Optimal Flow Matching: New method for generative modeling and optimal transport with straight trajectories in just one minimization round

MSc Program

Student: *Nikita Kornilov*
Research Advisor: Prof. *Alexander Gasnikov*
Co-Advisor : PhD *Alexander Korotin*

Skoltech

Introduction

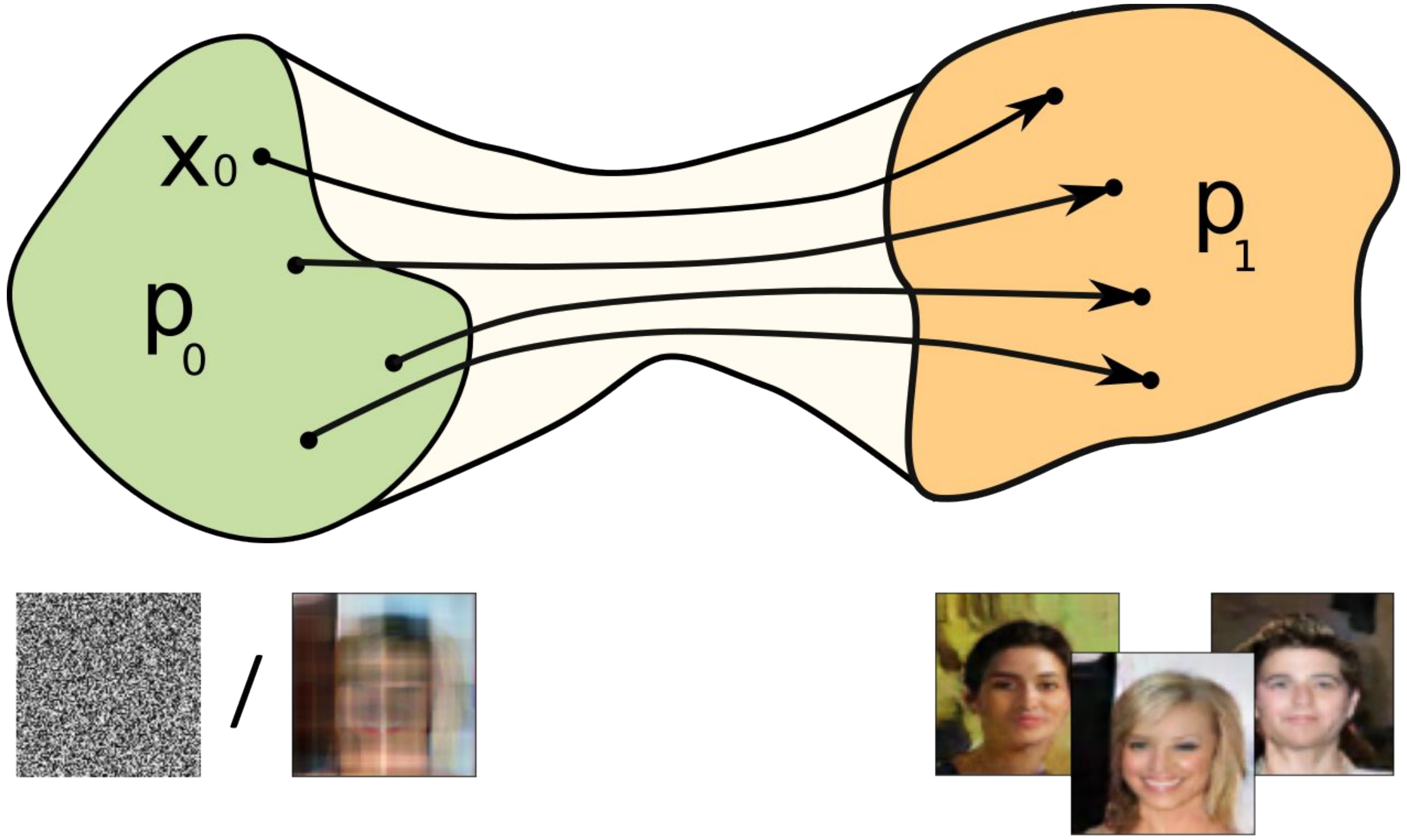
Generative modeling problem:

Map distribution p0 to distribution p1 via an ODE on [0,1]:

$$\begin{cases} \mathrm{d}x_t = v(x_t, t)\mathrm{d}t \\ x_0 \sim p_0 \end{cases}$$

Challenges:

Current methods obtain ODEs that have curved trajectories, resulting in time-consuming and ineffective sampling.



