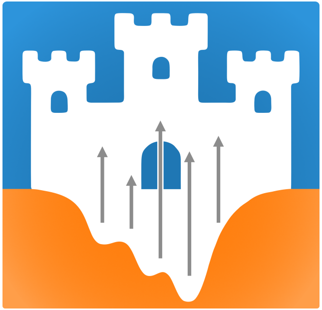


Flow Matching

Mikhail Persiianov, Grigoriy Ksenofontov

Moscow, 2026

Overview

Recap: Bridge Matching and Markovian projection

Ordinary Flows and Marginal Dynamics

Conditional Paths and Mixture Construction

Marginal Velocity as Conditional Average

Flow Matching and Conditional Flow Matching

BM $\rightarrow$ FM via the $\epsilon \rightarrow 0$ limit

Couplings and Rectified Flows: the IMF analog

Examples and Conclusions

Recap: Bridge Matching and Markovian projection

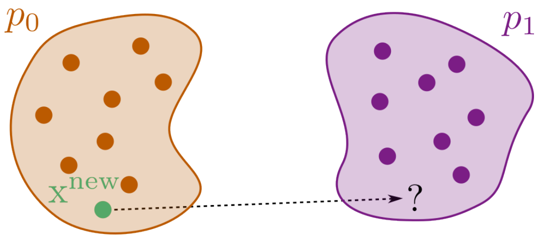
Unsupervised Domain Translation: Problem Setup

Unsupervised setting.

We observe two datasets

$$\{x^{(i)}\}_{i=1}^{M} \sim p_0, \qquad \{y^{(i)}\}_{i=1}^{M} \sim p_1,$$

with **no paired correspondences**.



Goal: learn a mapping

$$T : \mathcal{X} \to \mathcal{Y}$$

such that

$$T_\sharp p_0 \approx p_1.$$

Key difficulty: the pushforward constraint does not uniquely determine T.

Why Static Transport Models Are Not Enough

In practice, the discussed optimal transport corresponds to learning a **one-step generator**:

$$y = T(x, z), \qquad z \sim p(z).$$

Complex distributions are difficult to generate in a single step.

Thus, modern generative models are inherently **dynamic**, i.e.:

$$x \longrightarrow x_{t_1} \longrightarrow \cdots \longrightarrow x_{t_N} \longrightarrow y$$

Examples of state-of-the-art approaches:

- Diffusion models (DDPM, DDIM, NCSN, etc.);
- Flow matching / continuous normalizing flows (FM, Rectified Flows, etc.).

QUESTION: Can we bring dynamics into Optimal Transport?

Bridge Matching

Reference SDE $\mathbb{P}^{\mathrm{ref}}$ (e.g. scaled Wiener) and its **bridges** $\mathbb{P}^{\mathrm{ref}}_{|x,y}$ with transition density $q^{\mathrm{ref}}(y \mid x_t)$.

Given a coupling $\pi \in \Pi(p_0, p_1)$, the **mixture of bridges**

$$\mathbb{Q} \;=\; \int \mathbb{P}^{\mathrm{ref}}_{|x_0,x_1}\, \pi(x_0, x_1)\, \mathrm{d}x_0\, \mathrm{d}x_1$$

$\min_{P \in \Pi(p_0,p_1)} KL(P \,\|\, P^{\mathrm{ref}})$ " σW_t

is generally non-Markov. The **Markovian projection** chooses an SDE drift correction $u(x_t, t)$ by averaging bridge score targets:

$$u^*(x_t, t) \;=\; \mathbb{E}_{y \sim q^{\mathbb{Q}}(x_1 | x_t)}\left[\nabla_{x_t} \log q^{\mathrm{ref}}(y \mid x_t)\right].$$

Bridge Matching learns $u_\theta \approx u^*$ by regressing $\nabla_{x_t} \log q^{\mathrm{ref}}(y \mid x_t)$ on samples (x_0, x_1, t, x_t) from π and $q^{\mathrm{ref}}(x_t \mid x_0, x_1)$; see previous lecture for details[1,2].

[1] Stefano Peluchetti (2023). **"Diffusion bridge mixture transports, Schrödinger bridge problems and generative modeling".** In: *Journal of Machine Learning Research* 24.374, pp. 1–51.
[2] Yuyang Shi et al. (2023). **"Diffusion schrödinger bridge matching".** In: *Advances in Neural Information Processing Systems* 36, pp. 62183–62223.

From Stochastic Bridges to a Deterministic ODE?

What BM delivered.

- BM learns an **SDE** whose drift matches a *noisy bridge mean*; inference requires **many small SDE steps**.
- The mixture of bridges $\mathbb{Q}$ is non-Markov; its **Markovian projection** is an SDE with the *same marginals* $(p_t)_{t\in[0,1]}$.

Natural question. Can we keep those same marginals p_t, but transport mass by a **deterministic ODE**

$$\mathrm{d}x_t = v_t(x_t)\,\mathrm{d}t$$

instead of an SDE? That is **Flow Matching**[1]: the *deterministic analog* of BM, built on the same "mixture of conditional paths" idea.

