Well-posed generative flows via combined Wasserstein-1 and Wasserstein-2 proximals of f-divergences

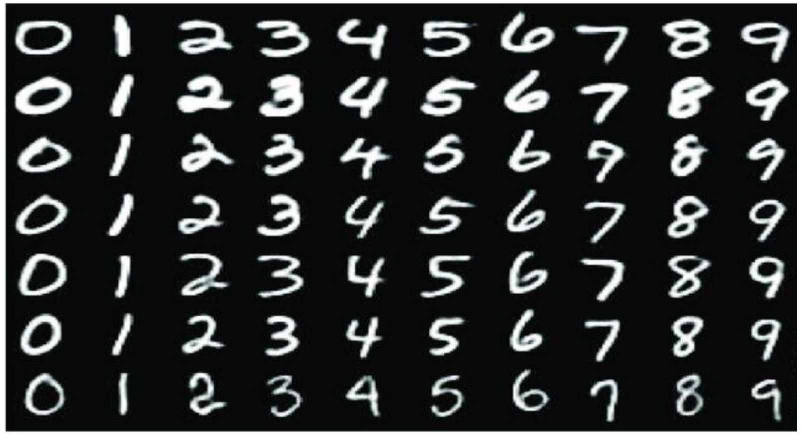
UMass Amherst, Department of Mathematics and Statistics

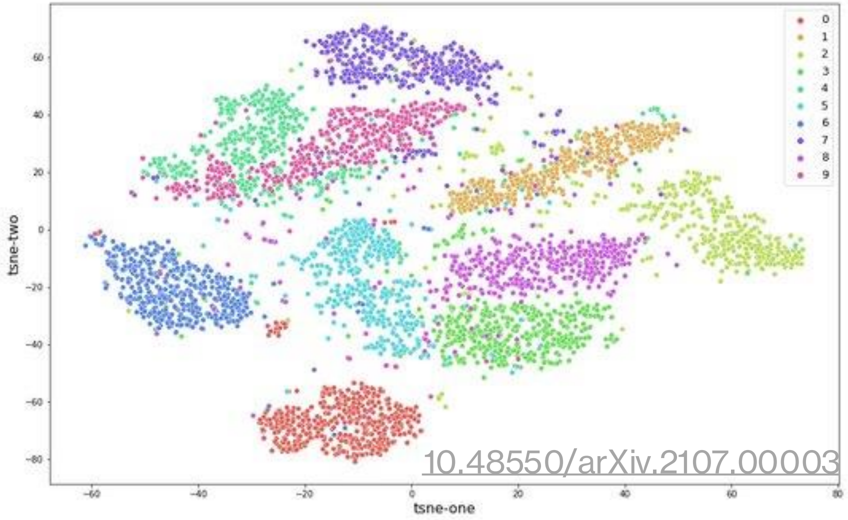
Hyemin Gu (joint work with Markos A. Katsoulakis, Luc Rey-Bellet, Benjamin J. Zhang)

SIAM MDS 2024

Problem Setup and Motivation

- **Objective**: Learn distributions supported on **low-dimensional manifolds** using flow-based generative models a.k.a. generative flows

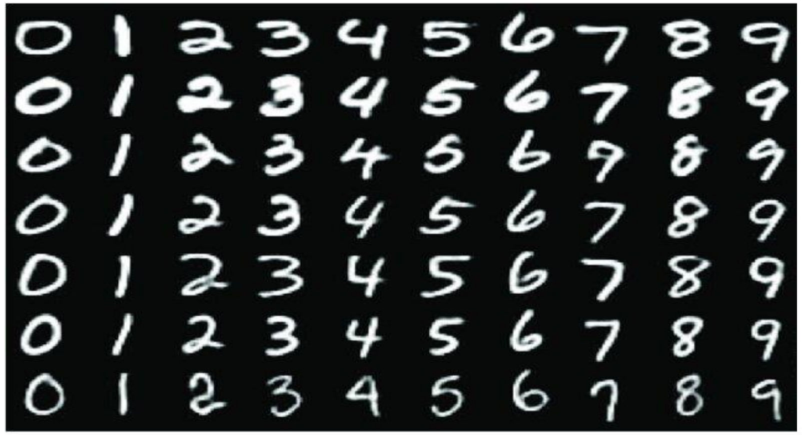


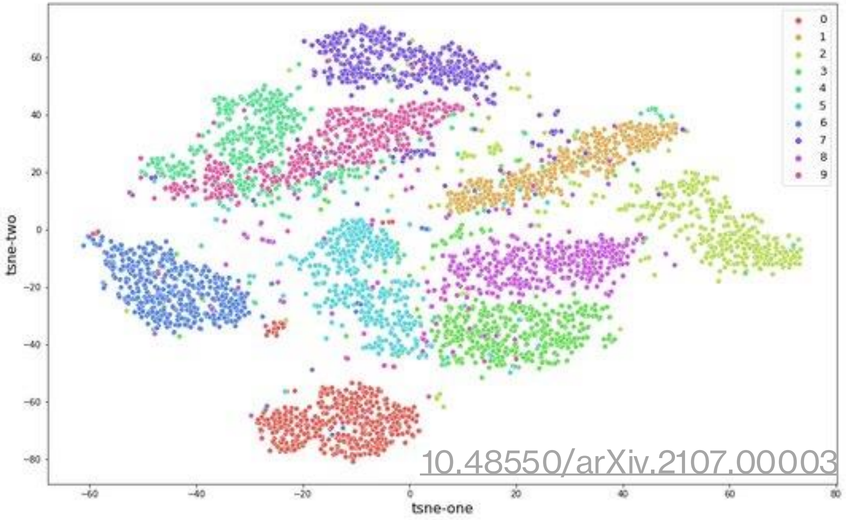


10.48550/arXiv.2107.00003

Problem Setup and Motivation

- **Objective**: Learn distributions supported on **low-dimensional manifolds** using flow-based generative models a.k.a. generative flows
- **Challenges**:
  - An optimization-friendly metric for comparing high-dimensional distributions with one of those supported on low-dimensional manifolds
  - Choosing among flows that push-forwards a prior distribution to a target distribution

