Temporal Difference Flows [FPT+25]

Kellen Kanarios

July 10, 2025

Outline

1. Motivation
2. Flow Matching
3. Application to Reinforcement Learning

Variational inference with normalizing flows [RM]

- Want to maximize log-likelihood $\log p_\theta(\mathbf{x})$. Introduce latent $\mathbf{z} \sim p$.

$$\log p_\theta(\mathbf{x}) \geq \mathbb{D}_{\mathrm{KL}}[q_\phi(\mathbf{z} \mid \mathbf{x}) \mid\mid p(\mathbf{z})] + \mathbb{E}_q[\log p_\theta(\mathbf{x} \mid \mathbf{z})]. \quad \text{(ELBO)}$$

- Simultaneously learn q_ϕ and p_θ

Variational inference with normalizing flows [RM]

- Want to maximize log-likelihood $\log p_\theta(\mathbf{x})$. Introduce latent $\mathbf{z} \sim p$.

$$\log p_\theta(\mathbf{x}) \geq \mathbb{D}_{\mathsf{KL}}[q_\phi(\mathbf{z} \mid \mathbf{x}) \mid\mid p(\mathbf{z})] + \mathbb{E}_q[\log p_\theta(\mathbf{x} \mid \mathbf{z})]. \quad \text{(ELBO)}$$

- Simultaneously learn q_ϕ and p_θ

Problem. Need family q_ϕ to contain $p_\theta(z \mid x)$.

- Rarely the case.

Variational inference with normalizing flows [RM]

- Want to maximize log-likelihood $\log p_\theta(\mathbf{x})$. Introduce latent $\mathbf{z} \sim p$.

$$\log p_\theta(\mathbf{x}) \geq \mathbb{D}_{\mathsf{KL}}[q_\phi(\mathbf{z} \mid \mathbf{x}) \,||\, p(\mathbf{z})] + \mathbb{E}_q[\log p_\theta(\mathbf{x} \mid \mathbf{z})]. \quad \text{(ELBO)}$$

- Simultaneously learn q_ϕ and p_θ

Problem. Need family q_ϕ to contain $p_\theta(z \mid x)$.

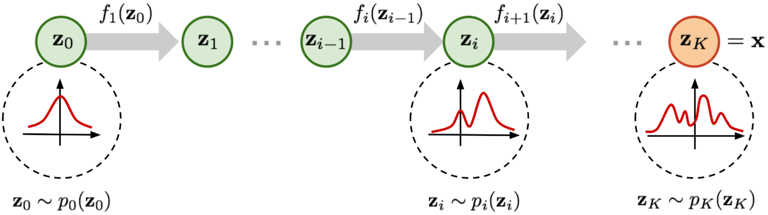
- Rarely the case.

How do we create a more expressive family that we can still optimize?

Given target distribution q. Pick some initial distribution p_0, where we can sample $\mathbf{z}_0 \sim p_0$.

Idea: *Learn sequence of functions $f_1, f_2, \ldots, f_K$, such that*

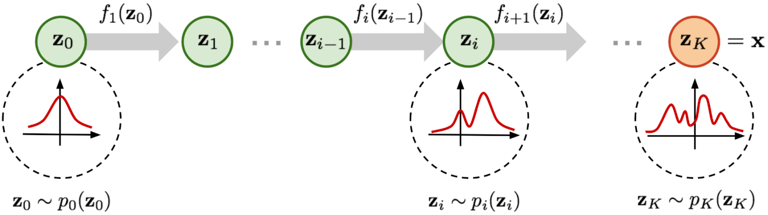
$$f_K \circ f_{K-1} \cdots \circ f_1(\mathbf{z}_0) \sim q$$



Given target distribution q. Pick some initial distribution p_0, where we can sample $\mathbf{z}_0 \sim p_0$.

Idea: *Learn sequence of functions $f_1, f_2, \ldots, f_K$, such that*

$$f_K \circ f_{K-1} \cdots \circ f_1(\mathbf{z}_0) \sim q$$

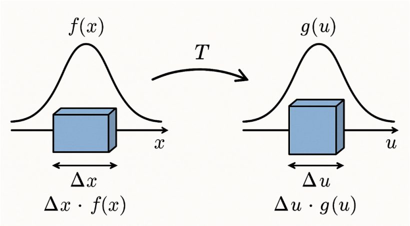


How do we learn such a sequence of functions f_i?

Given **invertible**, **smooth both ways** function $T : \mathbb{R}^d \to \mathbb{R}^d$.

Change of variables theorem

$$\int_{\mathbb{R}^d} p_0(\mathbf{x})d\mathbf{x} = \int_{\mathbb{R}^d} p_0(T^{-1}(\mathbf{y})) \left|\det \frac{\partial T}{\partial y}\right|^{-1} d\mathbf{y}.$$



$\Rightarrow$ Density of $f_1(\mathbf{z}_0) = p_0(\mathbf{z}_0) \left|\det \frac{\partial f_1}{\partial \mathbf{z}}\right|^{-1}$

$\Rightarrow \log p_K(\mathbf{z}_k) = \log p_0(\mathbf{z}_0) - \sum_{k=1}^{K} \left|\det \frac{\partial f_k}{\partial \mathbf{z}_{k-1}}\right|^{-1}$

$$\boxed{\log p_K(\mathbf{z}_k) = \log p_0(\mathbf{z}_0) - \sum_{k=1}^{K} \left|\det \frac{\partial f_k}{\partial \mathbf{z}_{k-1}}\right|^{-1}}$$

1. Given $\mathbf{x}^{(i)}$ from dataset, compute $\mathbf{x}_0^{(i)} = (f_K \circ \cdots \circ f_1)^{-1}(\mathbf{x}^{(i)})$

$$\boxed{\log p_K(\mathbf{z}_k) = \log p_0(\mathbf{z}_0) - \sum_{k=1}^{K} \left| \det \frac{\partial f_k}{\partial \mathbf{z}_{k-1}} \right|^{-1}}$$

