电子科技大学
University of Electronic Science and Technology of China

The ODE Form of the Generate Diffudion Model

A new direction for generative models

Data Mining Lab, Big Data Research Center, USETC

Rui Zhang, zhang_rui@std.uestc.edu.cn

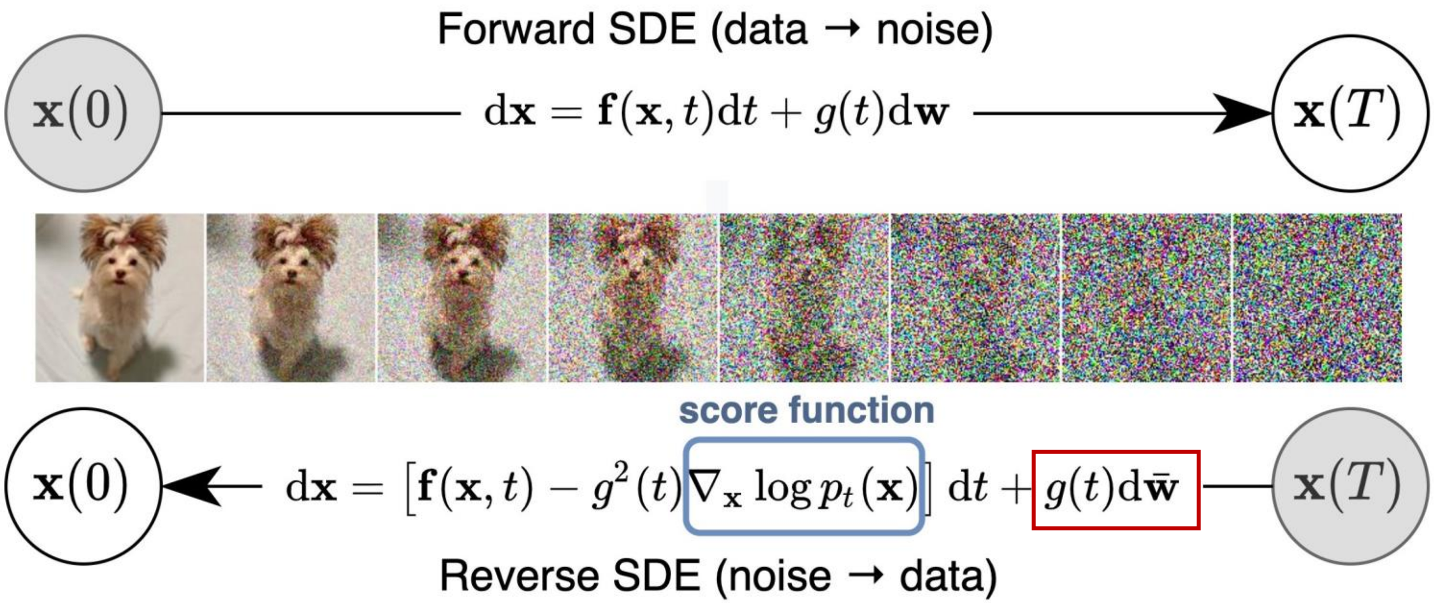
Table of Contents

数据挖掘实验室
Data Mining Lab

Background

1. Background

- Generative Model
  - **SDE form of Diffusion Model**



There is a noise term.

PFGM

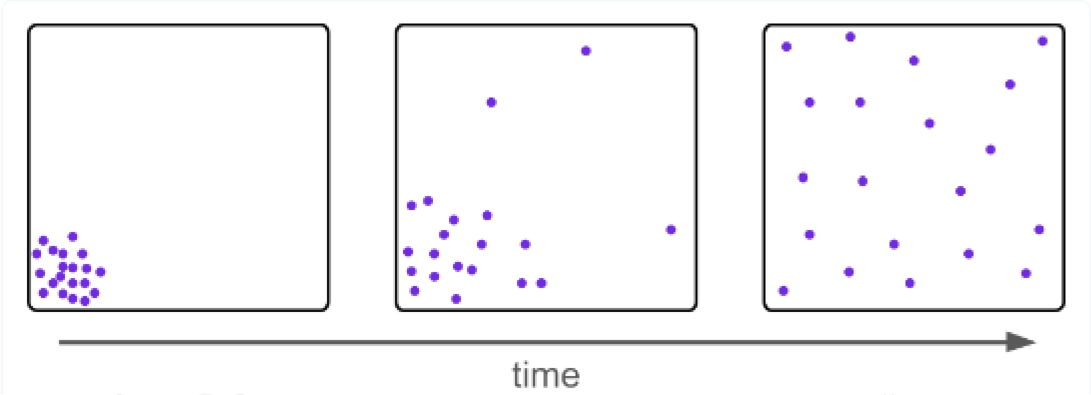
Xu Y, Liu Z, Tegmark M, et al. **Poisson flow generative models**[J]. Advances in Neural Information Processing Systems, 2022, 35: 16782-16795.

Xu Y, Liu Z, Tian Y, et al. **Pfgm++: Unlocking the potential of physics-inspired generative models**[J]. arXiv preprint arXiv:2302.04265, 2023

2. PFGM

- Overview

- Over the past couple of years, significant research efforts have been undertaken to develop Diffusion Models. Diffusion Models draw inspiration from physics.