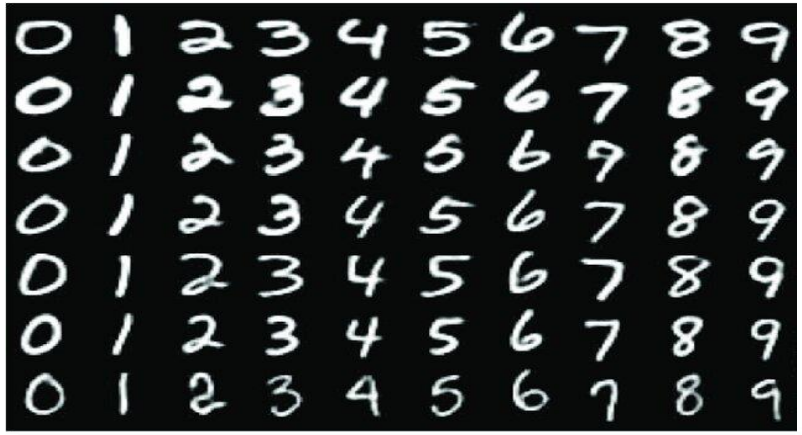


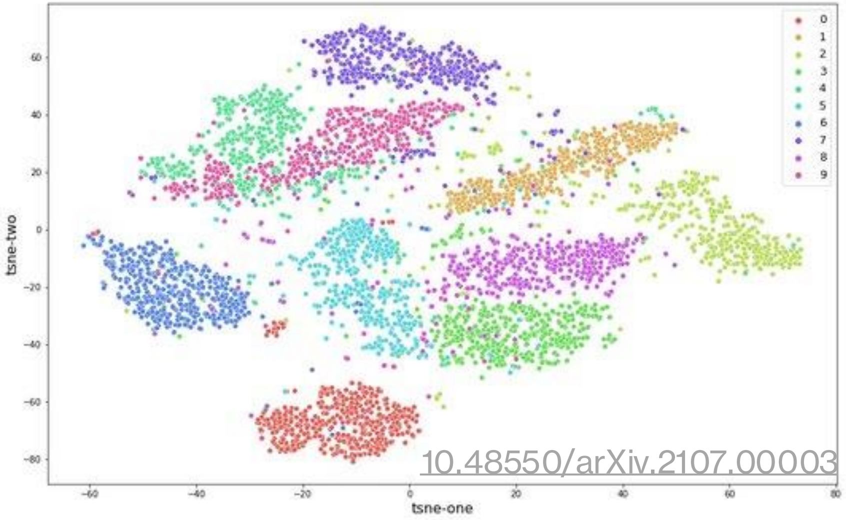


10.48550/arXiv.2107.00003

Problem Setup and Motivation

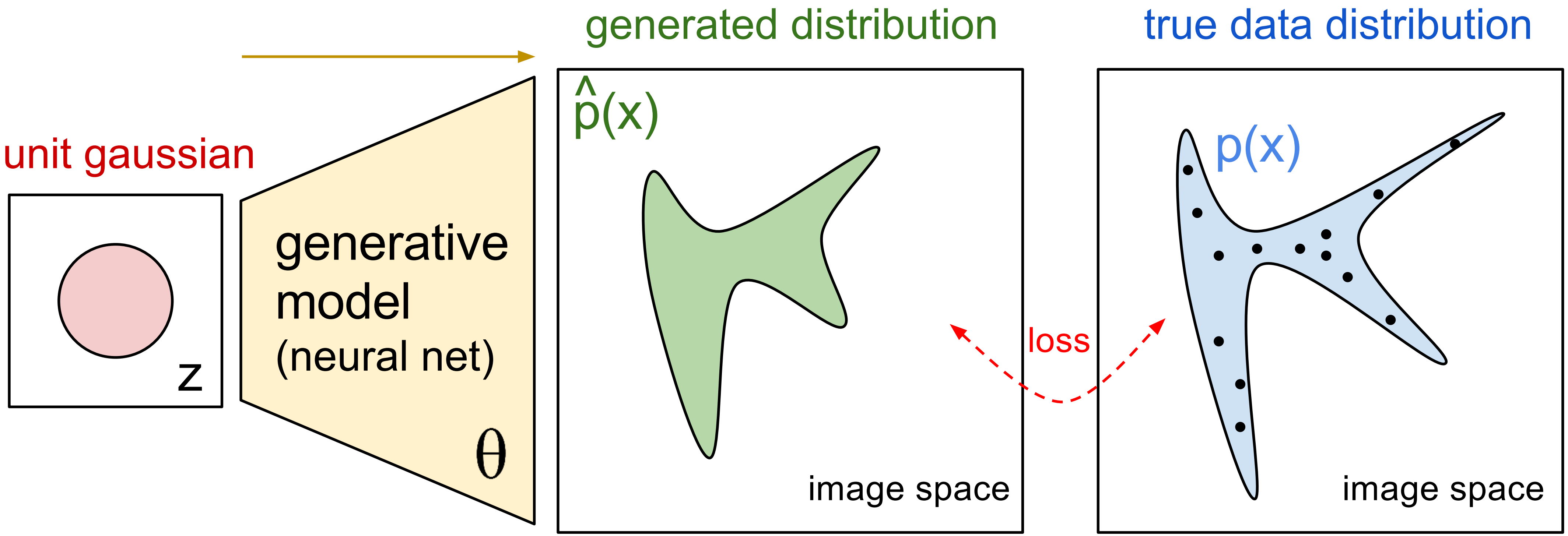
- **Objective**: Learn distributions supported on **low-dimensional manifolds** using flow-based generative models a.k.a. generative flows
- **Challenges**:
  - An optimization-friendly metric for comparing high-dimensional distributions with one of those supported on low-dimensional manifolds
  - Choosing among flows that push-forwards a prior distribution to a target distribution
- **Key Question**: How do we ensure that a learning problem for continuous-time generative flows to be **well-posed** and **robust** with respect to data submanifolds and time-discretization?





10.48550/arXiv.2107.00003

Concept of generative models



Generative Flow Formulation

Learning problem as a **transport between distributions** ρ_0 and ρ_T

- **Fokker-Planck equation** (eventually formulated as a Mean Field Game)

$$\inf_{v,\rho} \; J(v, \rho; \pi)$$

$v: \mathbb{R}^d \times [0, \infty) \to \mathbb{R}^d, \quad \rho: \mathcal{P}(\mathbb{R}^d) \times [0, \infty) \to \mathcal{P}(\mathbb{R}^d)$

$$s.t. \;\; \rho_t + \nabla \cdot (v\rho) = \frac{\sigma^2}{2} \Delta\rho, \qquad \rho_0 \; is \; given$$

Generative Flow Formulation

Learning problem as a **transport between distributions** ρ_0 and ρ_T

- **Fokker-Planck equation** (eventually formulated as a Mean Field Game)

$$\inf_{v,\rho} \; J(v,\rho;\pi)$$

$v: \mathbb{R}^d \times [0,\infty) \to \mathbb{R}^d, \quad \rho: \mathcal{P}(\mathbb{R}^d) \times [0,\infty) \to \mathcal{P}(\mathbb{R}^d)$

$$s.t. \;\; \rho_t + \nabla \cdot (v\rho) = \frac{\sigma^2}{2}\Delta\rho, \qquad \rho_0 \; is \; given$$

- **ODE/SDE**

$$\inf_{v,\rho} \; J(v,\rho;\pi)$$

$$s.t. \;\; X_t = v(X_t,t)dt + \sigma dW_t, \qquad X_0 \sim \rho_0$$

We consider deterministic flows, i.e. $\sigma = 0$.

f-Divergences and Their Challenges

Formal definition of f-divergences

$f: (0, \infty) \to \mathbb{R}$ convex, $f(1) = 0$, lower semi-continuous, super-linear

$$D_f(P||Q) := E_Q\left[f\left(\frac{dP}{dQ}\right)\right]$$

- ex) KL divergence $D_{KL}(P||Q)$ for $f(x) = x \log x$, α –divergence $D_\alpha(P||Q)$ for $f(x) = \frac{x^\alpha - 1}{\alpha(\alpha - 1)}$

f-Divergences and Their Challenges

Formal definition of f-divergences

$f: (0, \infty) \to \mathbb{R}$ convex, $f(1) = 0$, lower semi-continuous, super-linear

$$D_f(P||Q) := E_Q\left[f\left(\frac{dP}{dQ}\right)\right]$$

- ex) KL divergence $D_{KL}(P||Q)$ for $f(x) = x \log x$, α –divergence $D_\alpha(P||Q)$ for $f(x) = \frac{x^\alpha - 1}{\alpha(\alpha-1)}$

Variational formulation of f-divergences

$$D_f(P||Q) = \sup_{\phi \in C_b(\mathbb{R}^d)} \{E_P[\phi] - E_Q[f^*(\phi)]\}$$

where f^* is the Legendre transform of f.

f-Divergences and Their Challenges

Formal definition of f-divergences

$f: (0, \infty) \to \mathbb{R}$ convex, $f(1) = 0$, lower semi-continuous, super-linear

$$D_f(P||Q) := E_Q\left[f\left(\frac{dP}{dQ}\right)\right]$$

- ex) KL divergence $D_{KL}(P||Q)$ for $f(x) = x \log x$, α –divergence $D_\alpha(P||Q)$ for $f(x) = \frac{x^\alpha - 1}{\alpha(\alpha-1)}$

Variational formulation of f-divergences

$$D_f(P||Q) = \sup_{\phi \in C_b(\mathbb{R}^d)} \{E_P[\phi] - E_Q[f^*(\phi)]\}$$

where f^* is the Legendre transform of f.

- **Properties:**
  - $P \mapsto D_f(P||Q)$ is strictly convex and $(P, Q) \mapsto D_f(P||Q)$ is convex. **(convexity)**
  - $D_f(P||Q) < \infty$ only if $P \ll Q$. **(absolute continuity required)**

f-Divergences and Their Challenges

Formal definition of f-divergences

$f : (0, \infty) \to \mathbb{R}$ convex, $f(1) = 0$, lower semi-continuous, super-linear

$$D_f(P||Q) \coloneqq E_Q\left[f\left(\frac{dP}{dQ}\right)\right]$$

- ex) KL divergence $D_{KL}(P||Q)$ for $f(x) = x \log x$, α –divergence $D_\alpha(P||Q)$ for $f(x) = \frac{x^\alpha - 1}{\alpha(\alpha-1)}$

Variational formulation of f-divergences

$$D_f(P||Q) = \sup_{\phi \in C_b(\mathbb{R}^d)} \{E_P[\phi] - E_Q[f^*(\phi)]\}$$

where f^* is the Legendre transform of f.

- **Properties:**
  - $P \mapsto D_f(P||Q)$ is strictly convex and $(P, Q) \mapsto D_f(P||Q)$ is convex. **(convexity)**
  - $D_f(P||Q) < \infty$ only if $P \ll Q$. **(absolute continuity required)**
- **Challenge**: Mutually singular distributions P and Q make f-divergences ill-posed.

