UP.T1 — The Lemma of the Course

Flow-Based Generative Models · UFRJ · 2026.2

01 Why this session exists

Where this course is going

- **Discrete flows.** Build the map out of invertible layers. Exact likelihood, restricted map.
- **Continuous time.** Let the map be the solution of a differential equation. The restriction goes; training now runs a solver inside every step.
- **Flow Matching.** Stop asking for the likelihood. Regress against a target you can write down.

Today is none of those. Today is the foundation all three stand on.

The promise made today

We prove **one** result. Then we announce, on one slide, the **three** later sessions that use it.

THE PROMISE

> When each of those three moments arrives, I will point back at this board.

Everything else in this session exists so that the promise can be stated precisely.

02 | Conditional expectation, done properly

Three objects share one notation

1 · a number

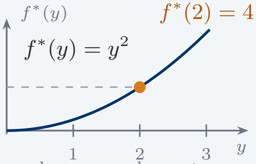
you fixed $y = 2$

$$\mathbb{E}[X \mid Y = 2] = 4$$

one point on the real line. It has no argument and no randomness.

2 · a function

you left y free



a rule, one number out per number in. Still not random.

3 · a random variable

you fed Y into rule 2

$$\boxed{Y} \longrightarrow \boxed{f^*} \longrightarrow \boxed{\mathbb{E}[X \mid Y]}$$

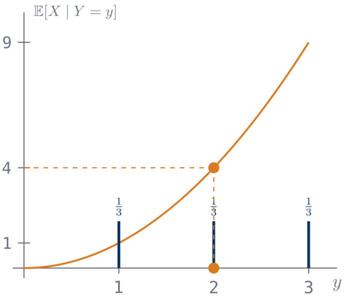
$$= f^*(Y) = Y^2$$

It takes the values 1, 4, 9, each with probability 1/3. Random, because Y is. This is the object the tower rule averages.

Ask it of every conditional expectation you meet: **which panel is this?**

The three objects, separated in time

Three objects, one notation — separated in time



$$\mathbb{E}[X \mid Y] = f^*(Y) = Y^2$$

```
Y = 1   ->   f*(Y) = 1   with prob 0.333
Y = 2   ->   f*(Y) = 4   with prob 0.333
Y = 3   ->   f*(Y) = 9   with prob 0.333
```

$$\mathbb{E}[f^*(Y)] = \tfrac{14}{3} = 4.6667$$

A RANDOM VARIABLE — random for one reason only: Y is.

Animated in the web deck — final frame shown.

03 | The projection Lemma

The statement

RESULT (THE PROJECTION LEMMA, UP.T1)

Write $f^*(y) = \mathbb{E}[X \mid Y = y]$.
Then for **every** square-integrable function f of Y:

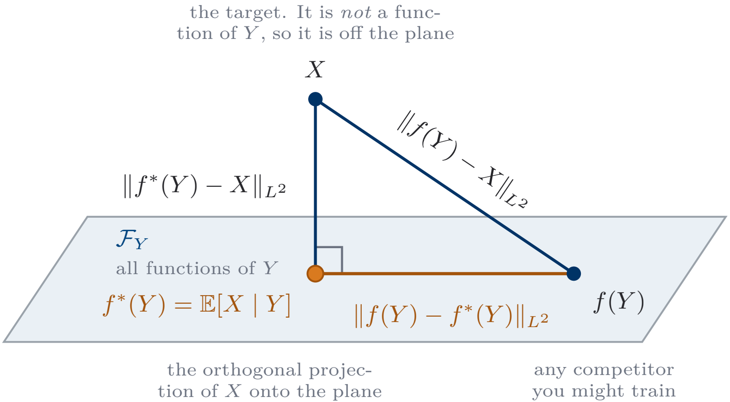
$$\mathbb{E}\|f(Y) - X\|^2 = \mathbb{E}\|f^*(Y) - X\|^2 + \mathbb{E}\|f(Y) - f^*(Y)\|^2.$$

So f^* minimizes the left side, and it is the only minimizer.

Square-integrable, throughout: $\mathbb{E}\|X\|^2 < \infty$ and $\mathbb{E}\|f(Y)\|^2 < \infty$.

The proof is on the board. It is one expansion, and one cross term that dies.

