# Understanding Generalization In Conditional Flow Matching

Minisymposium « Modèles génératifs, OT et restauration d’images »

2025-06-03

Rémi Emonet – Anne Gagneux – Ségolène Martin – Quentin Bertrand – Mathurin Massias

SMAI 2025

Laboratoire Hubert Curien
UMR ▪ CNRS ▪ 5516 ▪ Saint-Étienne

Université Jean Monnet Saint-Étienne

Institut d'Optique Graduate School ParisTech

# Overview

- Generative Modeling
- A focus on flow approaches
- Conditional Flow Matching (CFM)
- Stochasticity and Generalization in CFM
- Inductive Bias, Failure and Generalization in CFM

*More details:* CFM Blogpost (ICLR Blogpost track) including the CFM playground (at the bottom)

# Generative Modeling

# Generative Modeling[W] = Density Estimation[W]

Given some dataset $\{x_i\}_{i=1}^N$

supposed drawn i.i.d. from an unknown distribution $P(X)$ ... or $p(X)$ or $p(X=x)$ or $p(x)$

try to recover $p(X)$

# Generative Model vs Discriminative Model

- Discriminative:
  - $P(Y|X)$
  - E.g. classification, regression, least squares, etc.
  - *given an image, what is the probability that this is a picture of a cat?*
- Generative:
  - $P(X)$
  - *what is the likelihood of this image?*
  - *generate a realistic image!*
- Generative:
  - joint, with class: $P(X,Y)$
  - class-conditional: $P(X|Y)$
  - NB: not the same "conditional" as in CFM

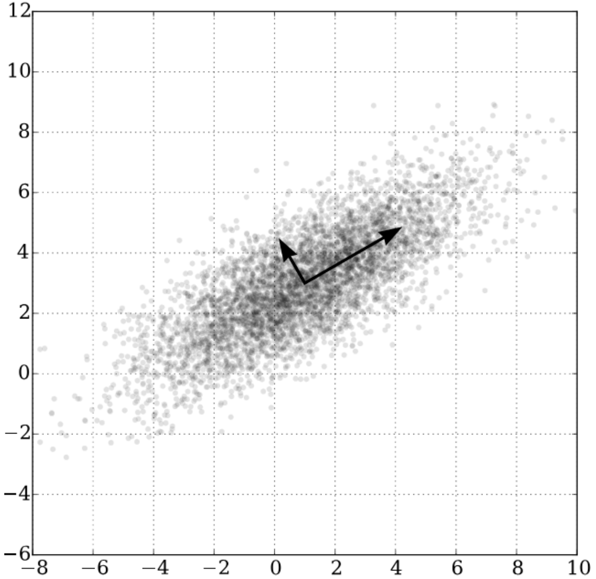
# Principal Components Analysis[W] (PCA)

Find an orthogonal subspace (lower dimension)

- maximizing the captured variance
- i.e. minimizing the "residual variance"

$$\arg\min_{\{z_i\}_i, W} \sum_i ||x_i - W z_i||_2^2$$

e.g. latent representation $z_i \in \mathbb{R}^5$, "unprojector" $W \in \mathbb{R}^{1024 \times 5}$




# Autoencoders (AE), Variational AE[W] (VAE)

PCA: $\arg\min_{\{z_i\}_i, W} \sum_i ||x_i - W z_i||_2^2$ … ⇒ "Reconstruction-error" minimizer

## Autoencoder (AE): A non-linear version of PCA

- replace $W z_i$ by a trained "non-linear model $Dec_\theta(z_i)$ … no simple projection ( $W^T$ ) to get $\{z\}_i$
- need to estimate all $\{z_i\}_i$ (as in any Bayesian Network) ⇒ trick: "amortize" (share the cost) by
  - replacing the estimation of all $\{z_i\}_i$
  - by a $z_i$ -guesser… $z_i = Enc_{\theta'}(x_i)$

$\text{we have} \quad ||x_i - W z_i||_2^2 = -\log(\exp(-||x_i - W z_i||_2^2)) = K - \lambda \cdot \log \mathcal{N}(\mu = W z_i, \sigma = 1)(x_i))$

PCA = Maximum Likelihood Estimator (minimizer of "constant minus log-likelihood")

VAE: A probabilistic version of non-linear PCA… $z_i \sim \mathcal{N}(Enc_{\theta'})$ + prior[1] (maximum a posteriori)

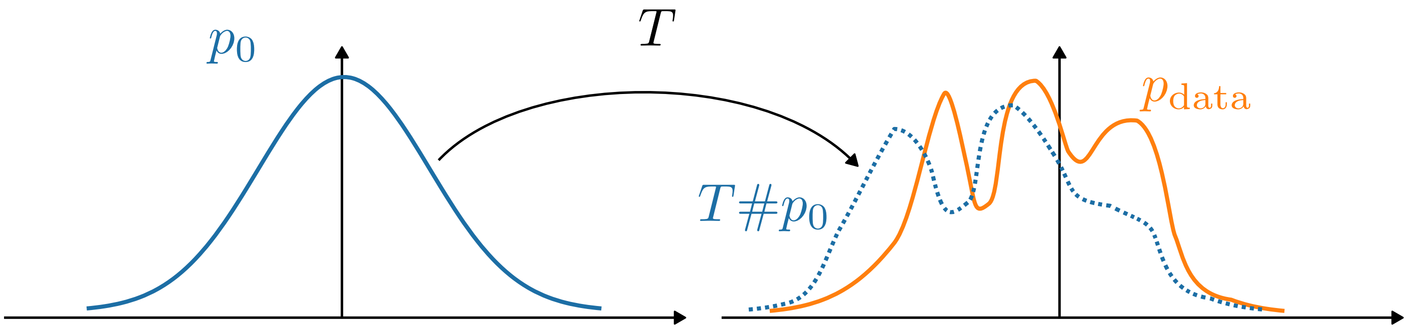
# Overview

- Generative Modeling
- A focus on flow approaches
- Conditional Flow Matching (CFM)
- Stochasticity and Generalization in CFM
- Inductive Bias, Failure and Generalization in CFM

*More details:* CFM Blogpost (ICLR Blogpost track) including the CFM playground (at the bottom)

# A focus on flow approaches

# Normalizing Flows



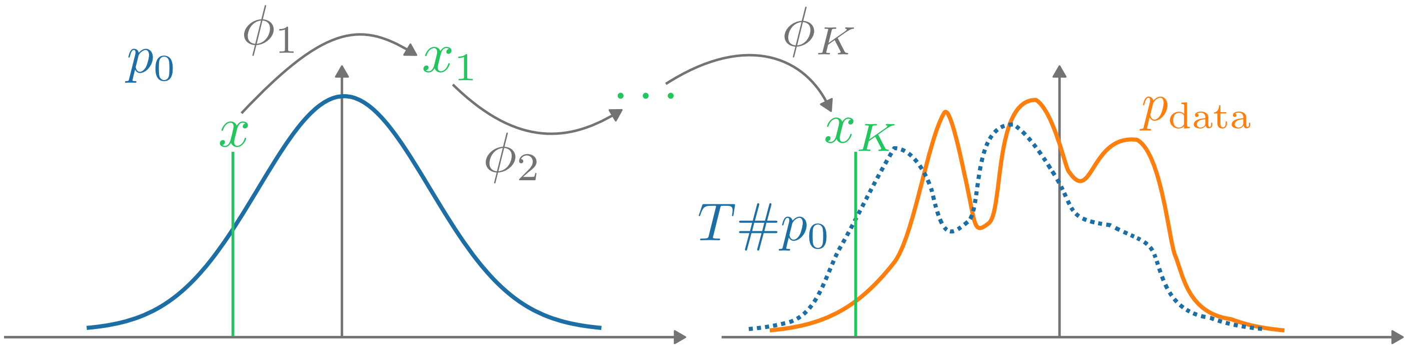

Definition (push-forward): if $x \sim p_0$ then $T(x) \sim T_{\#}p_0$

Normalizing flow (intuition):