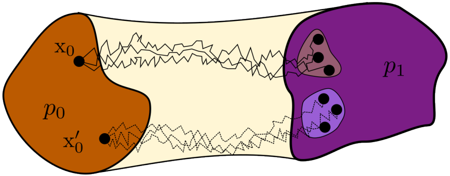


Figure 1: Bridge Matching

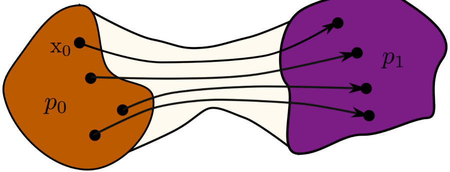


Figure 2: Flow Matching

[2]Yaron Lipman et al. (2023). **"Flow Matching for Generative Modeling".** In: *International Conference on Learning Representations*

Ordinary Flows and Marginal Dynamics

Ordinary Flows: ODE, Flow Map, Path Law

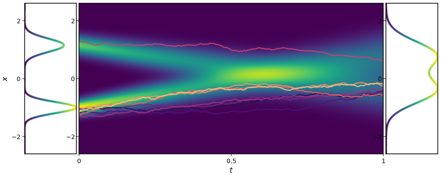
A time-dependent velocity field $v_t(x) \in \mathbb{R}^d$ defines an **ordinary flow**

$$\mathrm{d}x_t = v_t(x_t)\,\mathrm{d}t, \qquad t \in [0,1],$$

whose **flow map** $\Psi_{s\to t} : \mathbb{R}^d \to \mathbb{R}^d$ sends x_s to the unique solution x_t of the ODE with initial condition x_s (for locally Lipschitz v_t).

Given an initial law $x_0 \sim p_0$, the induced **path law** $\mathbb{P}^v$ is a probability measure on $C([0,1];\mathbb{R}^d)$ concentrated on *deterministic* integral curves $t \mapsto \Psi_{0\to t}(x_0)$. All randomness sits in the initial point.

Marginals satisfy the pushforward $p_t = \left(\Psi_{0\to t}\right)_\sharp p_0$.



SDE paths $\mathrm{d}x_t = b\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$: here $\sigma \equiv 0$, so paths are *smooth curves*, not samples of a noisy bridge.

Same goal as dynamic SB tools: specify time-dependent marginals p_t connecting p_0 at t=0 to p_1 at t=1, but the **generator is an ODE** (no reference Wiener measure).

Continuous Normalizing Flows: Likelihood is Expensive

Continuous Normalizing Flows (CNFs)[3] parameterize $v_t = v_\theta(\cdot, t)$ and train by **maximum likelihood** using the instantaneous change-of-variables formula along the ODE: for a trajectory $x_t = \Psi_{0\to t}(x_0)$,

$$\frac{\mathrm{d}}{\mathrm{d}t} \log p_t(x_t) \;=\; -\nabla \cdot v_\theta(x_t, t).$$

Integrating from 0 to 1 and combining with a tractable prior at $t{=}0$ yields $\log p_1(x_1)$.

Why this is costly in practice.

- Each training sample needs a full **ODE solve** (many function evaluations) to evaluate $\log p_1$.
- The divergence $\nabla \cdot v_\theta$ is $\mathcal{O}(d)$ to compute exactly (or noisy via Hutchinson's estimator).
- Gradient backpropagation through the ODE solver (adjoint method) compounds cost and can be memory-/numerically-fragile.

Take-away: We want the **ODE generator** from CNFs, but trained **simulation-free**, i.e. no ODE rollout inside the loss. Flow Matching achieves exactly this by regressing v_θ against tractable velocity targets.

[3] Ricky T. Q. Chen et al. (2018). **"Neural Ordinary Differential Equations".** In: *Advances in Neural Information Processing Systems*. Vol. 31.

Continuity Equation: the ODE Analog of KFP

If x_t solves $\mathrm{d}x_t = v_t(x_t)\,\mathrm{d}t$ with law p_t, then p_t satisfies the **continuity equation**

$$\partial_t p_t(x) \; + \; \nabla \cdot \big(p_t(x)\, v_t(x)\big) \; = \; 0, \qquad p_0 = p_0,\; p_1 = p_1.$$

This is the **zero-noise limit** of the Kolmogorov–Fokker–Planck equation: setting $\sigma \equiv 0$ in

$$\partial_t p_t \; = \; -\nabla \cdot (b\, p_t) \; + \; \tfrac{1}{2}\, \partial_{ij}(a_{ij}\, p_t) \qquad \xrightarrow{\sigma \to 0} \qquad \partial_t p_t \; = \; -\nabla \cdot (v_t\, p_t).$$

Note: Mass at x changes only because the velocity field v_t carries it in or out, i.e. no diffusion term.

Role in this lecture. The continuity equation is the ODE counterpart of KFP and will play *the same role* in L6 as KFP plays in L5: it is the tool that lets us **certify** that a candidate v_t transports p_0 to p_1 through prescribed marginals.

Non-uniqueness: Many v_t Realize the Same Marginals

Problem. The continuity equation is **under-determined**: given a family $(p_t)_{t\in[0,1]}$ with $p_0 = p_0$ and $p_1 = p_1$, there are *infinitely many* velocity fields v_t that transport these marginals. If v_t works and w_t is divergence-free w.r.t. p_t (i.e. $\nabla \cdot (p_t w_t) = 0$), then $v_t + w_t$ also transports them.

- **Curved.** Paths bend or loop; you can still hit the same marginals at $t = 0, 1$.
- **Straight.** Paths are segments between paired endpoints.

Both satisfy $\partial_t p_t + \nabla \cdot (p_t v_t) = 0$ with identical boundary conditions.