1. Given $\mathbf{x}^{(i)}$ from dataset, compute $\mathbf{x}_0^{(i)} = (f_K \circ \cdots \circ f_1)^{-1}(\mathbf{x}^{(i)})$
2. Maximize log-likelihood

$$\max_{\theta} \left[\log p_0(\mathbf{x}_0^{(i)}(\theta)) - \sum_{k=1}^{K} \left| \det \frac{\partial f_k(\theta)}{\partial \mathbf{x}_{k-1}^{(i)}} \right|^{-1} \right].$$

$$\boxed{\log p_K(\mathbf{z}_k) = \log p_0(\mathbf{z}_0) - \textstyle\sum_{k=1}^{K} \left|\det \frac{\partial f_k}{\partial \mathbf{z}_{k-1}}\right|^{-1}}$$

1. Given $\mathbf{x}^{(i)}$ from dataset, compute $\mathbf{x}_0^{(i)} = (f_K \circ \cdots \circ f_1)^{-1}(\mathbf{x}^{(i)})$
2. Maximize log-likelihood

$O(LD^3)$

$$\max_{\theta} \left[\log p_0(\mathbf{x}_0^{(i)}(\theta)) - \sum_{k=1}^{K} \left|\det \frac{\partial f_k(\theta)}{\partial \mathbf{x}_{k-1}^{(i)}}\right|^{-1}\right].$$

$$\boxed{\log p_K(\mathbf{z}_k) = \log p_0(\mathbf{z}_0) - \sum_{k=1}^{K} \left|\det \frac{\partial f_k}{\partial \mathbf{z}_{k-1}}\right|^{-1}}$$

1. Given $\mathbf{x}^{(i)}$ from dataset, compute $\mathbf{x}_0^{(i)} = (f_K \circ \cdots \circ f_1)^{-1}(\mathbf{x}^{(i)})$
2. Maximize log-likelihood

$O(LD^3)$

$$\max_{\theta} \left[\log p_0(\mathbf{x}_0^{(i)}(\theta)) - \sum_{k=1}^{K} \left|\det \frac{\partial f_k(\theta)}{\partial \mathbf{x}_{k-1}^{(i)}}\right|^{-1}\right].$$

Need to pick **simple** transformations i.e.

$$f(\mathbf{z}) = \mathbf{z} + \mathbf{u}h(\mathbf{w}^\top \mathbf{z} + b).$$

Require **very large** K to represent complex distributions.

$$\boxed{\log p_K(\mathbf{z}_k) = \log p_0(\mathbf{z}_0) - \sum_{k=1}^{K} \left| \det \frac{\partial f_k}{\partial \mathbf{z}_{k-1}} \right|^{-1}}$$

1. Given $\mathbf{x}^{(i)}$ from dataset, compute $\mathbf{x}_0^{(i)} = (f_K \circ \cdots \circ f_1)^{-1}(\mathbf{x}^{(i)})$
2. Maximize log-likelihood

$O(LD^3)$

$$\max_{\theta} \left[\log p_0(\mathbf{x}_0^{(i)}(\theta)) - \sum_{k=1}^{K} \boxed{\left| \det \frac{\partial f_k(\theta)}{\partial \mathbf{x}_{k-1}^{(i)}} \right|^{-1}} \right].$$

Need to pick **simple** transformations i.e.

$$f(\mathbf{z}) = \mathbf{z} + \mathbf{u}h(\mathbf{w}^\top \mathbf{z} + b).$$

Require **very large** K to represent complex distributions.

Why not take $K \to \infty$...

Neural Ordinary Differential Equations [CRBD]

Replace $\mathbf{z}_{t+1} = f_t(\mathbf{z}_t)$ with the **ODE**

$$\frac{\mathrm{d}\mathbf{z}}{\mathrm{d}t} = f_t(\mathbf{z}_t).$$

Remark. *Instantaneous COV*

$$\frac{\partial \log p_t(\mathbf{z}_t)}{\partial t} = -\mathrm{tr}\left(\frac{df}{d\mathbf{z}_t}\right)$$

Neural Ordinary Differential Equations [CRBD]

Replace $\mathbf{z}_{t+1} = f_t(\mathbf{z}_t)$ with the **ODE**

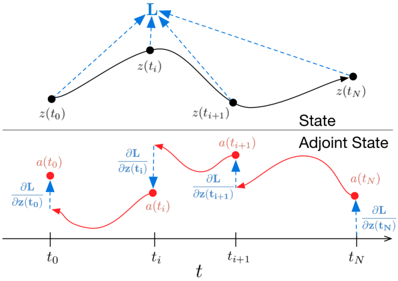
$$\frac{d\mathbf{z}}{dt} = f_t(\mathbf{z}_t).$$

Remark. *Instantaneous COV*

$$\frac{\partial \log p_t(\mathbf{z}_t)}{\partial t} = -\mathrm{tr}\left(\frac{df}{d\mathbf{z}_t}\right)$$

Given $\mathbf{x}^{(i)}$ from dataset, now must compute $\mathbf{x}_0^{(i)} = \int_1^0 f_t^{-1}(\mathbf{x}_t^{(i)})dt$

- Requires **exact** ODE solve for unbiased gradient.



Neural Ordinary Differential Equations [CRBD]

Replace $\mathbf{z}_{t+1} = f_t(\mathbf{z}_t)$ with the **ODE**

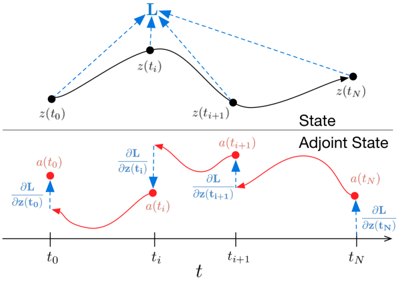
$$\frac{d\mathbf{z}}{dt} = f_t(\mathbf{z}_t).$$

Remark. *Instantaneous COV*

$$\frac{\partial \log p_t(\mathbf{z}_t)}{\partial t} = -\mathrm{tr}\left(\frac{df}{d\mathbf{z}_t}\right)$$

Given $\mathbf{x}^{(i)}$ from dataset, now must compute $\mathbf{x}_0^{(i)} = \int_1^0 f_t^{-1}(\mathbf{x}_t^{(i)})dt$

- Requires **exact** ODE solve for unbiased gradient.



