Generative Flow Maps: An overview of the math and methods behind them.

arXiv: 2406.07507 and 2505.18825

Agenda for this talk:

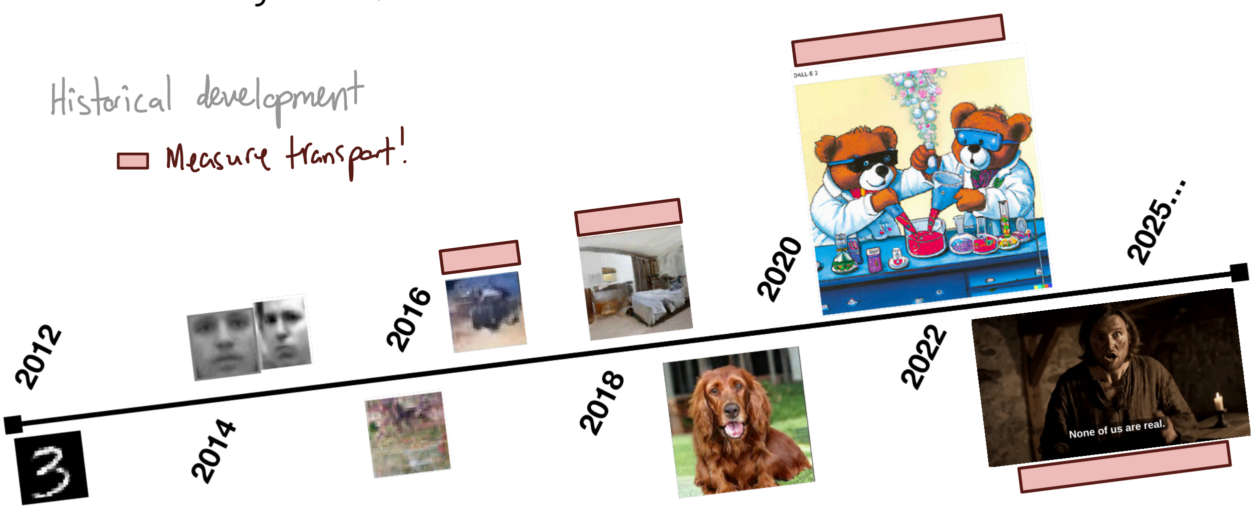
- Introduce dynamical measure transport for generative modeling
- Motivate the flow map as a computationally efficient method
- Illustrate how equations governing this map can be used to learn it, and categorize the recent efforts made in this direction

Generative Modeling

Goal: Estimate some unknown distribution with density ρ_1 through sample data $x_1 \sim \rho_1$.

Historical development

Measure transport!

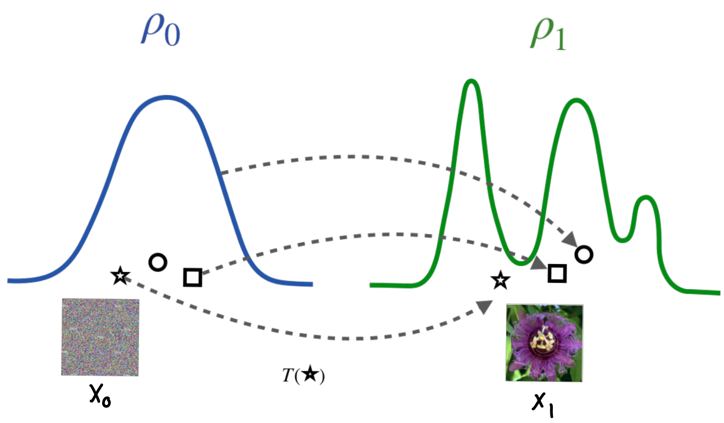


What do we mean by measure transport, and how can we can adapt the equations governing it to create more understandable and performant tools?

Measure Transport

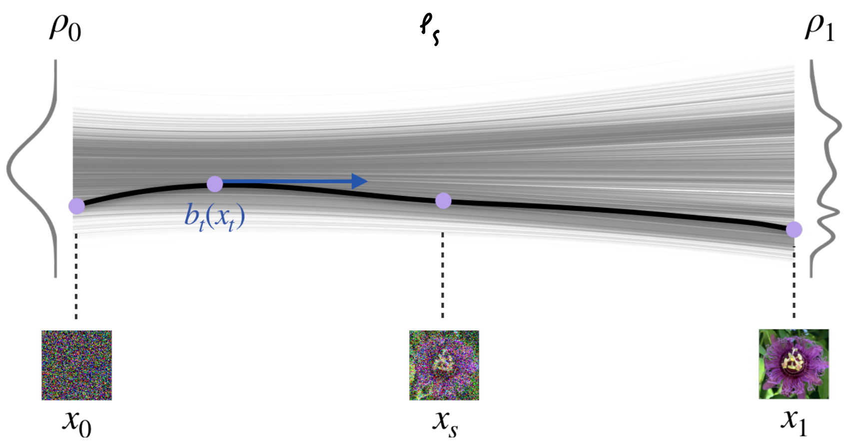
Building maps between distributions

- Sample base distribution $x_0 \sim \rho_0$
- Build a map $T: \Omega \to \Omega$
- Produce $x_1 \sim \rho_1$ via $T(x_0) = x_1$



Dynamical Measure Transport

This map can be constructed as the solution to a dynamical equation. Imagine that x_0 continually evolves over time $t \in [0,1]$ to some x_1.



