Entropy-Controlled Flow Matching

Chika Maduabuchi[1]

William & Mary, Williamsburg, VA, USA

Abstract. Modern vision generators transport a base distribution to data through time-indexed measures, implemented as deterministic flows (ODEs) or stochastic diffusions (SDEs). Despite strong empirical performance, standard flow-matching objectives do not directly control the information geometry of the trajectory, allowing low-entropy bottlenecks that can transiently deplete semantic modes. We propose *Entropy-Controlled Flow Matching* (ECFM): a constrained variational principle over continuity-equation paths enforcing a global entropy-rate budget $\frac{d}{dt}\mathcal{H}(\mu_t) \geq -\lambda$. ECFM is a convex optimization in Wasserstein space with a KKT/Pontryagin system, and admits a stochastic-control representation equivalent to a Schrödinger bridge with an explicit entropy multiplier. In the pure transport regime, ECFM recovers entropic OT geodesics and Γ-converges to classical OT as $\lambda \to 0$. We further obtain certificate-style mode-coverage and density-floor guarantees with Lipschitz stability, and construct near-optimal collapse counterexamples for unconstrained flow matching.

Keywords: Optimal transport · Schrödinger bridges · Generative modeling

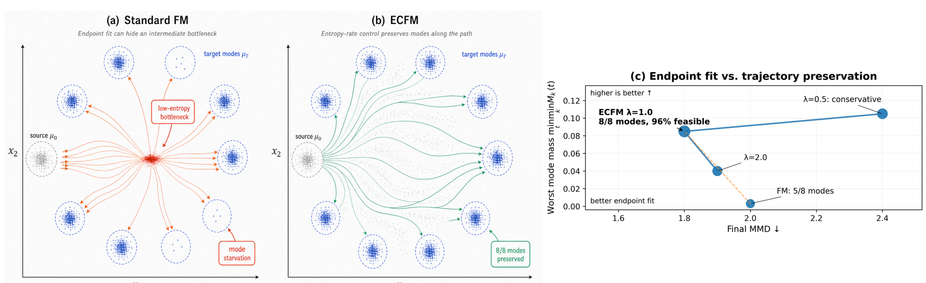

Fig. 1. Trajectory bottleneck mechanism. A controlled toy 8-Gaussian experiment illustrates that endpoint fit can hide harmful intermediate transport geometry. (a) Standard FM routes mass from the source distribution μ_0 through a transient low-entropy bottleneck, starving several target modes despite endpoint matching. (b) ECFM imposes an entropy-rate constraint, keeping trajectories distributed and preserving all target modes μ_T. (c) Quantitative diagnostics separate endpoint fidelity from trajectory preservation: ECFM with $\lambda = 1.0$ preserves all 8/8 modes and satisfies 96% of entropy-feasible time bins while maintaining better final MMD than unconstrained FM; smaller λ is more conservative.

1 Introduction

Modern vision generators transport a simple base distribution to a complex data distribution through a time-indexed family of measures $\{\mu_t\}_{t\in[0,T]}$, implemented either as a deterministic flow (ODE/probability-flow style) [4, 15, 28, 31, 41] or a stochastic diffusion (SDE) [9, 17, 20, 39, 42]. While these approaches achieve impressive sample quality, they lack a *constraint-level* mechanism that controls the *information geometry* of the entire trajectory—in particular, the evolution of entropy and the formation of low-entropy "bottlenecks" that can transiently deplete modes. The teaser experiment in Fig. 1 isolates this mechanism on a controlled 8-Gaussian task; the full protocol, diagnostic definitions, and λ sweep are reported in App. J.5 (Tables 1–2).

Such bottlenecks are a structural route to mode collapse (in the broad sense of intermediate-time mass depletion across semantic regions), a phenomenon historically associated with adversarial training [14] but also relevant for deterministic transport models when the learned velocity field permits strong compression, thereby reducing distributional coverage (low recall) and biasing samples toward a subset of modes [24].

Optimal transport as geometry, Schrödinger bridges as regularization. Optimal transport (OT) provides a canonical geometric notion of interpolation between distributions via Wasserstein geodesics [30], with the dynamic Benamou–Brenier formulation [2] giving a minimum-kinetic-energy transport under the continuity equation [2, 45]. However, OT geodesics can be brittle in high dimensions: they may route mass through narrow sets and are generally sensitive to perturbations, exhibiting fragility under sampling error and structural perturbations in practice [13, 27]. Entropic regularization replaces the hard OT problem by a strictly convex objective that yields smooth entropic interpolations [7]. At the path level, this regularization is captured by Schrödinger bridges (SB) [5, 8, 25, 26, 34]: KL projections of a reference path measure onto endpoint constraints, which admit a stochastic-control interpretation and connect deeply to OT through small-noise limits and large deviations [6,32]. Despite this conceptual alignment, the training objectives used in flow-based vision generators are typically posed as regression of velocities/scores [28, 31, 42] and do *not* explicitly impose an entropy-control principle that rules out collapse channels at the level of trajectories.

Our idea: entropy-controlled flow matching. We introduce *Entropy-Controlled Flow Matching (ECFM)*, a constrained variational principle for learning velocity fields in which the transport path μ_t must satisfy an *entropy-rate budget*

$$\frac{d}{dt}\mathcal{H}(\mu_t) \;\geq\; -\lambda, \tag{1}$$

where $\mathcal{H}$ is differential entropy (defined on the absolutely continuous class) and $\lambda \geq 0$ is a user-specified budget. In deterministic continuity-equation dynamics, the constraint has an immediate geometric meaning: since $\frac{d}{dt}\mathcal{H}(\mu_t) = \mathbb{E}_{\mu_t}[\nabla \cdot v(\cdot, t)]$, (1) directly limits average compressibility and prevents arbitrarily sharp concentration in finite time. We combine this constraint with a flow-matching-style objective that regresses a learned velocity field toward a reference field, yielding a principled mechanism that is *trajectory-level*, *model-agnostic* (ODE or SDE), and *certifiable* via entropy-rate diagnostics.

Contributions. Our main results provide a complete mathematical foundation for ECFM and connect it to OT/SB theory:

- **A constrained transport formulation.** We formalize ECFM as a constrained optimization in Wasserstein space over solutions of the continuity equation, with precise assumptions, admissible classes, and a boxed primal problem (Sec. 2.2).
- **Optimality and duality (KKT/Pontryagin form).** We derive first-order conditions with a nonnegative multiplier enforcing the entropy-rate constraint, and give a dual interpretation as a time-dependent barrier against compressive flows (Sec. 2.3).
- **Equivalence to Schrödinger bridges.** We prove that ECFM admits a Schrödinger-bridge-equivalent formulation: the induced objective is strictly convex in the path law (in the SB specialization), giving uniqueness at the level of trajectories (Sec. 2.5) [6, 32].
- **Entropic OT geodesics and the $\lambda \to 0$ limit.** For the pure transport specialization, we show that ECFM trajectories coincide with entropic OT interpolations (Sec. 2.6) and establish Γ-convergence to classical OT as $\lambda \downarrow 0$ (Sec. 2.7) [2, 7, 45].
- **Mode coverage and stability guarantees.** We introduce a formal collapse notion in terms of intermediate-time mode mass depletion and prove quantitative mode-floor and density-minimum bounds under the entropy budget, together with perturbation stability (Sec. 2.8–2.9).
- **Necessity: failure without entropy control.** We construct explicit unconstrained flow-matching sequences with near-optimal objective value but singular entropy bottlenecks and mode depletion, showing that the entropy constraint is structurally necessary for certificate-level non-collapse claims (Sec. 2.10).