A localized gas (purple particles) will spread out to evenly fill a room over time simply through random motion

2. PFGM

- Overview

- Images in a "high dimensional room", corrupting them with random noise in a similar way

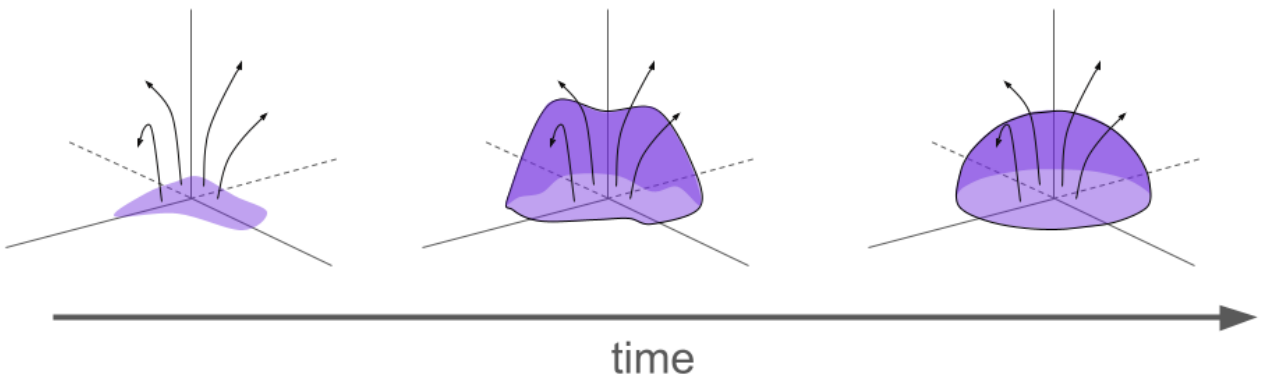
The pixels in an image can be viewed as a localized cloud of particles that we "diffuse" into random noise over time

2. PFGM

数据挖掘实验室
Data Mining Lab

- Overview

- Poisson Flow Generative Models are inspired in a very similar manner to Diffusion Models,



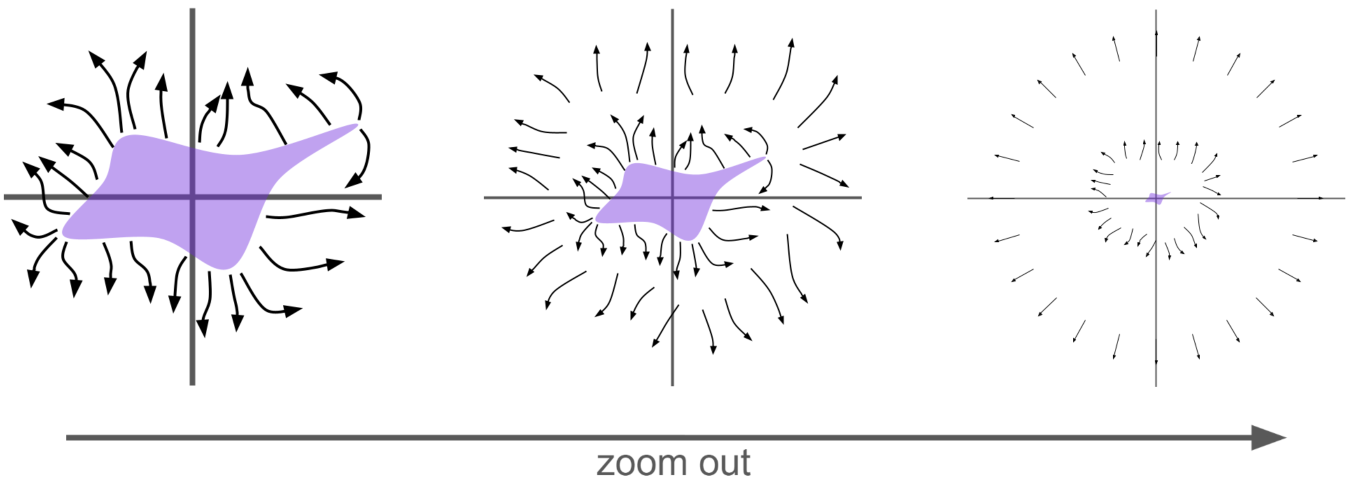
Treating a data distribution as a *charge* distribution defines an electric field that transforms the distribution into a uniform hemisphere over time

2. PFGM

数据挖掘实验室
Data Mining Lab

- Overview

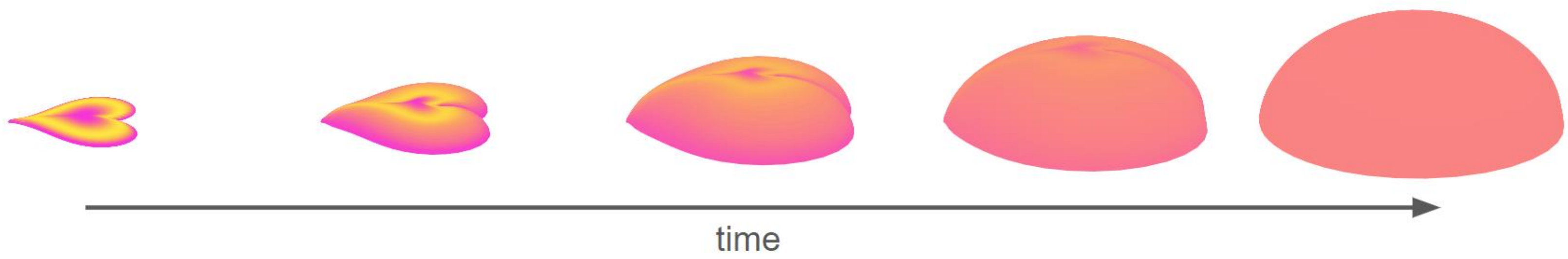
- Poisson field generated by (almost) any distribution in a hyperplane results in a uniform angular distribution at a far enough distance



The multi-source gravitational field at infinity is equivalent to the gravitational field of a particle located at the center of mass and superposition of mass.
(无穷远处的多源引力场，等价于位于质心、质量叠加的质点引力场。)

