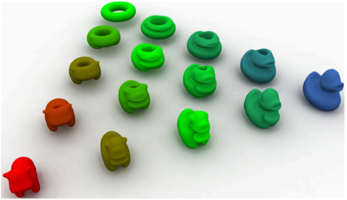
OAT-FM: Improved Flow Matching via Optimal Acceleration Transport

Hongteng Xu

Gaoling School of Artificial Intelligence, Renmin University of China

Jan. 23, 2026

Distribution Comparison and Transformation: Key Learning Tasks



Shape Matching and Interpolation

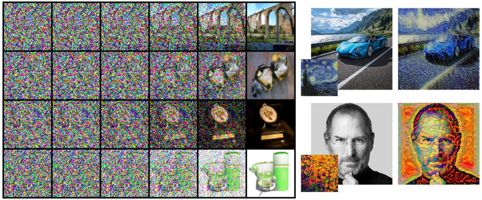
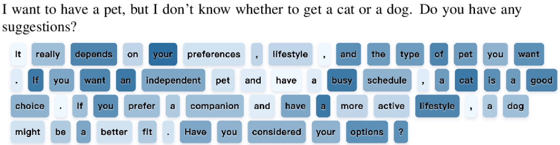


Image (Conditional) Generation

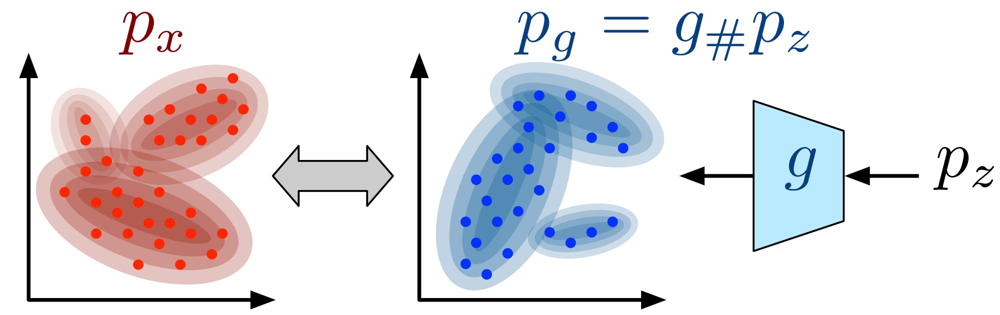


Textual Data Comparison and Generation

- Clustering, Domain Adaptation, Representation Learning, Generative Modeling, ...
- Optimal transport: A solid solution to distribution comparison and transformation.

Convolutional Wasserstein distances: Efficient optimal transportation on geometric domains. ToG, 2015.
Flow Matching for Generative Modeling. ICLR, 2023. Large language diffusion models. NeurIPS, 2025.

Generative Modeling = Distribution Fitting and Matching



- $g : \mathcal{Z} \mapsto \mathcal{X}$ is the generator/decoder.
- p_z is the (predefined) latent distribution, and $p_g = g_{\#} p_z$ is the model distribution.
- Learn g to fit data distribution p_x by p_g under a metric, and OT is a natural choice.

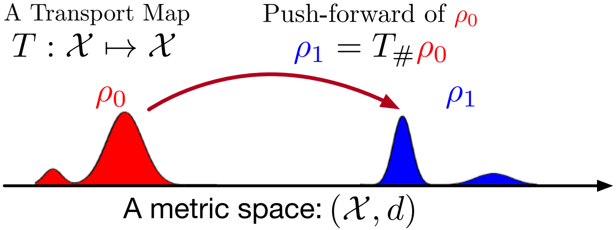
Outline

1. **A Quick Review of Generative Modeling Based on Static OT**
   - Optimal transport problem and Wasserstein distance
   - Wasserstein GAN (WGAN)
   - Wasserstein Autoencoder (WAE)

2. **Recent Generative Modeling Methods Based on Dynamic OT**
   - OT-based conditional flow matching
   - Improved flow matching based on Optimal Acceleration Transport (OAT)

Origin: The Monge-form of the Optimal Transport Problem

Gaspard Monge (1746-1818)

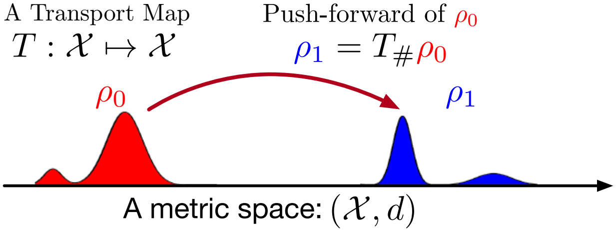


The Monge-form of OT problem proposed in 1781.

Key question: How to find a map that minimizes the transport cost?

Origin: The Monge-form of the Optimal Transport Problem

Gaspard Monge (1746-1818)



The Monge-form of OT problem proposed in 1781.

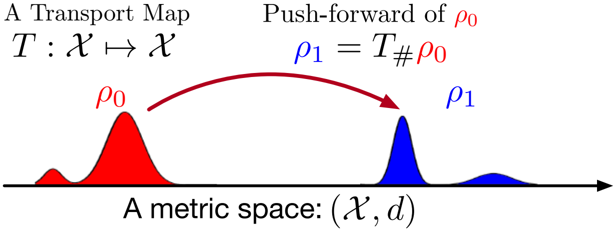
Key question: How to find a map that minimizes the transport cost?

- The p-order Monge problem:

$$\mathcal{M}_p(\rho_0, \rho_1) := \left(\inf_T \int_{x \in \mathcal{X}} \underbrace{d^p(x, T(x))}_{\text{cost per sample}} \mathrm{d}\rho_0(x)\right)^{1/p}, \quad s.t. \quad \underbrace{T_{\#}\rho_0 = \rho_1}_{\text{measure preserving}} \tag{1}$$

Origin: The Monge-form of the Optimal Transport Problem

Gaspard Monge (1746-1818)



The Monge-form of OT problem proposed in 1781.

Key question: How to find a map that minimizes the transport cost?

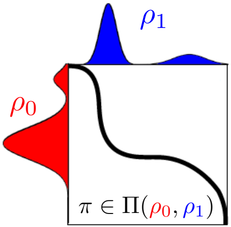
- The p-order Monge problem:

$$\mathcal{M}_p(\rho_0, \rho_1) := \left(\inf_T \int_{x \in \mathcal{X}} \underbrace{d^p(x, T(x))}_{\text{cost per sample}} \mathrm{d}\rho_0(x) \right)^{1/p}, \quad s.t. \quad \underbrace{T_\# \rho_0 = \rho_1}_{\text{measure preserving}} \tag{1}$$

- Notably, the minimizer of (1) may not exist, e.g., ρ_0 is a Dirac measure while ρ_1 is not.

From Transport Map to Transport Plan: The Kantorovich-form of OT

Leonid Kantorovich (1912-1986)

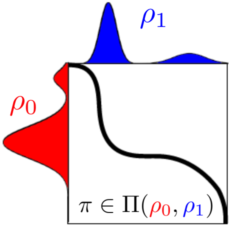


The Kantorovich-form of OT proposed in 1939

- Find a **transport plan/coupling** to minimize the expected cost.

From Transport Map to Transport Plan: The Kantorovich-form of OT

Leonid Kantorovich (1912-1986)



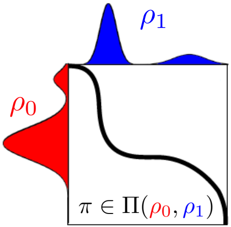
The Kantorovich-form of OT proposed in 1939

- Find a **transport plan/coupling** to minimize the expected cost.

$$\mathcal{W}_p(\rho_0, \rho_1) := \Big(\inf_{\pi} \underbrace{\int_{(x,y)\in\mathcal{X}^2} d^p(x,y)\pi(x,y)\mathrm{d}x\mathrm{d}y}_{\mathbb{E}_{x,y\sim\pi}[d^p(x,y)]}\Big)^{1/p}$$
$$s.t.\ \pi \in \Pi(\rho_0, \rho_1) = \Big\{\pi \geq 0 \Big| \int_{\mathcal{X}} \pi(x,\cdot)\mathrm{d}x = \rho_1, \int_{\mathcal{X}} \pi(\cdot, y)\mathrm{d}y = \rho_0\Big\}. \tag{2}$$

From Transport Map to Transport Plan: The Kantorovich-form of OT

Leonid Kantorovich (1912-1986)



The Kantorovich-form of OT proposed in 1939

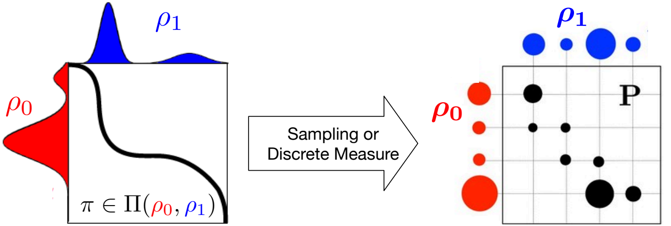
- Find a **transport plan/coupling** to minimize the expected cost.