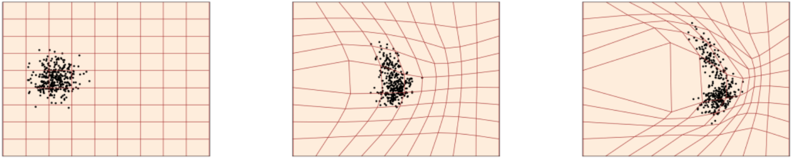
Can we learn f_t without backprop through an ODE?

Flow Matching for Generative Modeling [LCB+23]

Disclaimer: f_t will now often be referred to as the *flow*.

Book keeping. We say the *flow* f_t *generates* a probability path p_t if $X_t = f_t(X_0) \sim p_t$. Equivalently,

$$p_t(x) = [f_{t\sharp} p_0](x) \triangleq p_0(f_t^{-1}(y)) \left| \det \partial_y f_t^{-1}(y) \right| .$$



Key Idea: Do **not** learn the *flow* f_t, learn the **velocity** of the flow.

Def. This velocity is a *vector field* $u_t : \mathbb{R}^d \to \mathbb{R}^d$, such that

$$\frac{\mathrm{d}f_t(\mathbf{x})}{\mathrm{d}t} = u_t(f_t(\mathbf{x})).$$

Key Idea: Do **not** learn the *flow* f_t, learn the **velocity** of the flow.

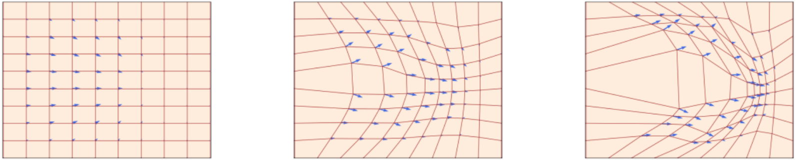
Def. This velocity is a *vector field* $u_t : \mathbb{R}^d \to \mathbb{R}^d$, such that

$$\frac{\mathrm{d}f_t(\mathbf{x})}{\mathrm{d}t} = u_t(f_t(\mathbf{x})).$$

To sample, solve ODE at **sample time** i.e.

$$f_{t+\Delta t}(\mathbf{x}) \approx f_t(\mathbf{x}) + \Delta t \cdot u_t(f_t(\mathbf{x})) \qquad \text{(Euler method)}$$

By uniqueness of ODE, (vector field u_t) $\leftrightarrow$ (flow f_t).



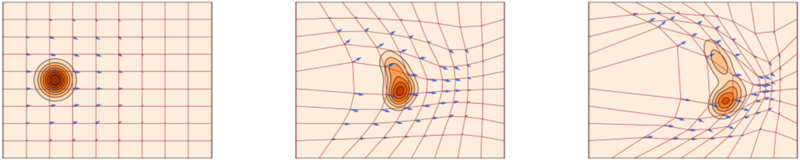
Def. A *vector field* u_t *generates* a probability path p_t if its corresponding flow f_t *generates* p_t.

Thm. A *vector field* u_t *generates* a probability path p_t if and only if it satisfies the *continuity equation*.

Continuity Equation.

$$\frac{\mathrm{d}}{\mathrm{d}t} p_t(x) + \mathrm{div}(p_t u_t)(x) = 0,$$

where $\mathrm{div}(v)(x) = \sum_{i=1}^{d} \partial_{x^i} v^i(x)$, and $v(x) = (v^1(x), \ldots, v^d(x))$.



Continuity Equation.

$$\frac{\mathrm{d}}{\mathrm{d}t}p_t(x) + \mathrm{div}(p_t u_t)(x) = 0,$$

where $\mathrm{div}(v)(x) = \sum_{i=1}^{d} \partial_{x^i} v^i(x)$, and $v(x) = (v^1(x), \ldots, v^d(x))$.

Proof: (If people care)

$$\frac{d}{dt}\mathbb{E}f(X_t) = \frac{d}{dt}\int_{\mathbb{R}^d} f(x)p_t(x)dx = \int_{\mathbb{R}^d} f(x)\partial_t p_t(x)dx,$$

$$\begin{aligned}
\mathbb{E}\frac{d}{dt}f(X_t) &= \int_{\mathbb{R}^d} \left(\nabla f(x_t)\cdot v_t(x_t)\right) p_t(x_t)dx_t \\
&= 0 - \int_{\mathbb{R}^d} f(x_t)\mathrm{div}(v_t(x_t)p_t(x_t))dx_t \qquad \text{(IBP)} \\
&= \int_{\mathbb{R}^d} -f(x)\mathrm{div}(v_t(x)p_t(x))dx..
\end{aligned}$$

The result follows from fundamental lemma of calculus of variations.

How do we actually learn the vector field u_t?

Flow Model Loss

$$\mathcal{L}_{\text{FM}}(\theta) = \mathbb{E}_{t,p_t(x)} \| v_t(x;\theta) - u_t(x) \|^2$$

- $p_t : \mathcal{X} \to [0,1]$: probability density path.
- $u_t : \mathbb{R}^d \to \mathbb{R}^d$: vector field that *generates* p_t.

Flow Model Loss

$$\mathcal{L}_{\mathsf{FM}}(\theta) = \mathbb{E}_{t, p_t(x)} \| v_t(x; \theta) - u_t(x) \|^2$$

- $p_t : \mathcal{X} \rightarrow [0, 1]$: probability density path.
- $u_t : \mathbb{R}^d \rightarrow \mathbb{R}^d$: vector field that *generates* p_t.

Intractable.

1. How to choose p_t?

Flow Model Loss

$$\mathcal{L}_{\text{FM}}(\theta) = \mathbb{E}_{t, p_t(x)} \left\| v_t(x; \theta) - u_t(x) \right\|^2$$

- $p_t : \mathcal{X} \to [0, 1]$: probability density path.
- $u_t : \mathbb{R}^d \to \mathbb{R}^d$: vector field that *generates* p_t.

