1.

Figure 3 reports the normalized profile $\widetilde{A}_v(t)$. Table 2 reports the coefficient of variation and the mass split between the two halves of the flow.

The perturbation is strongly non-uniform in flow time on both datasets. On MNIST the coefficient of variation is about 0.8 and 83% of the profile mass lies in the first half of the flow. On CIFAR-10 the coefficient of variation is about 0.43 and the first-half mass is about 68%. In both cases the profile

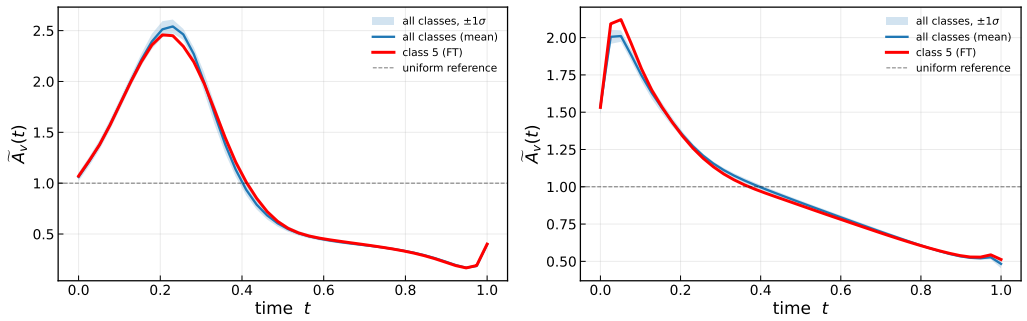

Figure 3: Normalized fine-tuning perturbation $\widetilde{A}_v(t)$ along pre-trained trajectories, at 500 fine-tuning steps on class 5. Left: MNIST. Right: CIFAR-10. Solid curve: mean across the ten classes. Shaded band: ± 1 standard deviation across classes. The fine-tuned class is shown in red.

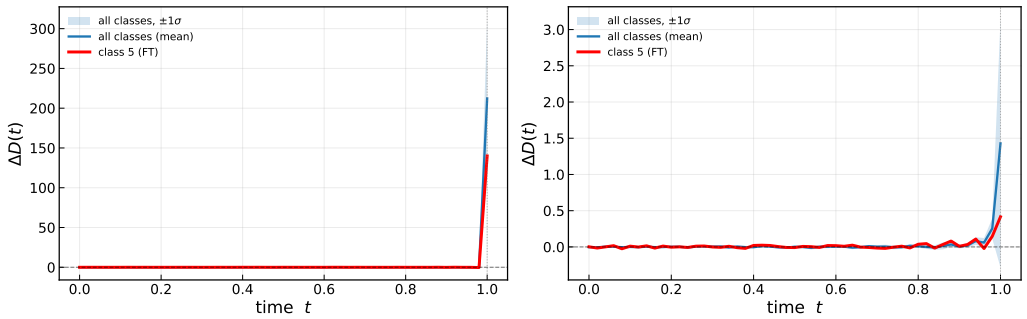

Figure 4: Fine-tuning change in the divergence, $\Delta(\operatorname{div} v)(t) = \operatorname{div} v_{\mathrm{FT}}(t) - \operatorname{div} v_{\mathrm{pre}}(t)$, along pre-trained trajectories at 500 fine-tuning steps on class 5. Left: MNIST. Right: CIFAR-10. Blue: mean across the ten classes with ± 1 standard deviation band. Red: the fine-tuned class.

decreases from a maximum near $t \approx 0.3$ through the interior of the flow, reaches a shallow minimum around $t \approx 0.9$, and rises again at the last grid point, where the velocity field does the most work as the trajectory approaches the target.

Figure 4 reports the divergence change $\Delta(\operatorname{div} v)(t)$ along pre-trained trajectories. The change is $O(1)$ throughout the interior of the flow and grows sharply as $t \to 1$, mirroring the single-model divergence.

The endpoint spike is present at every fine-tuning step we tested on both datasets. The two diagnostics agree on the temporal structure of fine-tuning: the velocity perturbation $A_v(t)$ concentrates at early flow times, while the divergence change $\Delta(\operatorname{div} v)(t)$ concentrates near the endpoint. The perturbation is therefore non-uniform in t in two distinct ways: the magnitude of the velocity change peaks early, and the change in the volume-rate diagnostic peaks late. This is the temporal structure that motivates the peak-based query distribution of Section 5.

C.4 Single-model SNR

This appendix reports the single-model signal-to-noise ratio $\mathrm{SNR}(t)$ defined in Section 4, for MNIST and CIFAR-10, averaged across the ten classes. The protocol is the one described in Appendix C.1.

Figure 5 reports $\mathrm{SNR}(t)$ on a logarithmic scale. The SNR is large through the interior of the flow and drops as $t \to 1$. In the interior the learned velocity tracks the straight-line reference closely; near the endpoint it departs from the reference substantially, which is consistent with the endpoint singularity of Corollary 3.2. The drop at the endpoint is a single-model diagnostic, distinct from the fine-tuning perturbation of Appendix C.3.