$$\mathcal{W}_p(\rho_0, \rho_1) := \Big(\inf_{\pi} \underbrace{\int_{(x,y)\in\mathcal{X}^2} d^p(x,y)\pi(x,y)\mathrm{d}x\mathrm{d}y}_{\mathbb{E}_{x,y\sim\pi}[d^p(x,y)]}\Big)^{1/p} \tag{2}$$

$$s.t.\ \pi \in \Pi(\rho_0, \rho_1) = \Big\{\pi \geq 0 \Big| \int_{\mathcal{X}} \pi(x,\cdot)\mathrm{d}x = \rho_1, \int_{\mathcal{X}} \pi(\cdot,y)\mathrm{d}y = \rho_0\Big\}.$$

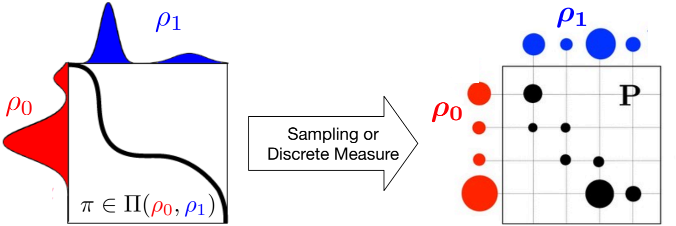
- When $d(x,y) = \|x-y\|_p$, $\mathcal{W}_p$ is p-**order Wasserstein distance**.

From Transport Map to Transport Plan: The Kantorovich-form of OT



Given $\boldsymbol{X} = \{x_m\}_{m=1}^M$, $\rho_0 = \sum_{m=1}^M \rho_{0,m}\delta_{x_m}$ and $\boldsymbol{Y} = \{y_n\}_{n=1}^N$, $\rho_1 = \sum_{n=1}^N \rho_{1,n}\delta_{y_n}$,

From Transport Map to Transport Plan: The Kantorovich-form of OT

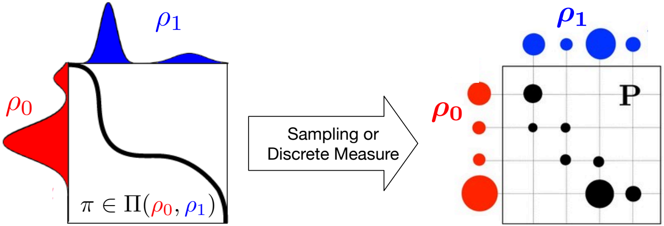


Given $\boldsymbol{X} = \{x_m\}_{m=1}^M$, $\rho_0 = \sum_{m=1}^M \rho_{0,m}\delta_{x_m}$ and $\boldsymbol{Y} = \{y_n\}_{n=1}^N$, $\rho_1 = \sum_{n=1}^N \rho_{1,n}\delta_{y_n}$,

$$\mathcal{W}_p(\boldsymbol{X}, \boldsymbol{Y}) := \Big(\min_{\boldsymbol{P}\in\Pi(\boldsymbol{\rho}_0,\boldsymbol{\rho}_1)} \sum_{m=1}^{M}\sum_{n=1}^{N} d^p(x_m, y_n)p_{mn}\Big)^{1/p} = \Big(\min_{\boldsymbol{P}\in\Pi(\boldsymbol{\rho}_0,\boldsymbol{\rho}_1)} \langle \boldsymbol{D}, \boldsymbol{P}\rangle\Big)^{1/p}, \quad (3)$$

where $\boldsymbol{D} = [d^p(x_m, y_n)]$, $\boldsymbol{P} = [p_{mn}]$, $\Pi(\boldsymbol{\rho}_0, \boldsymbol{\rho}_1) = \{\boldsymbol{P} > \boldsymbol{0} | \boldsymbol{P}\boldsymbol{1}_N = \boldsymbol{\rho}_0, \boldsymbol{P}^\top\boldsymbol{1}_M = \boldsymbol{\rho}_1\}$.

From Transport Map to Transport Plan: The Kantorovich-form of OT



Given $\boldsymbol{X} = \{x_m\}_{m=1}^M$, $\rho_0 = \sum_{m=1}^M \rho_{0,m}\delta_{x_m}$ and $\boldsymbol{Y} = \{y_n\}_{n=1}^N$, $\rho_1 = \sum_{n=1}^N \rho_{1,n}\delta_{y_n}$,

$$\mathcal{W}_p(\boldsymbol{X}, \boldsymbol{Y}) := \Big(\min_{\boldsymbol{P}\in\Pi(\boldsymbol{\rho}_0,\boldsymbol{\rho}_1)} \sum_{m=1}^M \sum_{n=1}^N d^p(x_m, y_n) p_{mn}\Big)^{1/p} = \Big(\min_{\boldsymbol{P}\in\Pi(\boldsymbol{\rho}_0,\boldsymbol{\rho}_1)} \langle \boldsymbol{D}, \boldsymbol{P}\rangle\Big)^{1/p}, \quad (3)$$

where $\boldsymbol{D} = [d^p(x_m, y_n)]$, $\boldsymbol{P} = [p_{mn}]$, $\Pi(\boldsymbol{\rho}_0, \boldsymbol{\rho}_1) = \{\boldsymbol{P} > \boldsymbol{0} | \boldsymbol{P}\boldsymbol{1}_N = \boldsymbol{\rho}_0, \boldsymbol{P}^\top \boldsymbol{1}_M = \boldsymbol{\rho}_1\}$.

- Applying the transport plan $\pi/\boldsymbol{P}$, we allow each sample $x \sim \rho_0$ to be split and mapped to multiple locations.
- If the optimal T^* exists, it determines an OT plan $\pi^*/\boldsymbol{P}^*$, so $\mathcal{W}_p \leq \mathcal{M}_p$.

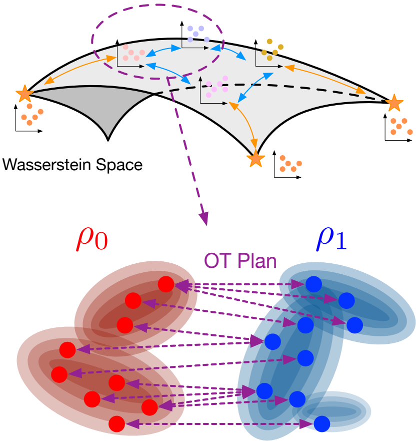
Advantages of Optimal Transport

A valid metric for probability measures

- $(\mathcal{P}(\mathcal{X}), \mathcal{W})$ is a metric space of probability measures defined in $\mathcal{X}$ (i.e., Wasserstein space).
- Apply to distribution comparison, fitting, and interpolation

OT plan indicates sample pairs

- Apply to point cloud/shape/graph matching



Classic OT-based Generative Modeling Paradigms

Solution 1: Minimize $\mathcal{W}_1$ approximately in its dual-form or its SW surrogates:

- **WGAN:** Wasserstein generative adversarial networks. ICML, 2017.
- **WGAN-GP:** Improved training of Wasserstein GANs. NeurIPS, 2017.
- **Max-SWG:** Max-sliced Wasserstein distance and its use for GANs. CVPR, 2019.
- **Amortized Max-SWG** Amortized projection optimization for sliced Wasserstein generative models. NeurIPS, 2022.

Classic OT-based Generative Modeling Paradigms

Solution 1: Minimize $\mathcal{W}_1$ approximately in its dual-form or its SW surrogates:

- **WGAN:** Wasserstein generative adversarial networks. ICML, 2017.
- **WGAN-GP:** Improved training of Wasserstein GANs. NeurIPS, 2017.
- **Max-SWG:** Max-sliced Wasserstein distance and its use for GANs. CVPR, 2019.
- **Amortized Max-SWG** Amortized projection optimization for sliced Wasserstein generative models. NeurIPS, 2022.

Solution 2: Minimize $\mathcal{W}_2$ approximately in its primal-form:

- **WAE:** Wasserstein Auto-Encoders. ICLR, 2018.
- **SinkDiff:** Learning generative models with Sinkhorn divergences. AISTATS, 2018.
- **SWAE:** Sliced Wasserstein auto-encoders. ICLR, 2018.
- **RAE:** Learning autoencoders with relational regularization. ICML, 2020.
- **Conditional Transport:** Exploiting Chain Rule and Bayes' Theorem to Compare Probability Distributions. NeurIPS, 2021.

