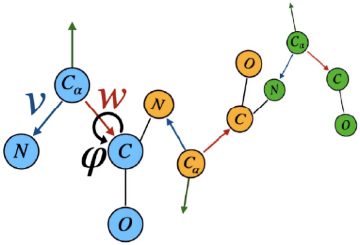
Constraints

Generative Models on Constrained Domains

- Flow / diffusion models usually live in unconstrained $\mathbb{R}^d$.
- Many applications demand **exact** constraint satisfaction:
- *collision-free robot trajectories*
- *valid 3D objects*
- *chemically valid molecules*
- *discrete, structured data*



Challenge: For continuous manifolds or discrete sets Ω (or a mixture of both), design processes to learn while ensuring

$$\mathbb{P}_\theta(X_1 \in \Omega) = 1,$$

regardless initialization and randomness. We call $\{X_t\}$ an Ω-**bridge** in this case.

Two Approaches

- **Process on manifold** ✗ define the ODE/SDE **directly on Ω**.
  - Guarantees feasibility, but new Ω requires bespoke math [BASH+23].
  - For discrete Ω, there's no differentiable structure [GRS+24].
- **Embedded and Relaxation** ✓ keep the process in $\mathbb{R}^d$ and guide it toward Ω via constraint-aware drifts, enabled by singular forces.
  - Flexible: works even when Ω is a finite set or defined via black-box constraints [UCE+25].
  - Leverages the full toolbox of continuous flow/diffusion methods.

Singular Forces

- Enforcing constraints using **singular forces**: these are deterministic drift terms that guarantee the terminal sample lies in the constrained domain Ω.

$$\mathrm{d}Z_t = \underbrace{v_t^{\Omega}(Z_t)\mathrm{d}t}_{\text{singular force}} + \underbrace{v_t^{\theta}(Z_t)}_{\text{trainable net}} \mathrm{d}t + \sigma_t \, \mathrm{d}W_t.$$

- v_t^{θ}: the trainable neural network that learns data-dependent dynamics.
- v_t^{Ω}: an analytical singular drift that ensures $Z_1 \in \Omega$ for any v_t^{θ}.

- As singular force v^{Ω} dominates near $t = 1$, it is possible to ensure constraints for all nicely behavioring v^{θ}:

$$\mathbb{P}(Z_1 \in \Omega) = 1.$$

Girsanov Theorem for Path Measures

Finite Perturbations Does not Change Support Given two SDEs:

$$\mathrm{d}Z_t = b_t(Z_t)\mathrm{d}t + \sigma_t \mathrm{d}W_t$$
$$\mathrm{d}\tilde{Z}_t = (b_t(\tilde{Z}_t) + \delta_t(\tilde{Z}_t))\mathrm{d}t + \sigma_t \mathrm{d}W_t,$$

with the same initialization $Z_0 = \tilde{Z}_0$ and bounded $\left|\sigma_t^{-1}\delta_t(x)\right| \leq M$. Then

$$\mathbb{P}(Z_1 \in \Omega) = 1 \qquad \Longrightarrow \qquad \mathbb{P}(\tilde{Z}_1 \in \Omega) = 1.$$

Girsanov Theorem for Path Measures

Finite Perturbations Does not Change Support Given two SDEs:

$$\mathrm{d}Z_t = b_t(Z_t)\mathrm{d}t + \sigma_t \mathrm{d}W_t$$
$$\mathrm{d}\tilde{Z}_t = (b_t(\tilde{Z}_t) + \delta_t(\tilde{Z}_t))\mathrm{d}t + \sigma_t \mathrm{d}W_t,$$

with the same initialization $Z_0 = \tilde{Z}_0$ and bounded $\left|\sigma_t^{-1}\delta_t(x)\right| \leq M$. Then

$$\mathbb{P}(Z_1 \in \Omega) = 1 \quad \Longrightarrow \quad \mathbb{P}(\tilde{Z}_1 \in \Omega) = 1.$$

Proof. Let P, $\tilde{P}$ the path measures of $\tilde{Z}_t$ and Z_t. By Girsanov's theorem,

$$\mathrm{KL}(\tilde{\mathsf{P}} \mid\mid \mathsf{P}) = \frac{1}{2}\int_0^1 \mathbb{E}\left[\|\sigma_t^{-1}\delta_t(\tilde{Z}_t)\|^2\right]\mathrm{d}t < \frac{M}{2} < +\infty.$$

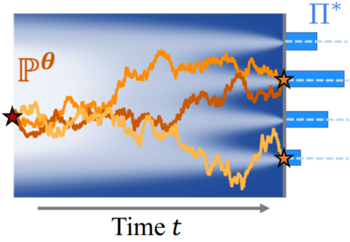
The same holds for $\mathrm{KL}(\mathsf{P} \mid\mid \tilde{P})$. Hence, P and $\tilde{\mathsf{P}}$ are absolutely continuous w.r.t. each other, meaning that they share the same support.