2. PFGM

- Overview



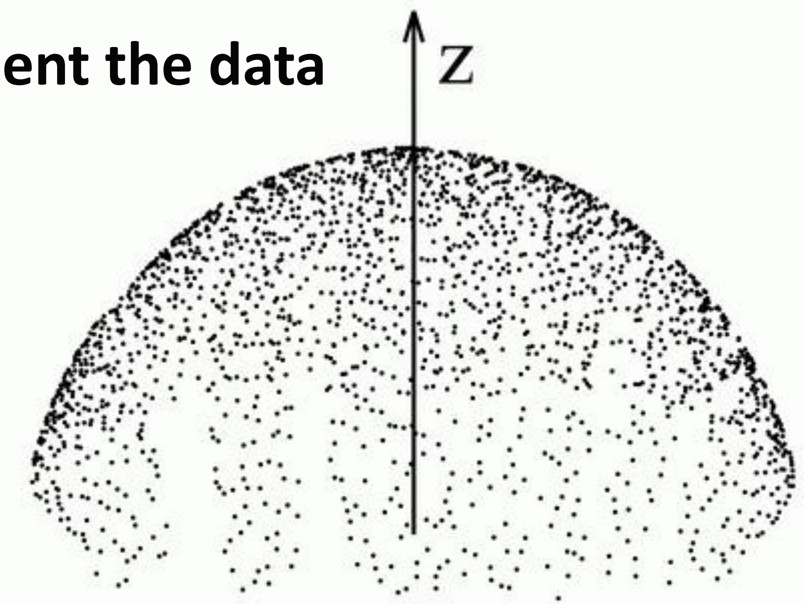
By following the electric field lines it generates, a heart-shaped distribution evolves into a uniform angular distribution over time

Uniformly sampled points on the hemisphere can be transformed into samples from the data distribution by evolving them backwards through the Poisson field generated by the data distribution

2. PFGM

- Learning the Poisson field

Step 1 - Augment the data



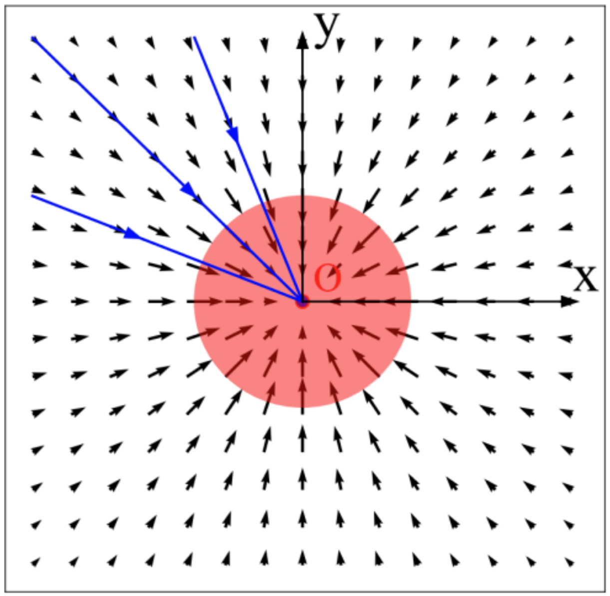
N-dimensional data are augmented with an additional dimension *z* and placed in the *z=0* hyperplane of the new (N+1)-dimensional space. The data are then mapped to an (N+1)-dimensional hemisphere

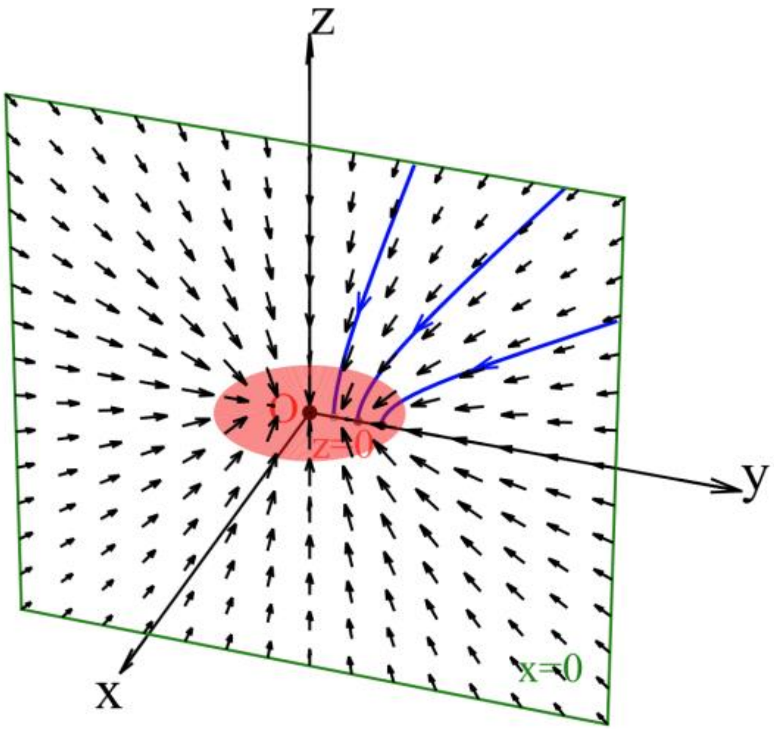
The reason for this spatial augmentation is to avoid mode collapse

2. PFGM

数据挖掘实验室
Data Mining Lab

- Learning the Poisson field





Without augmenting the data with an additional dimension, all trajectories converge to the origin resulting in mode collapse