Wasserstein-1 Proximal Regularization of f-divergence

Infimal convolution of D_f and W_1 provides **Wasserstein-1 proximal regularized f-divergence** [Birrell, Dupuis, Katsoulakis, Pantazis, Rey-Bellet (2022, JMLR)]

$$D_f^L(P||Q) = \inf_{R\in\mathcal{P}_1(\mathbb{R}^d)} \{D_f(R||Q) + L\cdot W_1(P,R)\}$$

$$= \sup_{\phi\in Lip_L(\mathbb{R}^d)} \{E_P[\phi] - E_Q[f^*(\phi)]\}$$

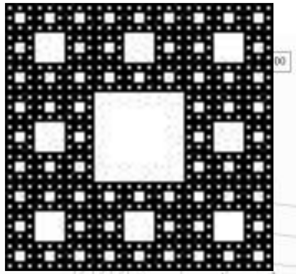
- Variational derivative $\frac{\delta D_f^L(P||Q))}{\delta P}$ exists **for all $P\in\mathcal{P}_1(\mathbb{R}^d)$ and Q**; It is the optimizer ϕ^*

$$\frac{\delta D_f^L(P||Q))}{\delta P} = \phi^*$$

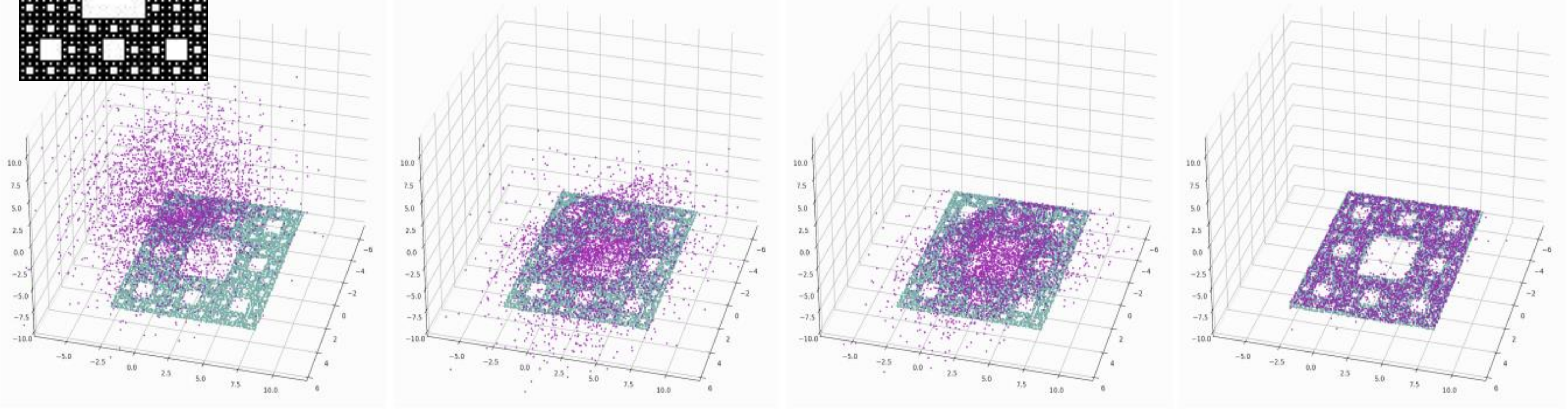
- $D_f^L(P||Q) \le \min(D_f(P||Q), L\cdot W_1(P,Q))$

- **Purpose**: comparison of mutually singular distributions

Wasserstein-1 Proximal Regularization of f-divergence



Wasserstein gradient flow learning Sierpinski carpet (cyan) from 3D gaussian prior (magenta) [Gu, Birmpa, Pantazis, Rey-Bellet, Katsoulakis (2024, SIAMODS)]



- **Purpose**: comparison of mutually singular distributions

Wasserstein-2 Proximal Regularization of terminal cost

Use **Dynamic (Bernamou-Brenier) formulation** of Wasserstein-2 divergence

$$W_2^2(P,Q) = \inf_{v,\rho} \int_0^1 \int_{\mathbb{R}^d} |v(x,t)|^2 \rho(x,t) dxdt \quad s.t. \quad \rho_t + \nabla \cdot (v\rho) = 0, \rho_0 = P, \rho_1 = Q$$

Infimal convolution of $\mathcal{F}$ and W_2^2 provides **Wasserstein-2 proximal regularized terminal cost**

$$\inf_{\rho,v} \left\{ \mathcal{F}(\rho_T) + \frac{\lambda}{2T} \cdot W_2^2(\rho_0, \rho_T) \right\} \quad s.t. \quad \rho_t + \nabla \cdot (v\rho) = 0, \quad \rho_0 = P$$

$$= \inf_{\rho,v} \left\{ \mathcal{F}(\rho_T) + \lambda \cdot \int_0^T \int_{\mathbb{R}^d} \frac{1}{2} |v(x,t)|^2 \rho(x,t) dxdt \right\} \quad s.t. \quad \rho_t + \nabla \cdot (v\rho) = 0, \quad \rho_0 = P$$

- **Interpretation:** Adds kinetic energy penalization to flow paths
- Unlike Wasserstein-1, it focus on path regularity

Wasserstein-2 Proximal Regularization of terminal cost

Use **Dynamic (Bernamou-Brenier) formulation** of Wasserst

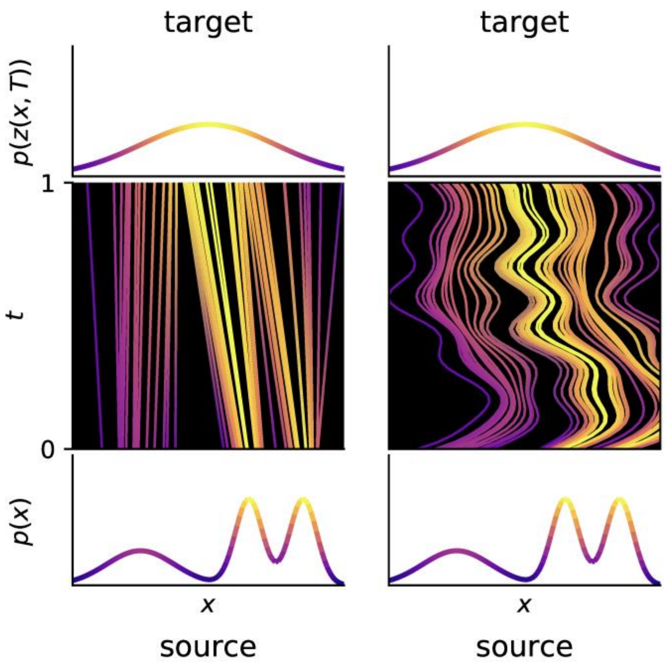
$$W_2^2(P,Q) = \inf_{v,\rho} \int_0^1 \int_{\mathbb{R}^d} |v(x,t)|^2 \rho(x,t) dxdt \qquad s.t. \qquad \rho_t + \nabla \cdot$$

Infimal convolution of $\mathcal{F}$ and W_2^2 provides **Wasserstein-2 pro terminal cost**

$$\inf_{\rho,v} \left\{ \mathcal{F}(\rho_T) + \frac{\lambda}{2T} \cdot W_2^2(\rho_0, \rho_T) \right\} \qquad s.t. \qquad \rho_t + \nabla \cdot (v\rho) = 0,$$

$$= \inf_{\rho,v} \left\{ \mathcal{F}(\rho_T) + \lambda \cdot \int_0^T \int_{\mathbb{R}^d} \frac{1}{2} |v(x,t)|^2 \rho(x,t) dxdt \right\} \qquad s.t. \qquad \rho_t \cdot$$

- **Interpretation:** Adds kinetic energy penalization to flow paths
- Unlike Wasserstein-1, it focus on path regularity



Left: Wasserstein-2 proximal regularized flow. **Right**: generic flow.

Formulating Generative Flows Using Mean-Field Game (MFG) Theory

Mean Field Game

$$\inf_{v,\rho}\left\{\mathcal{F}(\rho(\cdot,T)) + \int_0^T \mathcal{I}(\rho(\cdot,t))dt + \int_0^T \int_{\mathbb{R}^d} L(x,v(x,t))\rho(x,t)dxdt\right\}$$

$$s.t.\ \rho_t + \nabla\cdot(v\rho) = \frac{\sigma^2}{2}\Delta\rho, \rho_0 = \rho(\cdot,0)$$

Optimal solution satisfies the following coupled **PDE system**

- **Backward Hamilton-Jacobi-Bellman (HJB) equation**

$$-\partial_t U + H(x,\nabla U) - \frac{\sigma^2}{2}\Delta U = \frac{\delta\mathcal{I}(\rho)}{\delta\rho}(x), \qquad U(x,T) = \frac{\delta\mathcal{F}(\rho_T)}{\delta\rho_T}(x)$$

- **Forward Fokker-Planck equation**

$$\rho_t - \nabla\cdot\left(\nabla_p H(x,\nabla U)\rho\right) = \frac{\sigma^2}{2}\Delta\rho, \qquad \rho_0 = \rho(\cdot,0)$$

where the Hamiltonian $H(x,p) = \sup_v\{-p^T v - L(x,v)\}$.