Example

The following dynamics constraint Z_1 on a finite domain $\Omega = \{\mu^{(1)}, \ldots, \mu^{(K)}\}$:

$$\mathrm{d}Z_t = \underbrace{\sum_i \omega_i(Z_t)\frac{\mu^{(i)} - Z_t}{1-t}\mathrm{d}t}_{\text{singular force}} + \underbrace{v_t^\theta(Z_t)\mathrm{d}t}_{\text{trainable bounded force}} + \sigma \mathrm{d}W_t.$$

- The singular force is the rectified flow / Brownian bridge of uniform distribution on Ω.



General strategies for designing Ω-bridges:

- Derive analytic form of rectified flow / diffusion of a reference measure π_0 on Ω.
- Derive posterior processes conditioned on $X_1 \in \Omega$, using Doob's h-transform.
- More complex domains: Derive variants of gradient flow, or Langevin dynamics of a potential function.

Normalized Gradient Flow: Finite-Time Convergence

- Recall that $\mathrm{d}Z_t = \frac{x^* - Z_t}{1-t}\mathrm{d}t$ coincides with a normalized gradient flow.
- In general, normalized gradient flow **squeezes** gradient flow into a **finite time**.

$$\texttt{Normalized Gradient flow}: \quad \frac{\mathrm{d}}{\mathrm{d}t}x_t = -\eta \frac{\nabla f(x_t)}{\|\nabla f(x_t)\|}.$$

If f is strongly convex, then there exists a finite t^*, such that x_{t^*} reaches the minimum, that is, $f(x_{t^*}) = \min_x f(x)$.

Proof.

[RB20] Assume $\min_x f(x) = 0$. We have

$$\frac{\mathrm{d}}{\mathrm{d}t} f(x_t) = -\eta \|\nabla f(x_t)\| \leq -\mu\eta f(x_t)^{1/2}.$$

which gives that $2f(x_t)^{1/2} \leq 2f(x_0)^{1/2} - \mu\eta t$. Hence, we achieve $f(x_t) = 0$ within $t \leq 2f(x_0)^{1/2}/(\mu\eta)$. □

Singular ODE Guarantees Constraints

$$\mathrm{d}Z_t = \frac{e(Z_t, t) - Z_t}{1-t}\,\mathrm{d}t + v^\theta(Z_t, t)\,\mathrm{d}t.$$

If $\left\|v^\theta\right\|$ is bounded, $e_1(z) \in \Omega$, and e is continuous, then $Z_1 \in \Omega$.

Proof.

Computing the time derivative of $Z_t/(1-t)$ and integrating both sides:

$$\frac{Z_t}{1-t} - Z_0 = \int_0^t \frac{v^\theta(Z_\tau, \tau)}{1-\tau}\,\mathrm{d}\tau + \int_0^t \frac{e(Z_\tau, \tau)}{(1-\tau)^2}\,\mathrm{d}\tau.$$

As $t \to 1$, $(1-t)Z_0$ and $(1-t)\int_0^t \frac{v^\theta(Z_\tau,\tau)}{1-\tau}\,\mathrm{d}\tau$ vanish. Apply L'Hôpital's rule to the last term:

$$\lim_{t\to1} Z_t = \lim_{t\to1}(1-t)\int_0^t \frac{e(Z_\tau, \tau)}{(1-\tau)^2}\,\mathrm{d}\tau = \lim_{t\to1} \frac{\int_0^t \frac{e(Z_\tau,\tau)}{(1-\tau)^2}\,\mathrm{d}\tau}{\int_0^t \frac{1}{(1-\tau)^2}\,\mathrm{d}\tau} = \lim_{t\to1} e(Z_\tau, \tau) \in \Omega.$$

