Dynamic Schrödinger Bridge Problem

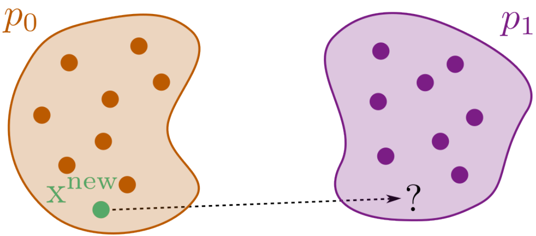
Grigoriy Ksenofontov, Mikhail Persiianov

Moscow, 2026

Overview

Recap: From Unsupervised Domain Translation to Entropic Optimal Transport

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.

Entropic Optimal Transport and Static Schrödinger Bridge

Entropy regularization[1] problem guarantees uniqueness and stochastic mappings.

Entropic Optimal Transport (EOT) solves:

$$\mathrm{EOT}_{c,\varepsilon}(p_0, p_1) = \min_{\pi \in \Pi(p_0,p_1)} \mathbb{E}_{x,y\sim\pi}[c(x,y)] - \varepsilon \mathbb{E}_{x\sim p_0} \mathrm{H}(\pi(\cdot|x)), \quad \text{(EOT)}$$

where $H(\pi)$ is the entropy of π and $\varepsilon > 0$ controls regularization.

Connection: Equivalent to the *Static Schrödinger Bridge* problem[2]:

$$\pi^* = \arg\min_{\pi \in \Pi(p_0,p_1)} \mathrm{KL}\left(\pi \| \pi^{\text{ref}}\right), \quad \text{(StaticSB)}$$

where the problem aims to find the transport plan π closest to π^{ref} in terms of the Kullback-Leibler (KL) divergence.

[1] Marco Cuturi (2013). **"Sinkhorn distances: Lightspeed computation of optimal transport".** In: *Advances in neural information processing systems* 26.

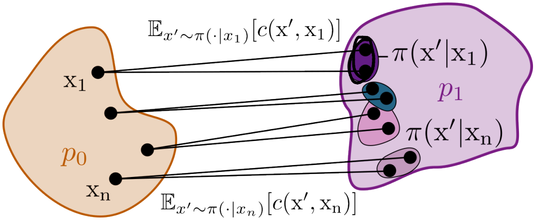
[2] Erwin Schrödinger (1931). **Über die umkehrung der naturgesetze.** Verlag der Akademie der Wissenschaften in Kommission bei Walter De Gruyter u ...

Optimal Conditional Transport Plans

Let π^* solve the equation EOT problem. By disintegration:

$$\pi^*(x, x') = \pi^*(x)\,\pi^*(x'|x) = p_0(x)\,\pi^*(x'|x).$$

The family $\{\pi^*(\cdot|x)\}_{x\in\mathcal{X}}$ are called **optimal conditional plans**.



Interpretation:

- For each source point x, we obtain a distribution over targets.
- Transport becomes **stochastic**.
- These conditionals define a generative mechanism: $x \longmapsto y \sim \pi^*(\cdot|x)$.

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?

Dynamic Formulation via Stochastic Processes

From Random Variables to Random Paths

Static OT: random variable x on state space $\mathbb{R}^d$

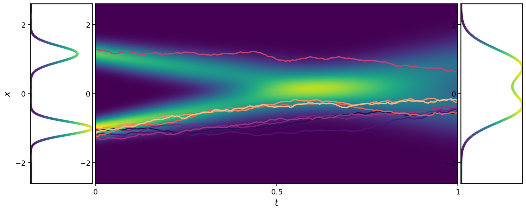
$$\Downarrow$$

Dynamic OT: random path[3] $x(t) \overset{\text{def}}{=} x_t$ on path space $C([0,1];\mathbb{R}^d)$

Evaluating the path at times $t \in [0,1]$ yields random variables $\{x_t\}_{t\in[0,1]}$, which together form a stochastic process, or formally:

$$(C([0,T];\mathbb{R}^d), \mathcal{B}(C([0,T];\mathbb{R}^d)), \mathbb{P}),$$

where $\mathbb{P}$ denotes path distribution.