- denoting $p_{gen} = T_{\#}p_0$
- e.g., locally, if $T$ compresses the space by a factor 42, then $p_{gen}(T(x)) = 42 \cdot p_0(x)$
- formally, change of variable, $p_{gen}(T(x)) = |det(J_{T^{-1}}(x))| \cdot p_0(x)$ *(determinant of the jacobian of* $T^{-1}$*)*

Principle: parametrize and learn $T$ ... so that its inverse exists (and has an easy jacobian det).

# Normalizing Flows, with composed functions




Learn a deep $T$, i.e.,

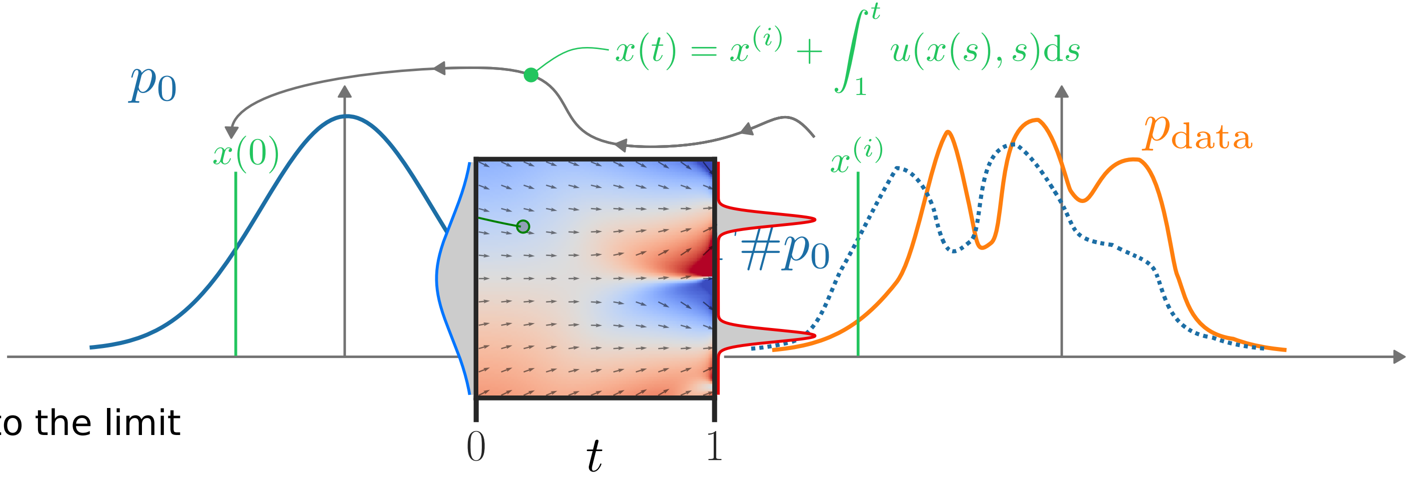
$$T = \phi_1 \circ \phi_2 \circ \ldots \circ \phi_K$$

Chain rule of change of variable,

$$|det(J_{T^{-1}}(x))| = \prod_k |det(J_{\phi_k^{-1}}(x))|$$

Principle: compose invertible blocks (with easy jacobian det)

# Continuous Normalizing Flows




Pushing to the limit

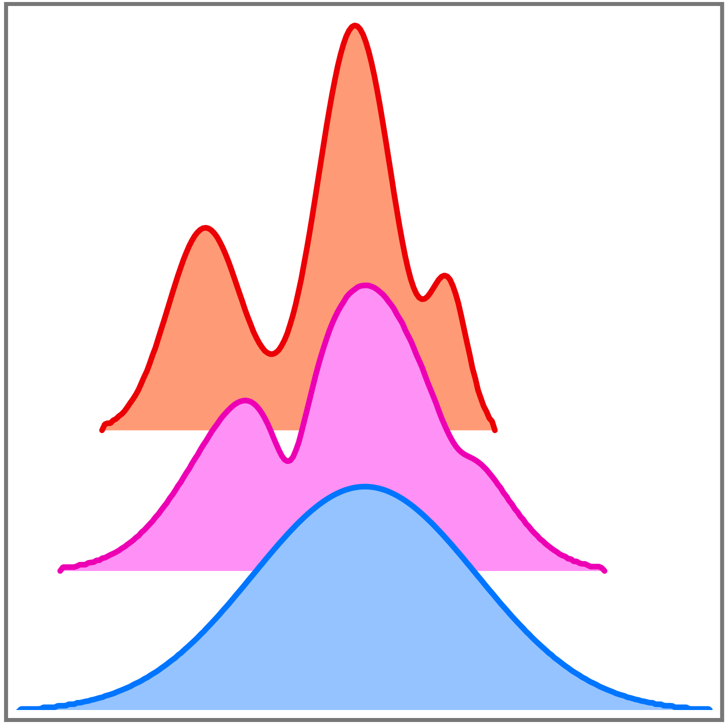
- infinitely many infinitely-small steps
- making depth continuous $k \mapsto t$
- replacing $\phi_k(x)$ by $u_t(x)$ , i.e., $u(x, t)$

Continuous Normalizing Flow

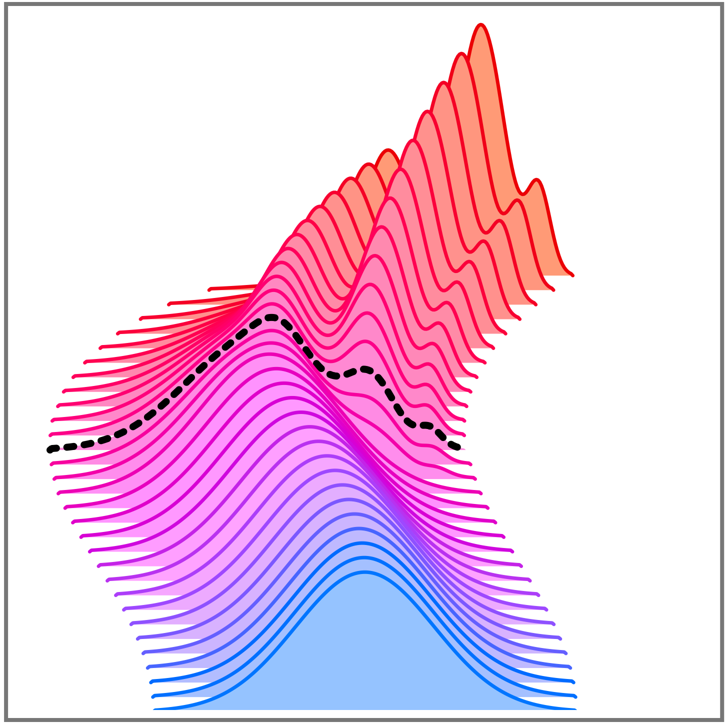
- easier: less constraints on $u$ than $\phi$