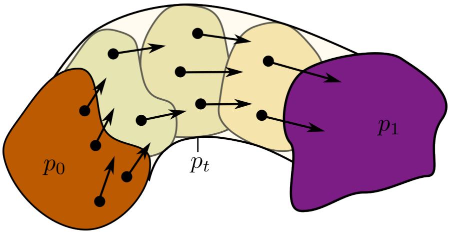


Figure 3: Curved Flow

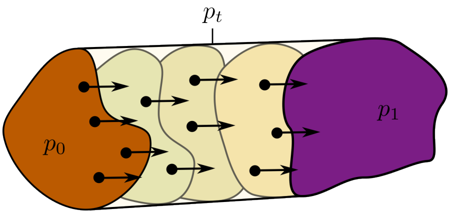


Figure 4: Straight Flow

Conditional Paths and Mixture Construction

Conditional Path: the ODE Analog of a Brownian Bridge

Fix a latent $z = (x_0, x_1) \in \mathbb{R}^d \times \mathbb{R}^d$ (pinned endpoints) and choose a **conditional path density**

$$p_t(x \mid x_0, x_1), \qquad t \in [0,1], \qquad p_0(\cdot \mid x_0, x_1) = \delta_{x_0}, \quad p_1(\cdot \mid x_0, x_1) = \delta_{x_1}.$$

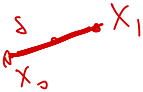
$$p_t(x) = \int p_t(x \mid x_0, x_1)\, \pi(x_0, x_1)\, dx_0\, dx_1$$

Require a **conditional velocity** $u_t(x \mid x_0, x_1)$ so that, for each fixed (x_0, x_1), the conditional density obeys the **conditional continuity equation**

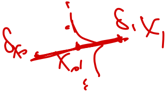
$$\partial_t p_t(x \mid x_0, x_1) \; + \; \nabla \cdot \big(p_t(x \mid x_0, x_1)\, u_t(x \mid x_0, x_1)\big) \; = \; 0.$$

This is the ODE analog of "the Brownian bridge $\mathbb{P}^{\mathrm{ref}}_{|x_0,x_1}$ pinned at (x_0, x_1)" from previous lecture: the SDE with endpoint pinning is replaced by a deterministic **flow pinned at** (x_0, x_1).

Examples:



- **Straight segment:** $p_t(\cdot \mid x_0, x_1) = \delta_{(1-t)x_0 + t x_1}$, $u_t(x \mid x_0, x_1) = x_1 - x_0$.
- **Gaussian tube:** $p_t(\cdot \mid x_0, x_1) = \mathcal{N}\big((1-t)x_0 + t x_1,\ \sigma^2 t(1-t)\mathbb{I}\big)$, u_t affine in x.



Mixture of Conditional Paths: ODE Analog of the Reciprocal Process

Given a coupling $\pi \in \Pi(p_0, p_1)$, define the **mixture marginals**

$$p_t(x) \;=\; \int p_t(x \mid x_0, x_1)\, \pi(\mathrm{d}x_0, \mathrm{d}x_1).$$

$p_t(x \mid x_0, x_1)$

$p_0(x \mid x_0, x_1) = \delta_{x_0}$

$p_1(x \mid x_0, x_1) = \delta_{x_1}$

Endpoint marginals. Since $p_0(\cdot \mid x_0, x_1) = \delta_{x_0}$ and $p_1(\cdot \mid x_0, x_1) = \delta_{x_1}$,

$$p_0(x) \;=\; \int \delta_{x_0}(x)\, \pi(\mathrm{d}x_0, \mathrm{d}x_1) \;=\; p_0(x), \qquad p_1(x) \;=\; \int \delta_{x_1}(x)\, \pi(\mathrm{d}x_0, \mathrm{d}x_1) \;=\; p_1(x).$$

Thus, $(p_t)_{t\in[0,1]}$ interpolates from p_0 to p_1 for any $\pi \in \Pi(p_0, p_1)$ and path family.

Note. Lecture 5 uses the reference SDE bridge $\mathbb{P}^{\mathrm{ref}}_{|x_0,x_1}$; here we use a chosen conditional law $p_t(\cdot \mid x_0, x_1)$. The same mixture pattern reads $\mathbb{Q} = \int \mathbb{P}^{\mathrm{ref}}_{|x_0,x_1}\, \pi(\mathrm{d}x_0, \mathrm{d}x_1)$ versus $p_t(x) = \int p_t(x \mid x_0, x_1)\, \pi(\mathrm{d}x_0, \mathrm{d}x_1)$; only the per-pair kernel changes.[4] Typically $\mathbb{Q}$ is not Markov in x_t, because knowing x_t does not determine which latent pair (x_0, x_1) generated the trajectory. Likewise the ODE-side mixture is typically not the law of any ODE $\mathrm{d}x_t = v_t(x_t)\,\mathrm{d}t$: several pairs consistent with the same observed (t, x) can require different conditional velocities $u_t(x \mid x_0, x_1)$, whereas an ODE assigns a single $v_t(x)$ at each (t, x).

[4] Stefano Peluchetti (2023). **"Diffusion bridge mixture transports, Schrödinger bridge problems and generative modeling".** In: *Journal of Machine Learning Research* 24.374, pp. 1–51.

Marginal Velocity as Conditional Average

Marginal Velocity Theorem: ODE Analog of Markovian Projection

Theorem

Suppose each $(p_t(\cdot \mid x_0, x_1), u_t)$ *satisfies the conditional continuity equation and some mild regularity conditions met.*[5] *For a.e.* (t, x) *with* $p_t(x) > 0$*, define*

$$v_t^*(x) \;:=\; \mathbb{E}\big[\, u_t(x \mid x_0, x_1) \mid x_t = x \,\big] \;=\; \frac{\int u_t(x \mid x_0, x_1)\, p_t(x \mid x_0, x_1)\, \pi(\mathrm{d}x_0, \mathrm{d}x_1)}{p_t(x)}.$$

Then $\partial_t p_t + \nabla \cdot (p_t v_t^*) = 0$ *with* $p_0 = p_0$*,* $p_1 = p_1$*. If* v_t^* *is locally Lipschitz in* x *uniformly in* t*, the flow of* $\mathrm{d}x_t = v_t^*(x_t)\,\mathrm{d}t$ *from* $x_0 \sim p_0$ *has marginals* p_t *and terminal law* p_1*.*

Note. Same closure idea as Markovian projection in Lecture 5: replace an intractable mixture object by averaging *tractable conditional generators* against the Bayesian posterior given the current state. BM averages bridge-score targets $\nabla_{x_t} \log q^{\mathrm{ref}}(x_1 \mid x_t)$ with respect to $q^{\mathbb{Q}}(x_1 \mid x_t)$; here one averages $u_t(x \mid x_0, x_1)$ with respect to $p(x_0, x_1 \mid x_t{=}x)$. Bridge drifts and KFP on the SDE side correspond to conditional velocities and the continuity equation on the ODE side.