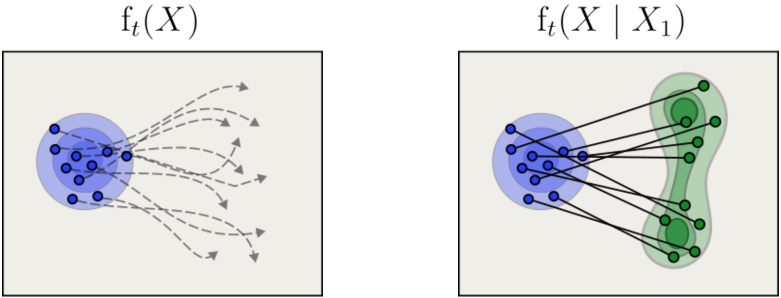
Intractable.

1. How to choose p_t?
2. Given p_t need to solve continuity equation for u_t, which is a **PDE** likely without close-form.

Key Idea: *Given the endpoint X_1, we can easily construct a path between X and X_1.*

Ex: *Consider*

$$f_t(x \mid x_1) = tx_1 + (1 - (1 - \sigma_{\min})t)x$$



What does the conditional flow being nice have to do with our problem?

Idea 1: Given target data distribution q, approximate q by marginalizing "nice" conditionals by q i.e.

$$p_t(\mathbf{x}) = \int p_t(\mathbf{x} \mid \mathbf{x}_1) q(\mathbf{x}_1) \mathrm{d}\mathbf{x}_1. \qquad \text{(Marginal Density)}$$

For $p_1(\mathbf{x} \mid \mathbf{x}_1) \sim N(x_1, \sigma_1)$, with $\sigma_1 \ll 1$, $p_1(\mathbf{x}) \approx q(\mathbf{x})$.

Idea 1: Given target data distribution q, approximate q by marginalizing "nice" conditionals by q i.e.

$$p_t(\mathbf{x}) = \int p_t(\mathbf{x} \mid \mathbf{x}_1) q(\mathbf{x}_1) \mathrm{d}\mathbf{x}_1. \qquad \text{(Marginal Density)}$$

For $p_1(\mathbf{x} \mid \mathbf{x}_1) \sim N(x_1, \sigma_1)$, with $\sigma_1 \ll 1$, $p_1(\mathbf{x}) \approx q(\mathbf{x})$.

Idea 2: Use this to define an approximate vector field

$$u_t(\mathbf{x}) = \int u_t(\mathbf{x} \mid \mathbf{x}_1) \frac{p_t(\mathbf{x} \mid \mathbf{x}_1) q(\mathbf{x}_1)}{p_t(\mathbf{x})} \mathrm{d}\mathbf{x}_1. \quad \text{(Marginal Vector Field)}$$

Here, $u_t(\mathbf{x} \mid \mathbf{x}_1)$ is the conditional vector field that generates $p_t(\mathbf{x} \mid \mathbf{x}_1)$.

Idea 1: Given target data distribution q, approximate q by marginalizing "nice" conditionals by q i.e.

$$p_t(\mathbf{x}) = \int p_t(\mathbf{x} \mid \mathbf{x}_1) q(\mathbf{x}_1) \mathrm{d}\mathbf{x}_1. \qquad \text{(Marginal Density)}$$

For $p_1(\mathbf{x} \mid \mathbf{x}_1) \sim N(x_1, \sigma_1)$, with $\sigma_1 \ll 1$, $p_1(\mathbf{x}) \approx q(\mathbf{x})$.

Idea 2: Use this to define an approximate vector field

$$u_t(\mathbf{x}) = \int u_t(\mathbf{x} \mid \mathbf{x}_1) \frac{p_t(\mathbf{x} \mid \mathbf{x}_1) q(\mathbf{x}_1)}{p_t(\mathbf{x})} \mathrm{d}\mathbf{x}_1. \qquad \text{(Marginal Vector Field)}$$

Here, $u_t(\mathbf{x} \mid \mathbf{x}_1)$ is the conditional vector field that generates $p_t(\mathbf{x} \mid \mathbf{x}_1)$.

Thm. *The marginal vector field u_t generates the marginal probability path p_t.* (Check continuity equation)

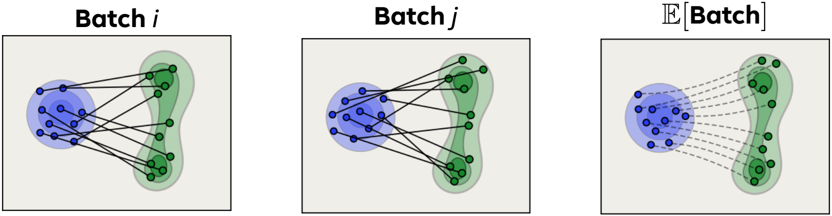
Conditional Flow Model Loss

$$\mathcal{L}_{\text{CFM}}(\theta) = \mathbb{E}_{t,q(x_1),p_t(x|x_1)}\|v_t(x;\theta) - u_t(x \mid x_1)\|^2$$

Conditional Flow Model Loss

$$\mathcal{L}_{\text{CFM}}(\theta) = \mathbb{E}_{t,q(x_1),p_t(x|x_1)} \left\| v_t(x;\theta) - u_t(x \mid x_1) \right\|^2$$

Thm. *Up to a constant independent of θ, $\mathcal{L}_{FM}(\theta) = \mathcal{L}_{CFM}(\theta)$. In particular, $\nabla\mathcal{L}_{FM}(\theta) = \nabla\mathcal{L}_{CFM}(\theta)$.*



How do we actually use this?