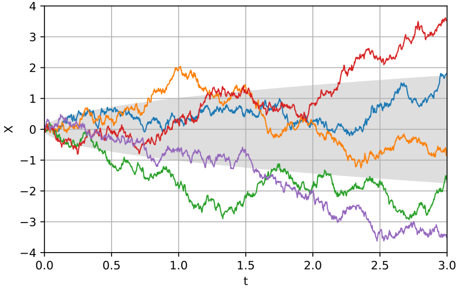
The target probabilistic object is now the path distribution.

[3] We abuse notation by using x_t both for the random path and for its realization, denoted as trajectories.

Wiener Process

Scaled Wiener process (Brownian motion), $\{W_t^\epsilon\}_{t\in[0,1]}$, is the basic source of randomness in dynamic models. Under the path distribution $\mathbb{W}^\epsilon$, it satisfies:

- the process starts from the origin, $W_0^\epsilon = 0$;
- trajectories are continuous in time;
- increments over disjoint time intervals are independent;
- the displacement $W_{t+\Delta t}^\epsilon - W_t^\epsilon \sim \mathcal{N}(0, \epsilon \Delta t \mathbb{I})$, i.e., its variance is proportional to the elapsed time and the noise level ϵ.



The random path of the scaled Wiener process $\{W_t^\epsilon\}_{t\in[0,1]}$ is given by the stochastic differential equation (SDE):

$$dx_t = \sqrt{\epsilon}\, dW_t, \qquad x_0 = 0.$$

Path Distributions Induced by SDEs

Given the scaled Wiener process, we define a new path distribution $\mathbb{P}$ by combining deterministic dynamics with Wiener noise:

$$T : dx_t = \underbrace{v(x_t, t)\, dt}_{\substack{\text{drift,} \\ \text{defines deterministic motion}}} + \underbrace{\sqrt{\epsilon}\, dW_t,}_{\substack{\text{scaled Wiener noise,} \\ \text{defines stochastic motion}}} \qquad x_0 \sim p_0,$$

Over a small time step Δt, the SDE can be interpreted as

$$\Delta X = x_{t+\Delta t} \approx x_t + v(x_t, t)\Delta t + \sqrt{\epsilon \Delta t} z_t, \qquad z_t \sim \mathcal{N}(0, \mathbb{I}).$$

The trajectories generated by this SDE induce a path distribution $\mathbb{P}$ on $C([0,1]; \mathbb{R}^d)$.

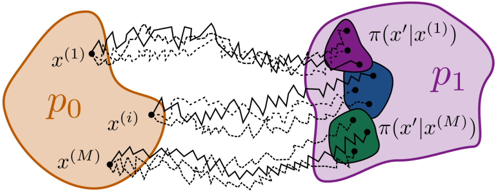
One useful notation is the bridge distribution $\mathbb{P}_{|x,y} \stackrel{\text{def}}{=} \mathbb{P}(\,\cdot\,|x_0 = x,\ x_1 = y)$, that is, the path distribution of the SDE conditioned on fixed endpoints.

Dynamic Schrödinger Bridge Formulation

Suppose we are given marginal distributions p_0 and p_1 and a reference scaled Wiener process $\mathbb{W}^\epsilon$. Then, the dynamic Schrödinger Bridge (SB) problem[4] is defined as

$$\mathbb{P}^* = \underset{\mathbb{P}\in\Pi(p_0,p_1)}{\arg\min}\ \mathrm{KL}\left(\mathbb{P}\|\mathbb{W}^\epsilon\right), \qquad \text{(DynSB)}$$

where $\Pi(p_0, p_1)$ is the set of all path distributions with initial and terminal marginals p_0 and p_1, respectively.



Intuitively, the problem seeks the path distribution $\mathbb{P}$ that is closest to the Wiener path distribution $\mathbb{W}^\epsilon$ in KL divergence, while matching the prescribed marginals p_0 and p_1 at times $t = 0$ and $t = 1$, respectively.

[4]Erwin Schrödinger (1931). **Über die umkehrung der naturgesetze.** Verlag der Akademie der Wissenschaften in Kommission bei Walter De Gruyter u ...