Wasserstein Generative Adversarial Network (WGAN)

Wasserstein Generative Adversarial Network (WGAN): Fit the model distribution p_g by minimizing its 1-Wasserstein distance to the data distribution p_x **in the dual-form**:

$$\mathcal{W}_1(p_x, p_g) = \inf_{\pi \in \Pi(p_x, p_g)} \mathbb{E}_{(x, g(z)) \sim \pi}[\|x - g(z)\|_1] = \sup_{f \in L_1} \mathbb{E}_x[f(x)] - \mathbb{E}_z[f(g(z))] \tag{4}$$

Wasserstein Generative Adversarial Network (WGAN)

Wasserstein Generative Adversarial Network (WGAN): Fit the model distribution p_g by minimizing its 1-Wasserstein distance to the data distribution p_x **in the dual-form**:

$$\mathcal{W}_1(p_x, p_g) = \inf_{\pi \in \Pi(p_x, p_g)} \mathbb{E}_{(x, g(z)) \sim \pi}[\|x - g(z)\|_1] = \sup_{f \in L_1} \mathbb{E}_x[f(x)] - \mathbb{E}_z[f(g(z))] \quad (4)$$

Therefore, we have

$$\inf_g \mathcal{W}_1(p_x, p_g) \iff \inf_g \sup_{f \in L_1} \mathbb{E}_x[f(x)] - \mathbb{E}_z[f(g(z))] \quad (5)$$

Wasserstein Generative Adversarial Network (WGAN)

Wasserstein Generative Adversarial Network (WGAN): Fit the model distribution p_g by minimizing its 1-Wasserstein distance to the data distribution p_x **in the dual-form**:

$$\mathcal{W}_1(p_x, p_g) = \inf_{\pi \in \Pi(p_x, p_g)} \mathbb{E}_{(x, g(z)) \sim \pi}[\|x - g(z)\|_1] = \sup_{f \in L_1} \mathbb{E}_x[f(x)] - \mathbb{E}_z[f(g(z))] \tag{4}$$

Therefore, we have

$$\inf_g \mathcal{W}_1(p_x, p_g) \Longleftrightarrow \inf_g \sup_{f \in L_1} \mathbb{E}_x[f(x)] - \mathbb{E}_z[f(g(z))] \tag{5}$$

Given a set of samples $X = \{x_n\}_{n=1}^N$ and a set of latent code $Z = \{z_n\}_{n=1}^N$, we have

$$\min_g \max_{f \in L_1} \sum_n [f(x_n)] - \sum_n [f(g(z_n))] \tag{6}$$

Wasserstein Autoencoder (WAE)

Wasserstein autoencoder (WAE): Fit the model distribution p_g by minimizing its W_2 distance to the data distribution p_x approximately.

$$\inf_g \mathcal{W}_2(p_x, p_g)$$

Wasserstein Autoencoder (WAE)

Wasserstein autoencoder (WAE): Fit the model distribution p_g by minimizing its W_2 distance to the data distribution p_x approximately.

$$\inf_{g} \mathcal{W}_2(p_x, p_g) \approx \inf_{g,f} \underbrace{\mathbb{E}_{p_x}\mathbb{E}_{q_{z|x;f}}[d_x(x, g(z))]}_{\text{reconstruction loss}} + \underbrace{\gamma d(\overbrace{\mathbb{E}_{p_x}[q_{z|x;f}]}^{q_{z;f}}, p_z)}_{\text{distance(posterior, prior)}}, \tag{7}$$

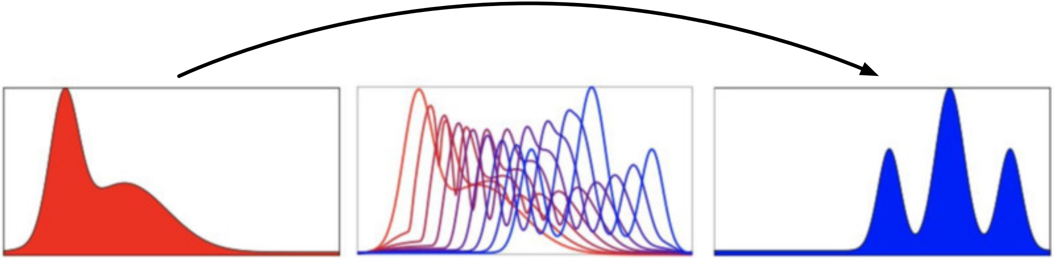
- $q_{z|x;f}$ is the posterior of z given x, parameterized by an **encoder** $f : \mathcal{X} \mapsto \mathcal{Z}$.
- $q_{z;f} = \mathbb{E}_{p_x}[q_{z|x;f}]$ is the expectation of the posterior distributions.
- p_z is the prior of z.
- d : MMD, OT distances, even GAN

The above methods are based on the static definition of OT (i.e., Kantorovich-form OT). The dynamic-form OT triggers more recent generative modeling methods — flow matching.

The Dynamic Definition of OT

The displacement interpolation determined by transport map T:

$$\rho_1 = T_{\#}\rho_0$$

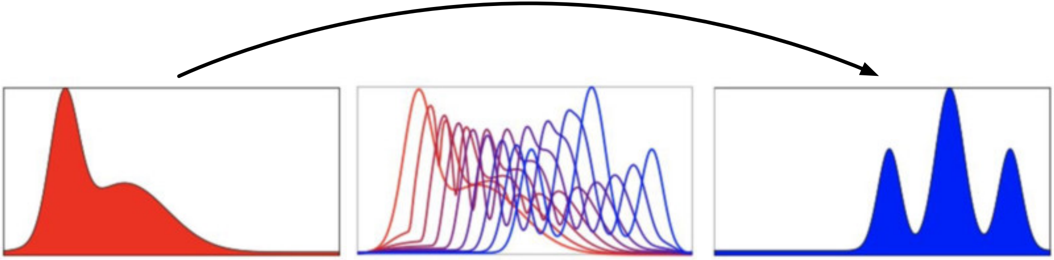


ρ_0 $\qquad$ $\rho_t = (tT + (1-t)Id)_{\#}\rho_0$ $\qquad$ ρ_1

The Dynamic Definition of OT

The displacement interpolation determined by transport map T:

$$\rho_1 = T_{\#}\rho_0$$



ρ_0 $\quad\quad$ $\rho_t = (tT + (1-t)Id)_{\#}\rho_0$ $\quad\quad$ ρ_1

What is the relationship between optimal transport and displacement interpolation?

The Dynamic Definition of OT

Definition 1 (Dynamic Formulation of Optimal Transport)

Let $\mathcal{X} \subset \mathbb{R}^d$ be the Euclidean sample space. For $\rho_0, \rho_1 \in \mathbb{P}(\mathcal{X})$, $\mathcal{W}_2^2(\rho_0, \rho_1)$ corresponds to seeking a unique least–kinetic–energy **flow (velocity field)** v:

$$\mathcal{W}_2^2(\rho_0, \rho_1) = \inf_{v(x,t)} \underbrace{\int_0^1 \int_{\mathcal{X}} \frac{1}{2}\rho(x,t)\|v(x,t)\|_2^2 \mathrm{d}x\mathrm{d}t}_{\textbf{Kinetic Energy}}, \quad s.t. \ \underbrace{\partial_t \rho + \nabla_x \cdot (v\rho) = 0}_{\textbf{Continuity Equation}} \tag{8}$$

A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. Numerische Mathematik, 2000.

The Dynamic Definition of OT

Definition 1 (Dynamic Formulation of Optimal Transport)

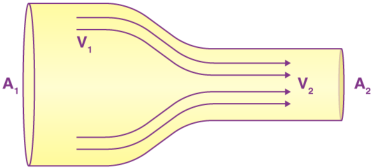
Let $\mathcal{X} \subset \mathbb{R}^d$ be the Euclidean sample space. For $\rho_0, \rho_1 \in \mathbb{P}(\mathcal{X})$, $\mathcal{W}_2^2(\rho_0, \rho_1)$ corresponds to seeking a unique least–kinetic–energy **flow (velocity field)** v:

$$\mathcal{W}_2^2(\rho_0, \rho_1) = \inf_{v(x,t)} \underbrace{\int_0^1 \int_{\mathcal{X}} \frac{1}{2}\rho(x,t)\|v(x,t)\|_2^2 \mathrm{d}x\mathrm{d}t}_{\textbf{Kinetic Energy}}, \quad s.t. \quad \underbrace{\partial_t \rho + \nabla_x \cdot (v\rho) = 0}_{\textbf{Continuity Equation}} \tag{8}$$

A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. Numerische Mathematik, 2000.

- Solving the continuity equation with the **optimal flow** v^* leads to the **optimal displacement interpolation** between ρ_0 and ρ_1.

Continuity Equation

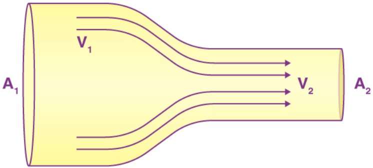


- Continuity equation describes the time rate of change of the fluid density ($\partial_t \rho(x,t)$) at a fixed point x in space.

$$\partial_t \rho + \nabla_x \cdot (v\rho) = 0 \qquad (9)$$

- The rate equals to the rate of change of density by convection ($\nabla_x \cdot (v\rho)$).

Continuity Equation



- Continuity equation describes the time rate of change of the fluid density ($\partial_t \rho(x,t)$) at a fixed point x in space.

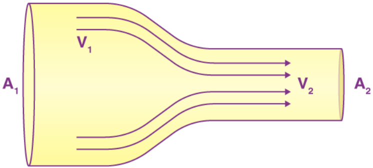
$$\partial_t \rho + \nabla_x \cdot (v\rho) = 0 \tag{9}$$

- The rate equals to the rate of change of density by convection ($\nabla_x \cdot (v\rho)$).