Example: Optimal Transport

For $f_t(x) = \mu_t(x_1) + x\sigma_t(x_1)$, we consider

$$\mu_t = tx_1, \quad \sigma_t(x) = 1 - (1 - \sigma_{\text{min}})t,$$

such that

Example: Optimal Transport

For $f_t(x) = \mu_t(x_1) + x\sigma_t(x_1)$, we consider

$$\mu_t = tx_1, \quad \sigma_t(x) = 1 - (1 - \sigma_{\min})t,$$

such that

$$f_t^{-1}(x) = \frac{x - tx_1}{1 - (1 - \sigma_{\min})t}, \quad \frac{\mathrm{d}}{\mathrm{d}t} f_t(x) = x_1 - (1 - \sigma_{\min})x.$$

Example: Optimal Transport

For $f_t(x) = \mu_t(x_1) + x\sigma_t(x_1)$, we consider

$$\mu_t = tx_1, \quad \sigma_t(x) = 1 - (1 - \sigma_{\text{min}})t,$$

such that

$$f_t^{-1}(x) = \frac{x - tx_1}{1 - (1 - \sigma_{\text{min}})t}, \quad \frac{\mathrm{d}}{\mathrm{d}t} f_t(x) = x_1 - (1 - \sigma_{\text{min}})x.$$

Recall

$$\frac{\mathrm{d}}{\mathrm{d}t} f_t(x) = u(f_t(x) \mid x_1) \implies u(x \mid x_1) = \frac{\mathrm{d}}{\mathrm{d}t} f_t(f_t^{-1}(x)).$$

Example: Optimal Transport

Recall

$$\frac{\mathrm{d}}{\mathrm{d}t} f_t(x) = u(f_t(x) \mid x_1) \implies u(x \mid x_1) = \frac{\mathrm{d}}{\mathrm{d}t} f_t(f_t^{-1}(x)).$$

Thus,

$$u_t(x \mid x_1) = \frac{x_1 - (1 - \sigma_{\text{min}})x}{1 - (1 - \sigma_{\text{min}})t}.$$

Example: Optimal Transport

Recall

$$\frac{\mathrm{d}}{\mathrm{d}t} f_t(x) = u(f_t(x) \mid x_1) \implies u(x \mid x_1) = \frac{\mathrm{d}}{\mathrm{d}t} f_t(f_t^{-1}(x)).$$

Thus,

$$u_t(x \mid x_1) = \frac{x_1 - (1 - \sigma_{\text{min}})x}{1 - (1 - \sigma_{\text{min}})t}.$$

$$\boxed{\mathcal{L}_{\text{CFM}}(\theta) = \mathbb{E}_{t, q(x_1), p_t(x|x_1)} \left\| v_t(x; \theta) - \frac{x_1 - (1-\sigma_{\text{min}})x}{1-(1-\sigma_{\text{min}})t} \right\|^2}$$

Diffusion Meets Flow Matching: Two Sides of the Same Coin [GHH$^+$24]

Recall that instead of an ODE, in diffusion, we have a **SDE**

$$\mathrm{d}\mathbf{x}_t = f_t(\mathbf{x}_t)\mathrm{d}t + \sigma(\mathbf{x}_t)\mathrm{d}B_t.$$

However, for *OU* process, we can actually absorb Brownian motion term and get vector field

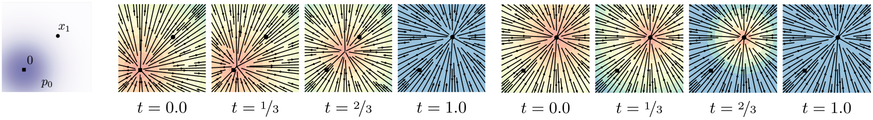
$$u_t(\mathbf{x}_t) = -(\mathbf{x}_t + \nabla \ln p_t(\mathbf{x}_t)).$$

Similarly, learning the score function $\nabla \ln p_t(\mathbf{x}_t)$ can be rewritten as the flow matching objective for certain choices of p_t see [LHH$^+$24].

Why Flows?

Anecdotally.

- *The OT path's conditional vector field has constant direction in time and is arguably simpler to fit with a parametric model. [LCB+23]*
- *The deterministic nature of ODEs equips flow-matching methods with simpler learning objectives and faster inference speed [ZPLE25]*



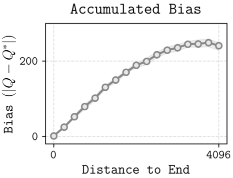
Conditional score

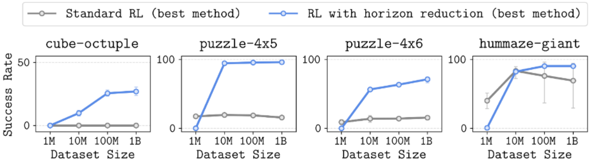
Conditional vector field

Q learning is not yet scalable [Par25]

Long horizon problems are hard.

$$\mathbb{E}_{(s,a,r,s')\sim\mathcal{D}}\left[\left(Q_\theta(s,a) - \underbrace{\left(r+\gamma\max_{a'} Q_{\bar{\theta}}(s',a')\right)}_{\texttt{Biased}}\right)^2\right].$$





Temporal Difference Flows [FPT+25]

Idea: Apply flow matching to learn the *successor measure* of an MDP.

Def. We define the *successor measure* as

$$m^\pi(\mathsf{X} \mid s, a) = (1-\gamma) \sum_{k=0}^{\infty} \gamma^k \Pr(S_{k+1} \in \mathsf{X} \mid S_0 = s, A_0 = a, \pi),$$

Recall. The *successor measure* is the unique fix point to the *Bellman* equation

$$\begin{aligned} m^\pi(\cdot \mid s, a) &= (\mathcal{T}^\pi m^\pi)(\cdot \mid s, a) \\ &:= (1-\gamma) P(\cdot \mid s, a) + \gamma (P^\pi m^\pi)(\cdot \mid s, a), \end{aligned}$$

where

$$(P^\pi m)(\mathrm{d}x \mid s, a) = \int_{s'} P(\mathrm{d}s' \mid s, a) m(\mathrm{d}x \mid s', \pi(s')).$$

The most straightforward idea is to just substitute m^π for q in flow matching i.e.

$$\mathcal{L}_{\text{MC-CFM}}(\theta) = \mathbb{E}_{\rho,t,Z,X_t}\left[\left\|\tilde{v}_t(X_t \mid S, A; \theta) - u_t(X_t \mid Z)\right\|^2\right],$$

where $Z = X_1 \sim m^\pi(\cdot \mid S, A), X_t \sim p_t(\cdot \mid Z)$.