Relevance to vision generators. ECFM complements the dominant paradigms for vision generation—diffusions/score models [17, 42], flow matching [28], and rectified flows [31]—by adding an explicit entropy-budget constraint that can be checked (or enforced via dual updates) during training, yielding theorem-backed anti-collapse guarantees without relying on SOTA comparisons.

Practical enforcement and certification. Training can be implemented with a projected primal–dual augmented-Lagrangian update: entropy-rate violations on a time grid increase nonnegative multipliers, while feasible bins incur no dual pressure. We give the full algorithm, estimator choices, and adaptive λ-scheduling details in App. J.3 and Alg. 1. Quantitative toy diagnostics and the λ sweep are reported in App. J.5.

Relation to Schrödinger bridges and entropic OT (what is new). Although Sec. 2.5 shows that ECFM admits a Schrödinger/KL representation under a Brownian reference law, this connection is *an analytical lens rather than the definition of our method*. Classical Schrödinger bridges (and entropic OT) [6, 8, 36, 40, 44] are posed as KL/entropy-*penalized* interpolations with a fixed stochastic reference, whereas ECFM is defined as a *flow-matching projection* onto an *entropy-rate feasible set* via the inequality constraint $\dot{\mathcal{H}}(\mu_t) \geq -\lambda$. Crucially, ECFM keeps an arbitrary FM reference drift $u^\star$ and yields the closest entropy-feasible trajectory; the KKT multiplier is *adaptive* (zero when the FM trajectory is feasible, active otherwise), inducing an *endogenous* entropic level $\varepsilon(\lambda)$ rather than choosing ε a priori. The coincidence with entropic OT/SB geodesics occurs only in the pure-transport specialization $u^\star \equiv 0$ (Sec. 2.6), serving as a sanity check and enabling the $\lambda \downarrow 0$ OT limit.

Roadmap. Sec. 2.2–2.3 presents the ECFM problem, duality, and optimality conditions. Sec. 2.5 proves equivalence to Schrödinger bridges. Sec. 2.6 and Sec. 2.7 establish convergence to entropic OT and Γ-limits to classical OT. Sec. 2.8–2.10 provide mode-coverage, stability, and failure-without-constraint results. Full proofs and measure-theoretic details are deferred to the Supplementary Material.

2 Theory

2.1 Setup and notation

Ambient space. Fix a horizon $T > 0$ and dimension $d \geq 1$. We work on $\mathbb{R}^d$ with its Borel σ-algebra. Let $\mathcal{P}_2(\mathbb{R}^d) := \{\mu \in \mathcal{P}(\mathbb{R}^d) : \int \|x\|^2 \, d\mu(x) < \infty\}$ and let W_2 denote the quadratic Wasserstein distance,

$$W_2^2(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^2 \, d\pi(x, y).$$

For measurable $T : \mathbb{R}^d \to \mathbb{R}^d$, $T_\# \mu(B) := \mu(T^{-1}(B))$, and $\langle f, \mu \rangle := \int f \, d\mu$. Additional OT preliminaries are in App. A.

Time-indexed measures and velocities. We consider Borel curves $t \mapsto \mu_t \in \mathcal{P}_2(\mathbb{R}^d)$ paired with Borel velocity fields $v : \mathbb{R}^d \times [0, T] \to \mathbb{R}^d$, and use the shorthand

$$\|v\|^2_{L^2(\mu)} := \int_0^T \int_{\mathbb{R}^d} \|v(x, t)\|^2 \, d\mu_t(x) \, dt.$$

Continuity equation. A pair (μ, v) satisfies the continuity equation (CE) on $[0, T]$ if for all $\varphi \in C_c^\infty(\mathbb{R}^d \times (0, T))$,

$$\int_0^T \int_{\mathbb{R}^d} \Big(\partial_t \varphi(x, t) + \nabla \varphi(x, t) \cdot v(x, t) \Big) \, d\mu_t(x) \, dt = 0. \tag{2}$$

When $\mu_t = \rho_t \, dx$, this is $\partial_t \rho_t + \nabla \cdot (\rho_t v_t) = 0$ in the distributional sense.

Entropy and information functionals. For $\mu = \rho\, dx$, define the differential entropy $\mathcal{H}(\mu) := -\int \rho \log \rho\, dx$ (and $\mathcal{H}(\mu) = -\infty$ if $\mu \not\ll dx$). For $\mu \ll \nu$, $\mathrm{KL}(\mu\|\nu) := \int \log\big(\frac{d\mu}{d\nu}\big)\, d\mu$. If $\rho > 0$ a.e. and ρ is weakly differentiable, the Fisher information is $\mathcal{I}(\mu) := \int \|\nabla \log \rho\|^2 \rho\, dx$. A key identity used throughout is the entropy-rate formula: if (μ, v) solves CE and ρ_t is regular enough, then

$$\frac{d}{dt}\mathcal{H}(\mu_t) = \int_{\mathbb{R}^d} \nabla\cdot v(x,t)\, d\mu_t(x) \qquad \text{for a.e. } t \in [0,T], \tag{3}$$

with precise hypotheses in Apps. A–B.

Admissible endpoints and paths. We consider endpoints $(\mu_0, \mu_T) \in \mathcal{P}_2(\mathbb{R}^d) \times \mathcal{P}_2(\mathbb{R}^d)$ and define

$$\mathcal{A}(\mu_0,\mu_T) := \Big\{(\mu, v):\ \mu_{|0} = \mu_0,\ \mu_{|T} = \mu_T,\ (\mu,v) \text{ satisfies (2)},\ \|v\|_{L^2(\mu)} < \infty\Big\}. \tag{4}$$

Entropy-rate budget (entropy control). Fix $\lambda \geq 0$. We impose the a.e. constraint

$$\frac{d}{dt}\mathcal{H}(\mu_t) \geq -\lambda \qquad \text{for a.e. } t \in [0,T], \tag{5}$$

interpreted under the regularity conditions of App. B.

Standing assumptions. We state assumptions in a model-agnostic form; a brief "how to read these in diffusion/FM/latent settings" mapping is given in App. J. Unless stated otherwise:

(S1) **Endpoints.** $\mu_0 = \rho_0 dx,\ \mu_T = \rho_T dx$ with $\mu_0, \mu_T \in \mathcal{P}_2(\mathbb{R}^d)$ and $\mathcal{H}(\mu_0), \mathcal{H}(\mu_T) > -\infty$.

(S2) **Feasibility.** $\mathcal{A}(\mu_0, \mu_T)$ contains at least one path satisfying (5).

