Google DeepMind

Schrödinger Bridge Flow for Unpaired Data Translation

Valentin De Bortoli*, Ira Korshunova*, Andriy Mnih, Arnaud Doucet

Introduction

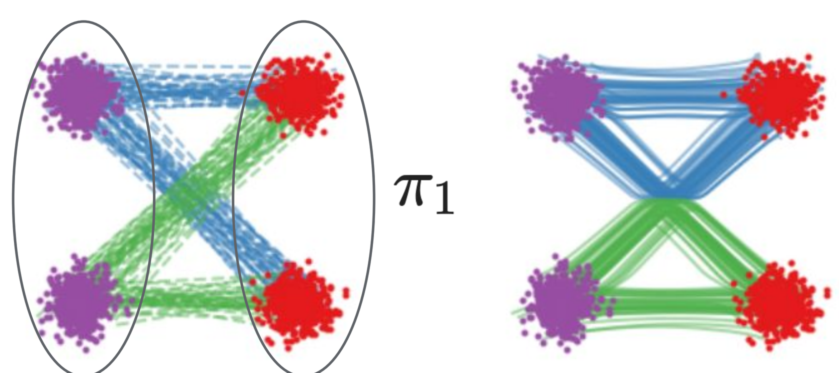
- Rectified flows [1] and Schrödinger Bridges [2] are extensions of **diffusion** (and **flow matching** models) for **unpaired data translation**.
- Existing algorithms involve costly **iterative procedures**. Can we do better?

Reflowing and retraining

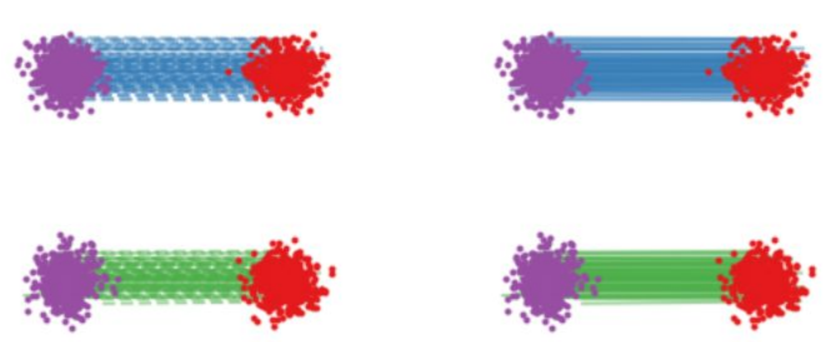
- Given data distributions π_0 and π_1
  - Find **optimal coupling**
  - "How to go from π_0 to π_1 optimally?"

$$\mathbf{X}_t = (1-t)\mathbf{X}_0 + t\mathbf{X}_1 + \sigma\sqrt{t(1-t)}\mathbf{Z}$$

- Rectified flow (Schrödinger Bridge = stochastic version)



- The **reflow step** (retraining using the data coupling from the trained model (!))



Efficient finetuning for unpaired data

- **Problem 1: ALIGNMENT VS QUALITY**
  - More reflows = better alignment, worse quality
  - Improvement: **forward/backward training** with half-batches

Method	NFE(↓)	FID (↓)
Euler solver		
1-Rectified Flow (+Distill)	1	378 (6.18)
2-Rectified Flow (+Distill)	1	12.21 (4.85)
3-Rectified Flow (+Distill)	1	8.15 (5.21)
Runge-Kutta (RK45)		
1-Rectified Flow	127	2.58
2-Rectified Flow	110	3.36
3-Rectified Flow	104	3.96