Introduction: Optimal Transport

Move one distribution to another and keep input-output similarity. Such transportation has ODE with straight paths generated by gradient of convex functions.

$$\inf_{T\#p_0=p_1} \int_{\mathbb{R}^D} \frac{\|x_0 - T(x_0)\|_2^2}{2} p_0(x_0)dx_0 \iff \min_{\text{convex } \Psi} \underbrace{\left[\int_{\mathbb{R}^D} \Psi(x_0)p_0(x_0)dx_0 + \int_{\mathbb{R}^D} \overline{\Psi}(x_1)p_1(x_1)dx_1\right]}_{=:\mathcal{L}_{OT}(\Psi)}.$$

Skoltech

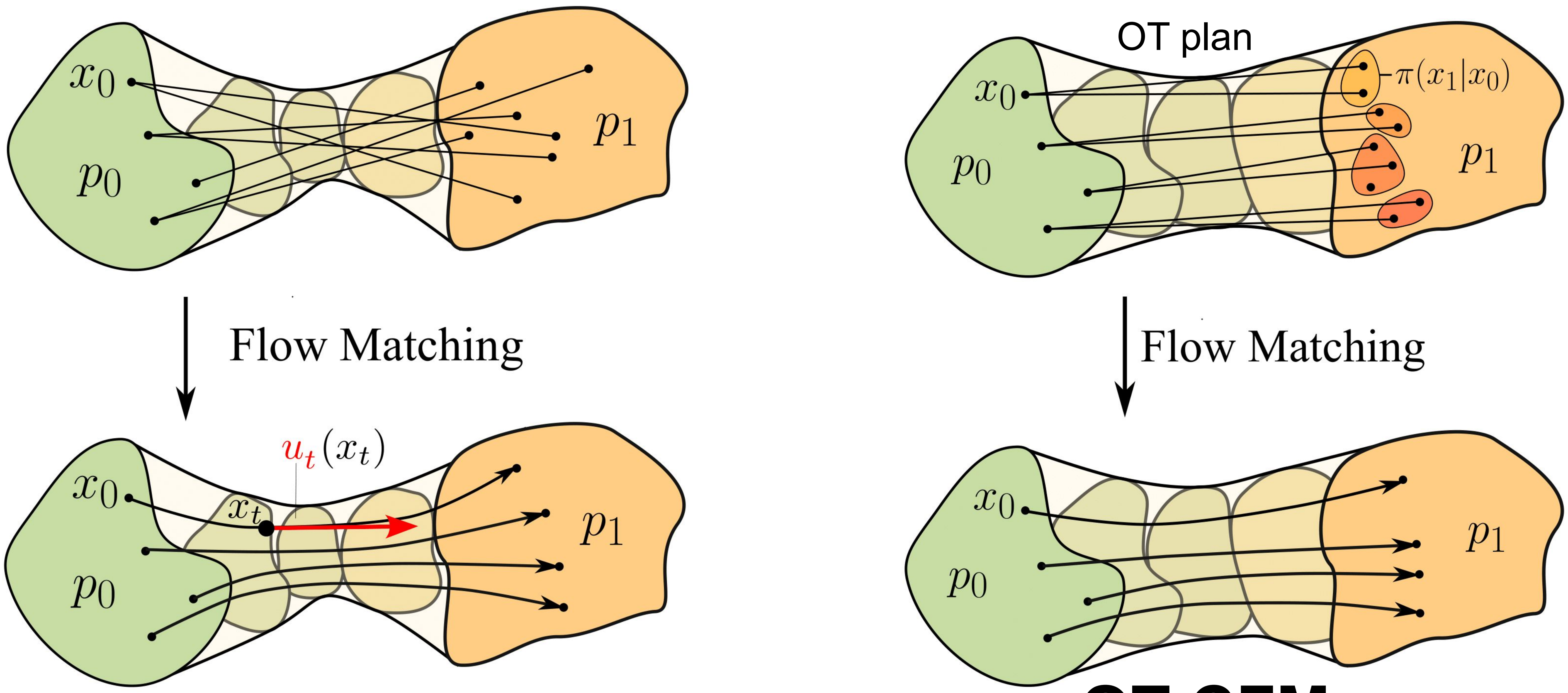
Previous approaches

Flow Matching: Learn the vector field of some linear interpolation between p0 and p1. The resulting paths depend on initial interpolation.