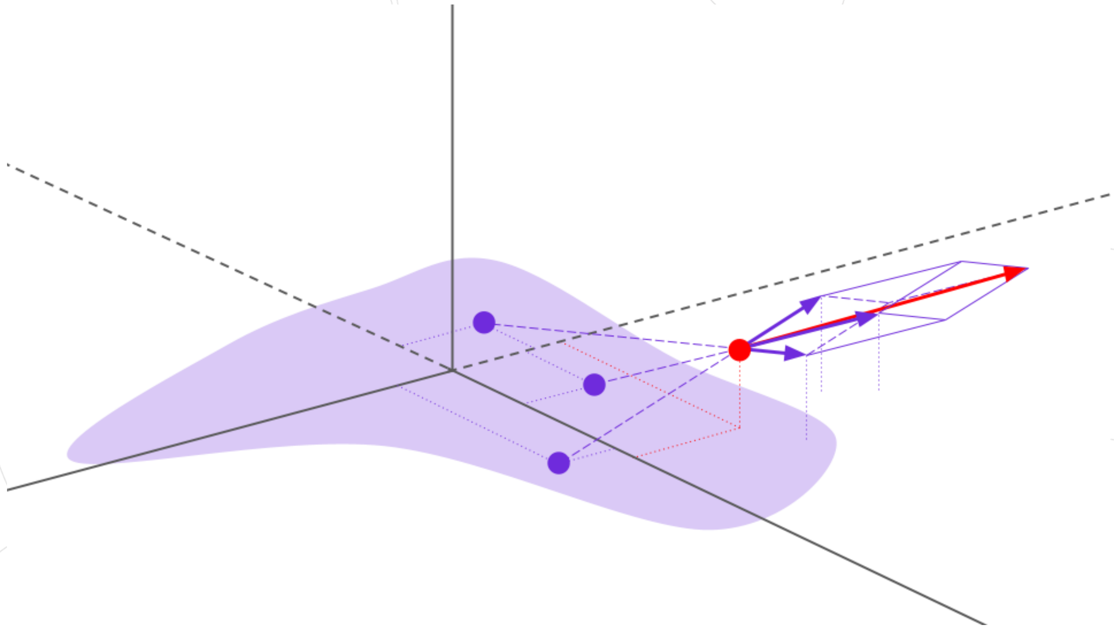
After augmenting the data with an additional dimension, many more trajectories are available that intersect with different points in the distribution, therefore avoiding mode

2. PFGM

数据挖掘实验室
Data Mining Lab

- Learning the Poisson field

Step 2 - Calculate the empirical field

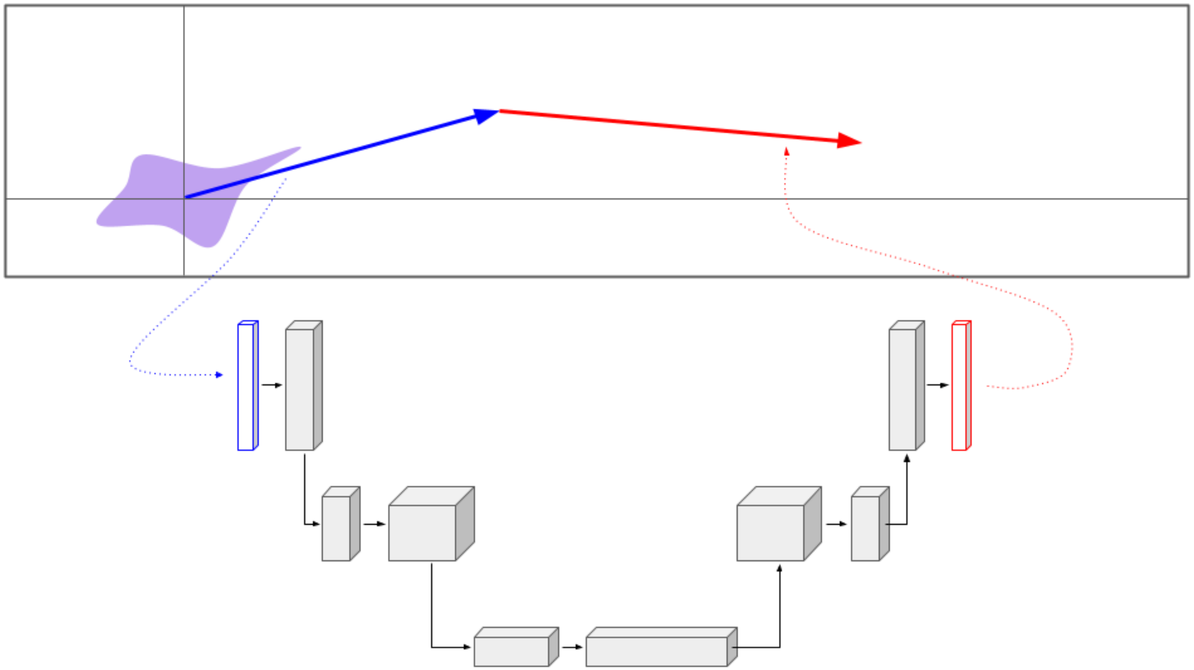


Due to the superposition principle, We calculate the empirical field for many randomly sampled points in the space

2. PFGM

- Learning the Poisson field

Step 3 - Calculate the loss and update the function approximator



A U-Net (block diagram) accepts a point in space (blue vector) and returns the approximate empirical field at that point (red vector) generated by data points sampled from the data/charge distribution (purple)

2. PFGM

- Sampling with PFGMs

Sample points from a uniform angular distribution, and then make them travel backwards along the Poisson field until we reach the z=0 hyperplane in which the data distribution sits.

The corresponding differential equation is:

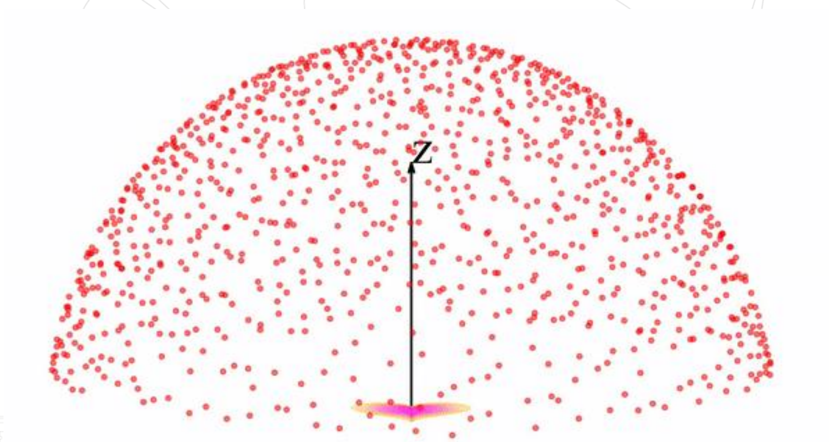
$$\mathrm{d}\mathbf{x} = -\mathbf{E}(\mathbf{x})\mathrm{d}t$$