Forward and reverse ODE

# Continuous Normalizing Flows: visual summary



# Continuous Normalizing Flows: "limitation"



The flow is unspecified!
(there is an infinity of equally good solutions)

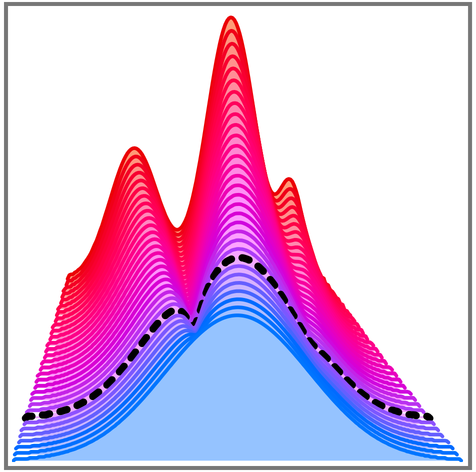
# Overview

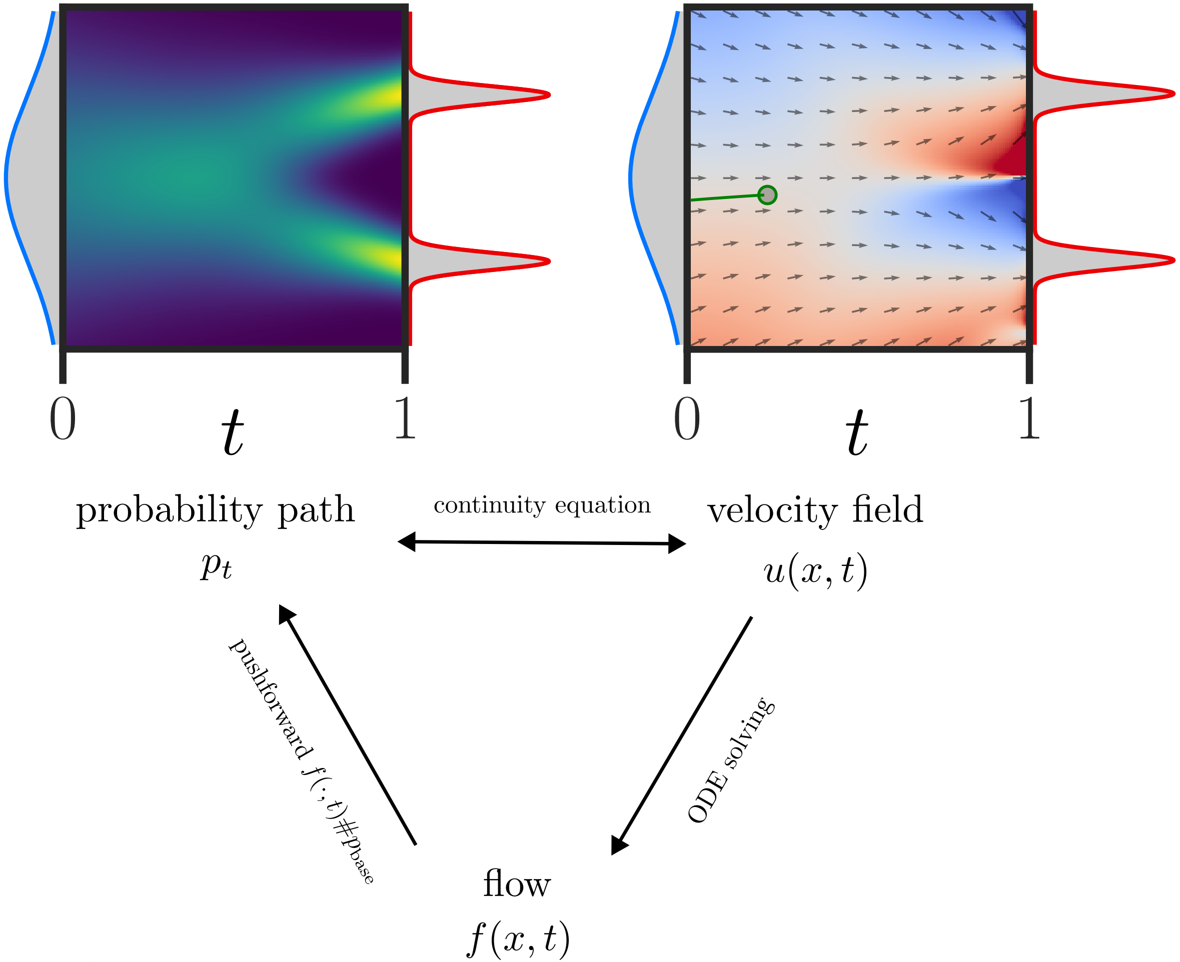
- Generative Modeling
- A focus on flow approaches
- Conditional Flow Matching (CFM)
- Stochasticity and Generalization in CFM
- Inductive Bias, Failure and Generalization in CFM

*More details:* CFM Blogpost (ICLR Blogpost track) including the CFM playground (at the bottom)

# Conditional Flow Matching (CFM)

# Visuals




0
t
1
0
t
1
probability path
$p_t$
continuity equation
velocity field
$u(x,t)$
pushforward $f(\cdot,t)\#p_{\text{base}}$
ODE solving
flow
$f(x,t)$

# Conditional Flow Matching (CFM) Principles

- Fully specify a probability path / velocity field / flow (like diffusion, unlike CNF)
- Use a ordinary (non-stochastic) differential equation (like CNF, unlike diffusion)

Solution ?

- introduce an arbitrary conditioning variables $z$
- specify the flow as an aggregation of conditional flows

Before diving into the details, let's look at one algorithm.