OT-Conditional Flow Matching: Apply FM with interpolation based on OT between discrete batches from the considered distributions. Such heuristic does not guarantee straight paths.

Rectified Flow: Iteratively solve FM and gradually rectify trajectories. In practice, it accumulates the error with each iteration and miss p1.

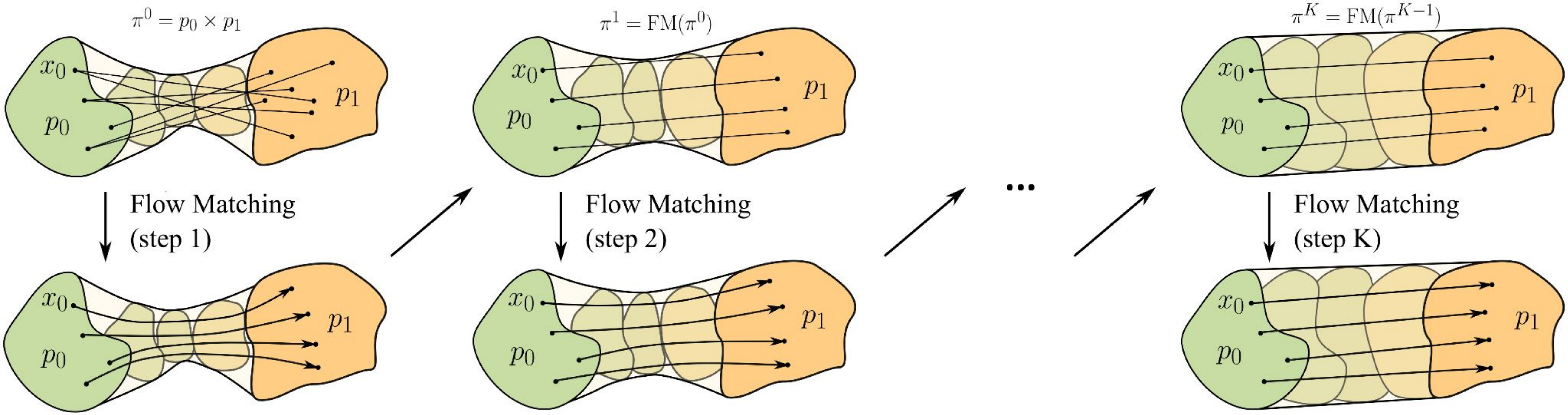
Previous approaches: FM and OT-CFM



FM

OT-CFM

Previous approaches: RF



Skoltech

Aim

I aim to propose and justify in theory and practice my own approach which fixes the problems of the the previous methods: iterative nature, non-straight paths, target miss.

Potential impact: Improvement of modern Flow Matching, Optimal Transport methods and theoretical contribution to the corresponding fields.

Contributions

I propose my theoretically justified novel **Optimal Flow Matching (OFM)** approach that after a single FM minimization for any initial plans obtains straight trajectories which can be simulated without ODE solving. Moreover, it recovers OT solution for the quadratic transport cost function.