Here we just use the optimal transport conditional vector field and corresponding probability density path.

The most straightforward idea is to just substitute m^π for q in flow matching i.e.

$$\mathcal{L}_{\text{MC-CFM}}(\theta) = \mathbb{E}_{\rho,t,Z,X_t}\left[\left\|\tilde{v}_t(X_t \mid S, A; \theta) - u_t(X_t \mid Z)\right\|^2\right],$$

where $Z = X_1 \sim m^\pi(\cdot \mid S, A), X_t \sim p_t(\cdot \mid Z)$.

Here we just use the optimal transport conditional vector field and corresponding probability density path.

This requires direct access to samples from m^π.

Can we learn from offline one-step transitions (S, A, S')?

The most straightforward idea is to just substitute m^π for q in flow matching i.e.

$$\mathcal{L}_{\text{MC-CFM}}(\theta) = \mathbb{E}_{\rho,t,Z,X_t}\left[\left\|\tilde{v}_t(X_t \mid S, A; \theta) - u_t(X_t \mid Z)\right\|^2\right],$$

where $Z = X_1 \sim m^\pi(\cdot \mid S, A), X_t \sim p_t(\cdot \mid Z)$.

Here we just use the optimal transport conditional vector field and corresponding probability density path.

This requires direct access to samples from m^π.

Can we learn from offline one-step transitions (S, A, S')?

Leverage **recursive** structure of Bellman equation i.e.

$$\begin{aligned} X_0 &\sim p_0 \\ Z = X_1 &\sim (1-\gamma)\delta_{S'} + \gamma\delta_{\tilde{f}_1(X_0 \mid S', \pi(S'))}. \end{aligned} \tag{TD-CFM}$$

Leverage **recursive** structure of Bellman equation i.e.

$$X_0 \sim p_0$$
$$Z = X_1 \sim (1-\gamma)\delta_{S'} + \gamma\delta_{\tilde{f}_1(X_0|S',\pi(S'))}.$$

- With probability $(1-\gamma)$, $X_1 = S'$
- With probability γ, sample from $\tilde{m}^{\pi}$ by integrating $\tilde{f}_t$.

Can we do better?

Lemma. Let v_t^1 and v_t^2 be vector fields that generate the probability paths p_t^1 and p_t^2, respectively. Then, for any $\gamma \in [0,1]$, the mixture probability path $p_t = (1-\gamma)p_t^1 + \gamma p_t^2$ is generated by the vector field.

$$v_t := \frac{(1-\gamma)p_t^1 v_t^1 + \gamma p_t^2 v_t^2}{(1-\gamma)p_t^1 + \gamma p_t^2}.$$

Lemma. Let v_t^1 and v_t^2 be vector fields that generate the probability paths p_t^1 and p_t^2, respectively. Then, for any $\gamma \in [0,1]$, the mixture probability path $p_t = (1-\gamma)p_t^1 + \gamma p_t^2$ is generated by the vector field.

$$v_t := \frac{(1-\gamma)p_t^1 v_t^1 + \gamma p_t^2 v_t^2}{(1-\gamma)p_t^1 + \gamma p_t^2}.$$

Remember The *successor measure* is the unique fix point to the *Bellman* equation

$$m^\pi(\cdot \mid s, a) = (1-\gamma)P(\cdot \mid s, a) + \gamma(P^\pi m^\pi)(\cdot \mid s, a),$$

where

$$(P^\pi m)(\mathrm{d}x \mid s, a) = \int_{s'} P(\mathrm{d}s' \mid s, a) m(\mathrm{d}x \mid s', \pi(s')).$$

Lemma. Let v_t^1 and v_t^2 be vector fields that generate the probability paths p_t^1 and p_t^2, respectively. Then, for any $\gamma \in [0,1]$, the mixture probability path $p_t = (1-\gamma)p_t^1 + \gamma p_t^2$ is generated by the vector field.

$$v_t := \frac{(1-\gamma)p_t^1 v_t^1 + \gamma p_t^2 v_t^2}{(1-\gamma)p_t^1 + \gamma p_t^2}.$$

Remember The *successor measure* is the unique fix point to the *Bellman* equation

$$m^\pi(\cdot \mid s, a) = (1-\gamma)P(\cdot \mid s, a) + \gamma(P^\pi m^\pi)(\cdot \mid s, a),$$

where

$$(P^\pi m)(\mathrm{d}x \mid s, a) = \int_{s'} P(\mathrm{d}s' \mid s, a) m(\mathrm{d}x \mid s', \pi(s')).$$

Idea: Learn v_t for $p_t^1 = P(\cdot \mid S, A)$, $p_t^2 = (P^\pi m)(\cdot \mid S, A)$

Lemma. Let v_t^1 and v_t^2 be vector fields that generate the probability paths p_t^1 and p_t^2, respectively. For $\gamma \in [0,1]$, the vector field $v_t = \frac{(1-\gamma)p_t^1 v_t^1 + \gamma p_t^2 v_t^2}{(1-\gamma)p_t^1 + \gamma p_t^2}$ satisfies

$$v_t = \underset{v:\mathbb{R}^d \to \mathbb{R}^d}{\arg\min} \Big\{ (1-\gamma)\mathbb{E}_{x_t \sim p_t^1} \left[||v_t(x_t) - v_t^1(x_t)||^2 \right] + \gamma \mathbb{E}_{x_t \sim p_t^2} \left[||v_t(x_t) - v_t^2(x_t)||^2 \right] \Big\}.$$

Lemma. Let v_t^1 and v_t^2 be vector fields that generate the probability paths p_t^1 and p_t^2, respectively. For $\gamma \in [0,1]$, the vector field $v_t = \frac{(1-\gamma)p_t^1 v_t^1 + \gamma p_t^2 v_t^2}{(1-\gamma)p_t^1 + \gamma p_t^2}$ satisfies

$$v_t = \arg\min_{v:\mathbb{R}^d \to \mathbb{R}^d} \Big\{ (1-\gamma)\mathbb{E}_{x_t \sim p_t^1}\left[||v_t(x_t) - v_t^1(x_t)||^2\right] + \gamma \mathbb{E}_{x_t \sim p_t^2}\left[||v_t(x_t) - v_t^2(x_t)||^2\right] \Big\}.$$