(S3) **Reference field.** A measurable drift $u^\star : \mathbb{R}^d \times [0,T] \to \mathbb{R}^d$ is given such that $\int_0^T\!\int \|u^\star(x,t)\|^2\, d\mu_t(x)\, dt < \infty$ for all considered admissible (μ, v), and $u^\star(\cdot, t)$ is locally Lipschitz for a.e. t.

Full technical conditions and variants are in Apps. A–B.

2.2 Entropy-Controlled Flow Matching (ECFM)

Reference drift. Let $u^\star : \mathbb{R}^d \times [0,T] \to \mathbb{R}^d$ be a prescribed reference drift (teacher / closed-form interpolation / model-implied target). ECFM learns a velocity field v and an induced path $(\mu_t)_{t\in[0,T]}$ that transports μ_0 to μ_T, stays close to $u^\star$ in mean square under μ_t, and respects a global entropy-rate budget.

Primal formulation. Define the entropy-feasible admissible set

$$\mathcal{A}_\lambda(\mu_0,\mu_T) := \Big\{(\mu,v)\in\mathcal{A}(\mu_0,\mu_T):\ \dot{\mathcal{H}}(\mu_t)\ge -\lambda \text{ for a.e. } t\in[0,T]\Big\},$$

where $\dot{\mathcal{H}}$ is interpreted via (3) under Apps. A–B. The ECFM problem is

$$\boxed{\mathrm{ECFM}_\lambda(\mu_0,\mu_T;u^\star) := \min_{(\mu,v)\in\mathcal{A}_\lambda(\mu_0,\mu_T)} \frac{1}{2}\int_0^T\int_{\mathbb{R}^d} \|v(x,t)-u^\star(x,t)\|^2\, d\mu_t(x)\, dt.} \tag{6}$$

We denote by (μ^λ, v^λ) an optimal pair when it exists.

Feasibility and choosing λ (practical rule). The budget λ controls the allowed entropy dissipation along the path. Trivially, $\lambda=+\infty$ recovers unconstrained FM and is always feasible. For finite λ, feasibility can be checked (and λ selected) using the entropy-rate diagnostics in App. J.2: on a discrete time grid $\{t_n\}$, estimate $\widehat{\dot{\mathcal{H}}}_n \approx \mathbb{E}_{\mu_{t_n}}[\nabla\cdot v(\cdot,t_n)]$ (for deterministic flows) or the Fokker–Planck/Fisher form (for diffusions), then set an *empirical effective budget*

$$\widehat{\lambda}_{\mathrm{eff}}^{\mathrm{LCB}} \ :=\ \max_n \big(-\widehat{\dot{\mathcal{H}}}_n^{\mathrm{LCB}}\big),$$

where $\widehat{\dot{\mathcal{H}}}_n^{\mathrm{LCB}}$ is a lower confidence bound (Sec. 1, App. J.4). Choosing $\lambda\ge\widehat{\lambda}_{\mathrm{eff}}^{\mathrm{LCB}}$ makes feasibility a measurable, certificate-compatible condition.

Flux form (Benamou–Brenier variables). When $\mu_t=\rho_t dx$, define momentum $m_t:=\rho_t v_t$. Then CE is $\partial_t\rho_t+\nabla\cdot m_t=0$ and

$$\frac{1}{2}\int_0^T\int \|v-u_t^\star\|^2\, d\mu_t dt = \frac{1}{2}\int_0^T\int \frac{\|m_t-\rho_t u_t^\star\|^2}{\rho_t}\, dx\, dt,$$

(with the convention $+\infty$ if $\rho_t=0$ and $m_t\neq 0$). We use this representation for convexity/coercivity arguments (App. C).

Entropy-budget interpretation. By (3), the constraint $\dot{\mathcal{H}}(\mu_t)\ge-\lambda$ is equivalently the divergence budget

$$\int_{\mathbb{R}^d} \nabla\cdot v(x,t)\, d\mu_t(x)\ \ge\ -\lambda \qquad \text{for a.e. } t\in[0,T]. \tag{7}$$

Thus ECFM forbids paths with transient low-entropy bottlenecks (large negative divergence spikes), the collapse mechanism made explicit in App. I.

Special cases.

1. **Unconstrained FM:** $\lambda=+\infty$ removes (7) and recovers classical flow matching.
2. **Pure transport:** $u^\star\equiv 0$ yields minimum-energy transport under an entropy-rate constraint, leading to entropic OT geodesics (App. E).
3. **Diffusions:** for Fokker–Planck dynamics, $\dot{\mathcal{H}}$ includes a Fisher-information term, enlarging feasibility; see Sec. J and App. B.

2.3 Optimality and duality

KKT structure (primal–dual form). ECFM admits a primal–dual characterization with a scalar potential ϕ (for CE) and a nonnegative multiplier $\eta(t)$ (for the entropy-rate constraint). We state the core KKT conditions; full derivations and the adjoint equation are in Apps. C.4–C.6.

Theorem 1 (KKT optimality system (core conditions)). *Assume* (μ^λ, v^λ) *solves* (6) *and* $\mu_t^\lambda = \rho_t^\lambda dx$ *satisfies the regularity conditions in Apps. A–C. Then there exist* ϕ^λ *and* $\eta^\lambda(t) \geq 0$ *such that:*

1. ***(Feasibility).*** $(\mu^\lambda, v^\lambda) \in \mathcal{A}_\lambda(\mu_0, \mu_T)$*, i.e. CE* (2) *holds with endpoints* μ_0, μ_T *and* $\dot{\mathcal{H}}(\mu_t^\lambda) \geq -\lambda$ *a.e.*
2. ***(Stationarity).*** *For a.e.* t *and* μ_t^λ*-a.e.* x*,*
$$v^\lambda(x,t) = u^\star(x,t) - \nabla\phi^\lambda(x,t) - \eta^\lambda(t)\,\nabla\log\rho_t^\lambda(x). \tag{8}$$
3. ***(Complementarity).*** *For a.e.* t*,*
$$\eta^\lambda(t) \geq 0, \qquad \dot{\mathcal{H}}(\mu_t^\lambda) + \lambda \geq 0, \qquad \eta^\lambda(t)\big(\dot{\mathcal{H}}(\mu_t^\lambda) + \lambda\big) = 0. \tag{9}$$

Dual viewpoint. In flux variables $m_t = \rho_t v_t$, the problem is a convex program in (ρ, m) with a convex quadratic perspective integrand and linear CE constraint; the associated dual is a concave maximization over (ϕ, η) whose maximizers recover v^λ via (8) (App. C.5).

Interpretation. $\eta^\lambda(t)$ acts as an *adaptive anti-collapse pressure*: it activates only when the trajectory attempts to dissipate entropy faster than λ, injecting a score-like correction $\nabla\log\rho_t^\lambda$ in (8) and thereby ruling out low-entropy bottlenecks (Apps. G, I).

2.4 Existence and uniqueness

Existence.