Given a sample at time t, i.e., $x_t \sim \rho_t$, we have

$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = v(x_t, t), \quad \underbrace{x_{t+\delta t} \approx x_t + \delta t \cdot v(x_t, t)}_{\text{Euler step}}. \tag{10}$$

Continuity Equation



- Continuity equation describes the time rate of change of the fluid density ($\partial_t \rho(x,t)$) at a fixed point x in space.

$$\partial_t \rho + \nabla_x \cdot (v\rho) = 0 \tag{9}$$

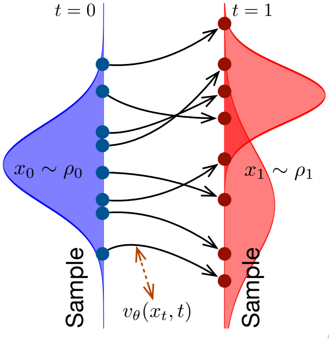
- The rate equals to the rate of change of density by convection ($\nabla_x \cdot (v\rho)$).

Given a sample at time t, i.e., $x_t \sim \rho_t$, we have

$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = v(x_t, t), \quad \underbrace{x_{t+\delta t} \approx x_t + \delta t \cdot v(x_t, t)}_{\text{Euler step}}. \tag{10}$$

Modeling the flow v leads to a new generative model strategy: Flow Matching.

Flow Matching (FM)

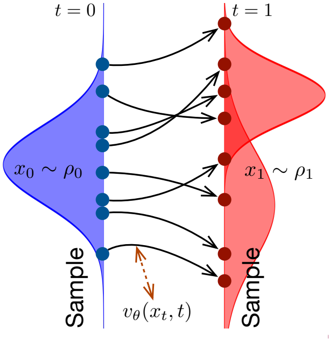


Flow Matching for Generative Modeling. ICLR, 2023.

Improving and generalizing flow-based generative models with minibatch optimal transport. TMLR, 2024.

Flow Matching (FM) (Sample Space): Learn a velocity field $v_\theta(x, t)$ capturing the transport of probability mass from a prior ρ_0 to a data ρ_1.

Flow Matching (FM)



Flow Matching for Generative Modeling. ICLR, 2023.

Improving and generalizing flow-based generative models with minibatch optimal transport. TMLR, 2024.

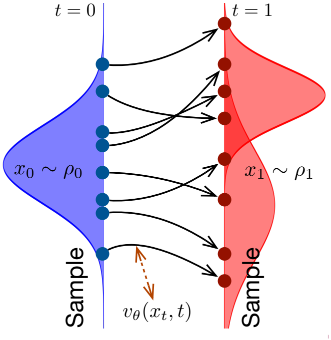
Flow Matching (FM) (Sample Space): Learn a velocity field $v_\theta(x, t)$ capturing the transport of probability mass from a prior ρ_0 to a data ρ_1.

- **Conditional FM (CFM)**: Set $\rho_0 = \mathcal{N}(0, 1)$, with an auxiliary variable $z \sim \pi$:

$$\min_\theta \ \mathbb{E}_{z\sim\pi,\ t,\ x}[\|v_\theta(x,t) - v_t(x|z)\|^2], \qquad (11)$$

  Generate new data by $\hat{x}_1 = x_0 + \int_0^1 v_\theta(x_t, t)\mathrm{d}t$. In practice, $x_{t+\Delta t} = x_t + \Delta t \cdot v_\theta(x_t, t)$.

Flow Matching (FM)



Flow Matching for Generative Modeling. ICLR, 2023.

Improving and generalizing flow-based generative models with minibatch optimal transport. TMLR, 2024.

Flow Matching (FM) (Sample Space): Learn a velocity field $v_\theta(x, t)$ capturing the transport of probability mass from a prior ρ_0 to a data ρ_1.

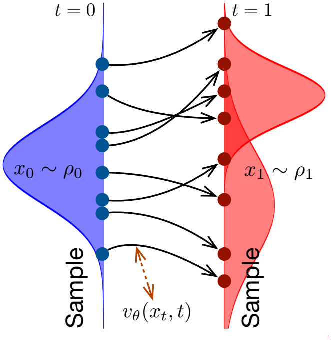
- **Conditional FM (CFM)**: Set $\rho_0 = \mathcal{N}(0, 1)$, with an auxiliary variable $z \sim \pi$:

$$\min_\theta \; \mathbb{E}_{z\sim\pi,\ t,\ x}[\|v_\theta(x,t) - v_t(x|z)\|^2], \tag{11}$$

  Generate new data by $\hat{x}_1 = x_0 + \int_0^1 v_\theta(x_t, t)\mathrm{d}t$. In practice, $x_{t+\Delta t} = x_t + \Delta t \cdot v_\theta(x_t, t)$.

- **FM (Lipman et al.)**: $p_t(x|z) = \mathcal{N}\big(tz, (t\sigma - t + 1)^2\big)$, $\pi = \rho_1$

Flow Matching (FM)



Flow Matching for Generative Modeling. ICLR, 2023.

Improving and generalizing flow-based generative models with minibatch optimal transport. TMLR, 2024.

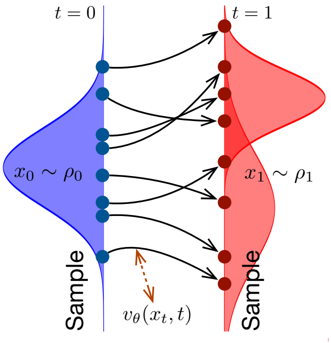
Flow Matching (FM) (Sample Space): Learn a velocity field $v_\theta(x, t)$ capturing the transport of probability mass from a prior ρ_0 to a data ρ_1.

- **Conditional FM (CFM)**: Set $\rho_0 = \mathcal{N}(0, 1)$, with an auxiliary variable $z \sim \pi$:

$$\min_\theta \; \mathbb{E}_{z\sim\pi,\ t,\ x}[\|v_\theta(x, t) - v_t(x|z)\|^2], \qquad (11)$$

  Generate new data by $\hat{x}_1 = x_0 + \int_0^1 v_\theta(x_t, t)\mathrm{d}t$. In practice, $x_{t+\Delta t} = x_t + \Delta t \cdot v_\theta(x_t, t)$.
- **FM (Lipman et al.)**: $p_t(x|z) = \mathcal{N}\big(tz, (t\sigma - t + 1)^2\big)$, $\pi = \rho_1$
- **I-CFM**: $x_t = (1 - t) \cdot x_0 + t \cdot x_1$, $\pi = \rho_0 \times \rho_1$

Flow Matching (FM)



Flow Matching for Generative Modeling. ICLR, 2023.

Improving and generalizing flow-based generative models with minibatch optimal transport. TMLR, 2024.

Flow Matching (FM) (Sample Space): Learn a velocity field $v_\theta(x,t)$ capturing the transport of probability mass from a prior ρ_0 to a data ρ_1.

- **Conditional FM (CFM)**: Set $\rho_0 = \mathcal{N}(0,1)$, with an auxiliary variable $z \sim \pi$:

$$\min_\theta \ \mathbb{E}_{z\sim\pi,\ t,\ x}[\|v_\theta(x,t) - v_t(x|z)\|^2], \qquad (11)$$

  Generate new data by $\hat{x}_1 = x_0 + \int_0^1 v_\theta(x_t,t)\mathrm{d}t$. In practice, $x_{t+\Delta t} = x_t + \Delta t \cdot v_\theta(x_t,t)$.
- **FM (Lipman et al.)**: $p_t(x|z) = \mathcal{N}\left(tz, (t\sigma - t + 1)^2\right),\ \pi = \rho_1$
- **I-CFM**: $x_t = (1-t)\cdot x_0 + t\cdot x_1,\ \pi = \rho_0 \times \rho_1$
- **OT-CFM**: **Optimal Transport (OT)** perspective...

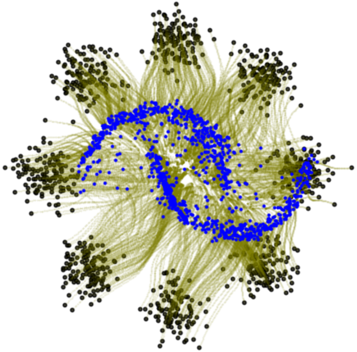
OT-CFM: Optimal Transport Perspective of FM

- **OT-CFM**: implements CFM by setting the distribution π in (11) as the OT plan corresponding to $\mathcal{W}_2^2(\rho_0, \rho_1)$ and $x_t = (1-t)\cdot x_0 + t\cdot x_1$.