Equivalence of Dynamic SB, Static SB, and EOT

The dynamic Schrödinger Bridge problem equation DynSB admits the following decomposition:

$$\mathrm{KL}\left(\mathbb{P}\|\mathbb{W}^{\epsilon}\right) = \mathrm{KL}\left(\pi^{\mathbb{P}}\|\pi^{\mathbb{W}^{\epsilon}}\right) + \int \mathrm{KL}\left(\mathbb{P}_{|x_0,x_1}\|\mathbb{W}^{\epsilon}_{|x_0,x_1}\right) d\pi^{\mathbb{P}}(x_0,x_1), \tag{1}$$

where $\pi^{\mathbb{P}}$ and $\pi^{\mathbb{W}^{\epsilon}}$ are the couplings of the corresponding path distributions.

Notably, for $\mathbb{P}^*$, the bridge distribution coincides with the Wiener bridge:

$$\mathbb{P}^*_{|x_0,x_1} = \mathbb{W}^{\epsilon}_{|x_0,x_1}.$$

Hence, the second term in equation 1 vanishes, and the problem reduces to

$$\pi^* = \underset{\pi\in\Pi(p_0,p_1)}{\arg\min}\ \mathrm{KL}\left(\pi\|\pi^{\mathbb{W}^{\epsilon}}\right).$$

By setting $\pi^{\mathrm{ref}} = \pi^{\mathbb{W}^{\epsilon}}$, we recover StaticSB.

This equivalence establishes the chain

$$EOT \iff StaticSB \iff DynSB.$$

KL Divergence on Path Space

In StaticSB, we work with couplings π, which can be modeled directly by one-step mappings T. In contrast, equation DynSB is formulated in terms of path distributions $\mathbb{P}$, where the mapping T is defined by an SDE. This leads to the key question:

How do we define the KL divergence between path distributions induced by SDEs?

By definition, the KL divergence between two path distributions is

$$\mathrm{KL}\left(\mathbb{P} \| \mathbb{W}^{\epsilon}\right) = \mathbb{E}_{\mathbb{P}}\left[\log \frac{d\mathbb{P}}{d\mathbb{W}^{\epsilon}}\right].$$

So, to answer the question, we need the Radon-Nikodym derivative $\frac{d\mathbb{P}}{d\mathbb{W}^{\epsilon}}$.

Girsanov's theorem tells us exactly how the path distribution of the drifted process $\mathbb{P}$ changes relative to the Wiener path distribution $\mathbb{W}^{\epsilon}$:

$$\frac{d\mathbb{P}}{d\mathbb{W}^{\epsilon}} = \exp\left(\frac{1}{\sqrt{\epsilon}}\int_0^T v^{\top}(x_t, t)dW_t - \frac{1}{2\epsilon}\int_0^T \|v(x_t, t)\|^2 dt\right). \tag{2}$$

Local Gaussian Reweighting Behind Girsanov's Formula

On path space, we compare two path distributions defined by the SDEs

$$dx_t = \sqrt{\epsilon}\, dW_t, \qquad \text{under } \mathbb{W}^\epsilon,$$
$$dx_t = v(x_t, t)\, dt + \sqrt{\epsilon}\, d\widetilde{W}_t, \qquad \text{under } \mathbb{P}.$$

Over a small time step Δt, the one-step transition densities of the stochastic process are

$$q^{\mathbb{W}^\epsilon}(x_{t+\Delta t}|x_t) = \mathcal{N}(x_t, \epsilon \Delta t\, \mathbb{I}), \qquad q^{\mathbb{P}}(x_{t+\Delta t}|x_t) = \mathcal{N}(x_t + v(x_t, t)\Delta t, \epsilon \Delta t\, \mathbb{I}).$$

Since these are Gaussian distributions, the corresponding density ratio is

$$\frac{q^{\mathbb{P}}(x_{t+\Delta t}|x_t)}{q^{\mathbb{W}^\epsilon}(x_{t+\Delta t}|x_t)} = \frac{\mathcal{N}(x_t + v(x_t, t)\Delta t, \epsilon \Delta t\, \mathbb{I})}{\mathcal{N}(x_t, \epsilon \Delta t\, \mathbb{I})} =$$
$$= \exp\left(\frac{1}{\epsilon} v(x_t, t)^\top (x_{t+\Delta t} - x_t) - \frac{1}{2\epsilon}\|v(x_t, t)\|^2 \Delta t\right) =$$
$$= \exp\left(\frac{1}{\sqrt{\epsilon}} v(x_t, t)^\top \Delta W_t - \frac{1}{2\epsilon}\|v(x_t, t)\|^2 \Delta t\right).$$