Skoltech

Objectives

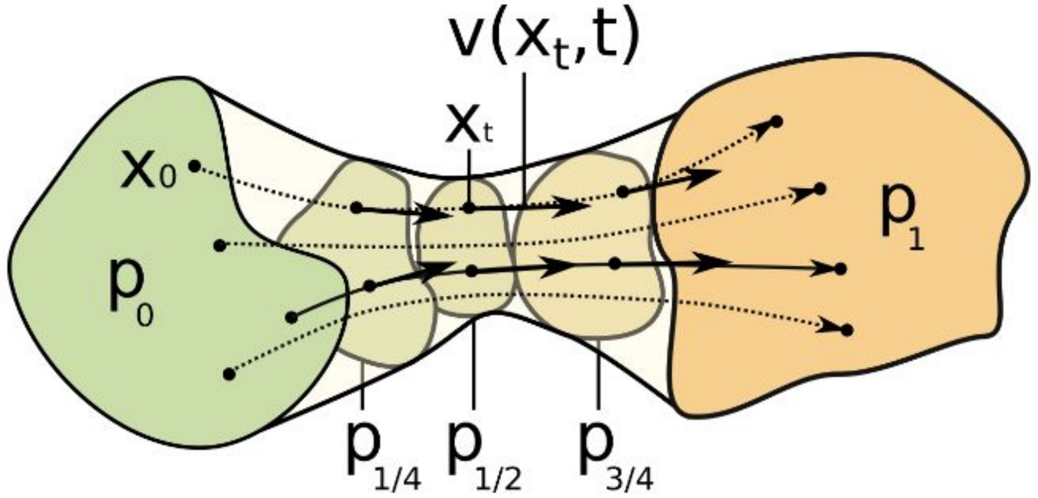
- To create theoretical foundation for my OFM: introduce specific vector fields and propose objective OFM loss.

- To prove that OFM retrieves OT solution via minimizing of the OT loss.

- Back up theory with the experiments, compare OFM with previous methods and test it on OT benchmark.

Skoltech

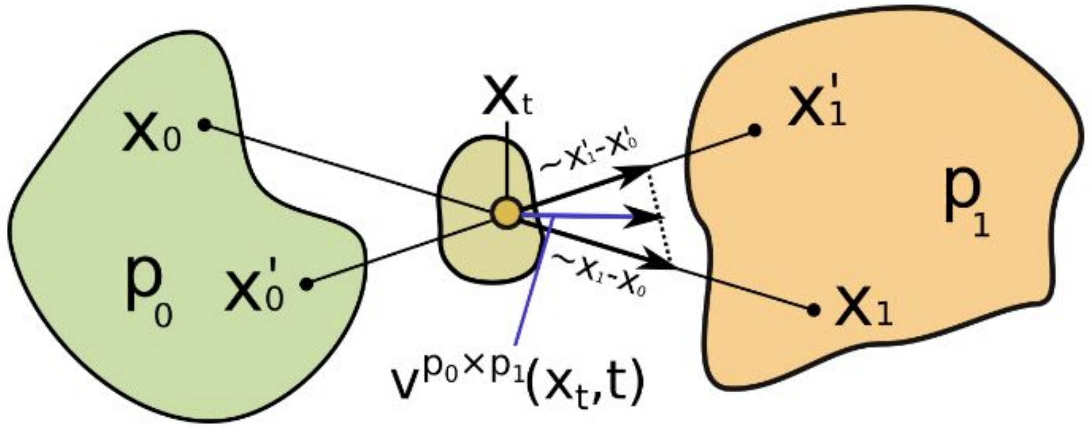
Flow Matching theory



$$\begin{cases} dx_t = v(x_t, t)t \\ x_0 \sim p_0 \end{cases}$$

Build linear interpolation probability path $p_t : x_t = x_0 \cdot (1 - t) + x_1 \cdot t$, where x_0, x_1 are sampled from transport plan $\pi \in \Pi(p_0, p_1)$.
This probability path is generated by a solution to Flow Matching (FM) objective:

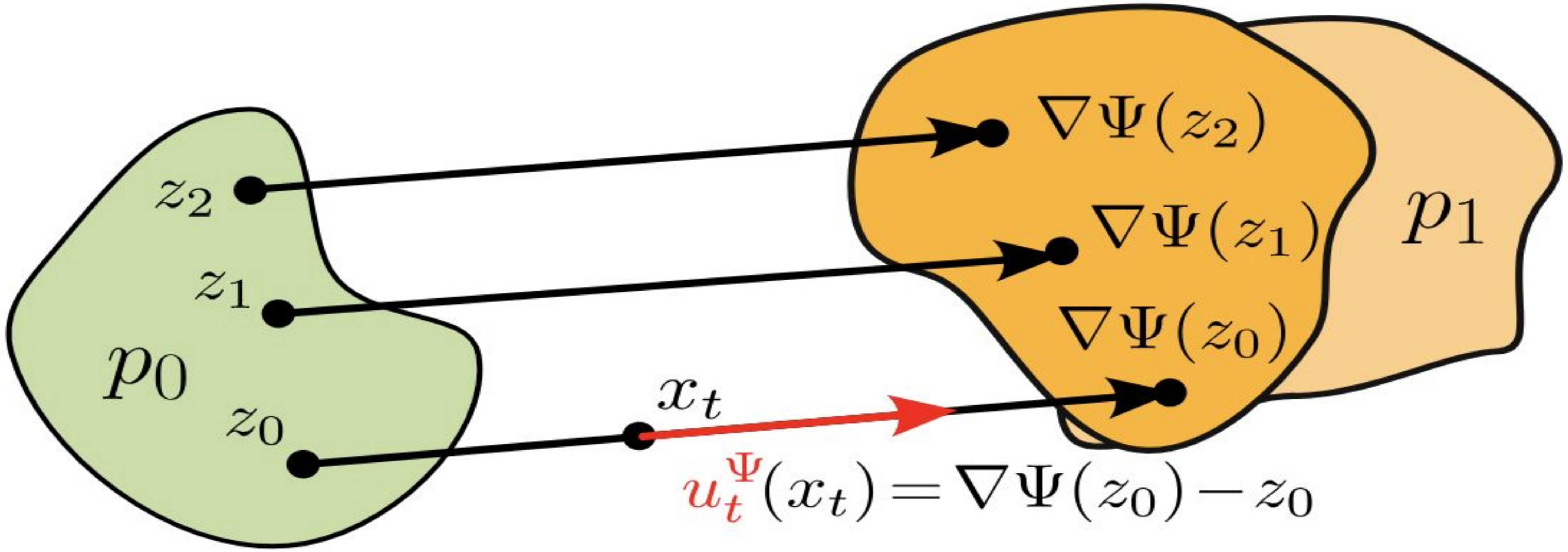
$$\boxed{\min_{v} \mathop{\mathbb{E}}_{x_0, x_1 \sim \pi} \left\{ \mathop{\mathbb{E}}_{t \sim [0,1]} \Big\| v(\underset{\substack{\| \\ x_0 \cdot (1-t) + x_1 \cdot t}}{x_t}, t) - (x_1 - x_0) \Big\|^2 \right\}.}$$



Optimal Vector Fields

Optimal vector field u^{Ψ} generates linear trajectories $\{\{z_t\}_{t\in[0,1]}\}$ s.t. there exist a convex function $\Psi : \mathbb{R}^D \to \mathbb{R}$, which for any path $\{z_t\}_{t\in[0,1]}$ pushes the initial point z_0 to the final one as $z_1 = \nabla\Psi(z_0)$, i.e.,

$$z_t = (1-t)z_0 + t\nabla\Psi(z_0), \quad t \in [0,1].$$



$$z_0^{\Psi}(x_t) = \arg\min_{z_0\in\mathbb{R}^D} \left[\frac{(1-t)}{2}\|z_0\|^2 + t\Psi(z_0) - \langle x_t, z_0\rangle \right].$$

Skoltech

Optimal Flow Matching

Main idea: during FM minimization, consider only optimal vector fields:

$$\mathcal{L}^{\pi}_{OFM}(\Psi) := \int_0^1 \left\{ \int_{\mathbb{R}^D \times \mathbb{R}^D} \|u_t^{\Psi}(x_t) - (x_1 - x_0)\|^2 \pi(x_0, x_1) dx_0 dx_1 \right\} dt.$$

Explicit gradient for convex functions parameterized by NN:

$$\frac{d\mathcal{L}^{\pi}_{OFM}}{d\theta} := \frac{d}{d\theta} \mathbb{E}_{t;x_0,x_1 \sim \pi} \left\langle \textbf{NO-GRAD} \left\{ 2 \left(t \nabla^2 \Psi_\theta(z_0) + (1-t) I \right)^{-1} \frac{(x_0 - z_0)}{t} \right\}, \nabla \Psi_\theta(z_0) \right\rangle,$$

where derivative is not taken from the terms under NO-GRAD{}.