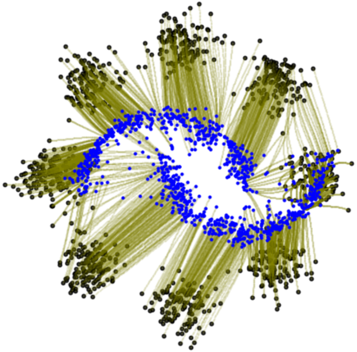
$$\min_\theta \overbrace{\mathbb{E}_{(x_0,x_1)\sim\pi^*,\ t\sim\text{Unif}[0,1]}[\|v_\theta(x_t,t) - (x_1 - x_0)\|^2]}^{\textbf{Upper-level}:\ \mathcal{L}_{\text{CFM}}},$$
$$\text{s.t. } \pi^* = \overbrace{\arg\min_{\pi\in\Pi(\rho_0,\rho_1)} \mathbb{E}_\pi[\|x_1 - x_0\|_2^2]}^{\textbf{Lower-level}:\ \mathcal{W}_2^2(\rho_0,\rho_1)}, \tag{12}$$

This is a **Bi-level Optimization Problem**.

OT-CFM: Optimal Transport Perspective of FM



(a) I-CFM, $\pi = \rho_0 \times \rho_1$



(b) OT-CFM, $\pi = \pi^*$

The objective of OT-CFM regresses $v_\theta(x_t, t)$ to the **constant velocity** $(x_1 - x_0)$, leading to the interpolation between ρ_0 and ρ_1 yielding OT, i.e., $\mathcal{W}_2(\rho_0, \rho_1)$.

OT-CFM: Optimal Transport Perspective of FM

Notably, constant velocity is sufficient but not necessary for straightening flows, which might introduce too strong inductive bias to the generative model.

Proposition 2 (Straightness Criterion)

The trajectory is straight if and only if the velocity direction is time invariant and the ***acceleration*** *is everywhere parallel to the velocity. The classical (first-order) dynamical optimal transport is recovered as the special case with zero acceleration.*

OT-CFM: Optimal Transport Perspective of FM

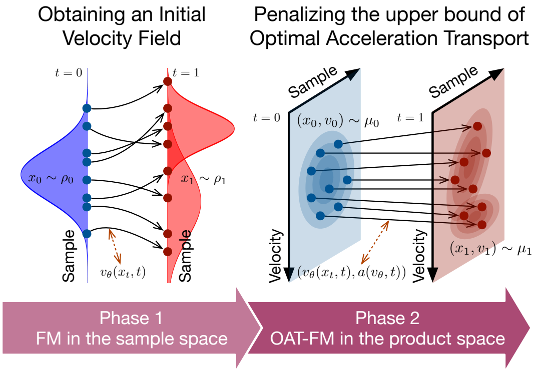
Notably, constant velocity is sufficient but not necessary for straightening flows, which might introduce too strong inductive bias to the generative model.

Proposition 2 (Straightness Criterion)

The trajectory is straight if and only if the velocity direction is time invariant and the ***acceleration*** *is everywhere parallel to the velocity. The classical (first-order) dynamical optimal transport is recovered as the special case with zero acceleration.*

How about pursue a flow with minimal acceleration instead of constant velocity?

OAT-FM: Two-Phase FM Based on Optimal Acceleration Transport



3) OAT-FM (Sample × Velocity Space): A novel two-phase FM based on **Optimal Acceleration Transport (OAT)**

- Refine an arbitrary pre-trained flow/diffusion generator
- Minimize the acceleration transport between μ_0 and μ_1

Optimal Acceleration Transport (OAT)

Key Idea: For μ_0 and μ_1 in the **product space** $(\mathcal{X} \times \mathcal{V})$, find a flow that minimizes total squared acceleration under second-order dynamics.

Optimal Acceleration Transport (OAT)

Key Idea: For μ_0 and μ_1 in the **product space** $(\mathcal{X} \times \mathcal{V})$, find a flow that minimizes total squared acceleration under second-order dynamics.

Definition 3 (Dynamic Formulation of Optimal Acceleration Transport [1])

Let $\mathcal{X} \subset \mathbb{R}^d$ be the sample space and $\mathcal{V} \subset \mathbb{R}^d$ the velocity space (by default $\mathcal{V} = \mathbb{R}^d$). For $\mu_0, \mu_1 \in \mathbb{P}(\mathcal{X} \times \mathcal{V})$, the optimal acceleration transport between them is defined as

$$\mathcal{A}_2^2(\mu_0, \mu_1) := \min_{\mu,\, a} \int_0^1 \int_{\mathcal{X}\times\mathcal{V}} \frac{1}{2}\, \mu(x, v, t)\, \|a(x, v, t)\|_2^2 \,\mathrm{d}x\,\mathrm{d}v\,\mathrm{d}t, \tag{13}$$

subject to the **Vlasov equation** $\partial_t \mu + v \cdot \nabla_x \mu + \nabla_v \cdot (a\,\mu) = 0$, with boundary conditions $\mu(\cdot, \cdot, 0) = \mu_0$ and $\mu(\cdot, \cdot, 1) = \mu_1$. Here, $a : \mathcal{X} \times \mathcal{V} \times [0, 1] \mapsto \mathbb{R}^d$ is the acceleration field, and the Vlasov equation expresses conservation of mass in the product space.

[1] Benamou et al., 2019. Second-order models for optimal transport and cubic splines on the Wasserstein space

Optimal Acceleration Transport (OAT)

Definition 4 (Kantorovich formulation of OAT [2,3,4])

Given $z_0 = (x_0, v_0) \sim \mu_0$ and $z_1 = (x_1, v_1) \sim \mu_1$, the OAT problem is equivalent to solving an optimal coupling w.r.t. squared acceleration cost, i.e.,

$$\begin{aligned}
&\mathcal{A}_2^2(\mu_0,\, \mu_1) \\
&= \min_{\pi \in \Pi(\mu_0,\mu_1)} \mathbb{E}_{(z_0,z_1)\sim\pi}[c_{\mathsf{A}}^2(z_0, z_1)] \\
&= \min_{\pi \in \Pi(\mu_0,\mu_1)} \mathbb{E}_{(z_0,z_1)\sim\pi}\Bigg[12 \underbrace{\left\| \frac{x_1 - x_0}{T} - \frac{v_1 + v_0}{2} \right\|^2}_{\text{velocity alignment}} + \underbrace{\|v_1 - v_0\|^2}_{\text{acceleration penalty}} \Bigg],
\end{aligned} \tag{14}$$

where $T > 0$ denotes the time horizon between μ_0 and μ_1, which is 1 in our case.

---

[2] Chen et al., 2018. Measure-valued spline curves: An optimal transport viewpoint

[3] Benamou et al., 2019. Second-order models for optimal transport and cubic splines on the Wasserstein space

[4] Brigati et al., 2025. Kinetic Optimal Transport (OTIKIN) – Part 1: Second-Order Discrepancies Between Probability Measures

OAT-FM: Refine Pre-trained v_θ Using OAT

Problem Setup:

- Trajectory endpoints: $z_0 = (x_0, v_\theta(x_0, 0))$ and $z_1 = (x_1, v_\theta(x_1, 1))$.
- Path x_t: Linear interpolation $x_t = (1-t)x_0 + tx_1$.
- Model state: $z_t(\theta) = (x_t, v_\theta(x_t, t))$.

OAT-FM: Refine Pre-trained v_θ Using OAT

Problem Setup:

- Trajectory endpoints: $z_0 = (x_0, v_\theta(x_0, 0))$ and $z_1 = (x_1, v_\theta(x_1, 1))$.
- Path x_t: Linear interpolation $x_t = (1-t)x_0 + tx_1$.
- Model state: $z_t(\theta) = (x_t, v_\theta(x_t, t))$.

Cost Function $\ell_{\mathcal{A}}$:

$$\begin{aligned}\ell_{\mathcal{A}}(z_0, z_1, t; \theta) = \alpha \underbrace{\left\| \frac{x_t - x_0}{t} - \frac{v_0 + v_\theta}{2} \right\|_2^2}_{\text{Velocity Alignment } (0 \to t)} + (1-\alpha) \underbrace{\|v_\theta - v_0\|_2^2}_{\text{Accel. Penalty } (0 \to t)} \\ + \alpha \underbrace{\left\| \frac{x_1 - x_t}{1-t} - \frac{v_\theta + v_1}{2} \right\|_2^2}_{\text{Velocity Alignment } (t \to 1)} + (1-\alpha) \underbrace{\|v_1 - v_\theta\|_2^2}_{\text{Accel. Penalty } (t \to 1)}\end{aligned} \tag{15}$$

Imitate the cost in OAT:

- Hyperparameter α balances *velocity alignment* vs. *acceleration minimization*.
- With $\alpha = \frac{12}{13}$, recovers OAT cost structure: $\ell_{\mathcal{A}} = \frac{1}{13}(c^2_{\mathcal{A}}(z_0, z_t) + c^2_{\mathcal{A}}(z_t, z_1))$.