Formulating Generative Flows Using Mean-Field Game (MFG) Theory

Mean Field Game

$$\inf_{v,\rho}\left\{\mathcal{F}\big(\rho(\cdot,T)\big)+\int_0^T \mathcal{I}(\rho(\cdot,t))dt+\int_0^T\int_{\mathbb{R}^d} L(x,v(x,t))\rho(x,t)dxdt\right\}$$

$$s.t.\ \rho_t+\nabla\cdot(v\rho)=0,\ \rho_0=\rho(\cdot,0)$$

Optimal solution satisfies the following coupled **PDE system**

- **Backward Hamilton-Jacobi (HJ) equation**

$$-\partial_t U+H(x,\nabla U)=\frac{\delta\mathcal{I}(\rho)}{\delta\rho}(x),\qquad U(x,T)=\frac{\delta\mathcal{F}(\rho_T)}{\delta\rho_T}(x)$$

- **Forward Fokker-Planck equation**

$$\rho_t-\nabla\cdot\big(\nabla_p H(x,\nabla U)\rho\big)=0,\qquad \rho_0=\rho(\cdot,0)$$

where the Hamiltonian $H(x,p)=\sup_v\{-p^T v-L(x,v)\}$.

Combining Wasserstein-1 and Wasserstein-2 Proximals

- $W_1 \oplus W_2$-flow [Gu, Katsoulakis, Rey-Bellet, Zhang (2024)]
  Combine $\boldsymbol{D_f^L = W_1}$ **proximal of** $\boldsymbol{D_f}$ and $\boldsymbol{W_2}$ **proximal of** $\boldsymbol{D_f^L}$

$$\inf_{\rho_T} \left\{ D_f^L(\rho_T || \pi) + \frac{\lambda}{2T} \cdot W_2^2(\rho_0, \rho_T) \right\}$$

Terminal cost $\mathcal{F}\big(\rho(\cdot, T)\big)$

Running cost $\int_0^T \int_{\mathbb{R}^d} L(x, v(x,t))\rho(x,t)\,dx\,dt$

Combining Wasserstein-1 and Wasserstein-2 Proximals

- $W_1 \oplus W_2$-flow [Gu, Katsoulakis, Rey-Bellet, Zhang (2024)]
  Combine $\boldsymbol{D_f^L = W_1}$ **proximal of** $\boldsymbol{D_f}$ and $\boldsymbol{W_2}$ **proximal of** $\boldsymbol{D_f^L}$

$$\inf_{\rho_T}\left\{D_f^L(\rho_T||\pi) + \frac{\lambda}{2T}\cdot W_2^2(\rho_0,\rho_T)\right\}$$

Terminal cost $\mathcal{F}(\rho(\cdot,T))$

Running cost $\int_0^T \int_{\mathbb{R}^d} L(x,v(x,t))\rho(x,t)dxdt$

$$= \inf_{\rho_T}\left\{\inf_{\sigma}\{D_f(\sigma||\pi) + L\cdot W_1(\rho_T,\sigma)\} + \frac{\lambda}{2T}\cdot W_2^2(\rho_0,\rho_T)\right\}$$

Composition of proximal operators

Combining Wasserstein-1 and Wasserstein-2 Proximals

- $W_1 \oplus W_2$-flow [Gu, Katsoulakis, Rey-Bellet, Zhang (2024)]
  Combine $\boldsymbol{D_f^L = W_1}$ **proximal of** $\boldsymbol{D_f}$ and $\boldsymbol{W_2}$ **proximal of** $\boldsymbol{D_f^L}$

$$\inf_{\rho_T}\left\{D_f^L(\rho_T\|\pi) + \frac{\lambda}{2T}\cdot W_2^2(\rho_0,\rho_T)\right\}$$

Terminal cost $\mathcal{F}(\rho(\cdot,T))$

Running cost $\int_0^T \int_{\mathbb{R}^d} L(x,v(x,t))\rho(x,t)dxdt$

$$= \inf_{\rho_T}\left\{\inf_{\sigma}\{D_f(\sigma\|\pi) + L\cdot W_1(\rho_T,\sigma)\} + \frac{\lambda}{2T}\cdot W_2^2(\rho_0,\rho_T)\right\}$$

Composition of proximal operators

$$= \inf_{v,\rho}\left\{\sup_{\phi\in Lip(L)}\{E_{\rho(\cdot,T)}[\phi] - E_\pi[f^*(\phi)]\} + \lambda\cdot\int_0^T\int_{\mathbb{R}^d}\frac{1}{2}|v(x,t)|^2\rho(x,t)dxdt\right\}$$

Dual formulation of D_f^L

Dynamical formulation of W_2^2

$$\text{s.t.}\quad \frac{dx}{dt} = v(x(t),t), \qquad x(0)\sim\rho_0,(x)\ t\in[0,T]$$

Main theorem

Theorem: $\inf_{v,\rho}\left\{\sup_{\phi\in Lip_L}\{E_{\rho(\cdot,T)}[\phi]-E_\pi[f^*(\phi)]\}+\lambda\cdot\int_0^T\int_{\mathbb{R}^d}\frac{1}{2}|v(x,t)|^2\rho(x,t)dxdt\right\}$

s.t. $\frac{dx}{dt}=v(x(t),t),\ x(0)\sim\rho_0, t\in[0,T]$

has the following optimality conditions:

- $D_f^L=W_1$ **proximal of** D_f provides a well-defined terminal condition of the HJ equation
$$U(x,T)=\frac{\delta D_f^L(\rho_T,\pi)}{\delta\rho_T}(x)=\phi^*(x)$$
- W_2 **proximal of** D_f^L provides a well-defined the HJ dynamics
$$-\partial_t U+\frac{1}{2\lambda}|\nabla U|^2=0$$
which leads to an optimal velocity field $v=-\frac{1}{\lambda}\nabla U$ and continuity equation
$$\partial_t\rho-\nabla\cdot\left(\rho\frac{\nabla U}{\lambda}\right)=0$$
- W_2 **proximal of** D_f^L provides a linear optimal trajectory
$$x(t)=x(T)+\frac{T-t}{\lambda}\nabla\phi^*(x(T))$$

Uniqueness of optimal $W_1 \oplus W_2$-flow

Theorem: If the backward-forward PDE system

$$\begin{cases} \partial_t U + \frac{1}{2\lambda}|\nabla U|^2 = 0 \\ \partial_t \rho - \nabla \cdot \left(\rho \frac{\nabla U}{\lambda}\right) = 0 \end{cases}$$

with terminal condition $U(x,T) = \frac{\delta D_f^L(\rho_T,\pi)}{\delta \rho_T}(x) = \phi^*(x)$ has smooth solutions (U,ρ) on the torus Ω, then they are **unique** and the solution to the optimization problem

$$\inf_{v,\rho}\left\{\sup_{\phi \in Lip_L}\{E_{\rho(\cdot,T)}[\phi] - E_\pi[f^*(\phi)]\} + \lambda \cdot \int_0^T \int_{\mathbb{R}^d} \frac{1}{2}|v(x,t)|^2 \rho(x,t)dxdt\right\}$$

s.t. $\frac{dx}{dt} = v(x(t),t),\ \ x(0) \sim \rho_0, t \in [0,T]$

is also unique.

Uniqueness of optimal $W_1 \oplus W_2$-flow

Theorem: If the backward-forward PDE system

$$\begin{cases} \partial_t U + \frac{1}{2\lambda}|\nabla U|^2 = 0 \\ \partial_t \rho - \nabla \cdot \left(\rho \frac{\nabla U}{\lambda}\right) = 0 \end{cases}$$

with terminal condition $U(x,T) = \frac{\delta D_f^L(\rho_T,\pi)}{\delta \rho_T}(x) = \phi^*(x)$ has smooth solutions (U,ρ) on the torus Ω, then they are **unique** and the solution to the optimization problem

$$\inf_{v,\rho}\left\{\sup_{\phi \in Lip_L}\{E_{\rho(\cdot,T)}[\phi] - E_\pi[f^*(\phi)]\} + \lambda \cdot \int_0^T \int_{\mathbb{R}^d} \frac{1}{2}|v(x,t)|^2 \rho(x,t)\,dxdt\right\}$$

$$\text{s.t.} \quad \frac{dx}{dt} = v(x(t),t),\ \ x(0) \sim \rho_0, t \in [0,T]$$

is also unique.