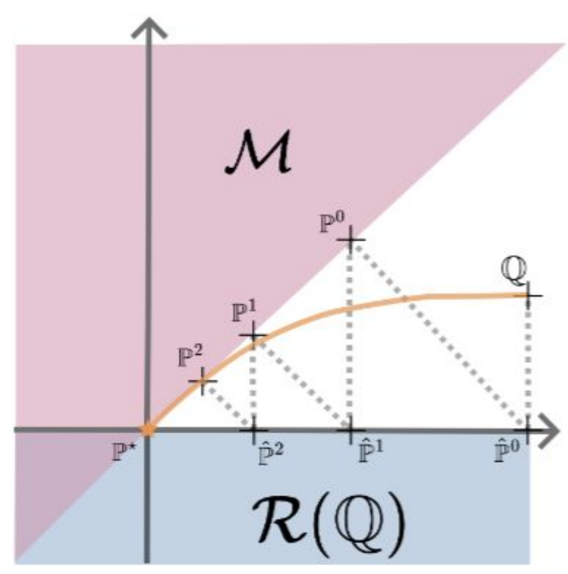
Generative modeling results from [1]

- **Problem 2: INTRICATE ITERATIVE IMPLEMENTATION**
  - Reflow iterations [1,2] require storing samples, choosing the storage size and number of reflow steps
  - Improvement: **online finetuning**

- **Problem 3: STOCHASTICITY AND NETWORKS**
  - Generative modeling = better **without** stochasticity (σ = 0)
  - Transfer tasks = better **with** stochasticity
  - Stochastic approaches, e.g. DSBM[2], require **two networks** (forward and backward)
  - Improvement: **joint architecture**

$$\mathcal{L}_{\text{fwd}} = E[\|v_\theta(s=1,t,\mathbf{X}_t) - (\mathbf{X}_1 - \mathbf{X}_0 - \sigma\sqrt{t/(1-t)}\mathbf{Z})\|^2]$$
$$\mathcal{L}_{\text{bwd}} = E[\|v_\theta(s=0,t,\mathbf{X}_t) - (\mathbf{X}_0 - \mathbf{X}_1 - \sigma\sqrt{(1-t)/t}\mathbf{Z})\|^2]$$

Theoretical results



Our **online procedure** is a discretisation of a **flow** of **path measures**

We call it the **Schrödinger Bridge flow**

$$\hat{\mathbb{P}}^0 = (\pi_0 \otimes \pi_1)\mathbb{Q}_{|0,1},$$
$$\partial_s \hat{\mathbb{P}}^s = \text{proj}_{\mathcal{R}}(\text{proj}_{\mathcal{M}}(\hat{\mathbb{P}}^s)) - \hat{\mathbb{P}}^s,$$

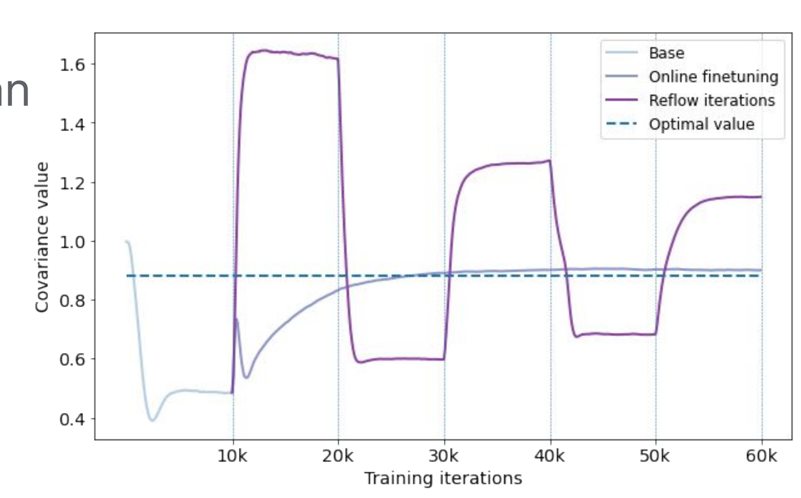
Algorithm 1 α-Diffusion Schrödinger Bridge Matching