[5] Yaron Lipman et al. (2023). **"Flow Matching for Generative Modeling".** In: *International Conference on Learning Representations*.

Proof sketch via the continuity equation

Step 1: Conditional Continuity Equation. For each latent (x_0, x_1),

$$\partial_t p_t(x \mid x_0, x_1) = -\nabla \cdot \big(p_t(x \mid x_0, x_1)\, u_t(x \mid x_0, x_1)\big).$$

Step 2: Integrate against π. Multiplying by $\pi(\mathrm{d}x_0, \mathrm{d}x_1)$ and integrating (exchanging $\partial_t, \nabla\cdot$ with the π-integral under mild regularity),

$$\partial_t \underbrace{\int p_t(x \mid x_0, x_1)\, \pi(\mathrm{d}x_0, \mathrm{d}x_1)}_{= p_t(x)} = -\nabla \cdot \underbrace{\int u_t(x \mid x_0, x_1)\, p_t(x \mid x_0, x_1)\, \pi(\mathrm{d}x_0, \mathrm{d}x_1)}_{= p_t(x)\, v_t^*(x) \text{ by definition}}.$$

$\partial_t p_t(x) = -\nabla \cdot p_t(x)\, v_t^*(x)$

Step 3: Recognize the conditional expectation. By Bayes' rule,

$$p(x_0, x_1 \mid x_t = x) = \frac{p_t(x \mid x_0, x_1)\, \pi(x_0, x_1)}{p_t(x)}.$$

so the averaged numerator in Step 2 is exactly $p_t(x)\, \mathbb{E}[u_t(x \mid x_0, x_1) \mid x_t = x] = p_t(x)\, v_t^*(x)$. Substituting yields

$$\partial_t p_t(x) + \nabla \cdot \big(p_t(x)\, v_t^*(x)\big) = 0,$$

with $p_0 = p_0$, $p_1 = p_1$ by the boundary conditions of the conditional family. □

Note. In Lecture 5, $\mathbb{E}_{x_1 \sim q^{\mathbb{Q}}(\cdot \mid x_t)}[\nabla_{x_t} \log q^{\mathrm{ref}}(x_1 \mid x_t)]$ emerged from averaging *bridge KFPs*; here, $\mathbb{E}_{(x_0, x_1) \sim p(\cdot \mid x_t)}[u_t(x \mid x_0, x_1)]$ emerges from averaging *conditional continuity equations*. In both cases one obtains a **Markovian** field consistent with the mixture marginals; neither side pins down a unique velocity among all fields with the same (p_t).

Flow Matching and Conditional Flow Matching

FM Regression: the Intractable Ideal Loss

If the marginal field v_t^* were known, a natural training objective would be the **flow-matching** loss

$$\mathcal{L}_{\mathrm{FM}}(\theta) = \mathbb{E}_{t\sim U[0,1]}\, \mathbb{E}_{x_t\sim p_t} \left\| v_\theta(x_t, t) - v_t^*(x_t) \right\|_2^2.$$

As in Lecture 5 for the projected drift $u^*(x_t, t) = \mathbb{E}_{x_1\sim q^{\mathbb{Q}}(\cdot|x_t)}[\nabla_{x_t} \log q^{\mathrm{ref}}(x_1 \mid x_t)]$, the target v_t^* is a **posterior integral**:

$$v_t^*(x) = \frac{\displaystyle\int u_t(x \mid x_0, x_1)\, p_t(x \mid x_0, x_1)\, \pi(\mathrm{d}x_0, \mathrm{d}x_1)}{p_t(x)}.$$

The mixture marginal p_t and the conditioning $p(x_0, x_1 \mid x_t{=}x)$ are generally not available in closed form, so drawing $x_t \sim p_t$ and evaluating $v_t^*(x_t)$ is impractical.

Remedy (as in Lecture 5). Do not sample the mixture marginal p_t. Draw $(x_0, x_1) \sim \pi$, $t \sim U[0, 1]$, and $x_t \sim p_t(\cdot \mid x_0, x_1)$, then regress $v_\theta(x_t, t)$ on the closed-form target $u_t(x_t \mid x_0, x_1)$ instead of on $v_t^*(x_t)$. For each (t, x_t), the L^2-best predictor of u_t given x_t is $\mathbb{E}[u_t \mid x_t] = v_t^*$, so among functions $v_\theta(x_t, t)$ this tractable loss has the same minimizer as $\mathcal{L}_{\mathrm{FM}}$.

CFM Regression: the Tractable Surrogate

Idea. Train v_θ against the **conditional velocity** $u_t(x \mid x_0, x_1)$, which is available in closed form for the usual path families (examples on the following slides).[6]

Conditional Flow Matching loss:

$$\mathcal{L}_{\mathrm{CFM}}(\theta) \;=\; \mathbb{E}_{t,\,(x_0,x_1),\,x_t} \left\| v_\theta(x_t, t) - u_t(x_t \mid x_0, x_1) \right\|_2^2, \qquad t \sim U[0,1], (x_0, x_1) \sim \pi, x_t \sim p_t(\cdot \mid x_0, x_1).$$

The sampling pattern matches Lecture 5's BM, with $\nabla_{x_t} \log q^{\mathrm{ref}}(x_1 \mid x_t)$ replaced by $u_t(x_t \mid x_0, x_1)$.