# Typical CFM algorithm

## Design choices

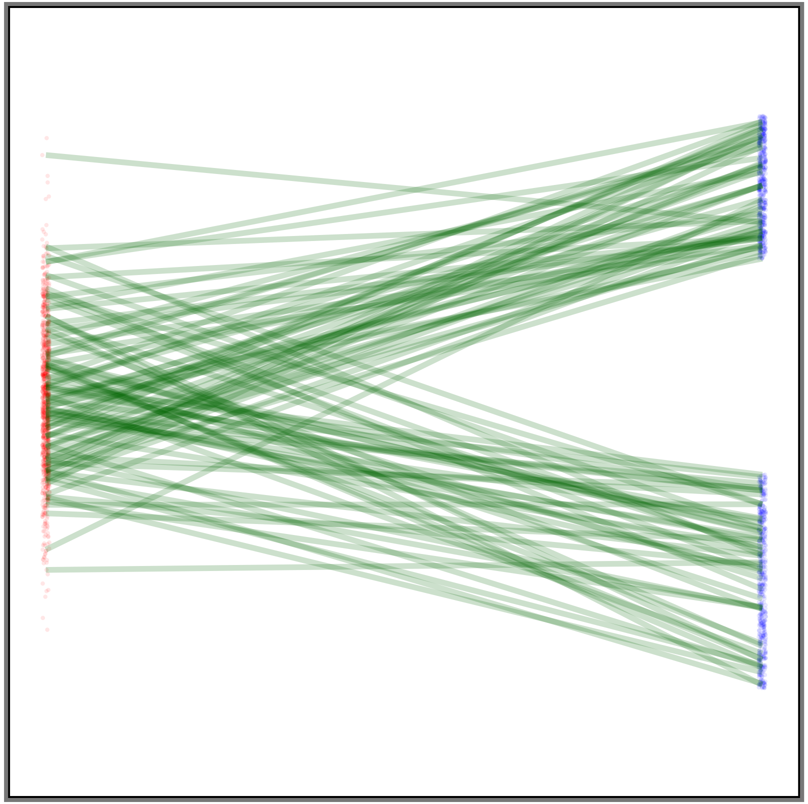
- conditioning variable $z$ is a pair
  - a *source* point, typically from $\mathcal{N}(0,1)$ *(but not necessarily, vs diffusion)*
  - a *target* point, typically form the (training) dataset
- conditional probability path/flow is a straight constant-velocity *(OT between two points)*

## Algorithm

$$
\begin{aligned}
z_0 &\sim \mathcal{N}(0, I) \\
z_1 &\sim Dataset \\
t &\sim Uniform([0,1]) \\
x &= t \cdot z_1 + (1-t) \cdot z_0 \\
\text{SGD step on } \theta \text{ with loss: } & \quad \|u_\theta(x,t) - (z_1 - z_0)\|_2^2
\end{aligned}
$$

That's it! *(up to practical hacks and a few days of training)*

# CFM: Does it works? the "inversion", path un-mixing



# CFM: Design choices

Decide on $p_0$ , typically $\mathcal{N}(0, I)$ *(but not necessarily, vs diffusion)*

Decide on the conditioning variable (and its distribution), e.g.

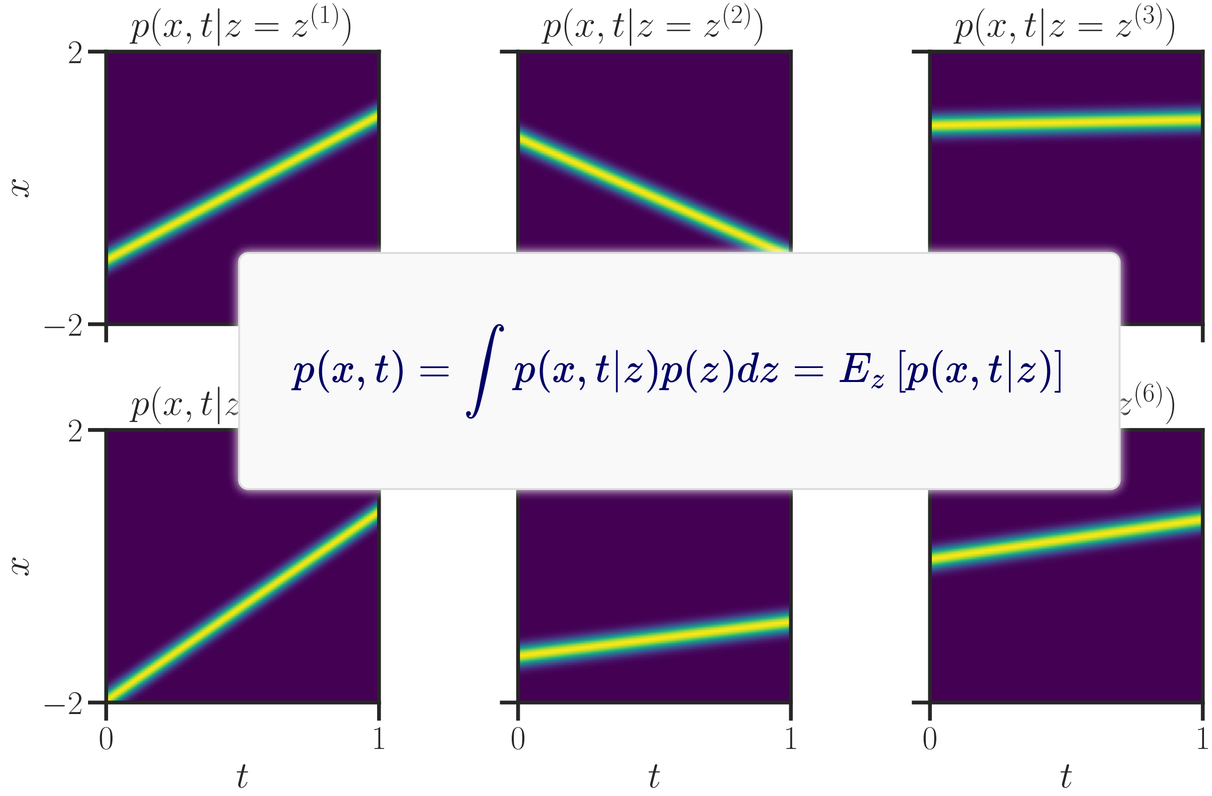
- $z$ is a pair $(x_0, x_1)$
- $z$ is a target point $x_1$
- $z$ is a minibatch of source and target
- $z$ is a pair, constrained by some clusters

Decide on the conditional "flow"

- conditional probability path $p_t(x|z)$ (or $p(x, t|z)$ )
- and the associated velocity field $u^{cond}(x, t)$