OAT-FM: Refine Pre-trained v_θ Using OAT

OAT-FM Problem: We fine-tune the flow model by solving the following **Bi-level Optimization Problem**:

$$\min_{\theta} \overbrace{\mathbb{E}_{(z_0,z_1)\sim\pi^*,\ t\sim\text{Unif}[0,1]}[\ell_{\mathcal{A}}(z_0, z_1, t;\ \theta)]}^{\textbf{Upper-level}:\ \mathcal{L}_{\text{OAT}}(\mu_0,\mu_1;\alpha)},$$
$$\text{s.t.}\quad \pi^* = \overbrace{\arg\min_{\pi\in\Pi(\mu_0,\mu_1)} \mathbb{E}_{(z_0,z_1)\sim\pi}[c_{\mathcal{A}}^2(z_0,\ z_1)]}^{\textbf{Lower-level}:\ \mathcal{A}_2^2(\mu_0,\mu_1)}. \tag{16}$$

- **Lower-level**: Finds the optimal coupling π^* that minimizes total acceleration in the product space.
- **Upper-level**: Aligns the learned flow with the OAT geodesics via $\ell_{\mathcal{A}}$.
- **Parameter** α: Balances *directional alignment* and *acceleration minimization*.

OAT-FM vs. OT-CFM

Component	**OT-CFM**	**OAT-FM (Proposed)**
Space	Sample Space $\mathcal{X}$	Product Space $\mathcal{X} \times \mathcal{V}$

OAT-FM vs. OT-CFM

Component	**OT-CFM**	**OAT-FM (Proposed)**
Space	Sample Space $\mathcal{X}$	Product Space $\mathcal{X} \times \mathcal{V}$
Dynamics	Continuity Equation $\partial_t \rho + \nabla_x \cdot (v\rho) = 0$	Vlasov Equation $\partial_t \mu + \nabla_x \cdot (v\mu) + \nabla_v \cdot (a\mu) = 0$

OAT-FM vs. OT-CFM

Component	**OT-CFM**	**OAT-FM (Proposed)**
Space	Sample Space $\mathcal{X}$	Product Space $\mathcal{X} \times \mathcal{V}$
Dynamics	Continuity Equation	Vlasov Equation
	$\partial_t \rho + \nabla_x \cdot (v\rho) = 0$	$\partial_t \mu + \nabla_x \cdot (v\mu) + \nabla_v \cdot (a\mu) = 0$
Lower-level	Optimal Transport (OT)	Optimal Acceleration Transport (OAT)
(Coupling)	$\pi^* = \arg\min \mathbb{E}[\|x_1 - x_0\|^2]$	$\pi^* = \arg\min \mathbb{E}[c_{\mathcal{A}}^2(z_0, z_1)]$

OAT-FM vs. OT-CFM

Component	OT-CFM	OAT-FM (Proposed)
Space	Sample Space $\mathcal{X}$	Product Space $\mathcal{X} \times \mathcal{V}$
Dynamics	Continuity Equation	Vlasov Equation
	$\partial_t \rho + \nabla_x \cdot (v\rho) = 0$	$\partial_t \mu + \nabla_x \cdot (v\mu) + \nabla_v \cdot (a\mu) = 0$
Lower-level	Optimal Transport (OT)	Optimal Acceleration Transport (OAT)
(Coupling)	$\pi^* = \arg\min \mathbb{E}[\|x_1 - x_0\|^2]$	$\pi^* = \arg\min \mathbb{E}[c_{\mathcal{A}}^2(z_0, z_1)]$
Upper-level	Velocity Matching	Velocity Alignment + Acceleration Minimization
(Objective)	$\min \|v_\theta - (x_1 - x_0)\|^2$	$\min \ell_{\mathcal{A}}(z_0, z_1, t; \theta)$

OAT-FM vs. OT-CFM

Component	**OT-CFM**	**OAT-FM (Proposed)**
Space	Sample Space $\mathcal{X}$	Product Space $\mathcal{X} \times \mathcal{V}$
Dynamics	Continuity Equation $\partial_t \rho + \nabla_x \cdot (v\rho) = 0$	Vlasov Equation $\partial_t \mu + \nabla_x \cdot (v\mu) + \nabla_v \cdot (a\mu) = 0$
Lower-level (Coupling)	Optimal Transport (OT) $\pi^* = \arg\min \mathbb{E}[\|x_1 - x_0\|^2]$	Optimal Acceleration Transport (OAT) $\pi^* = \arg\min \mathbb{E}[c^2_{\mathcal{A}}(z_0, z_1)]$
Upper-level (Objective)	Velocity Matching $\min \|v_\theta - (x_1 - x_0)\|^2$	Velocity Alignment + Acceleration Minimization $\min \ell_{\mathcal{A}}(z_0, z_1, t; \theta)$
Mechanism (Straightening)	**Constant Velocity** $\min \int_0^1 \|v_t\|^2 \mathrm{d}t \implies \ddot{x} = 0$	**Minimal Acceleration (Smooth Velocity)** $\min \int_0^1 \|a_t\|^2 \mathrm{d}t \implies \ddot{v} = 0$

Theoretical Guarantees of OAT-FM

Theorem 5 (OAT Bound of OAT-FM)

The OAT-FM objective $\mathcal{L}_{OAT}(\mu_0, \mu_1; \alpha)$ is lower-bounded by a scaled version of the true OAT second-order discrepancy, i.e.,

$$\mathcal{L}_{OAT}(\mu_0, \mu_1; \alpha) \geq \frac{2}{27}\mathcal{A}_2^2(\mu_0, \mu_1), \tag{17}$$

with $\alpha = 2/3$, and the equality held if and only if $v_1 = v_0$ for π^*-almost every pair.

Theoretical Guarantees of OAT-FM

Theorem 5 (OAT Bound of OAT-FM)

The OAT-FM objective $\mathcal{L}_{OAT}(\mu_0, \mu_1; \alpha)$ *is lower-bounded by a scaled version of the true OAT second-order discrepancy, i.e.,*

$$\mathcal{L}_{OAT}(\mu_0, \mu_1; \alpha) \geq \frac{2}{27}\mathcal{A}_2^2(\mu_0, \mu_1), \tag{17}$$

with $\alpha = 2/3$*, and the equality held if and only if* $v_1 = v_0$ *for* π^**-almost every pair.*

Theorem 6 (Straightening Flow via OAT)

Given two boundary distributions $\mu_0, \mu_1 \in \mathbb{P}(\mathcal{X} \times \mathcal{V})$*, OAT admits an optimal coupling* $\pi^* \in \Pi(\mu_0, \mu_1)$ *for the static problem in* (14)*. For every* $(x_0, v_0), (x_1, v_1) \sim \pi^*$*, the corresponding trajectory is straight iff* v_0 *and* v_1 *are collinear with* $x_1 - x_0$*. Otherwise, it bends exactly to match the endpoints' orthogonal components.*

Efficient Implementation via Decomposable Structure

$$\min_{\theta} \overbrace{\mathbb{E}_{(z_0,z_1)\sim\pi^*,\ t\sim\mathsf{Unif}[0,1]}\left[\ell_{\mathcal{A}}(z_0, z_1, t;\ \theta)\right]}^{\textbf{Upper-level}:\ \mathcal{L}_{\mathsf{OAT}}(\mu_0,\mu_1;\alpha)},$$
$$\text{s.t.}\quad \pi^* = \overbrace{\arg\min_{\pi\in\Pi(\mu_0,\mu_1)} \mathbb{E}_{(z_0,z_1)\sim\pi}\left[c^2_{\mathcal{A}}(z_0,\ z_1)\right]}^{\textbf{Lower-level}:\ \mathcal{A}_2^2(\mu_0,\mu_1)}. \tag{18}$$

The Challenge: Solving OAT requires the coupling π as a 4D tensor.

Efficient Implementation via Decomposable Structure

$$
\begin{aligned}
&\min_{\theta} \overbrace{\mathbb{E}_{(z_0,z_1)\sim\pi^*,\ t\sim\text{Unif}[0,1]}\big[\ell_{\mathcal{A}}(z_0, z_1, t;\ \theta)\big]}^{\textbf{Upper-level}:\ \mathcal{L}_{\text{OAT}}(\mu_0,\mu_1;\alpha)}, \\
&\text{s.t.}\ \ \pi^* = \overbrace{\arg\min_{\pi\in\Pi(\mu_0,\mu_1)} \mathbb{E}_{(z_0,z_1)\sim\pi}\big[c^2_{\mathcal{A}}(z_0,\ z_1)\big]}^{\textbf{Lower-level}:\ \mathcal{A}_2^2(\mu_0,\mu_1)}.
\end{aligned}
\tag{18}
$$

The Challenge: Solving OAT requires the coupling π as a 4D tensor.