From Local Reweighting to Path Space

Over a time grid $0 = t_0 < t_1 < \cdots < t_N = 1$, the density ratio of the whole discretized path factorizes into the product of the one-step conditional density ratios:

$$\frac{q^{\mathbb{P}}(x_{t_1}, \ldots, x_{t_N} | x_{t_0})}{q^{\mathbb{W}^\epsilon}(x_{t_1}, \ldots, x_{t_N} | x_{t_0})} = \prod_{n=0}^{N-1} \frac{q^{\mathbb{P}}(x_{t_{n+1}} | x_{t_k})}{q^{\mathbb{W}^\epsilon}(x_{t_{n+1}} | x_{t_k})}.$$

Substituting the Gaussian one-step ratio and passing to the continuous-time limit $\Delta t \to 0$ suggests the path-space Radon-Nikodym derivative

$$\frac{d\mathbb{P}}{d\mathbb{W}^\epsilon} = \exp\left(\frac{1}{\sqrt{\epsilon}} \int_0^1 v_t(x_t, t)^\top dW_t - \frac{1}{2\epsilon} \int_0^1 \|v_t(x_t, t)\|^2 dt\right).$$

The passage from local Gaussian reweighting to path space is the heuristic step.
Girsanov's theorem makes it exact.

From Path-Space KL to Drift Energy (I)

Before deriving the KL divergence, we need one additional fact:

Corollary (Shifted Wiener process under $\mathbb{P}$)

Assume Girsanov's formula holds. Then the Wiener process under $\mathbb{P}$ is related to the Wiener process under $\mathbb{W}^\epsilon$ by

$$\widetilde{W}_t \stackrel{\text{def}}{=} W_t - \frac{1}{\sqrt{\epsilon}} \int_0^t v(x_s, s)\, ds. \tag{3}$$

By definition,

$$\mathrm{KL}\left(\mathbb{P} \| \mathbb{W}^\epsilon\right) = \mathbb{E}_{\mathbb{P}}\left[\log \frac{d\mathbb{P}}{d\mathbb{W}^\epsilon}\right].$$

Using Girsanov's formula, the argument of the expectation becomes

$$\log \frac{d\mathbb{P}}{d\mathbb{W}^\epsilon} = \frac{1}{\sqrt{\epsilon}} \int_0^1 v(x_t, t)^\top dW_t - \frac{1}{2\epsilon} \int_0^1 \|v(x_t, t)\|^2 dt.$$

From Path-Space KL to Drift Energy (II)

We first focus on the stochastic term

$$\log \frac{d\mathbb{P}}{d\mathbb{W}^{\epsilon}} = \frac{1}{\sqrt{\epsilon}} \int_0^1 v(x_t, t)^\top dW_t - \frac{1}{2\epsilon} \int_0^1 \|v(x_t, t)\|^2 dt.$$

Using the relation with the shifted Wiener process $\widetilde{W}_t$ from equation 3, we rewrite the stochastic term as

$$\int_0^1 v(x_t, t)^\top dW_t = \int_0^1 v(x_t, t)^\top d\widetilde{W}_t + \frac{1}{\sqrt{\epsilon}} \int_0^1 \|v(x_t, t)\|^2 dt.$$

Substituting this back into the logarithm, we obtain

$$\log \frac{d\mathbb{P}}{d\mathbb{W}^{\epsilon}} = \frac{1}{\sqrt{\epsilon}} \left(\int_0^1 v(x_t, t)^\top d\widetilde{W}_t + \frac{1}{\sqrt{\epsilon}} \int_0^1 \|v(x_t, t)\|^2 dt \right) - \frac{1}{2\epsilon} \int_0^1 \|v(x_t, t)\|^2 dt =$$

$$= \frac{1}{\sqrt{\epsilon}} \int_0^1 v(x_t, t)^\top d\widetilde{W}_t + \frac{1}{2\epsilon} \int_0^1 \|v(x_t, t)\|^2 dt.$$