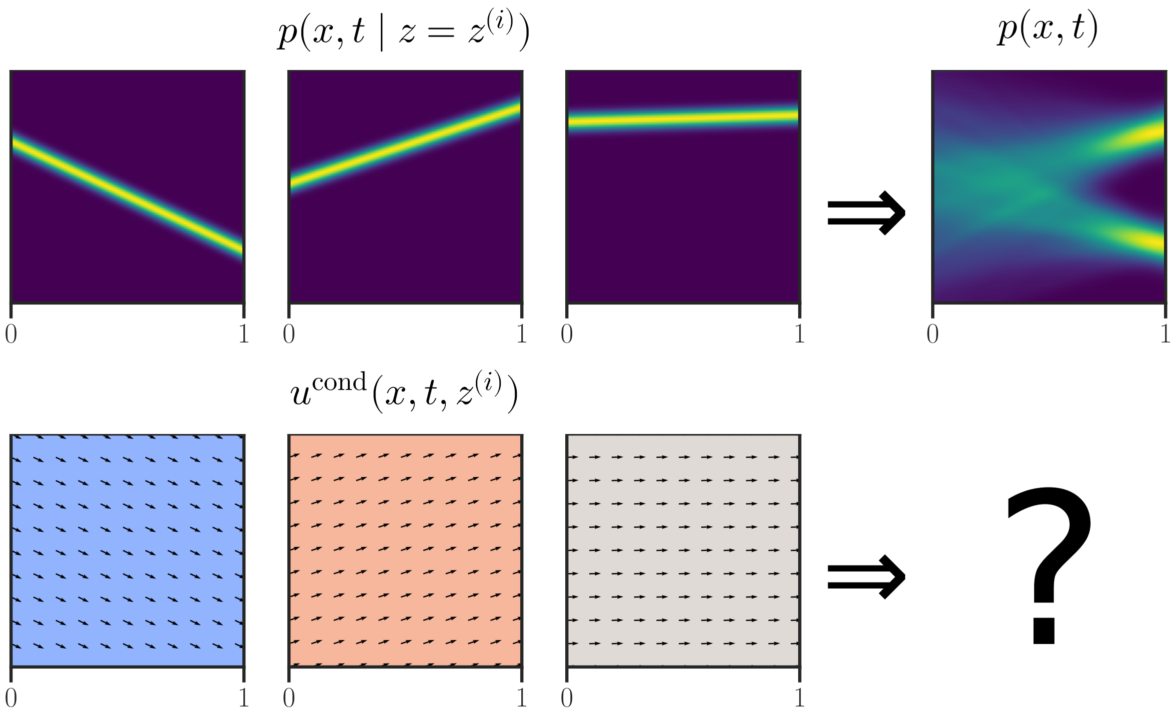
(under marginal constraints, on $p(x, t)$ )

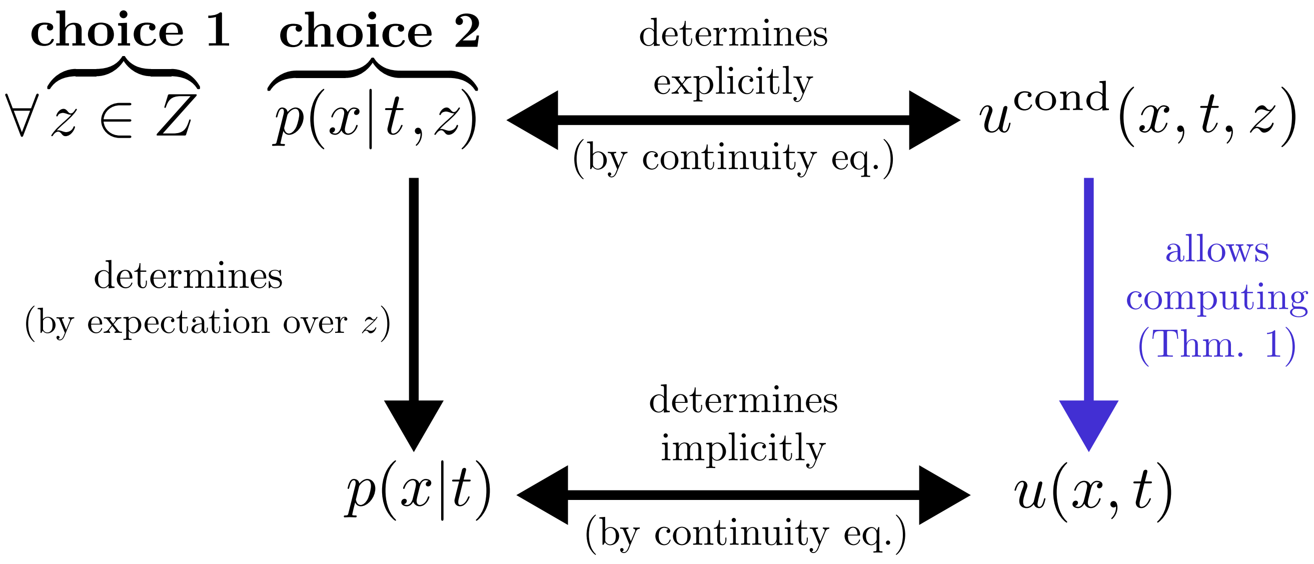
# CFM: $p(x,t|z)$ (conditional) to $p(x,t)$ is easy




$$p(x,t) = \int p(x,t|z)p(z)dz = E_z\left[p(x,t|z)\right]$$

# CFM: $u^{cond}(x,t,z)$ to $u(x,t)$ is less easy




choice 1
choice 2
$\forall z \in Z$
$p(x|t,z)$
determines
explicitly
(by continuity eq.)
$u^{\text{cond}}(x,t,z)$
determines
(by expectation over $z$)
allows
computing
(Thm. 1)
$p(x|t)$
determines
implicitly
(by continuity eq.)
$u(x,t)$

# CFM playground

R

$$p_0$$

G⇝.

U⇝.

G2⇝.

U2⇝.

$$p_1$$

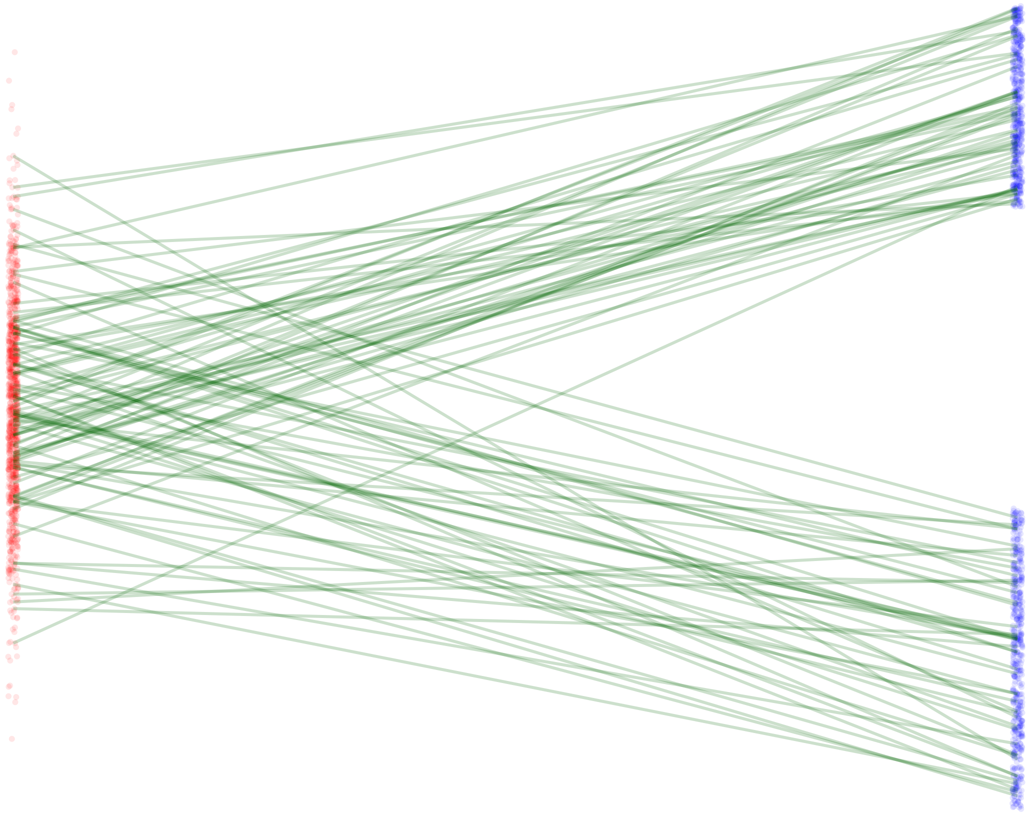
.⇝G

.⇝U

.⇝G2

.⇝U2

$$z$$



z

1

3

*

×

1s

.5s

.2s

# CFM: Closed form expression (Theorem 1 in the blog post)

$$\forall t, \forall x,$$

$$u^\star(x,t) = E_{z|x,t}[u^{cond}(x,t,z)]$$

(also written as)

$$\forall\, t, \forall\, x,$$

$$u^\star(x,t) = \int_z u^{cond}(x,t,z) p(z|x,t) = \sum_{i=1}^{N} u^{cond}(x,t,z=x_i) p(z=x_i|x,t)$$