The Simplification (Decomposition): In FM, velocities are deterministic given samples: $v = v_\theta(x, t)$. This implies a *decomposable structure* for the coupling:

$$
\pi(z_0, z_1) = \underbrace{\pi_x(x_0, x_1)}_{\text{Sample Coupling}} \cdot \underbrace{\delta_{v_\theta(x_0,0)}(v_0) \cdot \delta_{v_\theta(x_1,1)}(v_1)}_{\text{Deterministic Velocity Assignment}} . \tag{19}
$$

Efficient Implementation via Decomposable Structure

The Resulting Lower-Level Problem: We reduce the OAT problem to a classic OT problem on samples:

$$\arg\min_{\pi_x \in \Pi(\rho_0,\rho_1)} \mathbb{E}_{(x_0,x_1)\sim\pi_x}\left[12\|x_1 - x_0 - \bar{v}_{x_0,x_1}\|^2 + \|\tilde{v}_{x_0,x_1}\|_2^2\right], \quad (20)$$

where ρ_0, ρ_1 are marginals on $\mathcal{X}$, and velocities are fixed by the current model:

- $\bar{v}_{x_0,x_1} = \frac{1}{2}(v_\theta(x_0, 0) + v_\theta(x_1, 1))$ (Mean Velocity)
- $\tilde{v}_{x_0,x_1} = v_\theta(x_1, 1) - v_\theta(x_0, 0)$ (Velocity Difference)

Efficient Implementation via Decomposable Structure

The Resulting Lower-Level Problem: We reduce the OAT problem to a classic OT problem on samples:

$$\arg\min_{\pi_x \in \Pi(\rho_0, \rho_1)} \mathbb{E}_{(x_0,x_1)\sim\pi_x}\left[12\|x_1 - x_0 - \bar{v}_{x_0,x_1}\|^2 + \|\tilde{v}_{x_0,x_1}\|_2^2\right], \tag{20}$$

where ρ_0, ρ_1 are marginals on $\mathcal{X}$, and velocities are fixed by the current model:

- $\bar{v}_{x_0,x_1} = \frac{1}{2}(v_\theta(x_0, 0) + v_\theta(x_1, 1))$ (Mean Velocity)
- $\tilde{v}_{x_0,x_1} = v_\theta(x_1, 1) - v_\theta(x_0, 0)$ (Velocity Difference)

Computational Complexity Analysis:

- **Exact OT (Linear Program):** $\Omega(B^3)$.
- **Sinkhorn Algorithm (Approximation):** $\Omega(B^2)$.
  - Solved efficiently via iterative matrix scaling (highly parallelizable).
  - Recovers exact OT solution when $\epsilon \to 0$.

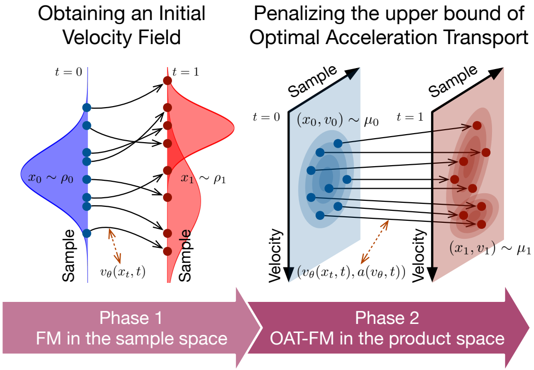
Algorithm Scheme: OAT-FM Training Loop

Algorithm 1 OAT-FM Training Procedure

Require: Pre-trained model v_{θ_0}, Dataset $\mathcal{D}$, Batch size B, EMA rate λ.
Ensure: Refined velocity field v_θ.

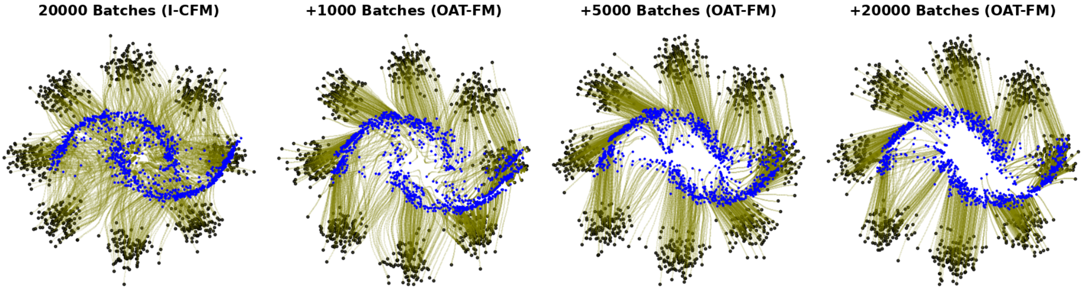
Initialize $v_\theta \leftarrow v_{\theta_0}$.
while training **do**
    // Step 1: Data Preparation
    Sample batch $\{x_{1,i}\}_{i=1}^{B} \sim \mathcal{D}$, $\{x_{0,i}\}_{i=1}^{B} \sim \mathcal{N}(0, I)$, $t \sim \mathcal{U}[0, 1]$.
    Estimate boundary velocities using current model:
    $\{v_{0,i} \leftarrow v_\theta(x_{0,i}, 0)\}_{i=1}^{B}, \quad \{v_{1,i} \leftarrow v_\theta(x_{1,i}, 1)\}_{i=1}^{B}.$
    // Step 2: Lower-Level (Coupling)
    Compute optimal coupling $\mathbf{T}^*$ by solving the reduced classic OT.
    Sample pairs $(x_1, x_0) \sim \mathbf{T}^*$ to get aligned batches.
    // Step 3: Upper-Level (Optimization)
    Interpolate $x_t \leftarrow (1-t)x_0 + tx_1$, predict $v_t \leftarrow v_\theta(x_t, t)$.
    Compute $\mathcal{L}_{\text{OAT}}$ and update: $\theta' \leftarrow \theta - \nabla_\theta \mathcal{L}_{\text{OAT}}$.
    Update EMA: $\theta \leftarrow \text{stopgrad}(\lambda\theta + (1-\lambda)\theta')$.
end while

Compare to Other 2-Phase FM Methods



- **ReFlow, Consistency Distillation:** Pursue straight flows and reduce sampling steps, but suffer from distribution drift inevitably.
- **OAT-FM:** Pursue smooth flows, may not reduce sampling steps but without distribution drift.

Application 1: Low-dimensional OT Benchmark



Experimental Setup:

- **Tasks:** 5 standard 2D distribution mapping tasks (e.g., 8gaussians $\rightarrow$ moons).
- **Evaluation Metric:** 2-Wasserstein distance and Normalized Path Energy (NPE):

$$\text{NPE}(v_\theta) = \frac{|\text{PE}(v_\theta) - \mathcal{W}_2^2(\rho_0, \rho_1)|}{\mathcal{W}_2^2(\rho_0, \rho_1)}, \quad \text{with PE}(v_\theta) = \mathbb{E}_{x_0} \int_0^1 \|v_\theta(x_t, t)\|^2 \mathrm{d}t. \quad (21)$$

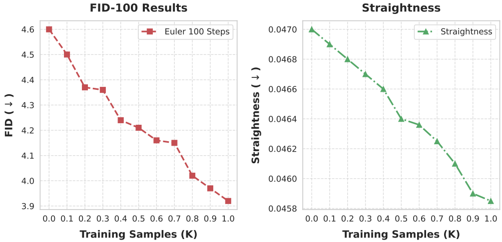
Application 1: Low-dimensional OT Benchmark