Simulation-free. If $p_t(\cdot \mid x_0, x_1)$ is sampleable without simulating the learned ODE (Dirac path, Gaussian bridge, ...), then $\mathcal{L}_{\mathrm{CFM}}$ avoids an inner ODE solve, unlike Jacobian-based CNF training.

Lecture 5 comparison. Bridge matching regresses an SDE drift to reference bridge scores along sampled reference bridges; CFM regresses an ODE velocity to u_t along samples from the chosen conditional path $p_t(\cdot \mid x_0, x_1)$. In both cases the target is *explicit* given (x_0, x_1, t, x_t).

[6] Yaron Lipman et al. (2023). **"Flow Matching for Generative Modeling".** In: *International Conference on Learning Representations*.

Equivalence lemma: CFM and FM share a minimizer

Lemma. Over networks $v_\theta(x_t, t)$ depending only on (x_t, t),

$$v^* \;=\; \arg\min_v \mathcal{L}_{\mathrm{FM}}(v) \;=\; \arg\min_v \mathcal{L}_{\mathrm{CFM}}(v) \;=\; \boxed{\mathbb{E}\big[\, u_t(x_t \mid x_0, x_1) \,\big|\, x_t = x \,\big].}$$

Proof. Condition $\mathcal{L}_{\mathrm{CFM}}$ on (t, x_t) using the tower property:

$$\mathcal{L}_{\mathrm{CFM}}(v) \;=\; \mathbb{E}_{t,x_t}\, \mathbb{E}_{(x_0,x_1)|x_t} \big\| v(x_t, t) - u_t(x_t \mid x_0, x_1) \big\|^2.$$

The inner expectation is the MSE of a *constant* (in (x_0, x_1)) predictor $v(x_t, t)$ against the random target $u_t(x_t \mid x_0, x_1)$. As in L5, MSE is minimized pointwise by the **conditional expectation**[7]

$$v(x_t, t)^\star \;=\; \mathbb{E}\big[\, u_t(x_t \mid x_0, x_1) \,\big|\, x_t \,\big] \;=\; v_t^*(x_t).$$

The classical bias/variance split then gives

$$\mathcal{L}_{\mathrm{CFM}}(v) \;=\; \underbrace{\mathbb{E}_{t,x_t} \big\| v(x_t, t) - v_t^*(x_t) \big\|^2}_{= \mathcal{L}_{\mathrm{FM}}(v)} + \underbrace{\mathbb{E}_{t,x_t} \mathrm{Var}\big[\, u_t(x_t \mid x_0, x_1) \,\big|\, x_t \,\big]}_{\text{constant in } v},$$

so the two losses differ by a v-independent constant; their gradients and minimizers coincide. □

> **Remark.** The proof is the same template as for bridge matching in Lecture 5, after substituting the bridge score $\nabla_{x_t} \log q^{\mathrm{ref}}(x_1 \mid x_t)$ for the conditional velocity $u_t(x \mid x_0, x_1)$ and the marginal drift u^* for v^*. The population objectives $\mathcal{L}_{\mathrm{FM}}$ and $\mathcal{L}_{\mathrm{CFM}}$ coincide up to an additive constant independent of v; only the regression target appearing inside the conditional expectation changes.

[7] General Bregman-divergence fact, identical to the L5 argument for u^*.

Specific case I: straight-line (OT) conditional path

Pinned endpoints. Fix (x_0, x_1) and interpolate linearly,

$$x_t = (1-t)\,x_0 + t\,x_1, \qquad p_t(\cdot \mid x_0, x_1) \;=\; \delta_{x_t}.$$

Conditional on the pair, all mass sits on **one** moving point; there is no diffusion.

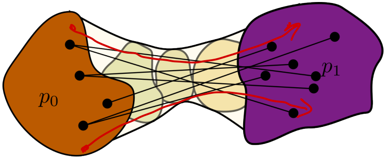


Figure 5: Straight-line conditional path

Conditional velocity. With $p_t = \delta_{x_t}$, the continuity equation $\partial_t p_t + \nabla \cdot (p_t u_t) = 0$ (given x_0, x_1) forces *constant translation*:

$$\boxed{u_t(x \mid x_0, x_1) \;=\; x_1 - x_0} \qquad \text{(independent of } t \text{ and } x\text{).}$$

CFM loss. Sample t and (x_0, x_1), set $x_t = (1-t)x_0 + tx_1$, and minimize

$$\mathcal{L}_{\mathrm{CFM}}(\theta) = \mathbb{E}_{t,(x_0,x_1)} \left\| v_\theta(x_t, t) - (x_1 - x_0) \right\|^2.$$

The regression target is deterministic given (x_0, x_1); no extra latent noise inside the loss.