Input: datasets π_0 and π_1, entropic regularisation ε, number of pretraining and finetuning steps $N_{\text{pretraining}}$ and $N_{\text{finetuning}}$, batch size B and half batch size $b = B/2$, EMA decay γ, initial parameters θ and initial EMA parameters $\theta^{\text{EMA}} = \theta$, $\alpha \in (0,1]$
for $n \in \{1, \dots, N_{\text{pretraining}}\}$ **do**
  Sample $(\mathbf{X}_0, \mathbf{X}_1) \sim (\pi_0 \otimes \pi_1)^{\otimes B}$
  Sample $t \sim \text{Unif}([0,1])^{\otimes B}$ and $\mathbf{Z} \sim \mathcal{N}(0, \text{Id})^{\otimes B}$ and compute $\mathbf{X}_t = \text{Interp}_t(\mathbf{X}_0, \mathbf{X}_1, \mathbf{Z})$
  Update θ with a gradient step on $\frac{1}{2}\left[\ell^{\rightarrow}(t^{1:b}, \mathbf{X}_1^{1:b}, \mathbf{X}_t^{1:b}) + \ell^{\leftarrow}(t^{b+1:B}, \mathbf{X}_0^{b+1:B}, \mathbf{X}_t^{b+1:B})\right]$
  Update EMA parameters: $\theta^{\text{EMA}} = \gamma\theta^{\text{EMA}} + (1-\gamma)\theta$
end for
for $n \in \{1, \dots, N_{\text{finetuning}}\}$ **do**
  Sample $(\mathbf{X}_0, \mathbf{X}_1) \sim (\pi_0 \otimes \pi_1)^{\otimes b}$
  Sample $\hat{\mathbf{X}}_1$ solving forward SDE (11)-(fwd) with $v_{\theta^{\text{EMA}}}(1, \cdot)$ or $v_\theta(1, \cdot)$ starting from $\mathbf{X}_0$
  Sample $\hat{\mathbf{X}}_0$ solving backward SDE (11)-(bwd) with $v_{\theta^{\text{EMA}}}(0, \cdot)$ or $v_\theta(0, \cdot)$ starting from $\mathbf{X}_1$
  Sample $t^{\rightarrow} \sim \text{Unif}([0,1])^{\otimes b}$ and $\mathbf{Z}^{\rightarrow} \sim \mathcal{N}(0, \text{Id})^{\otimes b}$ and compute $\mathbf{X}_t^{\rightarrow} = \text{Interp}_{t^{\rightarrow}}(\hat{\mathbf{X}}_0, \mathbf{X}_1, \mathbf{Z}^{\rightarrow})$
  Sample $t^{\leftarrow} \sim \text{Unif}([0,1])^{\otimes b}$ and $\mathbf{Z}^{\leftarrow} \sim \mathcal{N}(0, \text{Id})^{\otimes b}$ and compute $\mathbf{X}_t^{\leftarrow} = \text{Interp}_{t^{\leftarrow}}(\mathbf{X}_0, \hat{\mathbf{X}}_1, \mathbf{Z}^{\leftarrow})$
  Update θ with a gradient step on $\frac{1}{2}[\ell^{\rightarrow}(t^{\rightarrow}, \mathbf{X}_1, \mathbf{X}_t^{\rightarrow}) + \ell^{\leftarrow}(t^{\leftarrow}, \mathbf{X}_0, \mathbf{X}_t^{\leftarrow})]$ and stepsize α
  Update EMA parameters: $\theta^{\text{EMA}} = \gamma\theta^{\text{EMA}} + (1-\gamma)\theta$
end for
Output: $(\theta, \theta^{\text{EMA}})$ parameters of the finetuned model

- **Stage 1: pretraining**
  - Classical (stochastic) **flow matching**
- **Stage 2: finetuning**
  - Refinement on learned trajectories
  - **Optima transport**