Discrete Bridges If Ω is finite / discrete, this motivates another parameterization:

$$\mathrm{d}Z_t = \frac{e_t^\theta(Z_t) - Z_t}{1-t}\mathrm{d}t,$$

where

$$e_t^\theta(Z_t) = \sum_i \mu^{(i)} p^\theta(X_1 = \mu^{(i)} \mid Z_t).$$

- Train the probability $p^\theta(X_1 = \mu^{(i)} \mid Z_t)$, rather than the velocity.
- Cross entropy loss can be used
- Example: Dirichlet flow matching. [SJW+24]

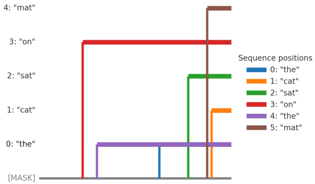
Discrete Flow / Diffusion: Two Approaches

Discrete Latents: Jump within a Discrete Set.

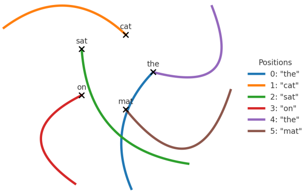
- Examples: D3PM [AJH+21], CTMC [CBDB+22], RADD [ONX+24], MDLM [SAS+24], LLaDA [NZY+25], discrete flow matching gat2024discrete.

Continuous Latents: Flow/Diffusion in Continuous or Embedding Space.

- Examples: Argmax Diffusion [HNJ+21], Diffusion-LM [LTG+22], Ω-bridges [LWYL22], Dirichlet Flow Matching [SJW+24].



Discrete Jump



Continuous Flow

Discrete vs Continuous Latents

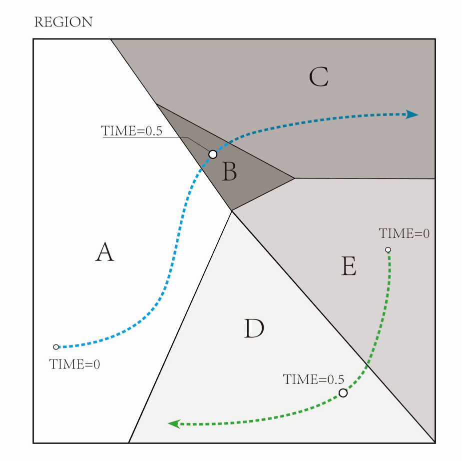
Discrete Latents:

- Curse of dimensionality: each jump must be factorized.
- Therefore, one-step generation is theoretically impossible.
- Ordering is key: essentially a randomly ordered autoregressive model.

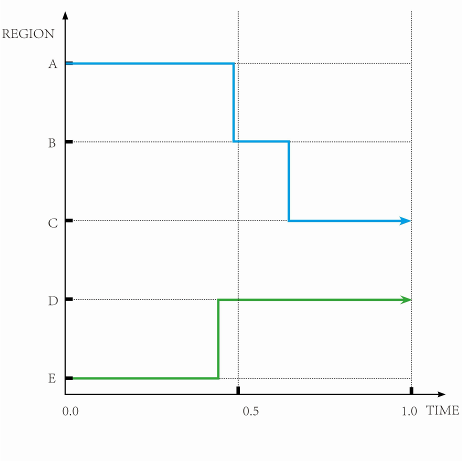
Continuous Latents:

- Traverse a more flexible, continuous space.
- One-step generation is theoretically possible (when the ODE is straight).
- Leverage a rich toolbox from the continuous domain: solvers, distillation, control, etc.

Discrete vs Continuous Latents



Continuous latent flow



Its argmax discretization

Two Paths to Rectified Discrete Flow

Rectify then Discretize: Build a continuous rectified flow, then discretize the trajectory to obtain a discrete jump process.

Discretize then Rectify: Directly construct a discrete jump process as the interpolant, then rectify (Markovize) it.

Under suitable conditions, both approaches yield the same jump processes [Liu24].

$$\texttt{Discretize}(\texttt{Rectify}(\{X_t\})) = \texttt{Rectify}(\texttt{Discretize}(\{X_t\})).$$

- Related: Diffusion Duality [SDG$^+$25].

Thank You!

$$\{Z_t\} = \texttt{Rectify}(\{X_t\})$$

A demon walks where paths cross,
It rewires time, and flows abide.

Continuity's gift: marginals stay.
Straightness cuts the transport way.

Gaussian blessings shape the score,
Noise refines what came before.

Consistency distills the past,
Reward reshapes the path so fast.
Singular forces carve the rule,
For constraints sharp and data dual.

All these threads, once intertwined,
Are straightened by the flow designed.

– ChatGPT