Theorem 2 (Existence of an ECFM minimizer). *Assume (S1)–(S3) from Sec. 2.1. Then the primal problem* (6) *admits at least one minimizer* $(\mu^\lambda, v^\lambda) \in \mathcal{A}_\lambda(\mu_0, \mu_T)$*. Moreover, one may choose a minimizer with* $\mu_t^\lambda \ll dx$ *for a.e.* t *and finite kinetic energy* $\int_0^T \|v_t^\lambda\|^2_{L^2(\mu_t^\lambda)}\,dt < \infty$*. Full proof and compactness topology are in App. C.*

Uniqueness (strict convex regimes). Uniqueness holds in settings where the induced objective is strictly convex in the (path) law, notably: (i) the Schrödinger/KL representation (Sec. 2.5) and (ii) the pure transport specialization $u^\star \equiv 0$ (App. E).

Theorem 3 (Uniqueness in strict convex regimes). *In either of the following regimes, the ECFM solution is unique:*

1. ***Schrödinger/KL form:*** *ECFM is represented as a KL minimization over path measures (Sec. 2.5);*
2. ***Pure transport:*** $u^\star \equiv 0$*, yielding entropic OT geodesics (App. E).*

In both cases, the path $t \mapsto \mu_t^\lambda$ *is unique and* $v^\lambda(\cdot, t)$ *is unique* μ_t^λ*-a.e. for a.e.* t*.*

Proof (Proof sketch). Both regimes reduce to a strictly convex objective in the path law (KL/entropic OT), hence uniqueness follows by strict convexity. Proofs are in App. D (Schrödinger/KL) and App. E.4 (geodesic case).

Scope and what is certified. Theorem 2 guarantees existence under (S1)–(S3). Uniqueness is claimed only in the strict convex regimes of Theorem 3 (SB/KL form or pure transport); in the general ECFM setting with arbitrary $u^\star$, multiple minimizers may exist. Importantly, *certification is trajectory-level rather than uniqueness-level*: every ECFM minimizer is feasible for the entropy-rate constraint by construction, and therefore all mode-floor and stability guarantees apply to *any* minimizer satisfying the stated regularity assumptions. In practice, the SB specialization (Sec. 2.5) recovers uniqueness at the level of the induced path law.

2.5 Equivalence to Schrödinger bridges

Schrödinger bridge (dynamic form). Let $\mathbf{R}$ be a Brownian reference path law on $C([0,T];\mathbb{R}^d)$ with diffusivity $\varepsilon > 0$ (App. D). The (dynamic) Schrödinger bridge between μ_0 and μ_T is the KL projection

$$\min_{\mathbf{P}:\ \mathbf{P}_0=\mu_0,\ \mathbf{P}_T=\mu_T} \mathrm{KL}\big(\mathbf{P} \,\|\, \mathbf{R}\big), \tag{10}$$

which is equivalent to the static endpoint formulation (App. D.1–D.2).

Control representation and current velocity. Any $\mathbf{P} \ll \mathbf{R}$ admits a controlled diffusion representation $dX_t = b_t(X_t)\,dt + \sqrt{2\varepsilon}\,dW_t$ whose marginals $\mu_t = \rho_t dx$ satisfy the Fokker–Planck equation. Define the *current velocity*

$$v_t := b_t - \varepsilon \nabla \log \rho_t, \tag{11}$$

which converts Fokker–Planck to the continuity equation $\partial_t \rho_t + \nabla\cdot(\rho_t v_t) = 0$. Moreover, the entropy-rate identity becomes

$$\frac{d}{dt}\mathcal{H}(\mu_t) = \mathbb{E}_{\mu_t}[\nabla\cdot b_t] \;+\; \varepsilon\,\mathcal{I}(\mu_t) = \mathbb{E}_{\mu_t}[\nabla\cdot v_t], \tag{12}$$

so the constraint $\dot{\mathcal{H}}(\mu_t) \geq -\lambda$ directly restricts admissible KL-controls.

Theorem 4 (Equivalence: ECFM as an entropy-controlled Schrödinger bridge). *Fix* $\varepsilon > 0$ *and* $\mathbf{R}$ *as above. Under (S1)–(S3) and the regularity assumptions of App. D, ECFM admits a Schrödinger-bridge representation in the following sense:*

1. ***(SB ⇒ ECFM).*** *If* $\mathbf{P}^\star$ *solves* (10) *with drift* b_t *and marginals* μ_t*, then the induced current velocity* v_t *in* (11) *yields a CE path* (μ, v) *that minimizes an ECFM objective* (6) *for an explicit reference field* $u^\star$ *determined by Schrödinger potentials, with multiplier* $\eta(t)$ *enforcing the entropy budget.*
2. ***(ECFM ⇒ SB).*** *Conversely, any ECFM minimizer* (μ^λ, v^λ) *induces a controlled drift* $b^\lambda := v^\lambda + \varepsilon\nabla\log\rho^\lambda$ *and hence a path law* $\mathbf{P}^\lambda \ll \mathbf{R}$ *whose KL objective equals the ECFM value up to a constant.*
3. ***(Uniqueness in path law).*** *The induced path law* $\mathbf{P}^\lambda$ *is unique whenever the admissible set is nonempty.*

Proof (Proof sketch). Combine the standard SB $\leftrightarrow$ stochastic control equivalence with the current-velocity transformation (11). Completing the square identifies the quadratic control cost with the ECFM integrand, while (12) links entropy-rate feasibility to admissible controls. Full proof and the explicit identification of $u^\star$ are in App. D.4.

Implication. ECFM is a Schrödinger bridge with an explicit entropy-dissipation budget: it inherits strict convexity (hence uniqueness in path law) in the SB specialization, and excludes the low-entropy bottleneck collapse channels of unconstrained deterministic flows (Apps. G, I).

2.6 Convergence to entropic OT geodesics

Pure transport specialization. For $u^\star \equiv 0$, ECFM reduces to minimum kinetic energy under the entropy-rate budget:

$$\min_{(\mu,v)\in\mathcal{A}_\lambda(\mu_0,\mu_T)} \frac{1}{2}\int_0^T\int \|v(x,t)\|^2\, d\mu_t(x)\, dt. \tag{13}$$

This defines an *entropy-controlled geodesic*. Via Sec. 2.5, this regime coincides with entropic OT / Schrödinger bridges.

Entropic OT. Fix $\varepsilon > 0$. Entropic OT admits a static regularized formulation and an equivalent dynamic (Schrödinger) KL formulation; we use the latter in the proofs (Apps. D–E).

Theorem 5 (Entropy-controlled geodesics are entropic OT geodesics). *Assume (S1)–(S2) and* $u^\star \equiv 0$*. Let* (μ^λ, v^λ) *minimize* (13)*. Then:*

1. ***(Identification).*** *There exists* $\varepsilon = \varepsilon(\lambda) > 0$ *and a Schrödinger bridge* $\mathbf{P}^\varepsilon$ *between* μ_0 *and* μ_T *such that* $\mu_t^\lambda = \mathbf{P}_t^\varepsilon$ *for a.e.* t*. Equivalently,* (μ^λ, v^λ) *is the entropic OT geodesic (current-velocity form) with regularization* $\varepsilon(\lambda)$*.*
2. ***(Uniqueness and stability).*** *The path* $t \mapsto \mu_t^\lambda$ *is unique, and there exists* $C = C(\lambda, T, \mu_0, \mu_T)$ *such that for any endpoints* $(\tilde\mu_0, \tilde\mu_T)$ *and corresponding minimizer* $(\tilde\mu^\lambda, \tilde v^\lambda)$,

$$\sup_{t\in[0,T]} W_2(\mu_t^\lambda, \tilde\mu_t^\lambda) \;\le\; C\big(W_2(\mu_0,\tilde\mu_0) + W_2(\mu_T,\tilde\mu_T)\big). \tag{14}$$

Proof (Proof sketch). Use the Schrödinger representation (Sec. 2.5) to identify (13) with a strictly convex entropic OT/KL functional. The entropy-rate constraint induces an effective regularization level $\varepsilon(\lambda)$ through the KKT multiplier (App. E). Uniqueness follows from strict convexity in the path law, and stability follows from strong convexity and stability of KL projections. Full proof is in App. E.