Skoltech

OFM Theory

Lemma 2 (Main Integration Lemma) *For any two points $x_0, x_1 \in \mathbb{R}^D$ and a convex function Ψ, the following equality holds true:*

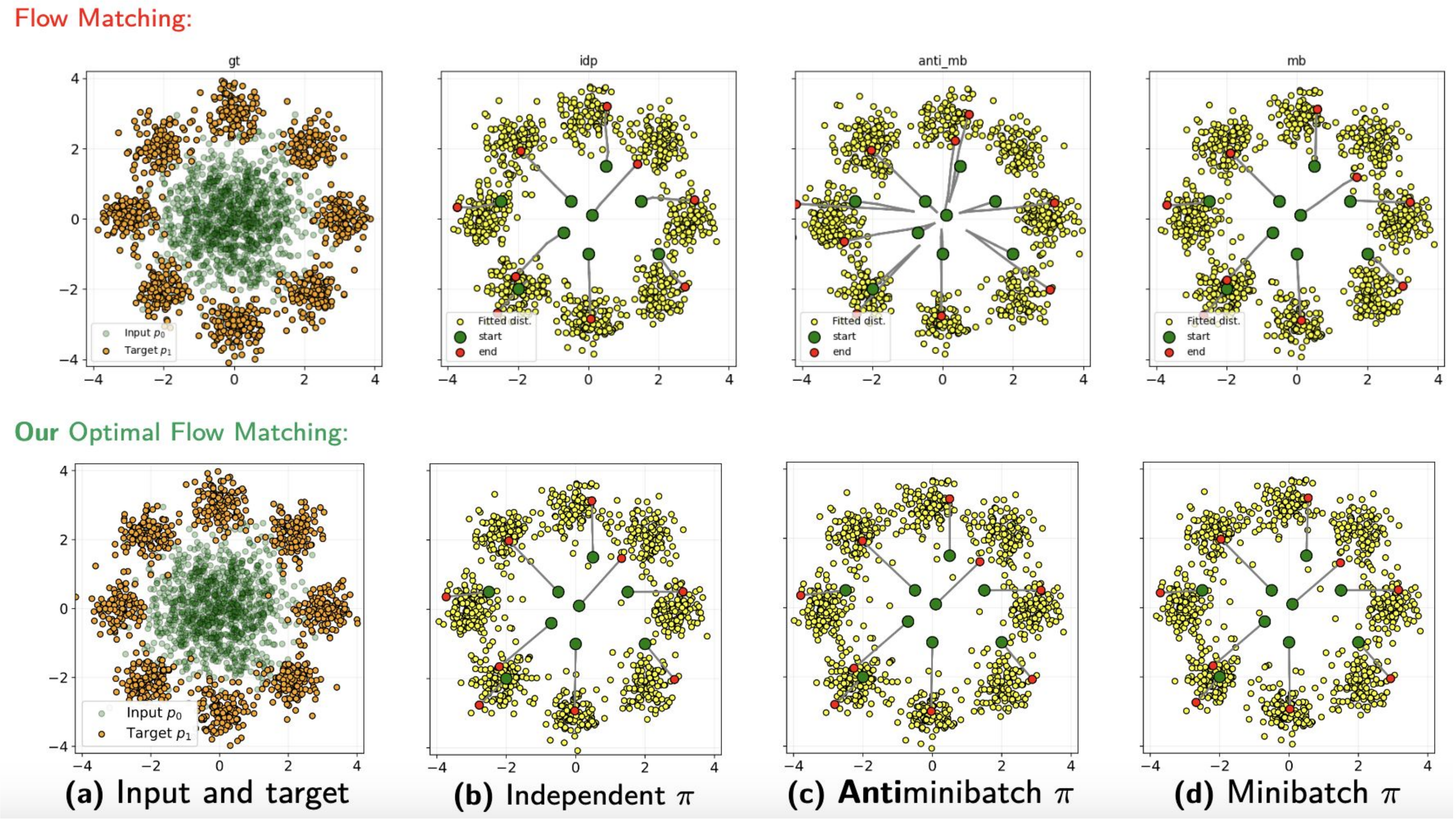
$$\int_0^1 \|u_t^{\Psi}(x_t) - (x_1 - x_0)\|^2 dt = 2 \cdot [\Psi(x_0) + \overline{\Psi}(x_1) - \langle x_0, x_1 \rangle],$$

where $x_t = tx_0 + (1-t)x_1$.