Probability flow ODE

(1) $\dot{x}_t = b_t(x_t), \quad x_0 \sim \rho_{t=0}$

- b_t is a velocity field which defines how x_t should instantaneously evolve

Continuity equation

(2) $\partial_t \rho_t + \nabla \cdot (b_t \rho_t) = 0, \quad \rho_{t=0} = \rho_0$

- Equation governing the evolution of ρ_t with b_t

Learning b_t via flow matching/stochastic interpolants

- To construct a ρ_{t}, stochastically combine x_0, x_1 via the interpolant:

$$(3) \quad I_t(x_0,x_1) = \alpha_t x_0 + \beta_t x_1 \,, \quad (x_0,x_1) \sim \rho(x_0,x_1) \qquad \text{eg} \quad \begin{array}{l} \alpha_t = 1-t \\ \beta_t = t \end{array}$$

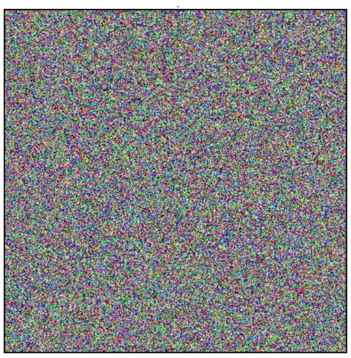
then $\rho_t = \mathrm{Law}(I_t)$ and b_t associated to (1), (2) is given by

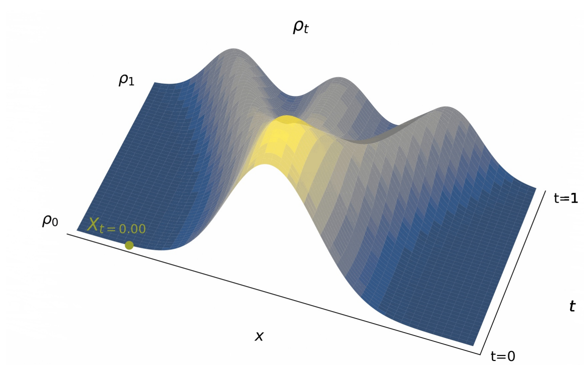
$$(4) \qquad b_t = \mathbb{E}\left[\dot{I}_t \,\middle|\, I_t = x\right] \qquad \text{Expectation over } \rho(x_0,x_1) \text{ conditional on } I_t = x.$$

- b_t can be learned over neural networks by minimizing

$$(5) \quad L_b[\hat{b}] = \int_0^1 \mathbb{E}_{x_0,x_1}\left[|\hat{b}_t(I_t) - \dot{I}_t|^2\right] dt$$

Then use b_t coming from (5) to generate samples by numerically solving (1)

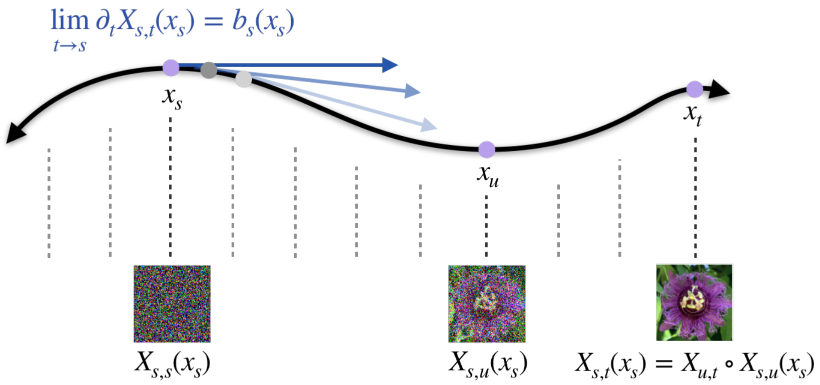




Powerful! But limitation: Sampling requires many evaluations of b_t to solve (1).
How can we avoid this?