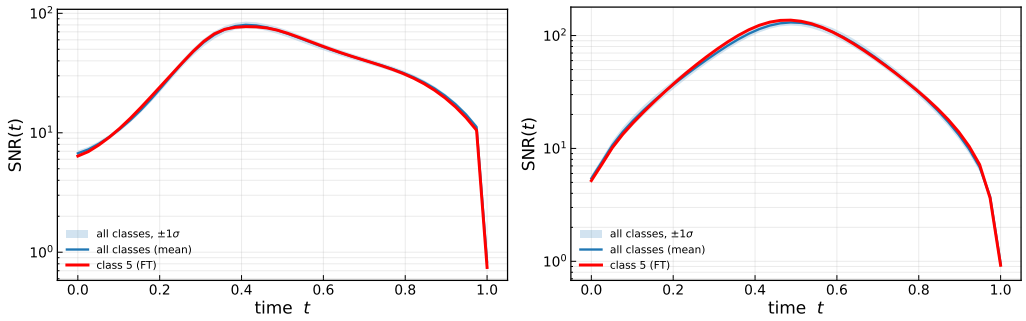

Figure 5: Single-model signal-to-noise ratio $\mathrm{SNR}(t)$ along pre-trained trajectories. Left: MNIST. Right: CIFAR-10. Solid curve: mean across the ten classes. Shaded band: ± 1 standard deviation across classes. The SNR is a diagnostic of the deviation of the learned velocity from the straight-line reference.

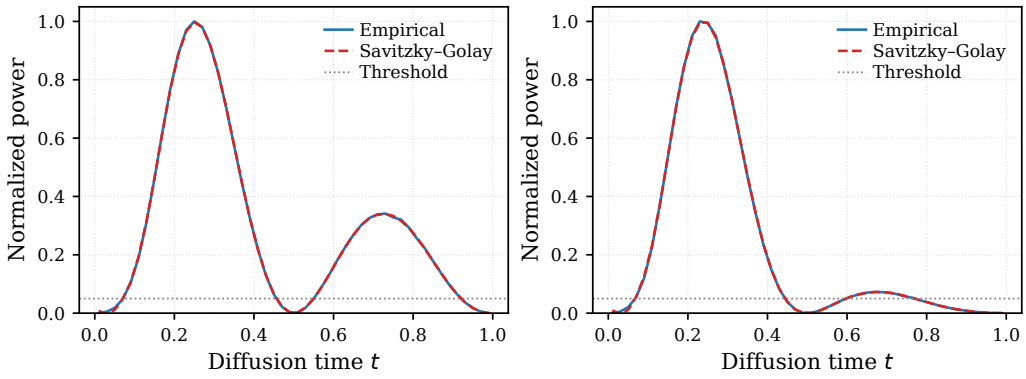

Figure 6: Power profile $P(t)$ estimated on the pre-trained watermarked model, normalized by its maximum. Left: MNIST. Right: CIFAR-10. Shaded region: the width of the sampling kernel around each detected peak. The profile has two local maxima whose locations match the envelope $a(t) = \sin(2\pi t)$.

C.5 Watermark detection

This appendix reports the detection accuracy of the peak-based query distribution against the uniform and fixed-mixture baselines, for MNIST and CIFAR-10, at fine-tuning steps from 50 to 500 on class 5. The embedding, the detector, and the sampling distributions are the ones defined in Section 5; the protocol is the one described in Appendix C.1.

Figure 6 shows the power profile $P(t)$ estimated on the pre-trained watermarked model. On both datasets the profile has two local maxima, one near $t \approx 0.25$ and one near $t \approx 0.7$, matching the envelope $a(t) = \sin(2\pi t)$ up to the modulation induced by the velocity field. The peak-based sampler places its queries at the locations of the two largest peaks, with weights proportional to the peak heights.

Figure 7 reports detection accuracy as a function of fine-tuning steps. The peak-based sampler retains higher detection accuracy than the uniform and fixed-mixture baselines over a wider range of fine-tuning steps. On MNIST it holds full accuracy through 200 steps while both baselines drop to chance by 200 steps; the advantage is a delay of roughly 100–200 steps in the collapse of the watermark. On CIFAR-10 all three methods track each other through 200 steps; the peak-based sampler separates at 300 steps (0.90 against 0.55 for the gaussian mixture and 0.35 for uniform) and retains its advantage through 400 steps, before all three fall to low accuracy at 500 steps. The gap on CIFAR-10 is narrower than on MNIST, consistent with the smaller height of the second power peak on that dataset.

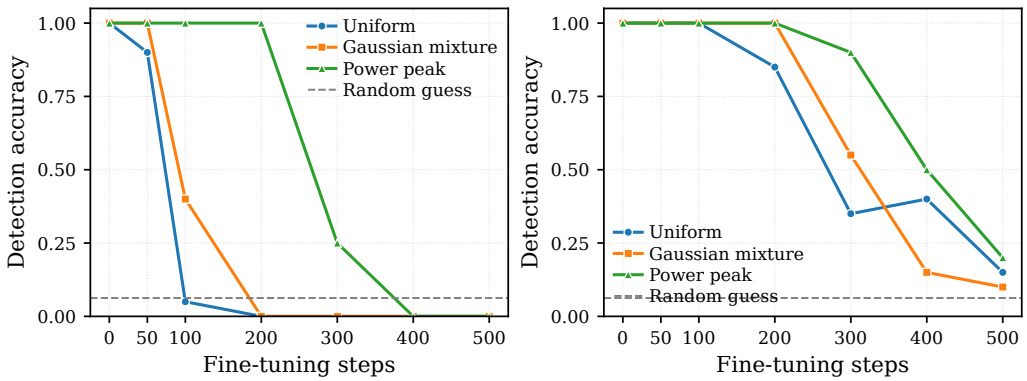

Figure 7: Detection accuracy as a function of fine-tuning steps. Left: MNIST, averaged over the messages. Right: CIFAR-10, averaged over the messages. Each curve is one sampling method. Values are means over messages and over the detection trials. The dashed line is chance level for a 4-bit codebook.

The result is the detection-level counterpart of the temporal structure of the fine-tuning perturbation reported in Appendix C.3: the same non-uniformity in flow time that makes the perturbation concentrate at certain t also makes the watermark signal concentrate at certain t, and the peak-based query distribution exploits this concentration.