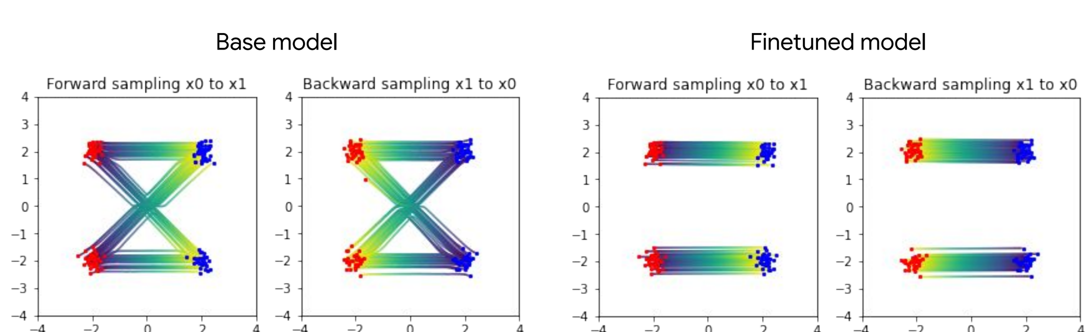
- Gaussian example
  - $\pi_0 = \mathrm{N}(\mu_0, \sigma_0^2)$ $\pi_1 = \mathrm{N}(\mu_0, \sigma_0^2)$
  - Schrödinger Bridge is Gaussian
  - $\sigma_\star^2 = (1/2)[(4\sigma_0^4 + \sigma^4)^{1/2} - \sigma^2]$

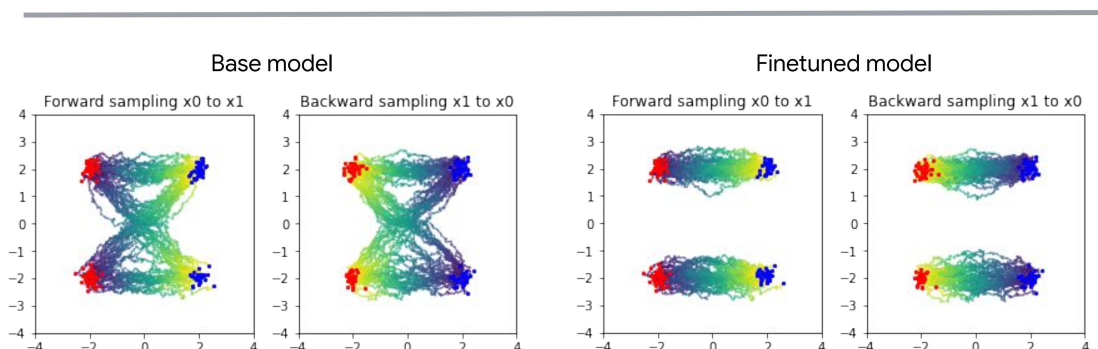
- **Advantage** of online
  - Reflow oscillates
  - Reflow converges slowly