(or through the bayes rule)

$$\forall\, t, \forall\, x,$$

$$u^\star(x,t) = \int_z u^{cond}(x,t,z) \frac{p(x,t|z)p(z)}{p(x,t)} = E_z\left[\frac{u^{cond}(x,t,z)p(x,t|z)}{p(x,t)}\right] = E_z\left[\frac{u^{cond}(x,t,z)p(x,t|z)}{\int_{z'} p(x,t|z')p(z')}\right]$$

# NB: CFM should only generate training points!

… but it does not, it generalizes

… yes, but why?

# Overview

- Generative Modeling
- A focus on flow approaches
- Conditional Flow Matching (CFM)
- Stochasticity and Generalization in CFM
- Inductive Bias, Failure and Generalization in CFM

*More details:* CFM Blogpost (ICLR Blogpost track) including the CFM playground (at the bottom)

# Stochasticity and Generalization in CFM

# Hypothesis: Generalization through variance?

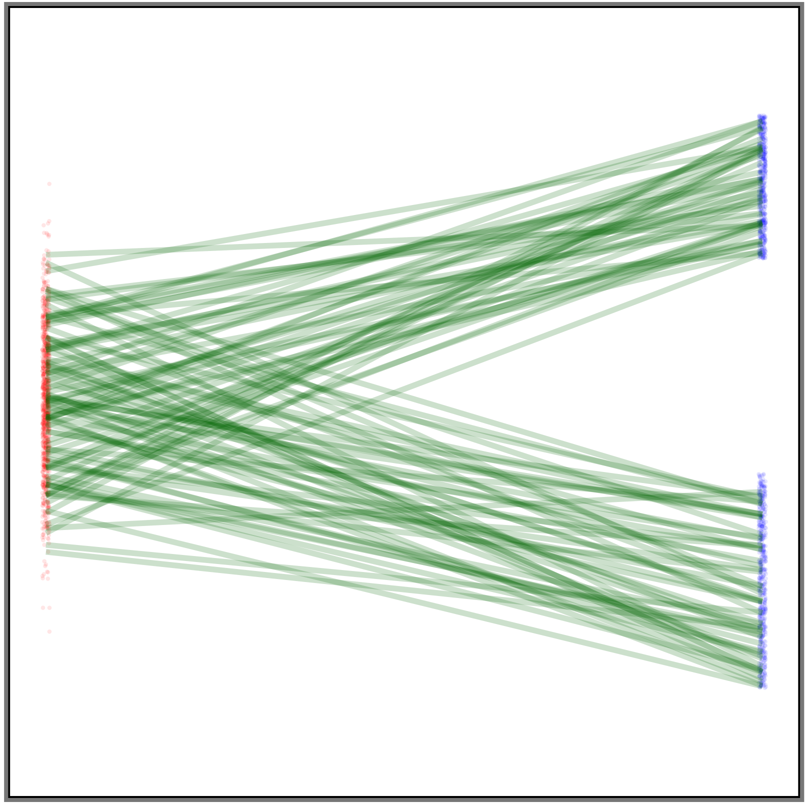
Algorithm (reminder)

$$z_0 \sim \mathcal{N}(0, I)$$
$$z_1 \sim Dataset$$
$$t \sim Uniform([0,1])$$
$$x = t \cdot z_1 + (1-t) \cdot z_0$$

SGD step on $\theta$ with loss:

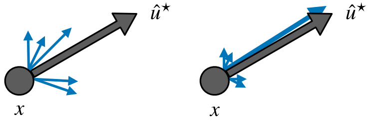
$$\|u_\theta(x,t) - (z_1 - z_0)\|_2^2$$

Maybe the noise in the target causes imperfect learning yielding generalization?



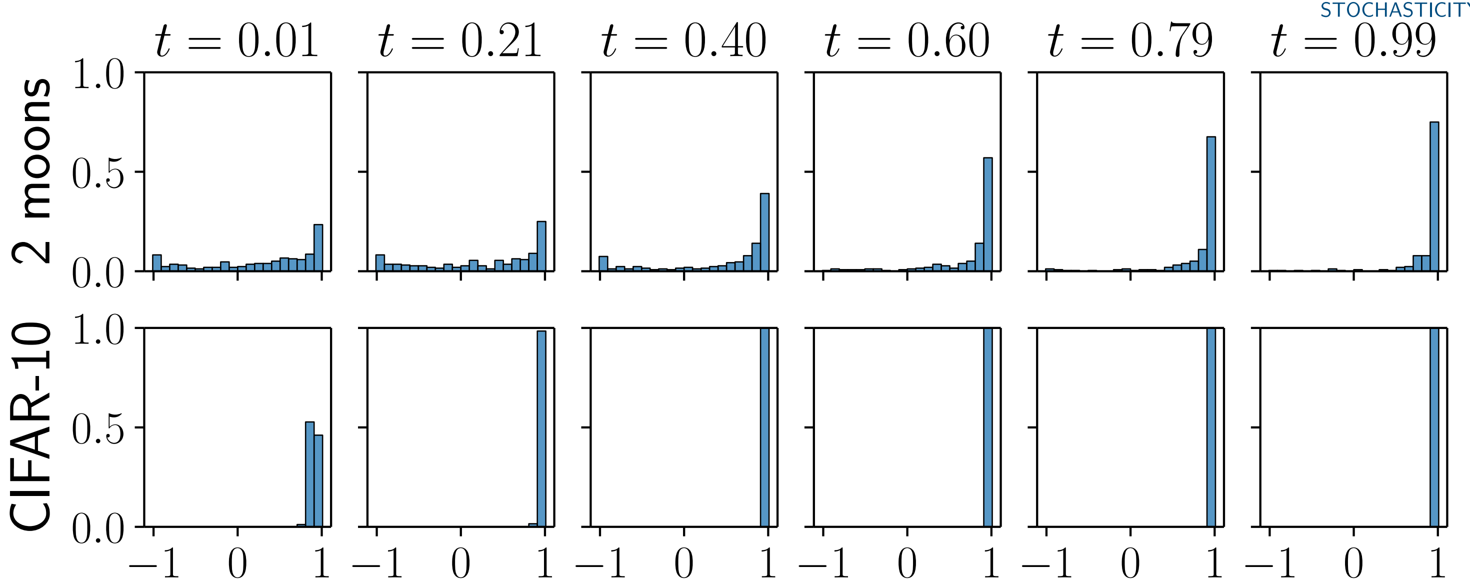
# CFM has a stochastic/noisy target?