From Path-Space KL to Drift Energy III

Taking expectation under $\mathbb{P}$ gives

$$\mathrm{KL}\left(\mathbb{P}\|\mathbb{W}^{\epsilon}\right) = \mathbb{E}_{\mathbb{P}}\left[\frac{1}{\sqrt{\epsilon}}\int_0^1 v(x_t,t)^\top d\widetilde{W}_t + \frac{1}{2\epsilon}\int_0^1 \|v(x_t,t)\|^2 dt\right] =$$

$$= \frac{1}{\sqrt{\epsilon}}\mathbb{E}_{\mathbb{P}}\left[\int_0^1 v(x_t,t)^\top d\widetilde{W}_t\right] + \frac{1}{2\epsilon}\mathbb{E}_{\mathbb{P}}\left[\int_0^1 \|v(x_t,t)\|^2 dt\right]$$

Since $\widetilde{W}_t$ is Wiener under $\mathbb{P}$, its increments are centered, so the accumulated stochastic fluctuation has zero mean:

$$\mathbb{E}_{\mathbb{P}}\left[\int_0^1 v(x_t,t)^\top d\widetilde{W}_t\right] = 0.$$

Thus, we obtain

$$\mathrm{KL}\left(\mathbb{P}\|\mathbb{W}^{\epsilon}\right) = \frac{1}{2\epsilon}\mathbb{E}_{\mathbb{P}}\left[\int_0^1 \|v(x_t,t)\|^2 dt\right].$$

Thus, the path-space KL divergence is exactly the expected quadratic drift energy.

Dynamic Schrödinger Bridge as Stochastic Control

The dynamic Schrödinger Bridge problem

$$\mathbb{P}^* = \underset{\mathbb{P}\in\Pi(p_0,p_1)}{\arg\min}\ \mathrm{KL}\left(\mathbb{P}\|\mathbb{W}^\epsilon\right)$$

is equivalent to the following **stochastic optimal control** problem[5]:

$$v^* = \arg\min_v \frac{1}{2\epsilon}\mathbb{E}_\mathbb{P}\left[\int_0^1 \|v(x_t,t)\|^2 dt\right]. \qquad \text{(SOC)}$$

p0

subject to

$$dx_t = v(x_t,t)\,dt + \sqrt{\epsilon}\,d\widetilde{W}_t, \qquad x_0 \sim p_0, \qquad x_1 \sim p_1.$$

In other words, dynamic SB asks:

Among all stochastic dynamics that transport p_0 to p_1, find the one that stays **closest to Wiener motion**, or equivalently, uses the **least drift energy**.

[5]The optimal drift v^* induces the path distribution $\mathbb{P}^*$, which is exactly the dynamic Schrödinger Bridge.

Entropic Neural Optimal Transport: Relaxed Dynamic SB

Why Static EOT Is Hard to Learn from Samples

In the recap, EOT was formulated as a **static one-step transport plan** $\pi \in \Pi(p_0, p_1)$, or equivalently, as a stochastic generator $x \longmapsto y \sim \pi(\cdot|x)$.

However, when p_0 and p_1 are accessible only through empirical samples, solving equation EOT directly is non-trivial:

- the marginal constraints $\pi \in \Pi(p_0, p_1)$ must be enforced,
- the entropy term $\mathrm{H}(\pi)$ must be estimated from samples,
- the resulting model remains static, i.e., a one-step generator.

The key idea of ENOT[6] is to replace the static coupling π by a dynamic path distribution $\mathbb{P}$ and rewrite EOT as a relaxed dynamic SB problem.

[6]Nikita Gushchin et al. (2023). **"Entropic neural optimal transport via diffusion processes"**. In: *Advances in Neural Information Processing Systems* 36, pp. 75517–75544.

Relaxed Dynamic Schrödinger Bridge Objective

Specifically, by introducing a Lagrange multiplier, the ENOT framework relaxes the terminal matching constraint $\mathbb{P}_1^v = p_1$ in the dynamic SB objective and considers the Lagrangian

$$\mathcal{L}(\beta, v) \stackrel{\text{def}}{=} \underbrace{\mathbb{E}_{\mathbb{P}^v}\left[\frac{1}{2\epsilon}\int_0^1 \|v(x_t,t)\|^2 dt\right]}_{=\mathrm{KL}(\mathbb{P}^v \| W^\epsilon)} + \underbrace{\int \beta(y)\, dp_1(y) - \int \beta(y)\, d\mathbb{P}_1^v(y)}_{\text{vanishes when } \mathbb{P}_1^v = p_1},$$

where $\mathbb{P}^v$ denotes the path distribution induced by the controlled SDE

$$dx_t = v(x_t,t)\, dt + \sqrt{\epsilon}\, d\widetilde{W}_t, \qquad x_0 \sim p_0,$$

and $\mathbb{P}_1^v$ is its terminal marginal at time $t = 1$.