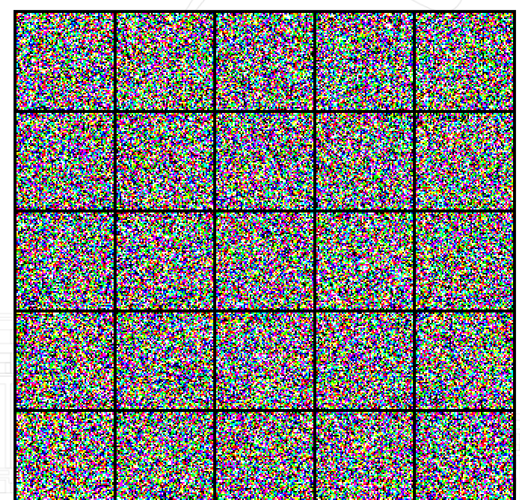
At each moment in time, the point should be displaced in the direction of the negative Poisson field.

2. PFGM

- Sampling with PFGMs

1. Uniformly sample data on a large hemisphere
2. Use an ODE solver to evolve the points backwards along the Poisson field
3. Evolve backwards until we reach z=0, at which point we have generated novel data from the training distribution





2. PFGM

- A deep dive - Anchoring the Backwards ODE

We use the following backward ODE for sampling:

$$d\tilde{\mathbf{x}} = \mathbf{v}(\tilde{\mathbf{x}})dt$$

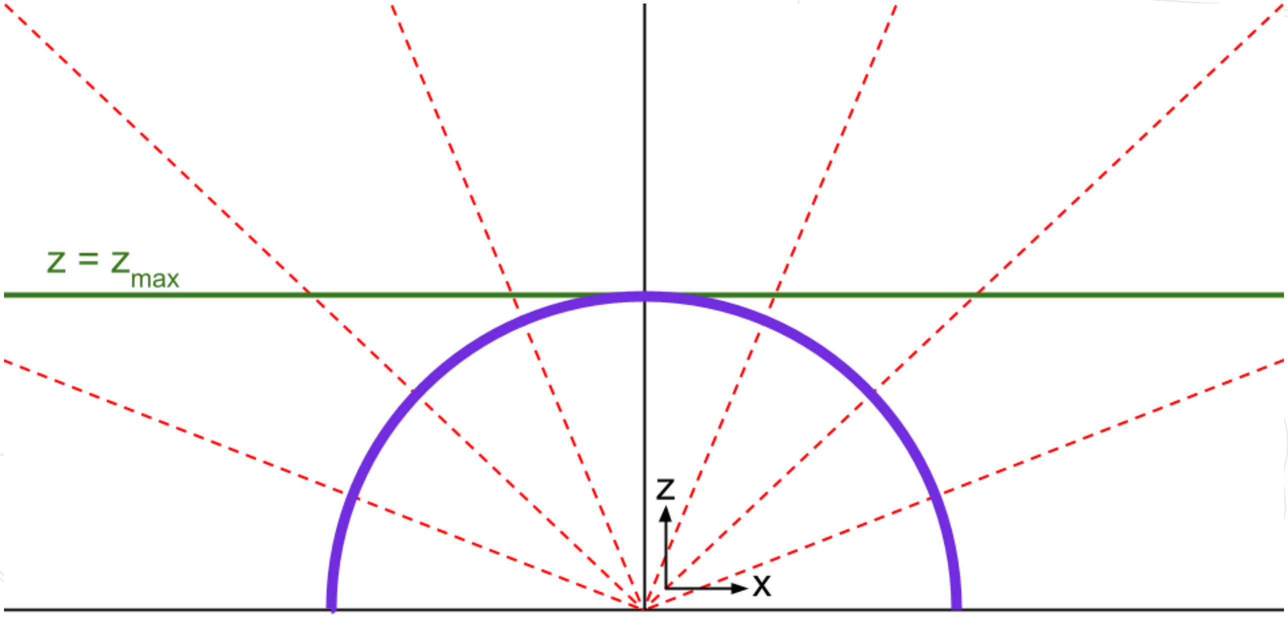
We do not know the starting and ending times of each trajectory.

$$\mathrm{d}(\mathbf{x}, z) = \left(\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}\frac{\mathrm{d}t}{\mathrm{d}z}\mathrm{d}z, \mathrm{d}z\right) = \left(\mathbf{v}(\mathbf{x})_{\mathbf{x}}\mathbf{v}(\mathbf{x})_{z}^{-1}, 1\right)\mathrm{d}z$$

2. PFGM

数据挖掘实验室
Data Mining Lab

- A deep dive - Determining the prior distribution



The uniform distribution on the hemisphere is **projected** to the hyperplane that coincides with the top of the hemisphere, z_max . The projection is a simple **radial** one given that the Poisson field is purely radial as **x** approaches infinity. We can see a schematic in 2D below, where the red lines connect purple points on the sphere to their projection points on the green line.