Uniqueness of optimal solution implies **well-posedness** of optimization problem.

Adversarial Training of Generative Flows

- Unlike **normalizing flows**, we bypass the need to **invert the flow** by **adversarial training** of the flow

$$\inf_{v,\rho}\left\{\sup_{\phi\in Lip(L)}\{E_{\rho(\cdot,T)}[\phi]-E_{\pi}[f^*(\phi)]\}+\lambda\cdot\int_0^T\int_{\mathbb{R}^d}\frac{1}{2}|v(x,t)|^2\rho(x,t)dxdt\right\}$$

Adversarial Training of Generative Flows

- Unlike **normalizing flows**, we bypass the need to **invert the flow** by **adversarial training** of the flow

$$\inf_{v,\rho}\left\{\sup_{\phi\in Lip(L)}\{E_{\rho(\cdot,T)}[\phi]-E_{\pi}[f^*(\phi)]\}+\lambda\cdot\int_0^T\int_{\mathbb{R}^d}\frac{1}{2}|v(x,t)|^2\rho(x,t)dxdt\right\}$$

- **Impact**: Our formulation resolves the ill-posedness issue of generative flows when learning distributions supported on **low-dimensional manifolds**.

Numerical experiment: Impact of Wasserstein-1 proximal regularization

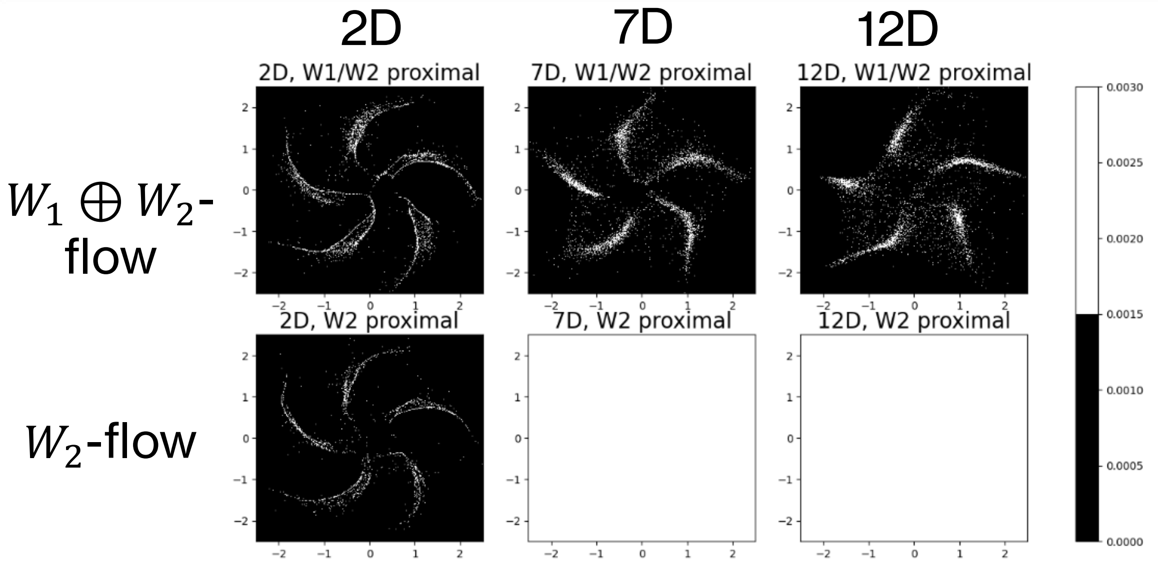


Fig. 1: Stable manifold learning via $\mathcal{W}_1$ proximal. $\mathcal{W}_1 \oplus \mathcal{W}_2$ flow (top), $\mathcal{W}_2$ flow (bottom).

Numerical experiment: Impact of Wasserstein-1 proximal regulari

Adversarial flow but no uniquely defined variational derivative

Normalizing flow

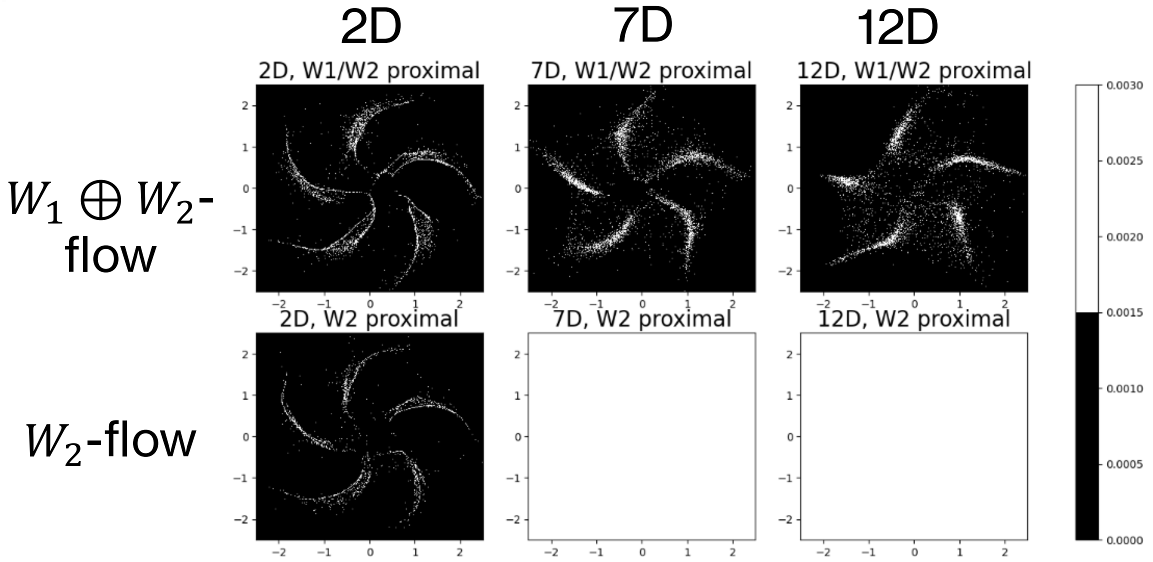


Fig. 1: Stable manifold learning via $\mathcal{W}_1$ proximal. $\mathcal{W}_1 \oplus \mathcal{W}_2$ flow (top), $\mathcal{W}_2$ flow (bottom).

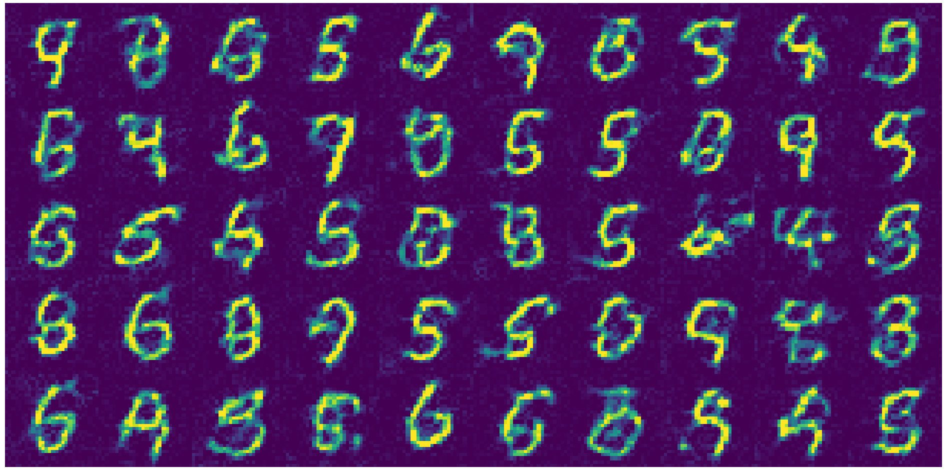
	$\mathcal{W}_1 \oplus \mathcal{W}_2$ flow	Potential Flow GAN	OT flow
2D	**8.0e-03**	1.3e-02	1.9e-01
7D	**1.0e-02**	1.6e+01	4.5e+09
12D	**1.6e-02**	3.7e+00	7.9e+26

Table 1: Comparison with Potential Flow GAN (Yang et al.) and OT flow (Onken et al.). $\mathcal{W}_2$ distance between original and generated data manifolds.