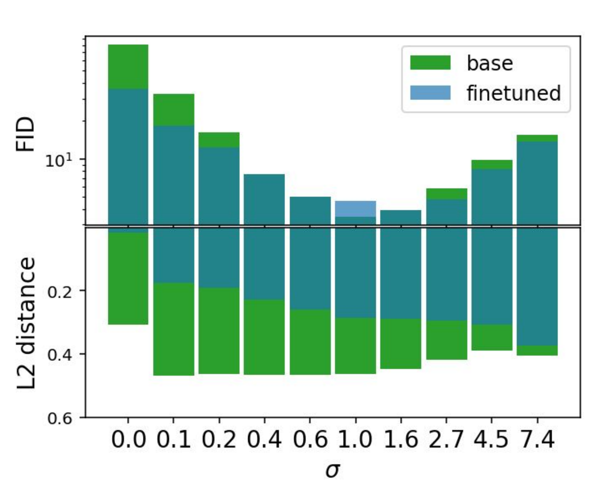
Experiments

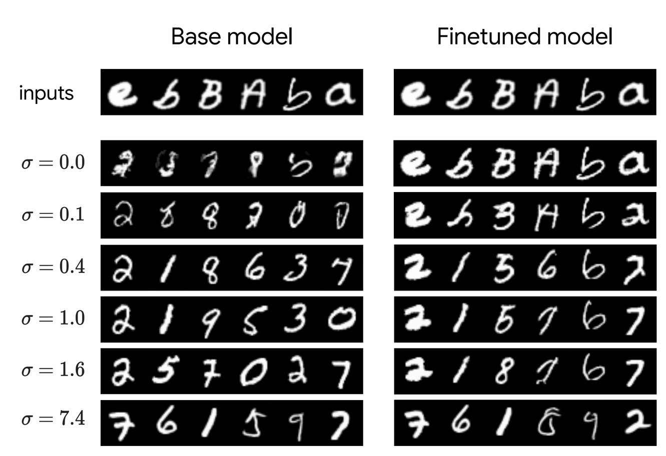
- Toy experiments





- MNIST ↔ EMNIST

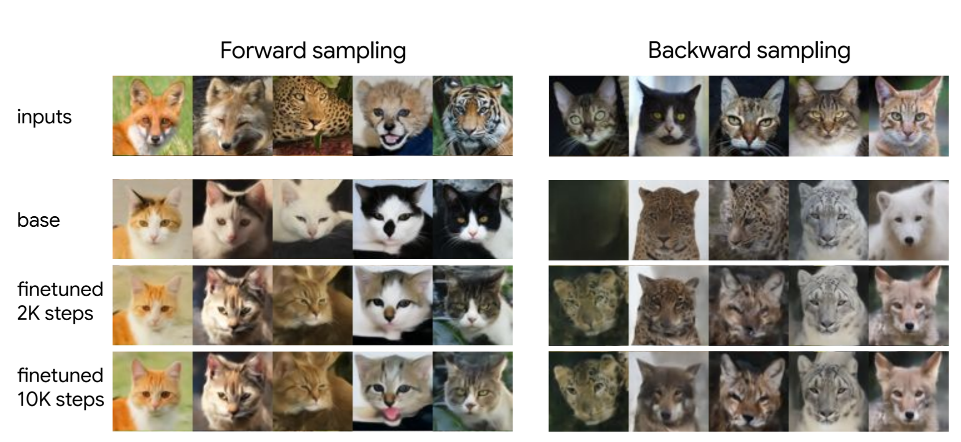




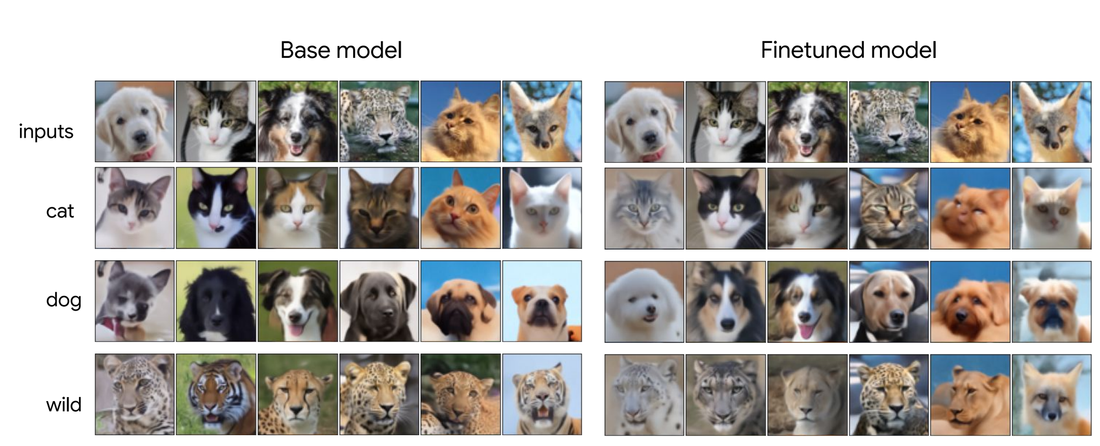
For the transport to work well, σ needs to be chosen carefully:

- **σ too small** → bad quality of the base model, and finetuning leads to a degenerate solution.
- **σ too large** → outputs become independent of the inputs, thus finetuning cannot improve upon the base model.

- AFHQ: cat ↔ wild



AFHQ (conditional): cat ↔ wild ↔ dog



What's next?

- Extension to **larger datasets**
  - **Multimodality**
  - Latent representation (same dimension)
  - Audio <-> Image <-> ???
- Applications in science
  - Downscaling in **climate science**

[1] Flow Straight and Fast, Liu et al. (2022)
[2] Diffusion Schrödinger Bridge Matching, Shi et al. (2023)