Substitute.

$$v_t = \arg\min_{v:\mathbb{R}^d \to \mathbb{R}^d} \Big\{ (1-\gamma)\overbrace{\mathbb{E}_{x_t \sim P}\left[||v_t(x_t) - v_t^1(x_t)||^2\right]}^{\mathcal{L}_{\mathrm{OS}}} + \gamma \underbrace{\mathbb{E}_{x_t \sim P^{\pi}m}\left[||v_t(x_t) - v_t^2(x_t)||^2\right]}_{\mathcal{L}_{\mathrm{GHM}}} \Big\}.$$

Lemma. Let v_t^1 and v_t^2 be vector fields that generate the probability paths p_t^1 and p_t^2, respectively. For $\gamma \in [0,1]$, the vector field $v_t = \frac{(1-\gamma)p_t^1 v_t^1 + \gamma p_t^2 v_t^2}{(1-\gamma)p_t^1 + \gamma p_t^2}$ satisfies

$$v_t = \arg\min_{v:\mathbb{R}^d \to \mathbb{R}^d} \Big\{ (1-\gamma)\mathbb{E}_{x_t \sim p_t^1}\left[||v_t(x_t) - v_t^1(x_t)||^2 \right] + \gamma \mathbb{E}_{x_t \sim p_t^2}\left[||v_t(x_t) - v_t^2(x_t)||^2 \right] \Big\}.$$

Substitute.

$$v_t = \arg\min_{v:\mathbb{R}^d \to \mathbb{R}^d} \Big\{ (1-\gamma)\overbrace{\mathbb{E}_{x_t \sim P}\left[||v_t(x_t) - v_t^1(x_t)||^2 \right]}^{\mathcal{L}_{\text{OS}}} + \gamma \underbrace{\mathbb{E}_{x_t \sim P^{\pi}m}\left[||v_t(x_t) - v_t^2(x_t)||^2 \right]}_{\mathcal{L}_{\text{GHM}}} \Big\}.$$

How do we get v_t^1 and v_t^2?

Lemma. Let v_t^1 and v_t^2 be vector fields that generate the probability paths p_t^1 and p_t^2, respectively. For $\gamma \in [0,1]$, the vector field $v_t = \frac{(1-\gamma)p_t^1 v_t^1 + \gamma p_t^2 v_t^2}{(1-\gamma)p_t^1 + \gamma p_t^2}$ satisfies

$$v_t = \underset{v:\mathbb{R}^d \to \mathbb{R}^d}{\arg\min} \Big\{ (1-\gamma)\mathbb{E}_{x_t \sim p_t^1}\left[||v_t(x_t) - v_t^1(x_t)||^2 \right] + \gamma \mathbb{E}_{x_t \sim p_t^2}\left[||v_t(x_t) - v_t^2(x_t)||^2 \right] \Big\}.$$

Substitute.

$$v_t = \underset{v:\mathbb{R}^d \to \mathbb{R}^d}{\arg\min} \Big\{ (1-\gamma)\overbrace{\mathbb{E}_{x_t \sim P}\left[||v_t(x_t) - v_t^1(x_t)||^2 \right]}^{\mathcal{L}_{\text{OS}}} + \gamma \underbrace{\mathbb{E}_{x_t \sim P^{\pi} m}\left[||v_t(x_t) - v_t^2(x_t)||^2 \right]}_{\mathcal{L}_{\text{GHM}}} \Big\}.$$

How do we get v_t^1 and v_t^2? Conditioning trick **separately!**

Case 1: $\mathcal{L}_{\text{OS}}$, $S' \sim P(\cdot \mid S, A)$

$$\vec{v}_t(x \mid s, a) = \int \vec{u}_t(x \mid x_1) \frac{\vec{p}_t(x \mid x_1) P(\mathrm{d}x_1 \mid s, a)}{\vec{p}_t(x \mid s, a)},$$

Case 1: $\mathcal{L}_{\text{OS}}$, $S' \sim P(\cdot \mid S, A)$

$$\vec{v}_t(x \mid s, a) = \int \vec{u}_t(x \mid x_1) \frac{\vec{p}_t(x \mid x_1) P(\mathrm{d}x_1 \mid s, a)}{\vec{p}_t(x \mid s, a)},$$

$$\boxed{\begin{aligned} \mathcal{L}_{COS}(\theta) = \mathbb{E}_{\rho, t, Z, \vec{X}_t}\left[\left\|\tilde{v}_t(\vec{X}_t \mid S, A; \theta) - \vec{u}_t(\vec{X}_t \mid Z)\right\|^2\right], \\ \text{where } Z = X_1 \sim P(\cdot \mid S, A),\ \vec{X}_t \sim \vec{p}_t(\cdot \mid Z) \end{aligned}}$$

- $\vec{u}_t(\cdot \mid Z)$, $\vec{p}_t(\cdot \mid Z)$ can be chosen as in **traditional** flow matching i.e. OT.

Case 2: We sample future state

$$\widehat{v}_t^{(n)}(x \mid s, a) = \int v_t^{(n)}(x \mid s', a') \frac{m_t^{(n)}(x \mid s', a') P(\mathrm{d}s' \mid s, a)}{\widehat{p}_t^{(n)}(x \mid s, a)},$$

where $\widehat{p}_t^{(n)}(x \mid s, a) = \int m_t^{(n)}(x \mid s', a') P(\mathrm{d}s' \mid s, a)$, and $a' = \pi(s')$.

Case 2: We sample future state

$$\widehat{v}_t^{(n)}(x \mid s, a) = \int v_t^{(n)}(x \mid s', a') \frac{m_t^{(n)}(x \mid s', a') P(ds' \mid s, a)}{\widehat{p}_t^{(n)}(x \mid s, a)},$$

where $\widehat{p}_t^{(n)}(x \mid s, a) = \int m_t^{(n)}(x \mid s', a') P(ds' \mid s, a)$, and $a' = \pi(s')$.