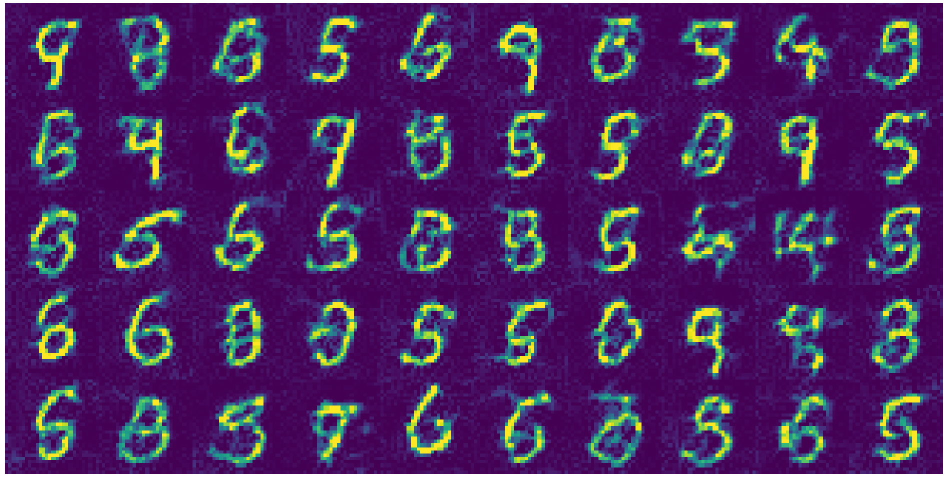
Unlike other generative flows, our proposed $W_1 \oplus W_2$-flow learns distributions supported on **low-dimensional manifolds** without autoencoders or specialized architectures.

Numerical experiment: Impact of Wasserstein-2 proximal regularization

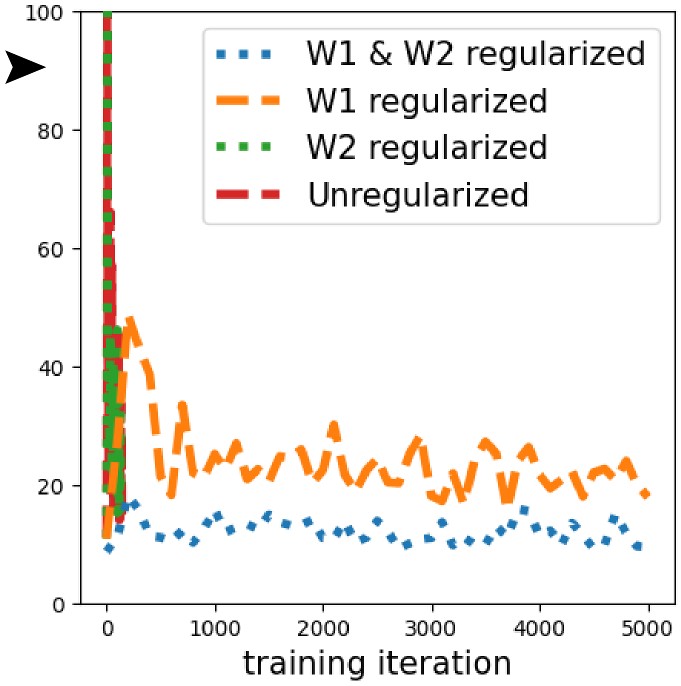
$\Delta t = 1$



$\Delta t = \frac{1}{64}$



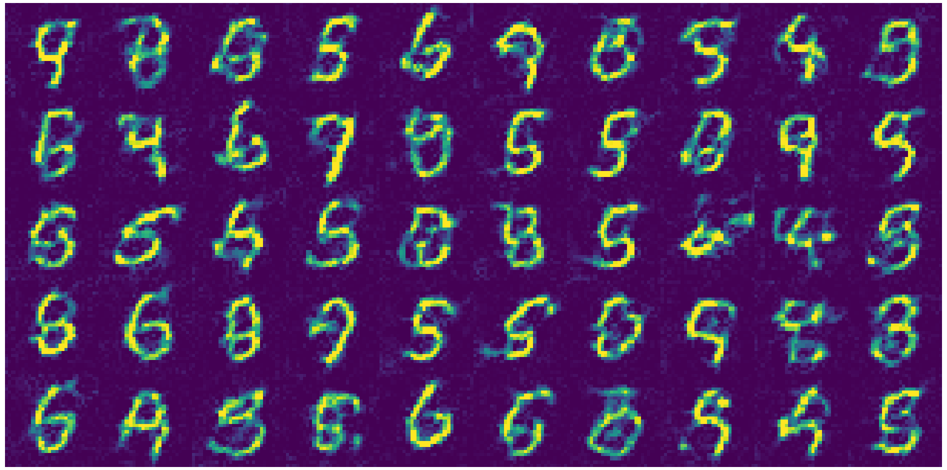
Total kinetic energy ➤



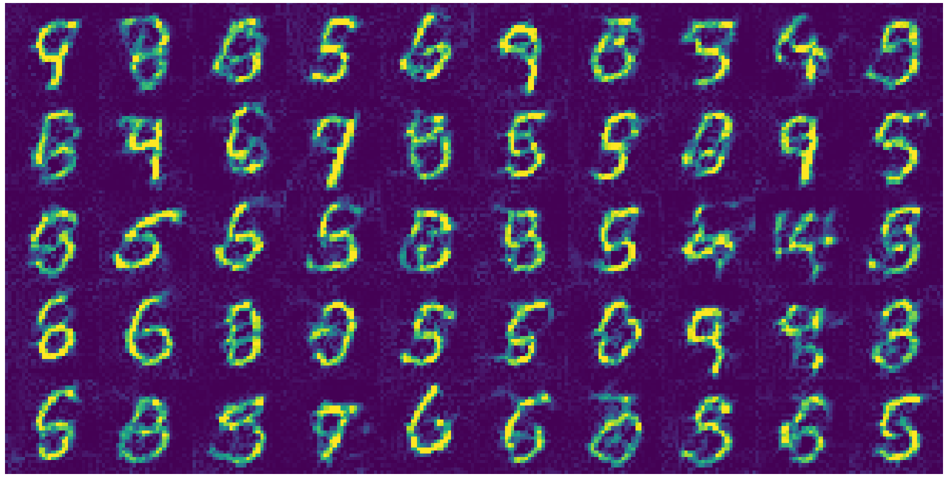
➤ $W_1 \oplus W_2$-flow implies **discretization invariance** in generative flows.

Numerical experiment: Impact of Wasserstein-2 proximal regularization

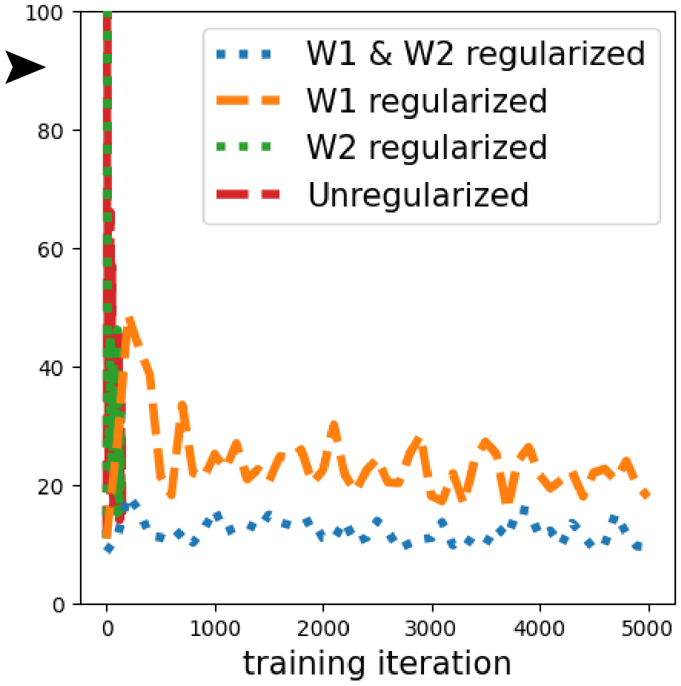
$\Delta t = 1$



$\Delta t = \frac{1}{64}$



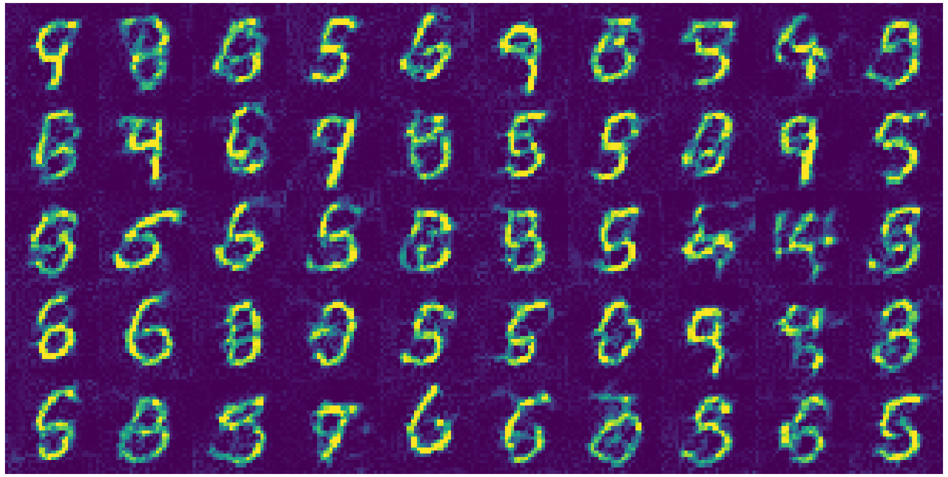
Total kinetic energy ➤



➤ $W_1 \oplus W_2$-flow implies **discretization invariance** in generative flows.