$$\hat{u}^\star(x,t) = \sum_{i=1}^{n} p\left(z = x^{(i)} \mid x, t\right) u^{\mathrm{cond}}\left(x, t, z = x^{(i)}\right)$$




Common belief — STOCHASTICITY

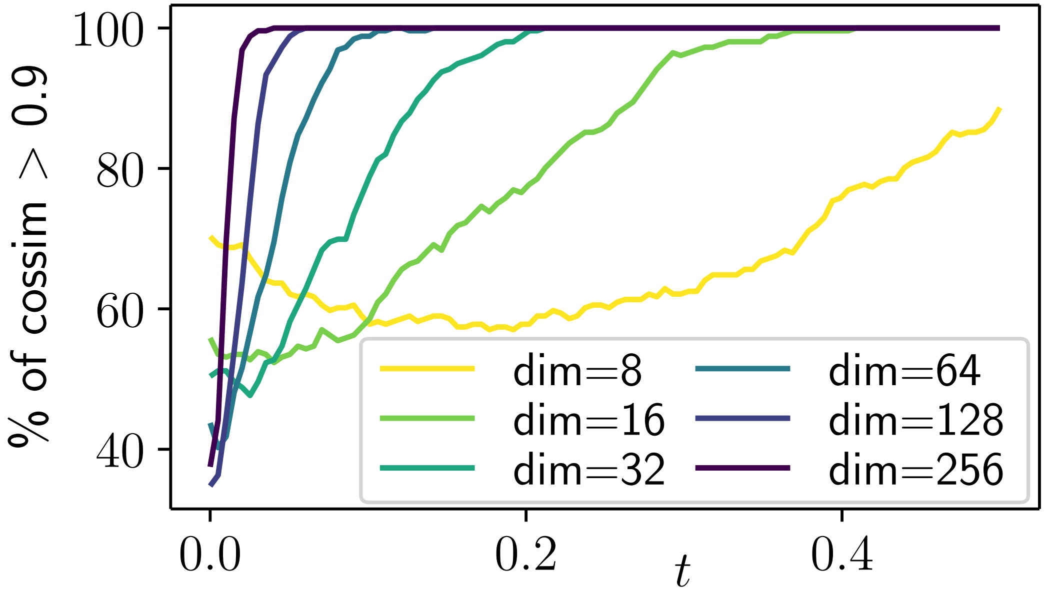
What really happens — NON-STOCHASTICITY




Histograms of cosine similarity between $u^\star$ and $u^{cond}$.

The target is not so stochastic (at most times).

# Beware of intuitions in small dimension



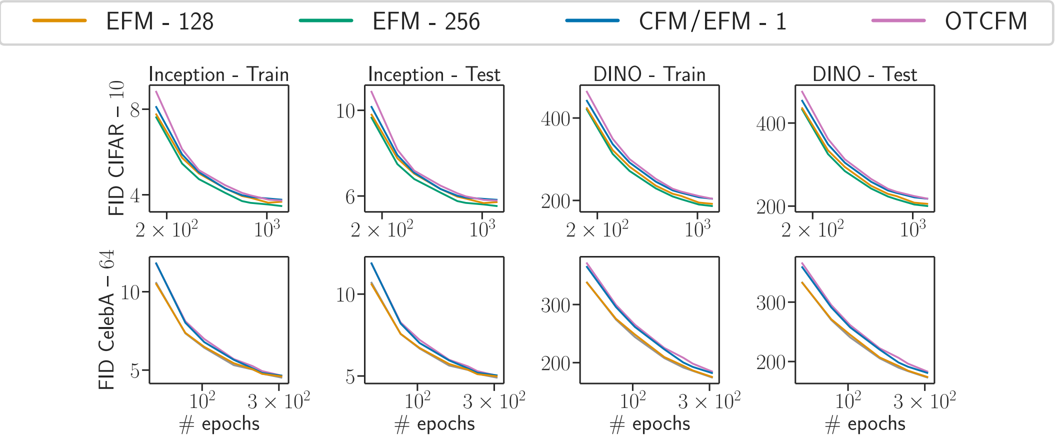

Aligment of $u^\star$ and $u^{cond}$, over time, for varying image dimensions $d$ on Imagenette.

Non-stochasticity happens even earlier in high dimension.

# Ruling out stochasticity: regressing against $u^\star$

$\dots$

SGD step on $\theta$ with loss: $\|u_\theta(x,t) - u^\star(x,t)\|_2^2$




FIDs (Frechet Inception Distance) across iterations, on CIFAR-10 and CelebA.

Training with a non stochastic target yields better/faster training.

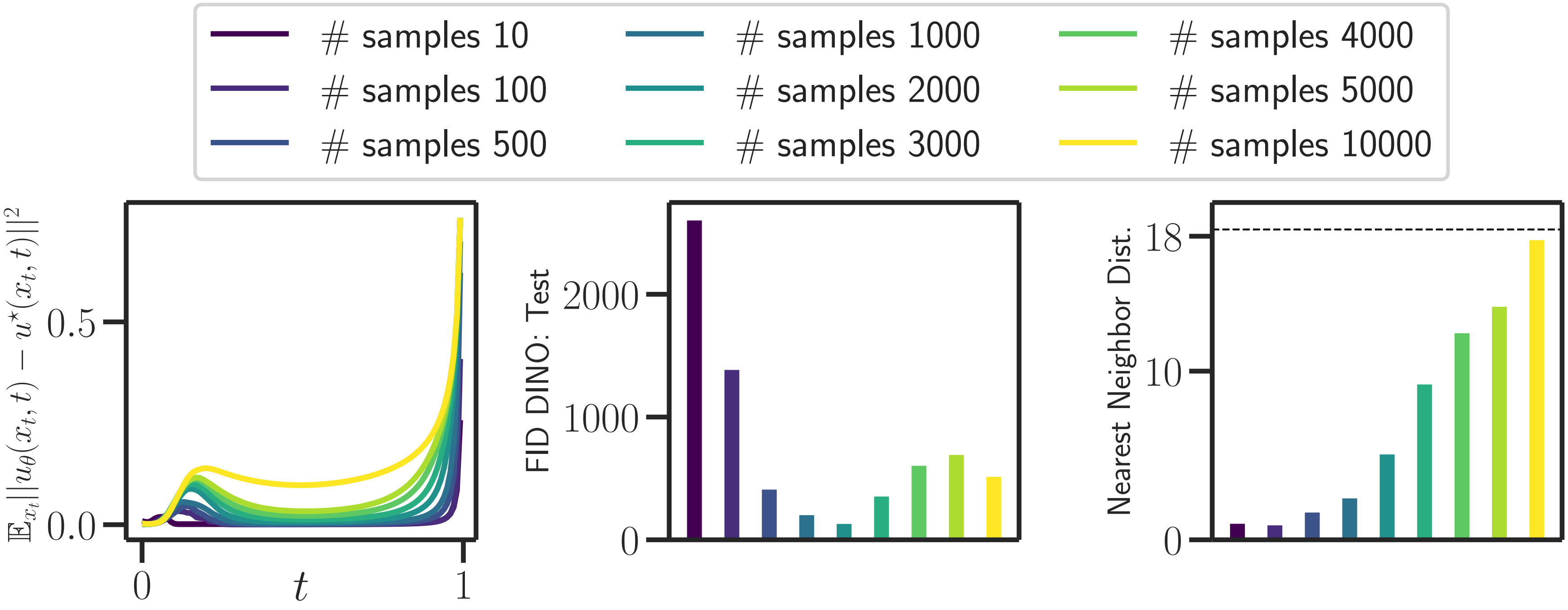
# Overview

- Generative Modeling
- A focus on flow approaches
- Conditional Flow Matching (CFM)
- Stochasticity and Generalization in CFM
- Inductive Bias, Failure and Generalization in CFM