This leads to the saddle-point problem

$$\sup_\beta \inf_v \mathcal{L}(\beta, v). \tag{4}$$

The optimization is performed over a drift field and a terminal potential, rather than a static coupling.

Recovering dynamic SB and EOT

Theorem (Relaxed dynamic SB formulation)

Let (β^*, v^*) be an optimal saddle point of

$$\sup_{\beta} \inf_{v} \mathcal{L}(\beta, v).$$

Then the induced path distribution $\mathbb{P}^{v^*}$ solves the dynamic Schrödinger Bridge problem.

Corollary (EOT as relaxed dynamic SB)

If (β^*, v^*) solves the above saddle-point problem, then the endpoint coupling induced by $\mathbb{P}^{v^*}$ is exactly the optimal entropic transport plan:

$$\pi^{\mathbb{P}^{v^*}} = \pi^*.$$

Solving the relaxed saddle-point problem recovers both the optimal dynamic SB path distribution and the optimal EOT coupling.

Practical Optimization of ENOT

To optimize the relaxed objective, the drift field and terminal potential are parameterized by neural networks drift field $v_\theta : \mathbb{R}^D \times [0,1] \to \mathbb{R}^D$, and terminal potential $\beta_\phi : \mathbb{R}^D \to \mathbb{R}$, i.e.,

$$\sup_{\phi} \inf_{\theta} \left\{ \frac{1}{2\epsilon} \mathbb{E}_{\mathbb{P}^{v_\theta}} \left[\int_0^1 \|v_\theta(x_t, t)\|^2 dt \right] - \int \beta_\phi(y)\, d\mathbb{P}_1^{v_\theta}(y) + \int \beta_\phi(y)\, dp_1(y) \right\}.$$

To sample trajectories from $\mathbb{P}^{v_\theta}$, the controlled SDE is simulated with the Euler-Maruyama:

$$x_{t_{n+1}} = x_{t_n} + v_\theta(x_{t_n}, t_n)\Delta t + \sqrt{\epsilon \Delta t}\, z_{t_n}, \qquad z_{t_n} \sim \mathcal{N}(0, \mathbb{I}) \text{ and } x_0 \sim p_0.$$

t X_t

The drift energy is approximated along simulated trajectories by

$$\int_0^1 \|v_\theta(x_t, t)\|^2 dt \approx \sum_{n=0}^{N+1} \|v_\theta(x_{t_n}, t_n)\|^2 \Delta t.$$

Unlike simulation-free objectives such as flow matching, ENOT requires explicit trajectory simulation from the controlled SDE.