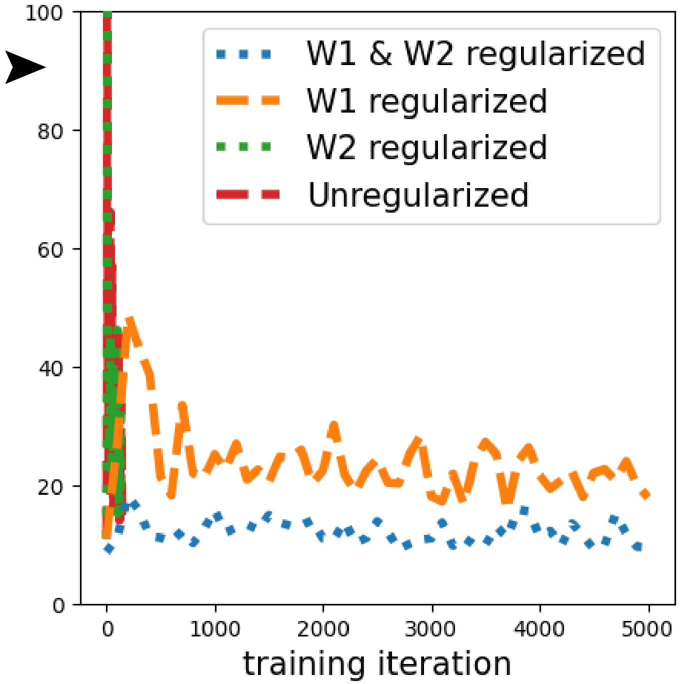
Numerical experiment: Impact of Wasserstein-2 proximal regularization

$\Delta t = 1$



$\Delta t = \frac{1}{64}$

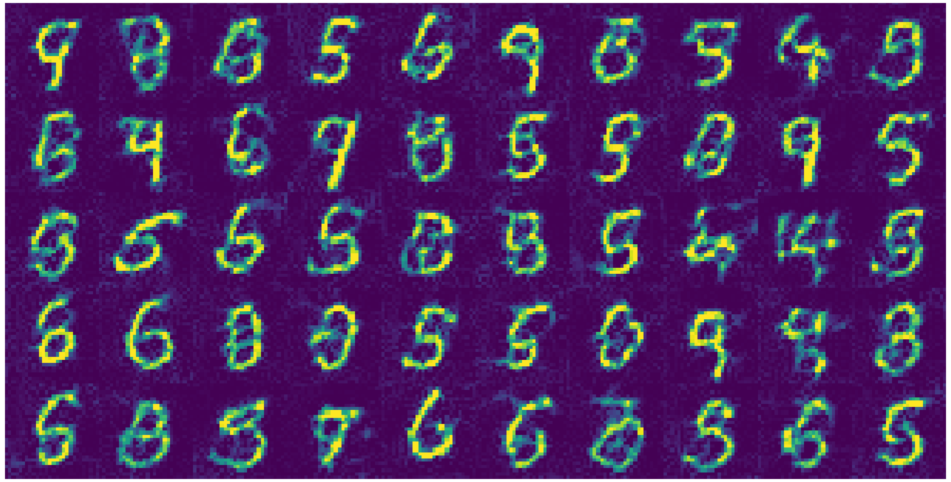
Total kinetic energy ➤



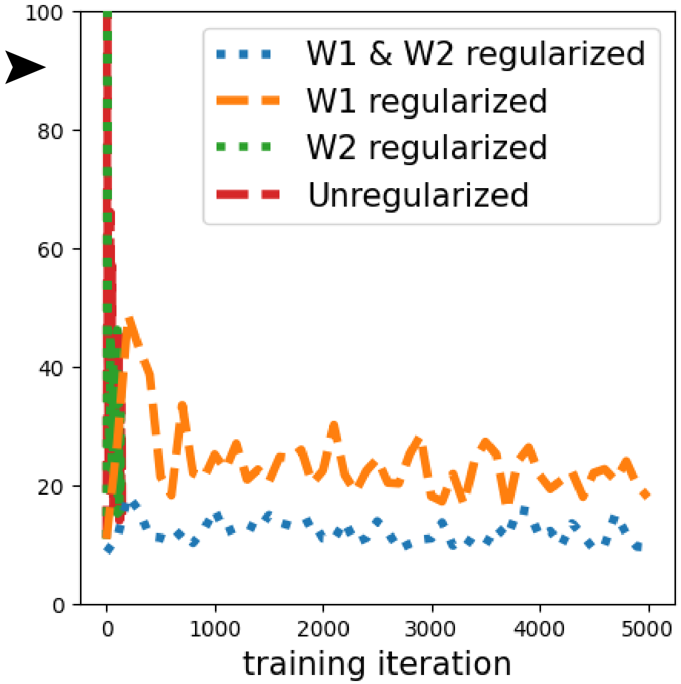
➤ $W_1 \oplus W_2$-flow implies **discretization invariance** in generative flows.

Numerical experiment: Impact of Wasserstein-2 proximal regularization

$\Delta t = 1$



Total kinetic energy ➤



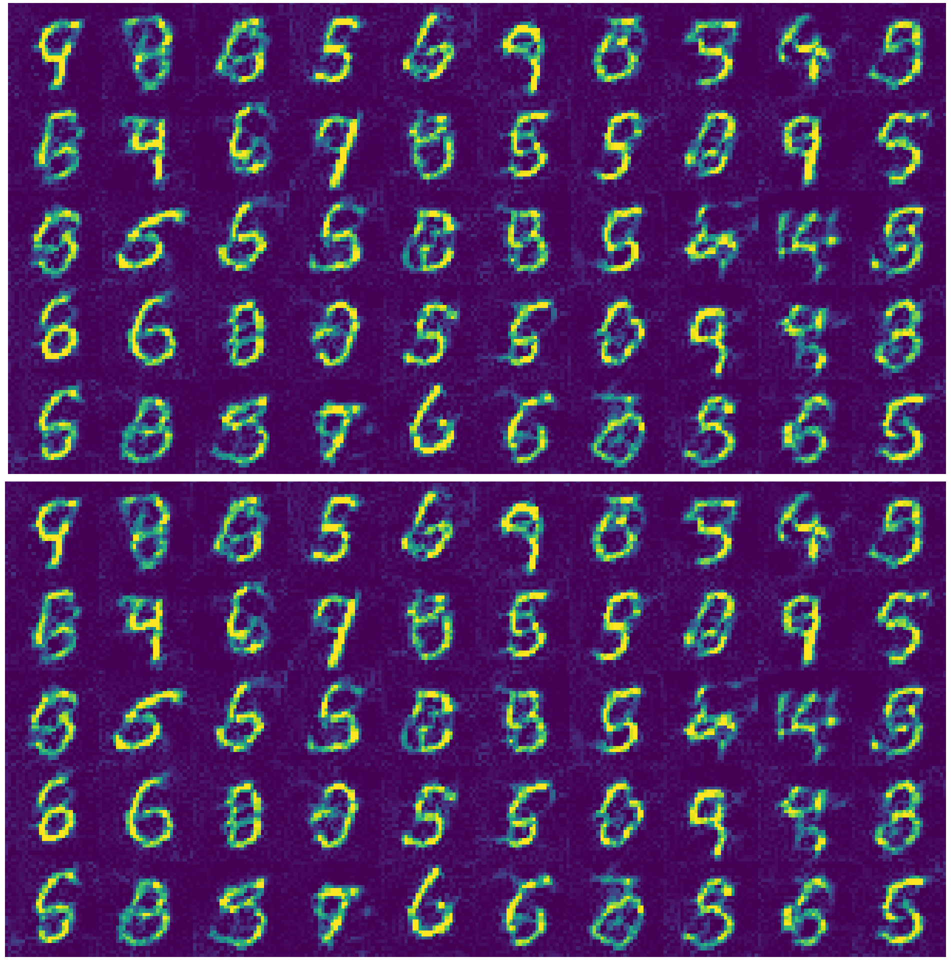
$\Delta t = \frac{1}{64}$

➤ $W_1 \oplus W_2$-flow implies **discretization invariance** in generative flows.