Role of λ. Smaller λ permits less entropy dissipation and corresponds to smaller $\varepsilon(\lambda)$, approaching classical OT as $\lambda \to 0$ (Sec. 2.7).

2.7 Γ-convergence as $\lambda \to 0$

Limit regime. We formalize the small-budget limit $\lambda \downarrow 0$ by Γ-convergence: entropy-controlled geodesics converge to classical OT (Benamou–Brenier) geodesics.

Functional setting. Work in the trajectory–velocity space $\mathsf{X} := \mathcal{A}(\mu_0, \mu_T)$ endowed with the standard weak topology for CE trajectories (App. F.1). For $u^\star \equiv 0$, define

$$\mathcal{F}_\lambda(\mu, v) := \begin{cases} \dfrac{1}{2} \displaystyle\int_0^T \int \|v(x,t)\|^2 \, d\mu_t(x) \, dt, & (\mu, v) \in \mathcal{A}_\lambda(\mu_0, \mu_T), \\ +\infty, & \text{otherwise}, \end{cases} \tag{15}$$

and let $\mathcal{F}_0$ be the classical Benamou–Brenier OT action (same expression on $\mathcal{A}(\mu_0, \mu_T)$; $+\infty$ otherwise).

Theorem 6 (Γ-convergence to classical OT). *Assume (S1)–(S2) and $u^\star \equiv 0$. Then $\mathcal{F}_\lambda$ Γ-converges to $\mathcal{F}_0$ on X as $\lambda \downarrow 0$. Consequently, $\inf \mathcal{F}_\lambda \to \inf \mathcal{F}_0$, and any accumulation point of minimizers (μ^λ, v^λ) is a Benamou–Brenier OT minimizer. If the OT geodesic is unique, then $\mu_t^\lambda \rightharpoonup \mu_t^0$ for all $t \in [0, T]$.*

Proof (Proof sketch). The liminf follows from weak lower semicontinuity of the kinetic action in flux form and closure of the CE constraint under the X-topology. A recovery sequence is obtained by smoothing an OT geodesic (strictly positive mollification) so the entropy-rate constraint is feasible for $\lambda_n \downarrow 0$ while the action changes by $o(1)$. Equicoercivity yields compactness and the fundamental theorem of Γ-convergence gives convergence of minimizers. Full proof is in App. F.

Consequence. ECFM provides a regularized geodesic family that recovers classical OT as $\lambda \to 0$, linking entropy control to Wasserstein geometry.

2.8 Mode coverage and anti-collapse guarantees

Modes and collapse. Let $\{A_k\}_{k=1}^K$ be measurable sets representing semantic modes (e.g., regions in an embedding space) and define $M_k(t) := \mu_t(A_k)$. Assume nontrivial endpoint mass:

$$\mu_0(A_k) \geq \alpha_k, \qquad \mu_T(A_k) \geq \alpha_k, \qquad k = 1, \dots, K, \tag{16}$$

for some $\alpha_k > 0$. We call (μ_t) *mode-collapsing* if $\inf_t \min_k \mu_t(A_k) = 0$.

Choosing modes $\{A_k\}$ in vision practice. Theorem 7 is stated for arbitrary measurable mode sets. In vision applications, a concrete and reproducible choice is to define modes in a fixed representation space: pick an embedding $f(\cdot)$ (e.g., the model's own latent space, or a frozen pretrained encoder), construct regions $\{C_k\}_{k=1}^K \subset \mathbb{R}^p$ (e.g., class-conditional regions, k-means clusters, or attribute partitions), and set $A_k := f^{-1}(C_k)$. Then $\mu_t(A_k)$ measures the probability mass retained in each semantic/cluster region along the generative trajectory. When the generator is trained in latent space, f is naturally the latent map and A_k should be defined in that same space.

Entropy barrier. Entropy-rate control rules out transient concentration: under $\dot{\mathcal{H}} \geq -\lambda$, the path cannot pass through arbitrarily low-entropy bottlenecks in finite time (App. G.2), which yields quantitative mode floors.

Theorem 7 (Mode coverage under entropy-rate control). *Assume (S1)–(S2) and let (μ^λ, v^λ) solve (6). Under the regularity conditions of App. G, there exist constants $\beta_k = \beta_k(\alpha_k, \lambda, T, \mu_0, \mu_T) > 0$ such that*

$$\mu_t^\lambda(A_k) \;\geq\; \beta_k \qquad \textit{for all } t \in [0,T] \textit{ and all } k = 1, \dots, K. \tag{17}$$

Proof (Proof sketch). Combine the entropy barrier (App. G.2) with a contradiction/compactness argument that transfers forbidden concentration into a uniform lower bound on set-masses $M_k(t)$; see App. G.3 for explicit constants.

Density floors on mode cores. If each mode contains a compact core $K_k \Subset A_k$ with endpoint density lower bounds $\rho_0, \rho_T \geq c_k > 0$ a.e. on K_k, then entropy control yields a uniform interior density floor.

Theorem 8 (Density floors on mode cores). *Assume (S1)–(S2) and the core condition above. Under App. G.4, there exists $\underline{\rho}_k > 0$ such that*

$$\rho_t^\lambda(x) \;\geq\; \underline{\rho}_k \qquad \textit{for a.e. } x \in K_k \textit{ and all } t \in [0,T]. \tag{18}$$

Moreover, these floors are perturbation-robust via Sec. 2.9.

Proof (Proof sketch). App. G.4 combines entropy-rate control with localization/regularity (Harnack-type) arguments to convert mode-mass floors into pointwise density minima on K_k.

Contrast. Without the entropy-rate constraint, unconstrained FM admits near-optimal collapsing paths (App. I); see Sec. 2.10.

2.9 Stability under perturbations

Perturbation model. Let (μ, v) be an ECFM solution (or any feasible CE trajectory), and consider perturbed data consisting of: (i) perturbed endpoints

$(\tilde{\mu}_0, \tilde{\mu}_T)$, and (ii) a perturbed reference field $\tilde{u}^\star$. Let $(\tilde{\mu}, \tilde{v})$ denote the corresponding ECFM solution under the same budget λ, when it exists. We quantify the size of perturbations by

$$\Delta_0 := W_2(\mu_0, \tilde{\mu}_0), \qquad \Delta_T := W_2(\mu_T, \tilde{\mu}_T), \qquad \Delta_u := \Big(\int_0^T \int \|u^\star - \tilde{u}^\star\|^2 \, d\mu_t \, dt \Big)^{1/2}. \tag{19}$$

We also allow additive noise in the learned velocity, modeled as

$$\tilde{v}(x,t) = v(x,t) + \xi(x,t), \qquad \int_0^T \int \|\xi(x,t)\|^2 \, d\mu_t(x) \, dt < \infty. \tag{20}$$

Stability of CE trajectories. The following theorem summarizes the Lipschitz-type stability of entropy-controlled trajectories with respect to endpoints and velocity perturbations; it consolidates Apps. H.1–H.4 into a main-text statement.