2. PFGM

数据挖掘实验室
Data Mining Lab

- A deep dive - Training

Algorithm 1 Learning the normalized Poisson Field

Input: Training iteration T, Initial model f_θ, dataset $\mathcal{D}$, constant γ, learning rate η.
for $t = 1 \ldots T$ **do**
Sample a large batch $\mathcal{B}_L$ from $\mathcal{D}$ and subsample a batch of datapoints $\mathcal{B} = \{\mathbf{x}_i\}_{i=1}^{|\mathcal{B}|}$ from $\mathcal{B}_L$
Simulate the ODE: $\{\tilde{\mathbf{y}}_i = \text{perturb}(\mathbf{x}_i)\}_{i=1}^{|\mathcal{B}|}$
Calculate the normalized field by $\mathcal{B}_L$: $\mathbf{v}_{\mathcal{B}_L}(\tilde{\mathbf{y}}_i) = -\sqrt{N}\hat{\mathbf{E}}_{\mathcal{B}_L}(\tilde{\mathbf{y}}_i)/(\| \hat{\mathbf{E}}_{\mathcal{B}_L}(\tilde{\mathbf{y}}_i) \|_2 + \gamma), \forall i$
Calculate the loss: $\mathcal{L}(\theta) = \frac{1}{|\mathcal{B}|} \sum_{i=1}^{|\mathcal{B}|} \| f_\theta(\tilde{\mathbf{y}}_i) - \mathbf{v}_{\mathcal{B}_L}(\tilde{\mathbf{y}}_i) \|_2^2$
Update the model parameter: $\theta = \theta - \eta \nabla \mathcal{L}(\theta)$
end for
return f_θ

Algorithm 2 perturb($\mathbf{x}$)

Sample the power $m \sim \mathcal{U}[0, M]$
Sample the initial noise $(\epsilon_{\mathbf{x}}, \epsilon_z) \sim \mathcal{N}(0, \sigma^2 I_{(N+1)\times(N+1)})$
Uniformly sample the vector from the unit ball $\mathbf{u} \sim \mathcal{U}(S_N(1))$
Construct training point $\mathbf{y} = \mathbf{x} + \| \epsilon_{\mathbf{x}} \| (1+\tau)^m \mathbf{u}$, $z = |\epsilon_z|(1+\tau)^m$
return $\tilde{\mathbf{y}} = (\mathbf{y}, z)$

2. PFGM

数据挖掘实验室
Data Mining Lab

- A deep dive - Others

- What is the connection between the continuous data distribution p(x) and the continuous charge source and electric field?
- How to establish a mathematical model for the motion process along the field line?
- How to randomly sample from infinity?
- Why physically moving along the field line to a point source corresponds to mathematically sampling the distribution?

https://zhuanlan.zhihu.com/p/630848560

https://www.assemblyai.com/blog/an-introduction-to-poisson-flow-generative-models/

2. PFGM

数据挖掘实验室
Data Mining Lab

- Highlights from applying PFGMs

- PFGMs attain the **best Inception scores** (9.68) and **best FID scores** (2.35) among normalizing flow models for CIFAR-10.
- PFGMs enjoy a **10-20x faster inference speed** than SDE methods that utilize similar architectures, while retaining **comparable sample quality**.
- The backward ODE if PFGM accommodates **many different architectures**.
- PFGMs demonstrate **scalability** to higher resolution image generation

Rectified Flow

Liu X, Gong C, Liu Q. **Flow straight and fast: Learning to generate and transfer data with rectified flow**[J]. arXiv preprint arXiv:2209.03003, 2022.

3 Rectified Flow

数据挖掘实验室
Data Mining Lab

- Extremely simple, one step generation

- Instead of complex derivation of the general diffusion model, a simple idea of "generating along a straight line" is used

- Through a method called "reflow", we can achieve the dream of "one-step generation“

- The method can convert any kind of data or noise (such as a photo of a cat's face) into another kind of data (such as a photo of a face).

3 Rectified Flow

数据挖掘实验室
Data Mining Lab

- Extremely simple, one step generation

DDPM

$$
\begin{aligned}
L_{\mathrm{VLB}} &= \mathbb{E}_{q(\mathbf{x}_{0:T})}\Big[\log \frac{q(\mathbf{x}_{1:T}|\mathbf{x}_0)}{p_\theta(\mathbf{x}_{0:T})}\Big] \\
&= \mathbb{E}_q\Big[\log \frac{\prod_{t=1}^T q(\mathbf{x}_t|\mathbf{x}_{t-1})}{p_\theta(\mathbf{x}_T)\prod_{t=1}^T p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)}\Big] \\
&= \mathbb{E}_q\Big[-\log p_\theta(\mathbf{x}_T) + \sum_{t=1}^T \log \frac{q(\mathbf{x}_t|\mathbf{x}_{t-1})}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)}\Big] \\
&= \mathbb{E}_q\Big[-\log p_\theta(\mathbf{x}_T) + \sum_{t=2}^T \log \frac{q(\mathbf{x}_t|\mathbf{x}_{t-1})}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} + \log \frac{q(\mathbf{x}_1|\mathbf{x}_0)}{p_\theta(\mathbf{x}_0|\mathbf{x}_1)}\Big] \\
&= \mathbb{E}_q\Big[-\log p_\theta(\mathbf{x}_T) + \sum_{t=2}^T \log \Big(\frac{q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} \cdot \frac{q(\mathbf{x}_t|\mathbf{x}_0)}{q(\mathbf{x}_{t-1}|\mathbf{x}_0)}\Big) + \log \frac{q(\mathbf{x}_1|\mathbf{x}_0)}{p_\theta(\mathbf{x}_0|\mathbf{x}_1)}\Big] \\
&= \mathbb{E}_q\Big[-\log p_\theta(\mathbf{x}_T) + \sum_{t=2}^T \log \frac{q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} + \sum_{t=2}^T \log \frac{q(\mathbf{x}_t|\mathbf{x}_0)}{q(\mathbf{x}_{t-1}|\mathbf{x}_0)} + \log \frac{q(\mathbf{x}_1|\mathbf{x}_0)}{p_\theta(\mathbf{x}_0|\mathbf{x}_1)}\Big] \\
&= \mathbb{E}_q\Big[-\log p_\theta(\mathbf{x}_T) + \sum_{t=2}^T \log \frac{q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} + \log \frac{q(\mathbf{x}_T|\mathbf{x}_0)}{q(\mathbf{x}_1|\mathbf{x}_0)} + \log \frac{q(\mathbf{x}_1|\mathbf{x}_0)}{p_\theta(\mathbf{x}_0|\mathbf{x}_1)}\Big] \\
&= \mathbb{E}_q\Big[\log \frac{q(\mathbf{x}_T|\mathbf{x}_0)}{p_\theta(\mathbf{x}_T)} + \sum_{t=2}^T \log \frac{q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} - \log p_\theta(\mathbf{x}_0|\mathbf{x}_1)\Big] \\
&= \mathbb{E}_q[\underbrace{D_{\mathrm{KL}}(q(\mathbf{x}_T|\mathbf{x}_0)\parallel p_\theta(\mathbf{x}_T))}_{L_T} + \sum_{t=2}^T \underbrace{D_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)\parallel p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))}_{L_{t-1}} \underbrace{- \log p_\theta(\mathbf{x}_0|\mathbf{x}_1)}_{L_0}]
\end{aligned}
$$