Remark. They show if $v_t^{(n)}$ is optimizer of proposed loss then $v_t^{(n)}$ generates $m_t^{(n)}$.

Case 2: We sample future state

$$\widehat{v}_t^{(n)}(x \mid s, a) = \int v_t^{(n)}(x \mid s', a') \frac{m_t^{(n)}(x \mid s', a') P(\mathrm{d}s' \mid s, a)}{\widehat{p}_t^{(n)}(x \mid s, a)},$$

where $\widehat{p}_t^{(n)}(x \mid s, a) = \int m_t^{(n)}(x \mid s', a') P(\mathrm{d}s' \mid s, a)$, and $a' = \pi(s')$.

Remark. They show if $v_t^{(n)}$ is optimizer of proposed loss then $v_t^{(n)}$ generates $m_t^{(n)}$.

$$\mathcal{L}_{CGHM}(\theta) = \mathbb{E}_{\rho, t, \widehat{X}_t} \left[\left\| \tilde{v}_t(\widehat{X}_t \mid S, A; \theta) - \tilde{v}_t^{(n)}(\widehat{X}_t \mid S', \pi(S')) \right\|^2 \right],$$

where $X_0 \sim p_0$, $S' \sim P(\cdot \mid S, A)$, $\widehat{X}_t = \widetilde{f}_t^{(n)}(X_0 \mid S', \pi(S'))$, Combine the **objectives**!!

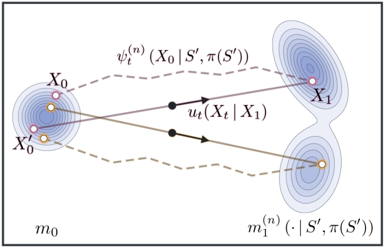
$$\boxed{\begin{gathered}\mathcal{L}_{COS}(\theta) = \mathbb{E}_{\rho,t,Z,\vec{X}_t}\left[\left\|\tilde{v}_t(\vec{X}_t \mid S, A; \theta) - \vec{u}_t(\vec{X}_t \mid Z)\right\|^2\right], \\ \text{where } Z = X_1 \sim P(\cdot \mid S, A),\ \vec{X}_t \sim \vec{p}_t(\cdot \mid Z)\end{gathered}}$$

$$+$$

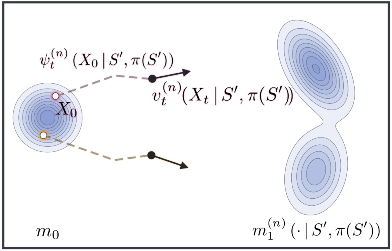
$$\boxed{\begin{gathered}\mathcal{L}_{CGHM}(\theta) = \mathbb{E}_{\rho,t,\widehat{X}_t}\left[\left\|\tilde{v}_t(\widehat{X}_t \mid S, A; \theta) - \tilde{v}_t^{(n)}(\widehat{X}_t \mid S', \pi(S')\right\|^2\right], \\ \text{where } X_0 \sim p_0, S' \sim P(\cdot \mid S, A), \widehat{X}_t = \widetilde{f}_t^{(n)}(X_0 \mid S', \pi(S')),\end{gathered}}$$

$$\mathcal{L}_{\mathrm{TD}^2-\mathrm{CFM}}(\theta) = (1-\gamma)\mathcal{L}_{COS}(\theta) + \gamma\mathcal{L}_{CGHM}(\theta) \implies \text{Lower } \mathbb{E}\nabla\mathcal{L}^2.$$

TD-CFM

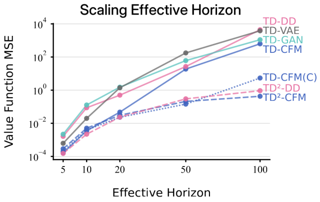


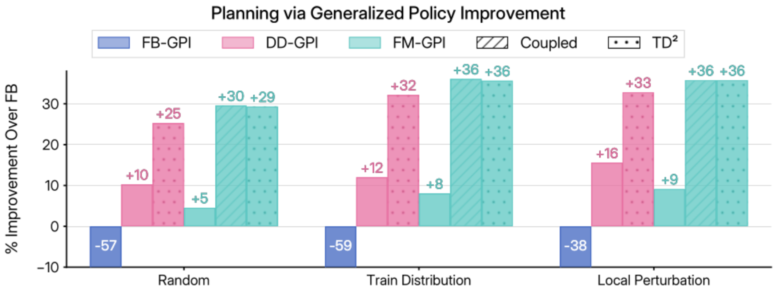
TD²-CFM



GPI:

- Train π_w with Forward backward.
- Do GPI as $w_t \in \arg\max_{w \sim D(W)} \underbrace{(1-\gamma)^{-1}\mathbb{E}_{X \sim m^{\pi_w}(\cdot|s_t, \pi_w(s_t))}[r(X)]}_{Q^{\pi_w}(s_t, \pi_w(s_t))}$.
- Averaged over 128 samples.





References

Tian Qi Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud.
Neural Ordinary Differential Equations.

Jesse Farebrother, Matteo Pirotta, Andrea Tirinzoni, Rémi Munos, Alessandro Lazaric, and Ahmed Touati.
Temporal Difference Flows, March 2025.

Ruiqi Gao, Emiel Hoogeboom, Jonathan Heek, Valentin De Bortoli, Kevin P. Murphy, and Tim Salimans.
Diffusion Meets Flow Matching: Two Sides of the Same Coin, December 2024.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le.
Flow Matching for Generative Modeling, February 2023.

Yaron Lipman, Marton Havasi, Peter Holderrieth, Neta Shaul, Matt Le, Brian Karrer, Ricky T. Q. Chen, David Lopez-Paz, Heli Ben-Hamu, and Itai Gat.
Flow Matching Guide and Code, December 2024.

Seohong Park.
Q-learning is not yet scalable, June 2025.

Danilo Jimenez Rezende and Shakir Mohamed.
Variational Inference with Normalizing Flows.

Chongyi Zheng, Seohong Park, Sergey Levine, and Benjamin Eysenbach.
Intention-Conditioned Flow Occupancy Models, June 2025.