Theorem 9 (Unified perturbation stability). *Assume (S1)–(S3) and the regularity conditions of App. H. Let (μ, v) and $(\tilde{\mu}, \tilde{v})$ be ECFM minimizers for data $(\mu_0, \mu_T, u^\star)$ and $(\tilde{\mu}_0, \tilde{\mu}_T, \tilde{u}^\star)$, respectively, under the same budget λ. Then there exists a constant $C = C(\lambda, T, \mathrm{Reg}) < \infty$ depending only on the stated regularity bounds such that*

$$\sup_{t \in [0,T]} W_2(\mu_t, \tilde{\mu}_t) \;\le\; C\Big(\Delta_0 + \Delta_T + \Delta_u + \|\xi\|_{L^2(\mu)} \Big), \tag{21}$$

where $\|\xi\|^2_{L^2(\mu)} := \int_0^T \int \|\xi\|^2 \, d\mu_t \, dt$. Moreover, mode masses and core density floors are stable: for any measurable $A \subset \mathbb{R}^d$ and any compact core K,

$$\big|\mu_t(A) - \tilde{\mu}_t(A)\big| \;\le\; C_A \, W_2(\mu_t, \tilde{\mu}_t), \tag{22}$$

and if $\rho_t(x) \ge \underline{\rho}$ a.e. on K for all t, then

$$\tilde{\rho}_t(x) \;\ge\; \underline{\rho} - C_\rho\Big(\Delta_0 + \Delta_T + \Delta_u + \|\xi\|_{L^2(\mu)} \Big) \quad \text{for a.e. } x \in K, \; \forall t \in [0,T], \tag{23}$$

for constants C_A, C_ρ depending on A, K and the same regularity envelope.

Proof (Proof sketch). App. H.1 provides Lipschitz stability for CE trajectories under perturbations of the velocity field in $L^2(\mu)$, while Apps. H.2–H.3 control the propagation of endpoint/initialization perturbations through the flow map. Combining these yields (21). The mode-mass bound (22) follows from regularity of indicator test sets under Wasserstein perturbations (approximating $\mathbf{1}_A$ by Lipschitz functions), and the density-floor stability (23) follows by composing the core lower bound of Theorem 8 with the perturbation bound (21). Full proofs and explicit constants are in App. H.

Interpretation. ECFM provides not only non-collapse guarantees, but also *robustness margins*: mode floors and density minima degrade at most linearly with deployment shifts in endpoints, target field, or learned velocity.

2.10 Failure without entropy control

Why a constraint is necessary. The previous sections establish that the entropy-rate budget prevents low-entropy bottlenecks and yields mode-mass/density floors. We now show that *without* entropy control, classical flow matching admits near-optimal trajectories that transiently concentrate mass, producing intermediate-time mode depletion.

Unconstrained FM objective. Consider the unconstrained counterpart of (6) obtained by removing (5):

$$\mathrm{FM}(\mu_0, \mu_T; u^\star) := \min_{(\mu,v)\in\mathcal{A}(\mu_0,\mu_T)} \frac{1}{2}\int_0^T \int \|v(x,t) - u^\star(x,t)\|^2 \, d\mu_t(x)\, dt. \quad (24)$$

Theorem 10 (Singular bottlenecks and explicit collapse in unconstrained FM). *There exist dimensions $d \geq 1$, endpoint measures (μ_0, μ_T) satisfying (S1), a reference field $u^\star$, and a sequence of feasible solutions $\{(\mu^n, v^n)\}_{n\geq 1} \subset \mathcal{A}(\mu_0, \mu_T)$ such that:*

1. ***(Near-optimality).*** *$\mathcal{J}_\infty(\mu^n, v^n) \downarrow \inf \mathrm{FM}(\mu_0, \mu_T; u^\star)$, where $\mathcal{J}_\infty$ denotes the objective in (6) without the entropy constraint.*
2. ***(Entropy collapse).*** *There exist times $t_n \in (0, T)$ with*

$$\mathcal{H}(\mu^n_{t_n}) \to -\infty, \qquad \textit{equivalently} \qquad \int \nabla \cdot v^n(\cdot, t_n)\, d\mu^n_{t_n} \to -\infty. \quad (25)$$

3. ***(Mode depletion).*** *For a suitable two-mode partition $\{A_1, A_2\}$ with $\mu_0(A_k), \mu_T(A_k) > 0$, the same sequence satisfies*

$$\inf_{t\in[0,T]} \min\{\mu^n_t(A_1), \mu^n_t(A_2)\} \;\to\; 0. \quad (26)$$

Consequently, unconstrained FM does not admit a uniform mode-coverage guarantee of the form (17) without additional assumptions beyond (S1)–(S3).

Proof (Proof sketch). App. I constructs an explicit “squeezing” family of CE trajectories that routes mass through a shrinking corridor (or collapsing mixture component), keeping the FM mismatch cost small while making the Jacobian determinant (hence divergence) very negative over a short time window. This forces $\mathcal{H}(\mu_t)$ to drop to $-\infty$ as in (25), and simultaneously starves one mode as in (26). Full construction and estimates are given in App. I.

Direct comparison with ECFM. Theorem 7 shows that ECFM forbids the collapse mechanism in Theorem 10 by enforcing $\dot{\mathcal{H}}(\mu_t) \geq -\lambda$, which rules out the divergence spikes responsible for (25). This separation is *structural* (constraint-level), not empirical.

2.11 Connections to vision generative models

Unified transport viewpoint. Vision generators induce time-indexed marginals $(\mu_t)_{t\in[0,T]}$ in pixel or latent space via either deterministic transport (CE) or stochastic diffusion (FP). ECFM supplies a single constraint-level control—an entropy-rate budget—that is compatible with both regimes and yields certificate-style anti-collapse guarantees.

Deterministic flows (FM / rectified flow). For $\dot{X}_t = v_\theta(X_t, t)$ with $X_0 \sim \mu_0$, the law solves $\partial_t \mu_t + \nabla\cdot(\mu_t v_\theta) = 0$. Under App. B, the entropy rate is

$$\frac{d}{dt}\mathcal{H}(\mu_t) = \mathbb{E}_{\mu_t}\big[\nabla\cdot v_\theta(\cdot, t)\big], \tag{27}$$

so $\dot{\mathcal{H}}(\mu_t) \geq -\lambda$ is an explicit restriction on expected divergence, excluding transient compressions that drive mode depletion (Sec. 2.10).