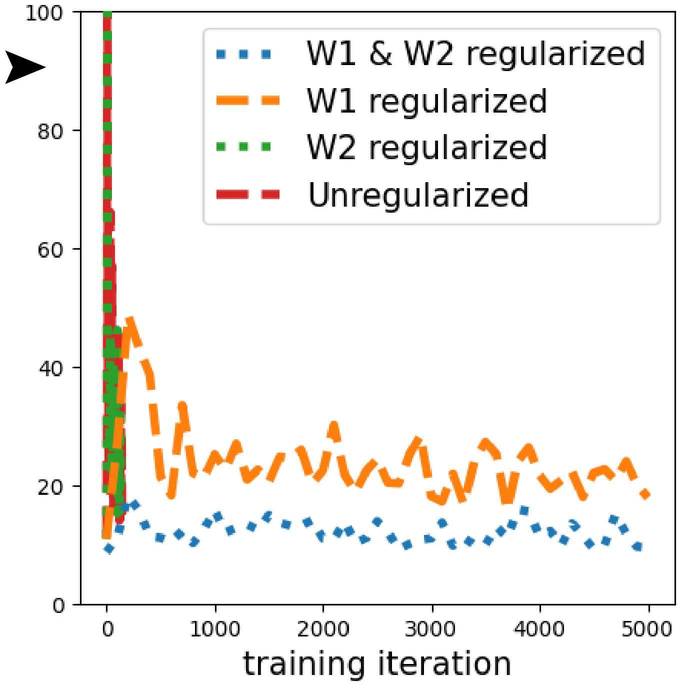
Numerical experiment: Impact of Wasserstein-2 proximal regularization

$\Delta t = 1$

$\Delta t = \frac{1}{64}$



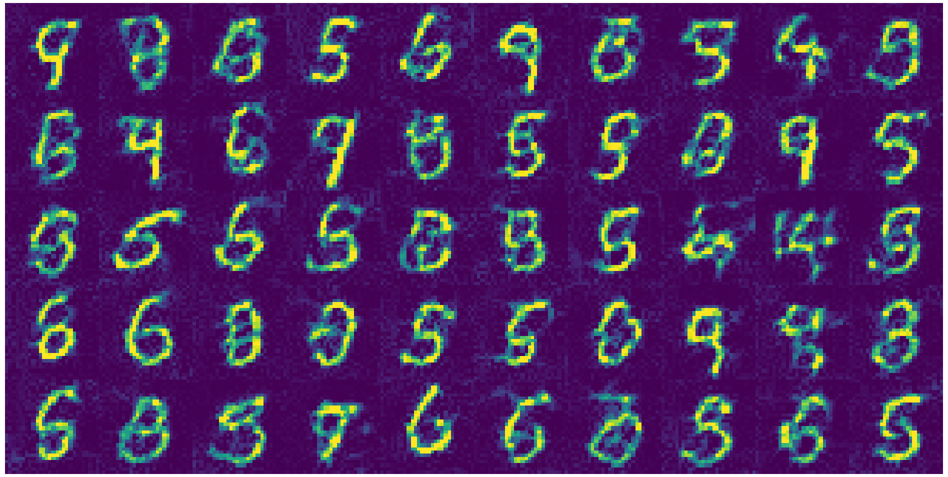
Total kinetic energy ➤



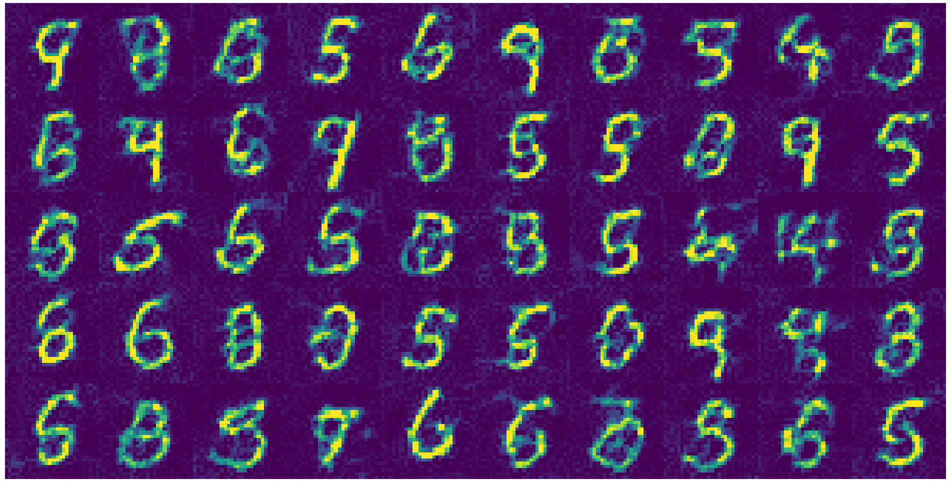
➤ $W_1 \oplus W_2$-flow implies **discretization invariance** in generative flows.

Numerical experiment: Impact of Wasserstein-2 proximal regularization

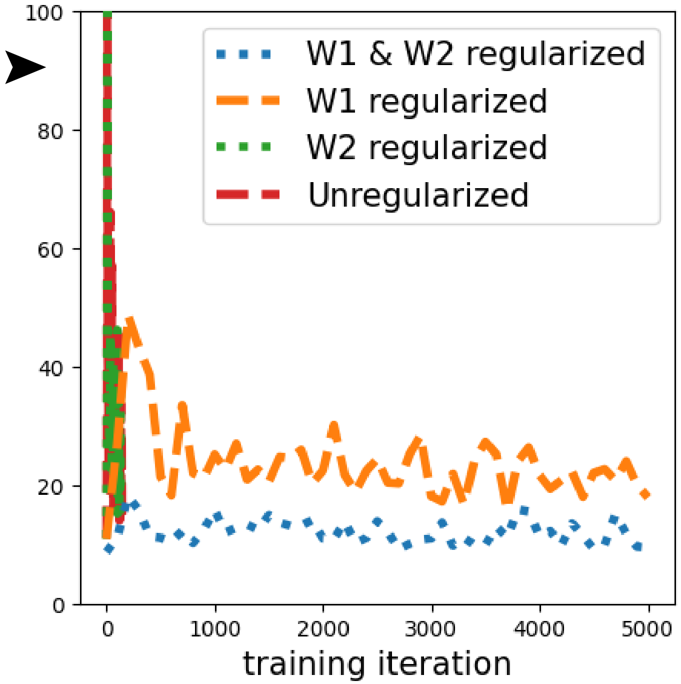
$\Delta t = 1$



$\Delta t = \frac{1}{64}$



Total kinetic energy ➤



➤ $W_1 \oplus W_2$-flow implies **discretization invariance** in generative flows.

$\Longrightarrow$ Try larger step size Δt to **train the flow faster**.

Summary and Future Work

Link to Paper

Learning objective

$$\inf_{v,\rho}\left\{\sup_{\phi\in Lip_L}\{E_{\rho(\cdot,T)}[\phi]-E_\pi[f^*(\phi)]\}+\lambda\cdot\int_0^T\int_{\mathbb{R}^d}\frac{1}{2}|v(x,t)|^2\rho(x,t)dxdt\right\}$$

$$\text{s.t.}\quad \frac{dx}{dt}=v(x(t),t),\ \ x(0)\sim\rho_0, t\in[0,T]$$

Optimality conditions

Wasserstein-2 Proximal regularization

$$\partial_t U-\frac{1}{\lambda}|\nabla U|^2=0,\qquad U(\cdot,T)=\phi^*$$

$$\partial_t\rho-\nabla\cdot\left(\rho\frac{\nabla U}{\lambda}\right)=0,\qquad \rho(\cdot,0)=\rho_0$$

Linear trajectories

Wasserstein-1 Proximal regularization

Uniqueness of optimal solution $\Longrightarrow$ Well-posed learning problem

Future work:

- Incorporate diffusion term and interaction cost to the model
- Efficient parametrization of the flow using the PDEs for the optimal solution