Rectified Flow

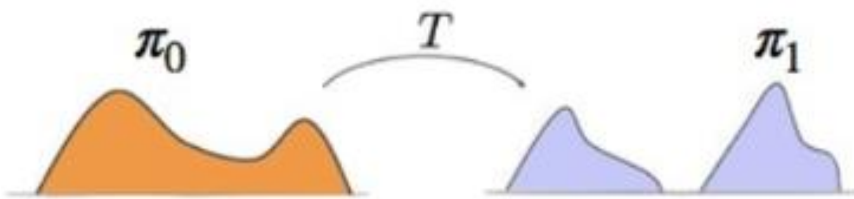
$$L = \int_0^1 \mathbb{E}_{X_0\sim\text{noise},\, X_1\sim\text{data}}\,[\|\,(X_1 - X_0) - v(X_t, t)\,\|^2]\,\mathrm{d}t,$$

$$\text{with } X_t = t\,X_1 + (1-t)X_0$$

3 Rectified Flow

数据挖掘实验室
Data Mining Lab

- Distribution Transformation



Given samples from the two distributions π_0 and π_1. we want to find a transport mapping T such that when $Z_0 \sim \pi_0$, $Z_1 = T(Z_0) \sim \pi_1$

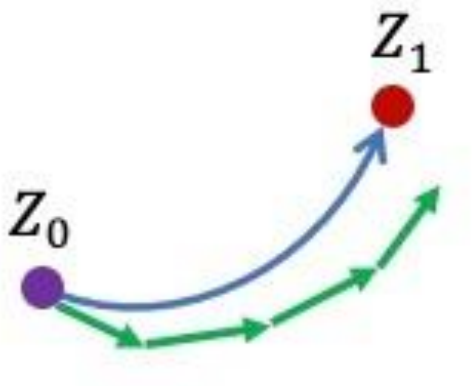
Mappings T are implicitly defined by the following system of continuous motion, an ordinary differential equation (ODE), or flow model

$$\frac{d}{dt} Z_t = v(Z_t, t), \quad Z_0 \sim \pi_0, \quad \forall t \in [0, 1].$$

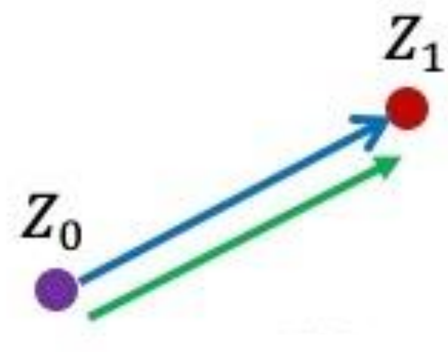
3 Rectified Flow

- Walk fast in a straight line

Euler solution $Z_{t+\epsilon} = Z_t + \epsilon v(Z_t, t)$



Multistep solution



One step solution

3 Rectified Flow

数据挖掘实验室
Data Mining Lab

- Learning generation model based on linear ODE

Connect X_0 and X_1 with a linear interpolation

$$X_t = tX_1 + (1-t)X_0, \quad t \in [0,1].$$

$$\frac{d}{dt}X_t = \boxed{X_1} - X_0, \quad \forall t \in [0,1].$$

$$\min_v \int_0^1 \mathbb{E}_{X_0 \sim \pi_0, X_1 \sim \pi_1} \left[||(X_1 - X_0) - v(X_t, t)||^2 \right] dt, \quad \text{where} \quad X_t = tX_1 + (1-t)X_0.$$

3 Rectified Flow

数据挖掘实验室
Data Mining Lab

- Learning generation model based on linear ODE

Connect X_0 and X_1 with a linear interpolation

Algorithm 2 Train(Data)

```
# Input:  Data={x0, x1}
# Output:  Model v(x,t) for the rectified flow
initialize Model
for x0, x1 in Data:  # x0, x1: samples from π0, π1
    Optimizer.zero_grad()
    t = torch.rand(batchsize)  # Randomly sample t ∈ [0,1]
    Loss = ( Model(t*x1+(1-t)*x0, t) - (x1-x0) ).pow(2).mean()
    Loss.backward()
    Optimizer.step()
return Model
```

3 Rectified Flow

数据挖掘实验室
Data Mining Lab

- Learning generation model based on linear ODE

Connect X_0 and X_1 with a linear interpolation

Algorithm 3 Sample(Model, Data)