The flow map

Instead of solving (1), we may be interested in learning an arbitrary integrator for the equation in terms of a flow map:



$$\boxed{X_{s,t}(x_s) = x_t} \qquad (6)$$

"Takes steps of arbitrary size $t-s$ along trajectories of the probability flow"

Properties of the flow map

Semigroup property: $\boxed{X_{u,t}\left(X_{s,u}(x_s)\right) = X_{s,t}(x_s)} = x_t$ (7)

↳ Invertibility $X_{s,t}(X_{t,s}(x_t)) = x_t$ (8)

Lagrangian eqn: $\partial_t X_{s,t}(x_s) = \underset{(1)}{\dot{x}_t} = b_t(x_t) = b_t(X_{s,t}(x))$

↳ $$\boxed{\partial_t X_{s,t}(x) = b_t\left(X_{s,t}(x)\right)} \qquad (9)$$

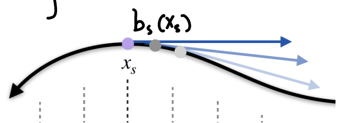
Eulerian eqn : Take a total derivative of (8)

$$\frac{d}{ds} X_{s,t}(X_{t,s}(x)) = \frac{\partial}{\partial s} X_{s,t}(X_{t,s}(x)) + \nabla X_{s,t}(X_{t,s}(x)) \cdot \underbrace{\frac{\partial}{\partial s} X_{t,s}(x)}_{b_s(X_{t,s}(x))} = 0$$

Evaluate it at $X_{t,s} = y$, so that

$$\boxed{\frac{\partial}{\partial s} X_{s,t}(x) + \nabla X_{s,t}(x) \cdot b_s(x) = 0} \qquad (10)$$

Tangent Condition :



$$\boxed{\lim_{s \to t} \partial_t X_{s,t}(x) = b_t(x)} \qquad (11)$$

Parameterizing the Flow map

Choose $\hat{X}_{s,t}(x) = x + (t-s)\hat{v}_{s,t}(x)$ (12), then, using (11) $\boxed{v_{t,t}(x) = b_t(x)}$ (13)

will be learned as an NN

Proposition: If $X_{s,t}$ is given by (12) and $v_{s,t}$ satisfies (13),

A: Lagrangian

$$\partial_t X_{s,t}(x) = v_{t,t}(X_{s,t}(x))$$

B: Eulerian

$$\frac{\partial}{\partial s} X_{s,t}(x) + \nabla X_{s,t}(x) \cdot b_s(x) = 0$$

C: consistency

$$X_{u,t}(X_{s,u}(x_s)) = X_{s,t}(x_s)$$

each characterize the flow map!

Let's use them, along with (13), to learn $X_{s,t}$ directly!

Flow maps via self-distillation

Self-distillation

Objective Function:

Learn $v_{t,t} = b_t$ on the diagonal using (5)

Learn the flow map on the off-diagonal w/ A, B, C or any combination

$$L_{SD}(\hat{v}) = L_b(\hat{v}) + L_D(\hat{v}) \qquad (14)$$

Example: Lagrangian Self-Distillation

- $L_b(\hat{v}) = \int_0^1 \mathbb{E}_{x_0,x_1}\left[\left| \hat{v}_{t,t}(I_t) - \dot{I}_t \right|^2 \right] dt$ Rewriting of (5) w/ $v_{t,t}$

- $L_D^{LSD}(\hat{v}) = \int_0^1 \int_0^t \mathbb{E}_{x_0,x_1}\left[\left| \partial_t \hat{X}_{s,t}(I_s) - \hat{v}_{t,t}(\hat{X}_{s,t}(I_s)) \right|^2 \right] ds\, dt$ PINN enforcing A

Others

- $L^{ESD}(\hat{v}) = \int_0^1 \int_0^t \mathbb{E}_{x_0,x_1}\left[\left| \partial_s \hat{X}_{s,t}(I_s) + \nabla \hat{X}_{s,t}(I_s)\, \hat{v}_{s,s}(I_s) \right|^2 \right] ds\, dt$ PINN enforcing B

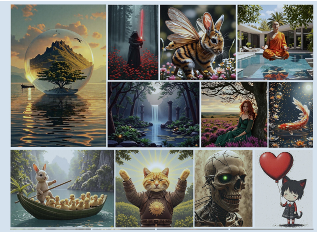
- $L^{PSD}(\hat{v}) = \int_0^1 \int_0^t \int_s^t \mathbb{E}_{x_0,x_1}\left[\left| \hat{X}_{s,t}(I_s) - \hat{X}_{u,t}(\hat{X}_{s,u}(I_s)) \right|^2 \right] du\, ds\, dt$ Enforcing C

Connection w/ the literature

Fixed point: Page

Distill from a known velocity field $b_t(x)$:

$$\mathbb{E}\,\frac{1}{2}\left|\partial_s X_{s,t}(I_s) + \underbrace{b_s(I_s)\cdot\nabla X_{s,t}(I_s)}_{\text{stop gradient on these terms}}\right|^2$$