Diffusions (FP / Fisher information). For $dX_t = b_\theta(X_t, t)\,dt + \sqrt{2\varepsilon(t)}\,dW_t$, ρ_t solves FP and (App. B.4)

$$\frac{d}{dt}\mathcal{H}(\mu_t) = \mathbb{E}_{\mu_t}\big[\nabla\cdot b_\theta(\cdot, t)\big] + \varepsilon(t)\,\mathcal{I}(\mu_t). \tag{28}$$

ECFM imposes $\dot{\mathcal{H}} \geq -\lambda$ directly (or via current velocity; Sec. 2.5), and the Fisher term provides an intrinsic anti-collapse contribution.

Certificate-style diagnostics. One can estimate $\dot{\mathcal{H}}(\mu_t)$ from minibatch trajectories (App. J.2):

$$\widehat{\dot{\mathcal{H}}}_{\mathrm{det}}(t) = \frac{1}{B}\sum_{i=1}^{B} \nabla\cdot v_\theta(x_i, t), \qquad \widehat{\dot{\mathcal{H}}}_{\mathrm{diff}}(t) = \frac{1}{B}\sum_{i=1}^{B} \nabla\cdot b_\theta(x_i, t) + \varepsilon(t)\,\frac{1}{B}\sum_{i=1}^{B} \|s_\theta(x_i, t)\|^2, \tag{29}$$

where $s_\theta \approx \nabla \log \rho_t$. Feasibility $\widehat{\dot{\mathcal{H}}}(t) \geq -\lambda$ yields a direct, SOTA-free certificate pathway to Theorems 7–8. Removing entropy control admits near-optimal collapse channels (Sec. 2.10).

3 Conclusion

ECFM augments flow matching with an explicit entropy-rate budget, turning trajectory compressibility into a controllable constraint. We characterize the resulting convex problem (KKT/Pontryagin), relate it to Schrödinger bridges, and in the transport specialization recover entropic OT and Γ-convergence to classical OT as $\lambda \to 0$. Entropy control yields verifiable mode-coverage and density-floor certificates (stable to perturbations), while unconstrained FM admits near-optimal collapsing paths.

Acknowledgements

The authors thank NSF and NCSA for computational support.

References

1. Bakry, D., Gentil, I., Ledoux, M.: Logarithmic Sobolev Inequalities, pp. 235–275. Springer International Publishing, Cham (2014). https://doi.org/10.1007/978-3-319-00227-9_5, https://doi.org/10.1007/978-3-319-00227-9_5
2. Benamou, J.D., Brenier, Y.: A computational fluid mechanics solution to the monge–kantorovich mass transfer problem. Numerische Mathematik **84**(3), 375–393 (2000). https://doi.org/10.1007/s002110050002, https://doi.org/10.1007/s002110050002
3. Brenier, Y.: Polar factorization and monotone rearrangement of vector-valued functions. Communications on Pure and Applied Mathematics **44**, 375–417 (1991), https://api.semanticscholar.org/CorpusID:123428953
4. Chen, R.T.Q., Rubanova, Y., Bettencourt, J., Duvenaud, D.K.: Neural ordinary differential equations. In: Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 31. Curran Associates, Inc. (2018), https://proceedings.neurips.cc/paper_files/paper/2018/file/69386f6bb1dfed68692a24c8686939b9-Paper.pdf
5. Chen, Y., Georgiou, T.T., Pavon, M.: Optimal transport over a linear dynamical system. IEEE Transactions on Automatic Control **62**, 2137–2152 (2015), https://api.semanticscholar.org/CorpusID:15303335
6. Chen, Y., Georgiou, T.T., Pavon, M.: On the relation between optimal transport and schrödinger bridges: A stochastic control viewpoint. Journal of Optimization Theory and Applications **169**(2), 671–691 (2016). https://doi.org/10.1007/s10957-015-0803-z, https://doi.org/10.1007/s10957-015-0803-z
7. Cuturi, M.: Sinkhorn distances: Lightspeed computation of optimal transport. In: Burges, C., Bottou, L., Welling, M., Ghahramani, Z., Weinberger, K. (eds.) Advances in Neural Information Processing Systems. vol. 26. Curran Associates, Inc. (2013), https://proceedings.neurips.cc/paper_files/paper/2013/file/af21d0c97db2e27e13572cbf59eb343d-Paper.pdf
8. De Bortoli, V., Thornton, J., Heng, J., Doucet, A.: Diffusion schrödinger bridge with applications to score-based generative modeling. In: Proceedings of the 35th International Conference on Neural Information Processing Systems. NIPS '21, Curran Associates Inc., Red Hook, NY, USA (2021)
9. Dhariwal, P., Nichol, A.Q.: Diffusion models beat GANs on image synthesis. In: Beygelzimer, A., Dauphin, Y., Liang, P., Vaughan, J.W. (eds.) Advances in Neural Information Processing Systems (2021), https://openreview.net/forum?id=AAWuCvzaVt
10. DiPerna, R.J., Lions, P.L.: Ordinary differential equations, transport theory and Sobolev spaces. Inventiones mathematicae **98**(3), 511–547 (Oct 1989). https://doi.org/10.1007/BF01393835, https://doi.org/10.1007/BF01393835
11. DiPerna, R.J., Lions, P.L.: Ordinary differential equations, transport theory and sobolev spaces. Inventiones Mathematicae **98**(3), 511–548 (1989). https://doi.org/10.1007/BF01393835

12. Donsker, M.D., Varadhan, S.R.S.: On a variational formula for the principal eigenvalue for operators with maximum principle. Proceedings of the National Academy of Sciences **72**(3), 780–783 (1975). https://doi.org/10.1073/pnas.72.3.780, https://www.pnas.org/doi/abs/10.1073/pnas.72.3.780
13. Goldfeld, Z., Greenewald, K.H.: Gaussian-smoothed optimal transport: Metric structure and statistical efficiency. In: International Conference on Artificial Intelligence and Statistics (2020), https://api.semanticscholar.org/CorpusID:210168147
14. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K. (eds.) Advances in Neural Information Processing Systems. vol. 27. Curran Associates, Inc. (2014), https://proceedings.neurips.cc/paper_files/paper/2014/file/f033ed80deb0234979a61f95710dbe25-Paper.pdf
15. Grathwohl, W., Chen, R.T.Q., Bettencourt, J., Duvenaud, D.: Scalable reversible generative models with free-form continuous dynamics. In: International Conference on Learning Representations (2019), https://openreview.net/forum?id=rJxgknCcK7
16. Gronwall, T.H.: Note on the derivatives with respect to a parameter of the solutions of a system of differential equations. Annals of Mathematics **20**(4), 292–296 (1919), http://www.jstor.org/stable/1967124
17. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H. (eds.) Advances in Neural Information Processing Systems. vol. 33, pp. 6840–6851. Curran Associates, Inc. (2020), https://proceedings.neurips.cc/paper_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf
18. Hoeffding, W.: Probability inequalities for sums of bounded random variables. Journal of the American Statistical Association **58**(301), 13–30 (1963). https://doi.org/10.1080/01621459.1963.10500830, https://doi.org/10.1080/01621459.1963.10500830
19. Kallianpur, G., Karandikar, R.L.: Girsanov's Theorem, pp. 95–101. Birkhäuser Boston, Boston, MA (2000). https://doi.org/10.1007/978-1-4612-0511-1_5, https://doi.org/10.1007/978-1-4612-0511-1_5
20. Karras, T., Aittala, M., Aila, T., Laine, S.: Elucidating the design space of diffusion-based generative models. In: Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., Oh, A. (eds.) Advances in Neural Information Processing Systems. vol. 35, pp. 26565–26577. Curran Associates, Inc. (2022), https://proceedings.neurips.cc/paper_files/paper/2022/file/a98846e9d9cc01cfb87eb694d946ce6b-Paper-Conference.pdf
21. Katanaev, M.O.: Complete separation of variables in the geodesic hamilton-jacobi equation in four dimensions. Physica Scripta **98**(10), 104001 (sep 2023). https://doi.org/10.1088/1402-4896/acf251, https://doi.org/10.1088/1402-4896/acf251
22. Klenke, A.: Probability Theory: A Comprehensive Course. Universitext, Springer London, London, 1 edn. (2008). https://doi.org/10.1007/978-1-84800-048-3
23. Kuhn, H.W., Tucker, A.W.: Nonlinear programming. In: Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950. pp. 481–492. Univ. California Press, Berkeley-Los Angeles, Calif. (1951)
24. Kynkäänniemi, T., Karras, T., Laine, S., Lehtinen, J., Aila, T.: Improved precision and recall metric for assessing generative models. In: Proceedings of the 33rd In-