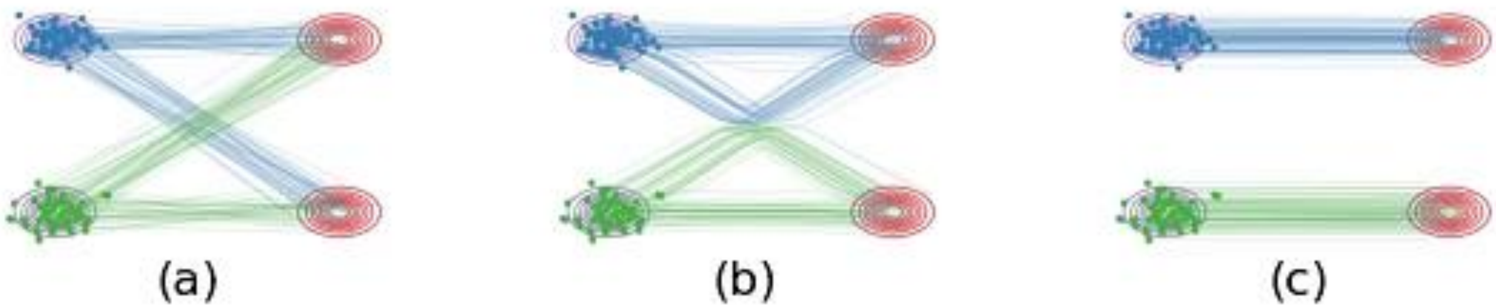
```
# Input:  Model v(x,t) of the rectified flow
# Output:  draws of the rectified coupling (Z_0, Z_1)
coupling = []
for x0, _ in Data:  # x0: samples from π_0 (batchsize×dim)
    x1 = model.ODE_solver(x0)
    coupling.append((x0, x1))
return coupling
```

3 Rectified Flow

数据挖掘实验室
Data Mining Lab

- Learning generation model based on linear ODE

Connect X_0 and X_1 with a linear interpolation



$$\min_v \int_0^1 \mathbb{E}_{X_0\sim\pi_0, X_1=\text{Flow}_1(X_0)}\left[||(X_1 - X_0) - v(X_t, t)||^2\right] dt, \quad \text{with} \quad X_t = tX_1 + (1-t)X_0.$$

3 Rectified Flow

- Learning generation model based on linear ODE

Connect A and B with a linear interpolation

Algorithm 4 Reflow(Data)

```
# Input:  Data={x0, x1}
# Output:  draws of the K-th rectified coupling
Coupling = Data
for k = 1,...,K:
    Model = Train(Coupling)
    Coupling = sample(Model, Data)
return Coupling
```

3 Rectified Flow

数据挖掘实验室
Data Mining Lab

- Reflow and Distillation

Optimization of the square error below appears to be possible with a direct "distillation" of a one-step model

Distillation $$\min_v \mathbb{E}\left[||X_1 - X_0 - v(X_0, 0)||^2\right].$$ Is that WORK?

$$\min_v \int_0^1 \mathbb{E}_{X_0\sim\pi_0, X_1\sim\pi_1}\left[||(X_1 - X_0) - v(X_t, t)||^2\right] dt, \quad \text{where} \quad X_t = tX_1 + (1-t)X_0.$$

Reflow

Distillation alone does not improve pairing, which is the essential difference between Reflow and distillation.

3 Rectified Flow

数据挖掘实验室
Data Mining Lab

- Reflow and Distillation

Reflow addresses these difficulties with Distillation:

- Given any (X_0, X_1) pairings, even random ones, it can learn a flow that gives the correct marginal distribution.
- Sampling from the ODE from Reflow, we can also get a better pairing, which gives a better flow
- Reflow and "Distillation" may also be used in combination: Reflow will be used to yield a good pairing, and then "Distillation" will be used with an already good pairing

3 Rectified Flow

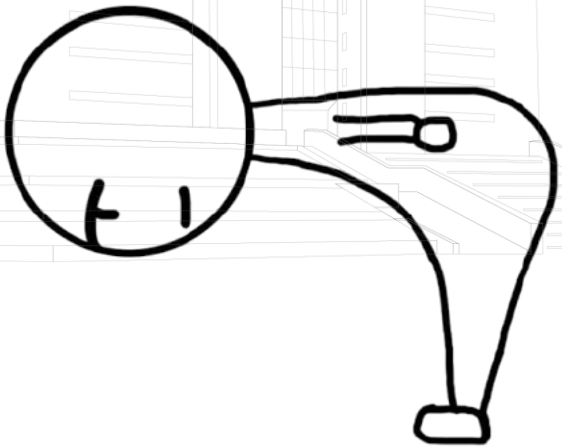
数据挖掘实验室
Data Mining Lab

- Result

Method	NFE(↓)	IS (↑)	FID (↓)	Recall (↑)
ODE	*One-Step Generation (Euler solver, N=1)*			
1-Rectified Flow (+*Distill*)	1	1.13 (***9.08***)	378 (*6.18*)	0.0 (*0.45*)
2-Rectified Flow (+*Distill*)	1	8.08 (*9.01*)	12.21 (***4.85***)	0.34 (*0.50*)
3-Rectified Flow (+*Distill*)	1	8.47 (*8.79*)	8.15 (*5.21*)	0.41 (***0.51***)
VP ODE [73] (+*Distill*)	1	1.20 (*8.73*)	451 (*16.23*)	0.0 (*0.29*)
sub-VP ODE [73] (+*Distill*)	1	1.21 (*8.80*)	451 (*14.32*)	0.0 (*0.35*)
ODE	*Full Simulation (Runge–Kutta (RK45), Adaptive N)*			
1-Rectified Flow	127	**9.60**	**2.58**	**0.57**
2-Rectified Flow	110	9.24	3.36	0.54
3-Rectified Flow	104	9.01	3.96	0.53
VP ODE [73]	140	9.37	3.93	0.51
sub-VP ODE [73]	146	9.46	3.16	0.55
SDE	*Full Simulation (Euler solver, N=2000)*			
VP SDE [73]	2000	9.58	2.55	0.58
sub-VP SDE [73]	2000	9.56	2.61	0.58

Discussion

Thanks



DM
LESS IS MORE

Data Mining Lab, Big Data Research Center, USETC

Rui Zhang, zhang_rui@std.uestc.edu.cn