Specific case II: Gaussian-tube conditional path (CFM)

Affine Gaussian tube. With $\xi \sim \mathcal{N}(0, \mathbb{I})$ and $\sigma > 0$,

$$x_t = (1-t)\,x_0 + t\,x_1 + \sigma\sqrt{t(1-t)}\,\xi,$$

$$p_t(\cdot \mid x_0, x_1) = \mathcal{N}\big((1-t)x_0 + tx_1,\ \sigma^2\, t(1-t)\,\mathbb{I}\big).$$

Writing $\mu_t = (1-t)x_0 + tx_1$ and $\gamma_t = \sigma\sqrt{t(1-t)}$, the conditional velocity is **affine in** x:

$$\boxed{u_t(x \mid x_0, x_1) \;=\; (x_1 - x_0) \;+\; \frac{\dot{\gamma}_t}{\gamma_t}\,(x - \mu_t)} \quad \text{with } \frac{\dot{\gamma}_t}{\gamma_t} = \frac{1-2t}{2\,t(1-t)}.$$

Flow Matching algorithm (FM / CFM training loop)

Parameterization.

$$v_\theta(x_t, t) \approx v_t^*(x_t) = \mathbb{E}_{(x_0,x_1)\sim p(\cdot|x_t)}\big[\, u_t(x_t \mid x_0, x_1)\,\big], \text{ with } u_t \text{ in closed form.}$$

Training (single stage, simulation-free).

1. Sample endpoints, bridge time, and interior state:

$$(x_0, x_1) \sim \pi, \quad t \sim U[0,1], \qquad x_t \sim p_t(\cdot \mid x_0, x_1).$$

2. Update network by a stochastic gradient step on the CFM loss:

$$\min_\theta \mathop{\mathbb{E}}_{(x_0,x_1),t,x_t} \big\| v_\theta(x_t, t) - u_t(x_t \mid x_0, x_1) \big\|^2.$$

Inference. Integrate the learned ODE $\mathrm{d}x_t = v_\theta(x_t, t)\,\mathrm{d}t$ from $x_0 \sim p_0$ to $t{=}1$.

- **Simulation-free** at training time (contrast DSBM): no SDE/ODE rollout inside the loss; $p_t(\cdot \mid x_0, x_1)$ is sampled in closed form (Dirac on the segment or Gaussian tube).
- **Single stage** with a *fixed* coupling π. The choice of π controls path geometry; iterative refinement (rectified flows, the FM twin of IMF) is covered later.

BM $\to$ FM via the $\epsilon \to 0$ limit

BM vs FM at a glance

x_0 x_1

Object	Bridge Matching	Flow Matching
Reference / paths	SDE $\mathbb{P}^{\mathrm{ref}}$ and bridges $\mathbb{P}^{\mathrm{ref}}_{\|x_0,x_1}$	**Chosen** conditional path $p_t(\cdot\|x_0,x_1)$
Per-sample target	**Bridge score** $\nabla_{x_t}\log q^{\mathrm{ref}}(x_1\|x_t)$	**Conditional velocity** $u_t(x_t\|x_0,x_1)$
Marginal field	$\mathbb{E}_{x_1\sim q^{\mathbb{Q}}(\cdot\|x_t)}[\nabla_{x_t}\log q^{\mathrm{ref}}(x_1\|x_t)]$ (Markovian projection)	$\mathbb{E}_{(x_0,x_1)\sim p(\cdot\|x_t)}[u_t(x_t\|x_0,x_1)]$ (marginal / FM velocity)
Generator	Projected **SDE** (reference noise + learned correction)	**ODE** $\mathrm{d}x_t = v_t\,\mathrm{d}t$
Couplings	π; refined by IMF / DSBM[8]	Independent / minibatch-OT[9] / rectified iterations[10]

Same "conditional $\to$ marginal" averaging; **different** reference object: BM fixes a noisy **SDE** bridge (noise scale ϵ in $q^{\mathrm{ref}}_{\epsilon}$), FM fixes p_t (often deterministic or a Gaussian tube).

[8] Yuyang Shi et al. (2023). **"Diffusion schrödinger bridge matching".** In: *Advances in Neural Information Processing Systems* 36, pp. 62183–62223.

[9] Alexander Tong et al. (2024). **"Improving and Generalizing Flow-Based Generative Models with Minibatch Optimal Transport".** In: *Transactions on Machine Learning Research.*

[10] Xingchao Liu, Chengyue Gong, and Qiang Liu (2022). **"Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow".** In: *arXiv preprint arXiv:2209.03003.*

From Scaled Wiener Bridge to Straight Line: $\epsilon \to 0$

Scaled Wiener bridge:

$$q_\epsilon^{\mathrm{ref}}(x_t \mid x_0, x_1) = \mathcal{N}((1-t)x_0 + tx_1,\ \epsilon\, t(1-t)\mathbb{I}), \quad \nabla_{x_t} \log q_\epsilon^{\mathrm{ref}}(x_1 \mid x_t) = \frac{x_1 - x_t}{\epsilon\,(1-t)}.$$

The BM **Doob-*h* drift** $\epsilon \nabla_{x_t} \log q_\epsilon^{\mathrm{ref}}(x_1 \mid x_t) = \dfrac{x_1 - x_t}{1-t}$ is ϵ-*independent*; only the diffusion shrinks.

Limit $\epsilon \to 0$. The bridge marginal $q_\epsilon^{\mathrm{ref}}(\cdot \mid x_0, x_1)$ concentrates on the **straight segment** $\mu_t = (1-t)x_0 + tx_1$. Evaluating the Doob-*h* drift on that segment,

$$\left.\frac{x_1 - x_t}{1-t}\right|_{x_t=\mu_t} = \frac{x_1 - (1-t)x_0 - tx_1}{1-t} = x_1 - x_0 = u_t(x_t \mid x_0, x_1) \quad \text{(linear CFM target)}.$$

Conclusion. The BM regression target for the scaled Wiener reference collapses, in the $\epsilon \to 0$ limit, to the **constant displacement** $x_1 - x_0$ used by linear / OT-conditional FM. Equivalently:

Flow Matching = Bridge Matching at zero noise, with ODE sampling.

Sampling Trade-off: Few-Step ODE vs Many-Step SDE

BM (SDE generator).