Theorem 3.1 (OFM and OT connection) *Consider two distributions $p_0, p_1 \in \mathcal{P}_{ac,2}(\mathbb{R}^D)$ and* ***any*** *transport plan $\pi \in \Pi(p_0, p_1)$ between them. Then, the dual Optimal Transport loss $\mathcal{L}_{OT}$ and Optimal Flow Matching loss $\mathcal{L}_{OFM}^{\pi}$ have* ***the same minimizers****, i.e.,*

$$\arg\min_{\text{convex } \Psi} \mathcal{L}_{OFM}^{\pi}(\Psi) = \arg\min_{\text{convex } \Psi} \mathcal{L}_{OT}(\Psi).$$

2D Exp (plan independence showcase)



OT Benchmark Exp

Solver	Solver type	$D=2$	$D=4$	$D=8$	$D=16$	$D=32$	$D=64$	$D=128$	$D=256$
MMv1* [13]		0.2	1.0	1.8	1.4	6.9	8.1	2.2	2.6
Amortization, ICNN** [20]	Dual OT solver	0.26	0.78	1.6	1.1	1.9	4.2	1.6	2.0
Amortization, MLP** [20]		0.03	0.22	0.6	0.8	2.0	2.1	0.67	0.59
Linear* [21]	Baseline	14.1	14.9	27.3	41.6	55.3	63.9	63.6	67.4
OT-CFM [9]		0.16	0.73	2.27	4.33	7.9	11.4	12.1	27.5
RF [6]		8.58	49.46	51.25	63.33	63.52	85.13	84.49	83.13
c-RF [5]	Flow Matching	1.56	13.11	17.87	35.39	48.46	66.52	68.08	76.48
OFM Ind		0.19	0.61	1.4	1.1	1.47	8.35	1.96	3.96
OFM MB		**0.15**	**0.52**	**1.2**	**1.0**	**1.2**	**7.2**	**1.5**	**2.9**

Table 4.1: $\mathcal{L}^2$−UVP values of solvers fitted on high-dimensional benchmarks in dimensions $D = 2, 4, 8, 16, 32, 64, 128, 256$.

OFM beats all FM-based methods and loses only to the best OT solver.

Skoltech

Unpaired Adult-Child OT Exp

All methods are applied in 512-dimensional latent space of the pretrained autoencoder.

OFM solutions does not depend on plan and visually keep input-output similarity.

Skoltech

Discussion of results

1) Unlike RF, OFM requires only single minimization round.

2) Unlike OT-CFM, OFM returns exactly straight trajectories.

3) Only OFM solves OT with the quadratic cost function.

4) Experiments uphold the theory.

Research limitations:

1) Input Convex Neural Networks parametrization.

2) Expensive training with hessian inversion and strongly convex minimization - can be fixed with advanced optimizers or amortization.

Scientific novelty

I propose completely new approach to generative modeling based on combination of OT and FM. Such combination has not been seen before in literature, and it builds a novel way to connect these two fields.

Conclusions

1. Alternative to the previous solutions was proposed and tested in theory and practice.

2. OFM is competitive with the previous solutions and does not have their fundamental drawbacks.

3. OFM paves a novel theoretical bridge between OT and FM.

Outlook

The achieved results indicate that the OFM has potential in generative modeling, and the future work must be dedicated to acceleration of practical performance:

1. Speed up training.

2. Find better parametrization for convex functions and soften dimensional dependency.

Acknowledgements

I would like to express my gratitude to Petr Mokrov, Alexander Gasnikov and Alexander Korotin who made this work possible.

The paper based on this research was accepted to NeurIPS 2024.

Thank you for your attention. Now I’m ready to answer your questions.