Task	$\mathcal{N} \to$8gs		8gs$\to$moons		$\mathcal{N} \to$moons		$\mathcal{N} \to$scurve		moons$\to$8gs	
Method	$\mathcal{W}_2^2 \downarrow$	NPE$\downarrow$	$\mathcal{W}_2^2 \downarrow$	NPE$\downarrow$	$\mathcal{W}_2^2 \downarrow$	NPE$\downarrow$	$\mathcal{W}_2^2 \downarrow$	NPE$\downarrow$	$\mathcal{W}_2^2 \downarrow$	NPE$\downarrow$
FM	$0.58_{\pm 0.16}$	$0.24_{\pm 0.01}$	$5.80_{\pm 0.06}$	$0.05_{\pm 0.02}$	$0.15_{\pm 0.07}$	$0.27_{\pm 0.05}$	$\mathbf{0.81}_{\pm 0.39}$	$0.08_{\pm 0.04}$	$7.39_{\pm 0.45}$	$0.96_{\pm 0.05}$
+OAT-FM	$\mathbf{0.31}_{\pm 0.09}$	$\mathbf{0.02}_{\pm 0.01}$	$\mathbf{0.08}_{\pm 0.03}$	$\mathbf{0.01}_{\pm 0.01}$	$\mathbf{0.08}_{\pm 0.03}$	$\mathbf{0.03}_{\pm 0.01}$	$0.90_{\pm 0.18}$	$\mathbf{0.03}_{\pm 0.02}$	$\mathbf{0.28}_{\pm 0.10}$	$\mathbf{0.04}_{\pm 0.02}$
I-CFM	$0.45_{\pm 0.18}$	$0.30_{\pm 0.01}$	$0.18_{\pm 0.05}$	$1.40_{\pm 0.05}$	$0.11_{\pm 0.03}$	$0.52_{\pm 0.06}$	$1.16_{\pm 0.47}$	$0.03_{\pm 0.03}$	$0.74_{\pm 0.12}$	$1.19_{\pm 0.06}$
+OAT-FM	$\mathbf{0.32}_{\pm 0.10}$	$\mathbf{0.04}_{\pm 0.01}$	$\mathbf{0.15}_{\pm 0.03}$	$\mathbf{0.13}_{\pm 0.01}$	$\mathbf{0.07}_{\pm 0.02}$	$\mathbf{0.04}_{\pm 0.04}$	$\mathbf{1.12}_{\pm 0.45}$	$\mathbf{0.03}_{\pm 0.02}$	$\mathbf{0.50}_{\pm 0.11}$	$\mathbf{0.44}_{\pm 0.03}$
VP-CFM	$0.43_{\pm 0.14}$	$0.24_{\pm 0.01}$	$0.15_{\pm 0.02}$	$1.24_{\pm 0.05}$	$0.10_{\pm 0.03}$	$0.31_{\pm 0.07}$	$\mathbf{1.05}_{\pm 0.41}$	$0.22_{\pm 0.04}$	$1.39_{\pm 0.35}$	$1.22_{\pm 0.05}$
+OAT-FM	$\mathbf{0.31}_{\pm 0.12}$	$\mathbf{0.03}_{\pm 0.01}$	$\mathbf{0.09}_{\pm 0.01}$	$\mathbf{0.02}_{\pm 0.01}$	$\mathbf{0.07}_{\pm 0.02}$	$\mathbf{0.04}_{\pm 0.01}$	$1.10_{\pm 0.34}$	$\mathbf{0.03}_{\pm 0.02}$	$\mathbf{0.32}_{\pm 0.10}$	$\mathbf{0.10}_{\pm 0.02}$
SB-CFM	$0.51_{\pm 0.10}$	$\mathbf{0.01}_{\pm 0.01}$	$0.13_{\pm 0.04}$	$0.03_{\pm 0.01}$	$\mathbf{0.08}_{\pm 0.03}$	$\mathbf{0.04}_{\pm 0.03}$	$\mathbf{0.79}_{\pm 0.29}$	$0.04_{\pm 0.02}$	$0.36_{\pm 0.14}$	$\mathbf{0.03}_{\pm 0.02}$
+OAT-FM	$\mathbf{0.34}_{\pm 0.08}$	$0.03_{\pm 0.01}$	$\mathbf{0.07}_{\pm 0.01}$	$\mathbf{0.01}_{\pm 0.01}$	$0.09_{\pm 0.04}$	$0.10_{\pm 0.04}$	$0.80_{\pm 0.18}$	$\mathbf{0.02}_{\pm 0.02}$	$\mathbf{0.25}_{\pm 0.08}$	$\mathbf{0.03}_{\pm 0.02}$
OT-CFM	$0.35_{\pm 0.09}$	$\mathbf{0.01}_{\pm 0.01}$	$\mathbf{0.07}_{\pm 0.02}$	$\mathbf{0.01}_{\pm 0.01}$	$0.07_{\pm 0.02}$	$\mathbf{0.04}_{\pm 0.02}$	$0.87_{\pm 0.33}$	$\mathbf{0.03}_{\pm 0.03}$	$0.31_{\pm 0.10}$	$\mathbf{0.02}_{\pm 0.02}$
+OAT-FM	$\mathbf{0.32}_{\pm 0.10}$	$0.04_{\pm 0.01}$	$\mathbf{0.07}_{\pm 0.01}$	$\mathbf{0.01}_{\pm 0.01}$	$\mathbf{0.06}_{\pm 0.01}$	$\mathbf{0.04}_{\pm 0.01}$	$\mathbf{0.83}_{\pm 0.34}$	$0.04_{\pm 0.02}$	$\mathbf{0.29}_{\pm 0.09}$	$0.10_{\pm 0.02}$

Application 2: Unconditional Image Generation (CIFAR-10)

Method	#Batch	NFE↓	FID↓
FM	400K	147	3.71
FM + OAT-FM	+1K	135	**3.54**
I-CFM	400K	149	3.67
I-CFM + OAT-FM	+1K	138	**3.48**
OT-CFM	400K	132	3.64
OT-CFM + OAT-FM	+1K	126	**3.46**
DDPM*		1K	3.17
Score SDE*		2K	2.38
LSGM*		147	2.10
2-ReFlow++*		35	2.30
EDM		35	1.96
EDM + OAT-FM	+12K	35	**1.93**

Lower-level Problem	Upper-level Problem	Phase-1 Method FM	EDM
Without Phase-2 Training		3.71	1.96
$\mathcal{W}_2^2$ in (12)	$\mathcal{L}_{\text{CFM}}$ in (12)	3.75	8.77
$\mathcal{W}_2^2$ in (12)	$\mathcal{L}_{\text{OAT}}$ in (20)	3.55	8.68
$\mathcal{A}_2^2$ in (20)	$\mathcal{L}_{\text{CFM}}$ in (12)	3.81	1.95
$\mathcal{A}_2^2$ in (20)	$\mathcal{L}_{\text{OAT}}$ in (20)	**3.54**	**1.93**



Application 3: Large-scale Conditional Image Generation

(a) SiT-XL (Left) v.s. + OAT-FM (Right)

(b) SiT-XL (Left) v.s. + OAT-FM (Right)

(c) SiT-XL (Left) v.s. + OAT-FM (Right)

(d) SiT-XL (Left) v.s. + OAT-FM (Right)

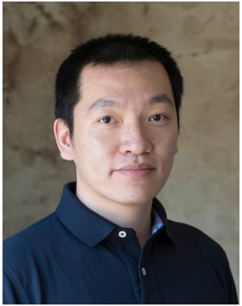
Application 3: Large-scale Conditional Image Generation

Method	#Epochs	FID↓	sFID↓	IS↑	P↑	R↑
BigGAN-deep		6.95	7.36	171.4	0.87	0.28
StyleGAN-XL		2.30	4.02	265.1	0.78	0.53
Mask-GIT		6.18	-	182.1	-	-
ADM-G/U		3.94	6.14	215.8	0.83	0.53
CDM		4.88	-	158.7	-	-
RIN		3.42	-	182.0	-	-
Simple Diffusion$_{\text{U-ViT, L}}$		2.77	-	211.8	-	-
VDM++		2.12	-	267.7	-	-
DiT-XL$_{\text{CFG}=1.5}$		2.27	4.60	278.2	0.83	0.57
SiT-XL$_{\text{CFG}=1.5,\ \text{Sampler}=\text{ODE}}$	1,400	2.11	**4.62**	256.0	**0.81**	**0.61**
SiT-XL$_{\text{CFG}=1.5,\ \text{Sampler}=\text{ODE}}$ + OAT-FM	+5	**2.05**	**4.62**	**259.4**	0.80	**0.61**
SiT-XL$_{\text{CFG}=2.5,\ \text{Sampler}=\text{ODE}}$	1,400	6.91	6.42	391.5	**0.89**	0.47
SiT-XL$_{\text{CFG}=2.5,\ \text{Sampler}=\text{ODE}}$ + OAT-FM	+5	**6.57**	**5.98**	**394.8**	**0.89**	**0.49**
SiT-XL$_{\text{CFG}=1.5,\ \text{Sampler}=\text{SDE}}$	1,400	2.05	4.50	269.6	**0.82**	**0.59**
SiT-XL$_{\text{CFG}=1.5,\ \text{Sampler}=\text{SDE}}$ + OAT-FM	+5	**2.00**	**4.43**	**275.1**	**0.82**	**0.59**
SiT-XL$_{\text{CFG}=2.5,\ \text{Sampler}=\text{SDE}}$	1,400	7.75	6.64	405.0	**0.90**	0.45
SiT-XL$_{\text{CFG}=2.5,\ \text{Sampler}=\text{SDE}}$ + OAT-FM	+5	**7.44**	**5.77**	**409.9**	**0.90**	**0.46**

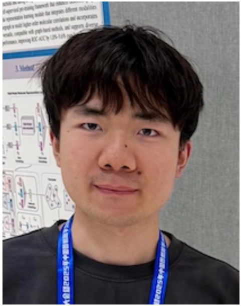
Summary

- OT-CFM shows the potential dynamic OT in generative modeling.
- Propose OAT-FM to straighten flow trajectories by minimizing acceleration in the joint sample-velocity space.
- Introduce an efficient two-phase fine-tuning paradigm that improves pre-trained models without distribution drift.
- Achieve superior generation quality on high-dimensional tasks (e.g., CIFAR-10, ImageNet) with minimal training overhead.

Acknowledgment



Hongteng Xu
RUC



Angxiao Yue
KTH

Anqi Dong
RUC

- *Paper:* `https://arxiv.org/pdf/2509.24936`
- *Code:* `https://github.com/AngxiaoYue/OAT-FM`

Thank you!

https://hongtengxu.github.io

https://github.com/HongtengXu

hongtengxu@ruc.edu.cn

AAAI'22 Tutorial on Gromov-Wasserstein Learning
IJCAI'23 Tutorial on OT-based Machine Learning
AAAI'26 Tutorial on OT-based Machine Learning
https://hongtengxu.github.io/talks.html