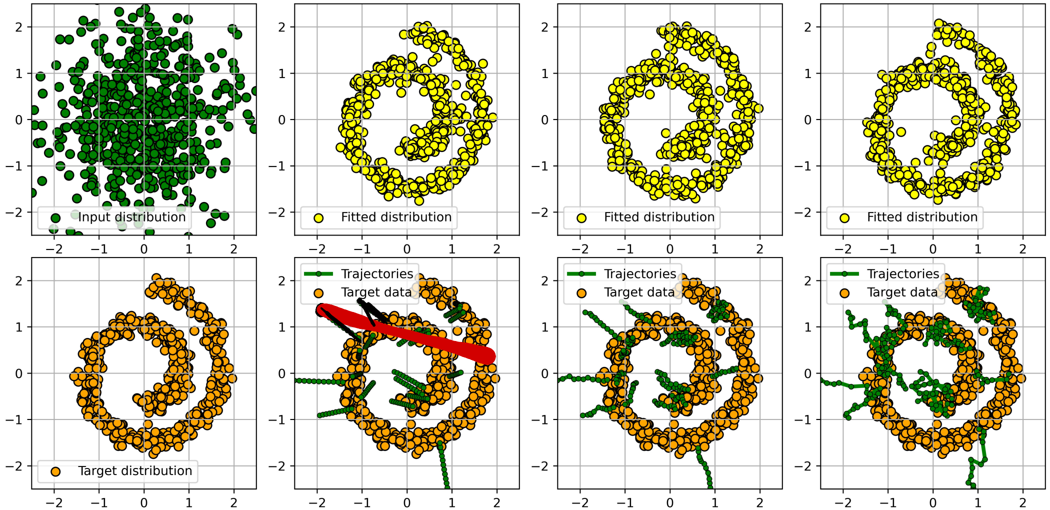
ENOT Training Algorithm

Input: samples from p_0 and p_1, batch size M, Wiener noise variance ϵ, drift field v_θ, terminal potential β_ϕ, and number of Euler-Maruyama steps N.

Optimization loop:

1. Sample batch $x_0^{(i)} \sim p_0$ and $x_1^{(i)} \sim p_1$.
2. Simulate trajectories from $\mathbb{P}^{v_\theta}$ using Euler-Maruyama.
3. Update the terminal potential β_ϕ by maximizing
$$\mathcal{L}_\beta = \frac{1}{M}\sum_{i\in[1,M]} \beta_\phi(x_1^{(i)}) - \frac{1}{M}\sum_{i\in[1,M]} \beta_\phi(x_0^{(i)}).$$
4. Update the drift field v_θ by minimizing
$$\mathcal{L}_v = \frac{1}{2\epsilon}\sum_{n=0}^{N-1} \frac{1}{M}\sum_{i\in[1,M]} \|v_\theta(x_{t_n}^{(i)}, t_n)\|^2 \Delta t - \frac{1}{M}\sum_{i\in[1,M]} \beta_\phi(x_0^{(i)}).$$
5. Repeat until convergence.

Gaussian to Swiss Roll Translation



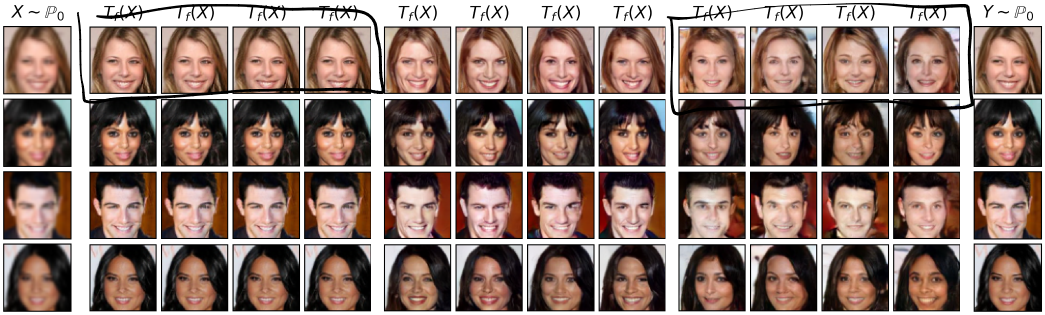
(a) Source and target distributions p_0 and p_1

(b) Fitted conditional plans $\pi_\theta(\cdot|x)$; $\epsilon = 0$.

(c) Fitted conditional plans $\pi_\theta(\cdot|x)$; $\epsilon = 0.01$.

(d) Fitted conditional plans $\pi_\theta(\cdot|x)$; $\epsilon = 0.1$.

Male to Female Translation on CelebA



(a) **(b)** ENOT, $\epsilon = 0$ **(c)** ENOT, $\epsilon = 1$ **(d)** ENOT, $\epsilon = 10$ **(e)**

Conclusions & Takeaways

- EOT resolves this by learning a **stochastic one-step coupling**, but this remains a static model.
- Dynamic Schrödinger Bridge lifts the problem from a static coupling to a path distribution, making transport **dynamic**.
- Girsanov's theorem shows that dynamic SB is equivalent to a **minimum drift-energy** stochastic control problem.
- ENOT relaxes the terminal constraint and turns dynamic SB into a **sample-based saddle-point objective**.
- Solving the ENOT objective recovers both the dynamic SB path distribution and the optimal EOT coupling.

Main message: EOT can be reformulated as a dynamic stochastic control problem and learned from samples through ENOT.