*More details:* CFM Blogpost (ICLR Blogpost track) including the CFM playground (at the bottom)

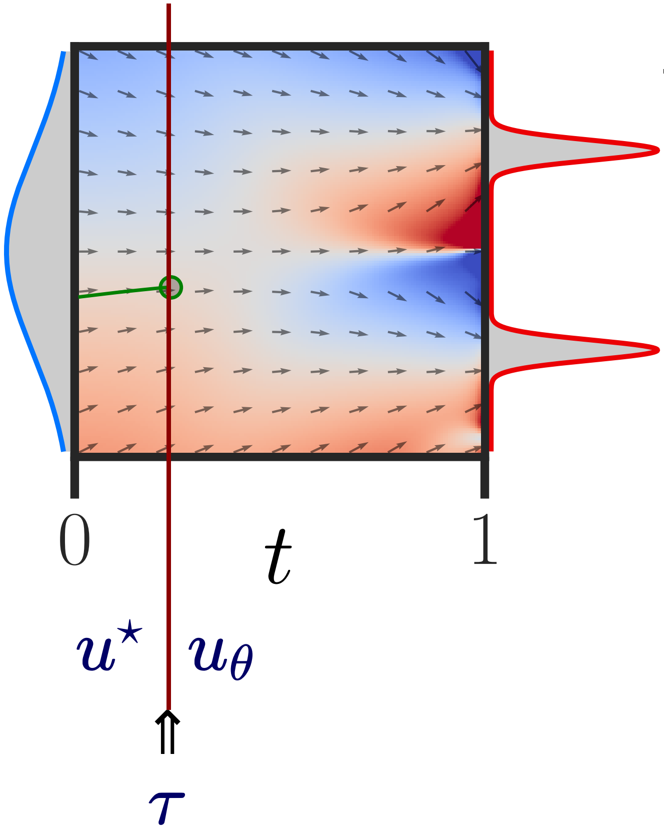
# Inductive Bias, Failure and Generalization in CFM

# CFM Works Because it Fails: comparing $u_\theta$ and $u^\star$




Generalization comes from not estimating $u^\star$ perfectly.

# Understanding generalization: using $u^\star$ then $u_\theta$ (CIFAR-10)



# Understanding generalization: using $u^\star$ then $u_\theta$ (CIFAR-10)

# Understanding generalization: using $u^\star$ then $u_\theta$ (CelebA)

# Overview

- Generative Modeling
- A focus on flow approaches
- Conditional Flow Matching (CFM)
- Stochasticity and Generalization in CFM
- Inductive Bias, Failure and Generalization in CFM

*More details:* CFM Blogpost (ICLR Blogpost track) including the CFM playground (at the bottom)

# Understanding Generalization In Conditional Flow Matching

Minisymposium « Modèles génératifs, OT et restauration d'images »

2025-06-03

Rémi Emonet – Anne Gagneux – Ségolène Martin – Quentin Bertrand – Mathurin Massias

SMAI 2025

Laboratoire Hubert Curien
UMR • CNRS • 5516 • Saint-Étienne
Université Jean Monnet Saint-Étienne
Institut d'Optique Graduate School ParisTech

# CFM playground

R

$$p_0$$

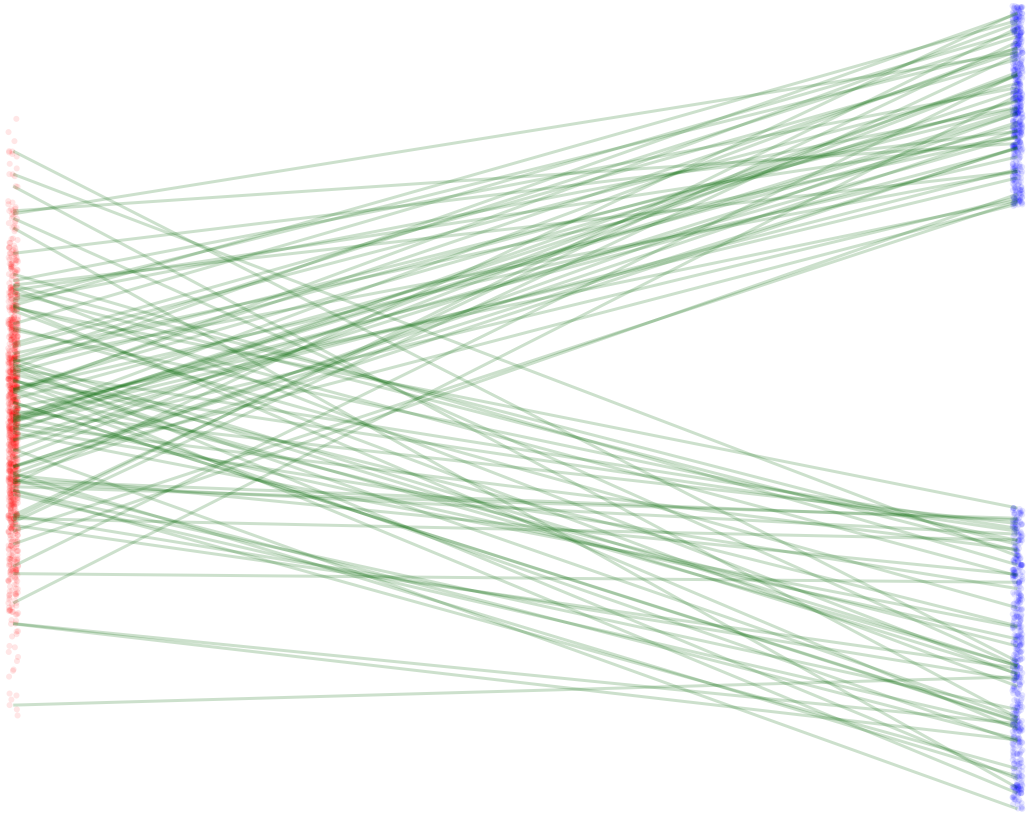
G⇝.
U⇝.
G2⇝.
U2⇝.

$$p_1$$

.⇝G
.⇝U
.⇝G2
.⇝U2

$$z$$

≈
☹
☺
⚛



1
3
*

×
1s
.5s
.2s