- Inference integrates an SDE $\mathrm{d}x_t = u_\theta(x_t, t)\,\mathrm{d}t + \sqrt{\epsilon}\,\mathrm{d}W_t$.
- **Many small steps** required: discretization error of an SDE solver scales like $\sqrt{\Delta t}$ (strong) vs Δt (ODE).
- Diffusion provides *built-in regularization* (entropic smoothing, $\mathbb{P}^{\mathrm{ref}}$-consistent paths).

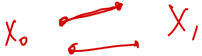
FM (ODE generator).

- Inference integrates the ODE $\mathrm{d}x_t = v_\theta(x_t, t)\,\mathrm{d}t$.
- **Few steps** often suffice: with straight conditional paths the marginal velocity is nearly constant along trajectories, so Euler with 5–20 steps is competitive.
- No noise injection: trajectories are deterministic, easier to analyse and to invert (likelihoods via change-of-variables).

Trade-off. Shrinking ϵ trades *path-measure regularization* (BM keeps $\mathbb{P}^{\mathrm{ref}}$-bridge structure) for *simulation efficiency* (FM needs far fewer NFE). Pick BM/DSBM when the reference path measure or SB optimality matters; pick FM when fast ODE sampling and straight paths are the priority.

Couplings and Rectified Flows: the IMF analog

Why the Coupling π Matters: Path Geometry



Recall: the FM marginal velocity $v_t^*(x) = \mathbb{E}_{(x_0,x_1)\sim p(\cdot|x_t=x)}[u_t(x|x_0,x_1)]$ averages *which* per-sample targets $u_t(\cdot|x_0,x_1)$ pass through (t,x). The coupling $\pi \in \Pi(p_0, p_1)$ controls which endpoint pairs enter that average, and therefore the **geometry of integral curves** of the learned ODE.

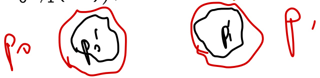
- **Independent**, $\pi(x_0,x_1) = p_0(x_0)p_1(x_1)$: endpoints paired arbitrarily; many straight segments $(1-t)x_0 + tx_1$ cross in (t,x). The marginal v_t^* must **bend** to be consistent with all crossings $\Rightarrow$ *curved* learned trajectories, *many* ODE steps.
- **Identity / known pairs** (e.g. paired data $\delta_{x_1=T(x_0)}$): no crossings; segments stay straight; few-step Euler suffices.
- **Minibatch OT / OT-CFM**[11,12]: each batch uses the OT plan between empirical source/target points. Reshapes which pairs enter the average $\Rightarrow$ *straighter* integral curves and lower-variance velocity targets.

[11] Alexander Tong et al. (2024). **"Improving and Generalizing Flow-Based Generative Models with Minibatch Optimal Transport".** In: *Transactions on Machine Learning Research*.
[12] Aram-Alexandre Pooladian et al. (2023). **"Multisample Flow Matching: Straightening Flows with Minibatch Couplings".** In: *Proceedings of the 40th International Conference on Machine Learning*. Vol. 202. Proceedings of Machine Learning Research. PMLR, pp. 28100–28127.

Rectified flow: iterative coupling refinement (FM analog of IMF)

Algorithm[13]**.** Initialize $\pi^{(0)} = p_0 \otimes p_1$. Given $\pi^{(k)}$, train $v_\theta^{(k)}$ by FM/CFM. Let $\Psi_{0\to 1}^{(k)}$ be the time-1 map of the ODE $\dot{x}_t = v_\theta^{(k)}(x_t, t)$ (well-defined under standard regularity). Define the **pushforward coupling**

$$\pi^{(k+1)} := \left(\mathrm{id} \times \Psi_{0\to 1}^{(k)}\right)_\sharp p_0 = \mathrm{Law}\left(X_0, \Psi_{0\to 1}^{(k)}(X_0)\right), \qquad X_0 \sim p_0.$$



Train $v_\theta^{(k+1)}$ on $\pi^{(k+1)}$ and repeat.

Marginals. The first marginal of $\pi^{(k+1)}$ is p_0; the second is $(\Psi_{0\to 1}^{(k)})_\sharp p_0$, which equals p_1 only if $v_\theta^{(k)}$ realizes exact transport $p_0 \mapsto p_1$.

Cost monotonicity $\mathbb{E}_{\pi^{(k+1)}}\|x_1 - x_0\|^2 \le \mathbb{E}_{\pi^{(k)}}\|x_1 - x_0\|^2$; successive iterates empirically[14] **straighten** trajectories (fewer crossings in (t, x)), so the learned field admits accurate **few-NFE** sampling.

[13] Xingchao Liu, Chengyue Gong, and Qiang Liu (2022). **"Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow".** In: *arXiv preprint arXiv:2209.03003*.
[14] Xingchao Liu, Chengyue Gong, and Qiang Liu (2022). **"Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow".** In: *arXiv preprint arXiv:2209.03003*.

RF is to FM what IMF is to BM/DSBM

	IMF / DSBM (L5)	**Rectified Flow (L6)**
Inner step	Bridge Matching with coupling $\pi^{(k)}$	Flow Matching with coupling $\pi^{(k)}$
Coupling update	$\pi^{(k+1)} = \mathrm{Law}(X_0, X_1)$ under the projected **SDE** $\mathbb{P}^{(k)}$	$\pi^{(k+1)} = (\mathrm{id} \times \Psi_{0\to1}^{(k)})_\sharp p_0$ from the learned **ODE**
Path measure constraint	Reciprocal w.r.t. $\mathbb{P}^{\mathrm{ref}}$ + Markov	None (v_t free)
Limit object	Schrödinger Bridge $\mathbb{P}^\star$[15]	Straight transport (typically not OT exactly)
Sampling at inference	SDE rollout (many steps)	ODE rollout (few steps after rectification)

Take-away. RF imports the **iterative refit-the-coupling** idea of IMF into the deterministic / ODE world. The *outer loop* is structurally identical (refit on a coupling that comes from simulating the previous generator); the *inner regression target* switches from bridge scores to conditional velocities.