"Align your flow"
Sabour, Fidler, Kreis
Thm 3.2/3.3

Take ∇_θ gradient of objective:

$$\mathbb{E}\,\nabla_\theta \frac{1}{2}\left[\partial_s X_{s,t}(I_s) + \text{stopgrad}\left(b_s(I_s)\cdot\nabla X_{s,t}(I_s)\right)\right]^2$$

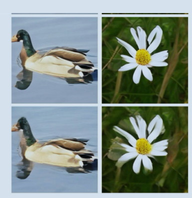
$$= \mathbb{E}\left[\nabla_\theta \partial_s X_{s,t}(I_s)\right]\left(\partial_s X_s(I_s) + b_s(I_s)\cdot\nabla_x X_{s,t}(I_s)\right)$$

Eulerian Map Distillation with stopgrad

Thm 3.2 in AYF Paper

Shortcut Models is Progressive Self Distillation

$$L^{PSD}(\hat{v}) = \int_0^1\int_0^t\int_s^t \mathbb{E}_{x_0,x_1}\left[\left|\hat{X}_{s,t}(I_s) - \hat{X}_{u,t}(\hat{X}_{s,u}(I_s))\right|^2\right] du\,ds\,dt$$



"Shortcut Models"
Frans et al

Connection w/ Mean Flow

Geng et al 2505.13447

- Take the Eulerian Self-Distillation term again

$$\min_{v} \int_{[0,1]^2} \mathbb{E}\, \tfrac{1}{2} \Big| \underbrace{\partial_s X_{s,t}(I_s) + v_{s,s}(I_s) \cdot \nabla X_{s,t}(I_s)}_{\star} \Big|^2 \, ds\, dt$$

- Plug in $X_{s,t}(x) = x + (t-s)\, v_{s,t}(x)$, which means that some terms simplify:

$$\partial_s X_{s,t}(x) = -v_{s,t}(x) + (t-s)\, \partial_s v_{s,t}(x)$$

- Plug this into $\star$:

$$\mathbb{E}\, \tfrac{1}{2} \Big| -v_{s,t}(I_s) + \underbrace{(t-s)\, \partial_s v_{s,t}(I_s) + v_{s,s}(I_s) \cdot \nabla X_{s,t}(I_s)}_{\text{stopgrad this whole term}} \Big|^2$$

- Then the gradient reads

$$\mathbb{E} -\nabla_\theta v_{s,t}(I_s) \cdot \Big[-v_{s,t}(I_s) + (t-s)\, \partial_s v_{s,t}(I_s) + v_{s,s}(I_s) \cdot \nabla X_{s,t}(I_s) \Big]$$

- Expand the last gradient $\nabla X_{s,t}$ using the definition of $X_{s,t}$:

$$b_t = \mathbb{E}[\dot{I}_t \mid I_t]$$

$$= \mathbb{E} -\nabla_\theta v_{s,t}(I_s) \cdot \Big[-v_{s,t}(I_s) + (t-s)\, \partial_s v_{s,t}(I_s) + v_{s,s}(I_s) + (t-s)\, v_{s,s}(I_s) \cdot \nabla v_{s,t}(I_s) \Big]$$

- Now, because this is linear in $v_{s,s}(I_s)$, it can be replaced with $\dot{I}_s$ here

$$= \mathbb{E} -\nabla_\theta v_{s,t}(I_s) \cdot \Big[-v_{s,t}(I_s) + \underline{(t-s)\, \partial_s v_{s,t}(I_s)} + v_{s,s}(I_s) + \underline{(t-s)\, \dot{I}_s \cdot \nabla v_{s,t}(I_s)} \Big]$$

- And this means finally that terms in blue can be collected as $\frac{d}{ds} v_{s,t}(I_s)$:

$$= \mathbb{E}_{s,t}\, -\nabla_\theta v_{s,t}(I_s) \cdot \left[- v_{s,t}(I_s) + (t-s) \frac{d}{ds} v_{s,t}(I_s) \right]$$

Directly learning the flow map solely in terms of $\hat{v}_{s,t}$!

Why? (4)

Take derivative of $\rho_t(x) = \int_{\mathbb{R}^d \times \mathbb{R}^d} \delta(x - I_t)\, \rho(x_0, x_1)\, dx_0\, dx_1$

$$\partial_t \rho_t(x) = -\nabla \cdot \int \dot{I}_t\, \delta(x - I_t)\, \rho(x_0, x_1)\, dx_0\, dx_1$$

$$= -\nabla \cdot \underbrace{j_t}_{\text{current } j_t} = -\nabla \cdot (b_t \rho_t)$$

$$\text{so } b_t = \frac{\int \dot{I}_t\, \delta(x - I_t)\, \rho(x_0, x_1)\, dx_0\, dx_1}{\int \delta(x - I_t)\, \rho(x_0, x_1)\, dx_0\, dx_1} = \mathbb{E}[\dot{I}_t \mid I_t = x]$$