ternational Conference on Neural Information Processing Systems. pp. 3927–3936. Curran Associates, Inc. (2019)
25. L'eonard, C.: From the schrödinger problem to the monge-kantorovich problem. arXiv: Optimization and Control (2010), https://api.semanticscholar.org/CorpusID:119683072
26. Léonard, C.: A survey of the schrödinger problem and some of its connections with optimal transport. Discrete and Continuous Dynamical Systems **34**(4), 1533–1574 (2014). https://doi.org/10.3934/dcds.2014.34.1533
27. Lin, C.H., Azabou, M., Dyer, E.: Making transport more robust and interpretable by moving data through a small number of anchor points. In: Meila, M., Zhang, T. (eds.) Proceedings of the 38th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 139, pp. 6631–6641. PMLR (18–24 Jul 2021), https://proceedings.mlr.press/v139/lin21a.html
28. Lipman, Y., Chen, R.T.Q., Ben-Hamu, H., Nickel, M., Le, M.: Flow matching for generative modeling. In: The Eleventh International Conference on Learning Representations (2023), https://openreview.net/forum?id=PqvMRDCJT9t
29. Liu, S., Li, W., Zha, H., Zhou, H.: Neural parametric fokker–planck equation. SIAM Journal on Numerical Analysis **60**(3), 1385–1449 (2022). https://doi.org/10.1137/20M1344986, https://doi.org/10.1137/20M1344986
30. Liu, S., Ma, S., Chen, Y., Zha, H., Zhou, H.: Learning high dimensional wasserstein geodesics (2021), https://arxiv.org/abs/2102.02992
31. Liu, X., Gong, C., qiang liu: Flow straight and fast: Learning to generate and transfer data with rectified flow. In: The Eleventh International Conference on Learning Representations (2023), https://openreview.net/forum?id=XVjTT1nw5z
32. Léonard, C.: A survey of the schrödinger problem and some of its connections with optimal transport. Discrete and Continuous Dynamical Systems **34**(4), 1533–1574 (2014). https://doi.org/10.3934/dcds.2014.34.1533, https://www.aimsciences.org/article/id/d5bcf817-901d-4104-b7da-eade7847c53e
33. McCann, R.J.: A convexity principle for interacting gases. Advances in Mathematics **128**(1), 153–179 (1997). https://doi.org/https://doi.org/10.1006/aima.1997.1634, https://www.sciencedirect.com/science/article/pii/S0001870897916340
34. Mikami, T., Thieullen, M.: Duality theorem for the stochastic optimal control problem. Stochastic Processes and their Applications **116**(12), 1815–1835 (2006). https://doi.org/https://doi.org/10.1016/j.spa.2006.04.014, https://www.sciencedirect.com/science/article/pii/S0304414906000627
35. Nikodym, O.: Sur une généralisation des intégrales de M. J. Radon. Errata et remarques. Fundam. Math. **15**, 358 (1930), https://eudml.org/doc/212339
36. Peluchetti, S.: Diffusion bridge mixture transports, schrÃ¶dinger bridge problems and generative modeling. Journal of Machine Learning Research **24**(374), 1–51 (2023), http://jmlr.org/papers/v24/23-0527.html
37. Poincaré, H.: Sur les equations aux dérivées partielles de la physique mathématique. American Journal of Mathematics **12**(3), 211–294 (1890), http://www.jstor.org/stable/2369620
38. Reshetnyak, Y.G.: General theorems on semicontinuity and on convergence with a functional. Siberian Mathematical Journal **8**(5), 801–816 (Sep 1967). https://doi.org/10.1007/BF01040656, https://doi.org/10.1007/BF01040656
39. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10684–10695 (June 2022)

40. Shi, Y., De Bortoli, V., Campbell, A., Doucet, A.: Diffusion schrödinger bridge matching. In: Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., Levine, S. (eds.) Advances in Neural Information Processing Systems. vol. 36, pp. 62183–62223. Curran Associates, Inc. (2023), https://proceedings.neurips.cc/paper_files/paper/2023/file/c428adf74782c2092d254329b6b02482-Paper-Conference.pdf
41. Song, Y., Dhariwal, P., Chen, M., Sutskever, I.: Consistency models. In: Krause, A., Brunskill, E., Cho, K., Engelhardt, B., Sabato, S., Scarlett, J. (eds.) Proceedings of the 40th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 202, pp. 32211–32252. PMLR (23–29 Jul 2023), https://proceedings.mlr.press/v202/song23a.html
42. Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based generative modeling through stochastic differential equations. In: International Conference on Learning Representations (2021), https://openreview.net/forum?id=PxTIG12RRHS
43. Stummer, W.: The Novikov and entropy conditions of multidimensional diffusion processes with singular drift. Probability Theory and Related Fields **97**(4), 515–542 (Dec 1993). https://doi.org/10.1007/BF01192962, https://doi.org/10.1007/BF01192962
44. Tong, A.Y., Malkin, N., Fatras, K., Atanackovic, L., Zhang, Y., Huguet, G., Wolf, G., Bengio, Y.: Simulation-free Schrödinger bridges via score and flow matching. In: Dasgupta, S., Mandt, S., Li, Y. (eds.) Proceedings of The 27th International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research, vol. 238, pp. 1279–1287. PMLR (02–04 May 2024), https://proceedings.mlr.press/v238/tong24a.html
45. Villani, C.: Optimal Transport: Old and New, Grundlehren der mathematischen Wissenschaften, vol. 338. Springer-Verlag Berlin Heidelberg (2009). https://doi.org/10.1007/978-3-540-71050-9
46. Young, W.H.: On Classes of Summable Functions and their Fourier Series. Proceedings of the Royal Society of London Series A **87**(594), 225–229 (Aug 1912). https://doi.org/10.1098/rspa.1912.0076