[15] Yuyang Shi et al. (2023). **"Diffusion schrödinger bridge matching".** In: *Advances in Neural Information Processing Systems* 36, pp. 62183–62223.

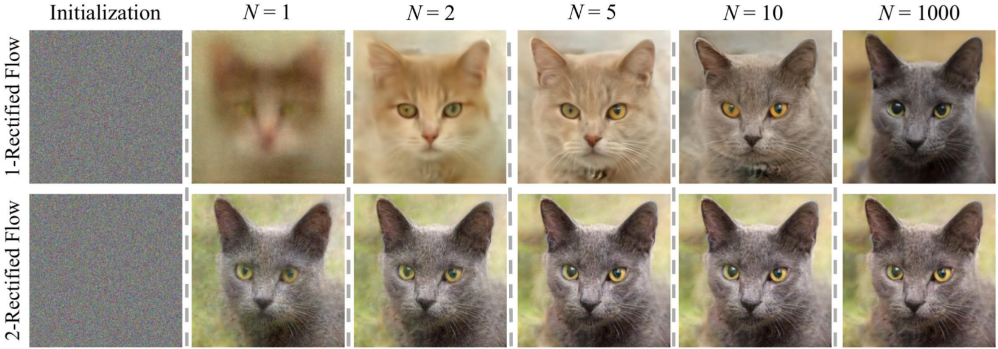
Caveats: what RF gives up compared to IMF

- **Not simulation-free after iteration 1.** Building $\pi^{(k+1)}$ requires **simulating** the ODE $\mathrm{d}x_t = v_\theta^{(k)}\,\mathrm{d}t$ for every x_0 in the training set: each rectification step costs an *ODE rollout* per sample. Only the very first stage ($k = 0$ with independent $\pi^{(0)}$) is genuinely simulation-free.
- **Error compounds across iterations.** $\pi^{(k+1)}$ is built from the *learned* (not exact) ODE, so approximation errors at step k enter the regression target at step $k+1$. The marginal $\Psi^{(k)}_{0\to1\sharp}p_0$ is only *approximately* p_1, slightly violating the boundary condition $p_1 = p_1$.
- **No SB structure.** Unlike IMF/DSBM, RF imposes *no reference path measure* and *no entropy regularization*: the limit (when it exists) is a deterministic transport map, but **not** the entropic Schrödinger Bridge, and convergence to the *Monge* OT map is not guaranteed in general.
- **When to prefer what.** Use OT-CFM / minibatch OT when one or two passes suffice (cheap, single-stage). Use RF iterations when few-NFE inference is critical and you can afford the rollout cost. Use DSBM/IMF when entropic SB structure is part of the modeling goal.

Examples and Conclusions

Rectified Flows: Image Generation Examples

Columns - different number of integration steps, **rows** – different FM iterations.



For the 2-iteration FM the trajectories are rather straight which can be deduced from the fact that the model works reasonable with just 1 Euler trajectory integration step (see 2-Rectified flow, $N = 1$).

When FM vs. SB / BM / DSBM?

Flow matching is often enough when

- endpoint marginals p_0, p_1 are the main object and a **simple conditional bridge** is acceptable;
- you want **simulation-free** conditional sampling in the loss;
- an ODE generator is compatible with downstream needs (differentiability, fast integration, few-NFE inference).

Prefer SB / BM / DSBM when

- a **reference path measure** (entropy, $\mathbb{P}^{\mathrm{ref}}$) is part of the modeling assumptions;
- uniqueness / regularization via entropic OT matters, or the reciprocal / bridge structure of the optimizer is essential;
- you are willing to pay for **SDE-level** modeling and iterative coupling refinement[16].

[16] Yuyang Shi et al. (2023). **"Diffusion schrödinger bridge matching"**. In: *Advances in Neural Information Processing Systems* 36, pp. 62183–62223.

Conclusions & Takeaways

- **ODE / marginals.** Mass transport is $\mathrm{d}x_t = v_t(x_t)\,\mathrm{d}t$ with the **continuity equation** (zero-noise KFP). A mixture $p_t = \int p_t(\cdot \mid x_0, x_1)\,\pi$ is generally not an ODE path measure; the **marginal velocity** $v_t^*(x) = \mathbb{E}[u_t(x \mid x_0, x_1) \mid x_t{=}x]$ is the unique ODE with those marginals (ODE analog of Markovian projection). **CFM** fits v_θ by regressing to closed-form u_t (segment / Gaussian tube).
- **BM → FM; couplings.** Scaled Wiener bridges pass to linear CFM as $\epsilon{\to}0$. Beyond fixed π: **OT-CFM**[17] straightens within one stage; **rectified flow**[18] refits π from the previous ODE (IMF-style outer loop, no SB path measure).

Pattern. As in BM: average **conditional** fields into a **marginal** generator; here conditional **velocities** replace bridge **scores**, and the reference path is **chosen** rather than $\mathbb{P}^{\mathrm{ref}}$.

[17] Alexander Tong et al. (2024). **"Improving and Generalizing Flow-Based Generative Models with Minibatch Optimal Transport".** In: *Transactions on Machine Learning Research.*
[18] Xingchao Liu, Chengyue Gong, and Qiang Liu (2022). **"Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow".** In: *arXiv preprint arXiv:2209.03003.*