What the picture says

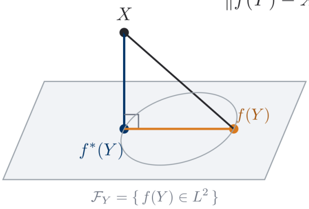


$$\|f(Y) - X\|_{L^2}^2 = \|f^*(Y) - X\|_{L^2}^2 + \|f(Y) - f^*(Y)\|_{L^2}^2$$

Pythagoras, written out: $\mathbb{E}\|f(Y) - X\|^2 = \mathbb{E}\|f^*(Y) - X\|^2 + \mathbb{E}\|f(Y) - f^*(Y)\|^2$

The angle that never opens

The cross term is zero for EVERY competitor



$$\|f(Y) - X\|^2 = \|f^*(Y) - X\|^2 + \|f(Y) - f^*(Y)\|^2$$

the floor

no function of Y goes below it

bars: L2 norms, Monte Carlo, n = 400,000
cross term across all 241 positions: at worst 0.71 standard errors
from zero — and a deliberately wrong projection reaches 65

Animated in the web deck — final frame shown.

04 | The announcement

This Lemma will be used exactly three times

1. **U3.T1** — to *define* the marginal velocity field: it is a conditional expectation of conditional velocities.
2. **U3.T2** — to *prove* the CFM theorem: swapping the unknown marginal target for a conditional surrogate leaves θ-gradients unchanged (corollary (b)).
3. **U3.T2, detour D2** — to *derive* denoising score matching: same three lines, different random variables.

Tattoo it somewhere.

05 Pushforwards, and what a model is

Pushforward

DEFINITION (PUSHFORWARD)

$\psi_\# p$ is the distribution of $\psi(x)$ when $x \sim p$:

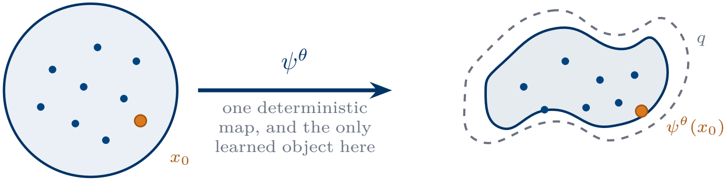
$$\mathbb{E}_{y \sim \psi_\# p}\left[g(y)\right] = \mathbb{E}_{x \sim p}\left[g(\psi(x))\right] \quad \text{for every bounded } g.$$

Read it as an instruction: **to sample from $\psi_\# p$, sample from p and apply ψ.**

You built one in the primer: $x = \mu + \sigma\varepsilon$ is $\mathcal{N}(\mu, \sigma^2) = \psi_\# \mathcal{N}(0, 1)$.

The symbol $\psi_\#$ is the course standard, after (Lipman et al. 2024) §3.3.

A generative model is a pushforward of noise



$p_0 = \mathcal{N}(0, I_d)$

the source. Fixed, standard, and never learned. Here $t = 0$.

$p_\theta = (\psi^\theta)_\# p_0$

what the map does to the source. Here $t = 1$, and q is the data the model has to reach.

Sampling is one draw from p_0, then one evaluation of ψ^θ.

Training is the job of closing the gap between the solid outline and the dashed one.

$$p_\theta = (\psi^\theta)_\# p_0, \qquad p_0 = \mathcal{N}(0, I_d)$$

Maximum likelihood is a KL

RESULT (MAXIMUM LIKELIHOOD IS A KL)

$$\arg\max_{\theta} \mathbb{E}_{x_1 \sim q}\left[\log p_\theta(x_1)\right] = \arg\min_{\theta} \mathrm{KL}(q \,\|\, p_\theta).$$

Two lines: expand the divergence, drop the term with no θ in it.

The data sits in the **first** argument — the mass-covering direction of the primer's §3.3.

But maximum likelihood needs $\log p_\theta(x_1)$**, and the definition above only gives you samples.**

06 What comes next

Why likelihoods will be hard

Unit	Strategy	What it costs
U1	invertible by construction	what the map is allowed to be
U2	volume change by calculus	a solver inside every training step
U3	refuse to play	nothing yet — this is where the Lemma returns

And one thing that cannot even be written down yet: comparing two whole *paths* of distributions, rather than two distributions.

PS1.2 — the Lemma, in closed form and on a machine

$X \sim \mathcal{N}(\mu_0, \sigma_0^2 I_d)$, and $Y = X + \sigma\varepsilon$.

THREE PARTS

(a) Compute $\mathbb{E}[X \mid Y = y]$ in closed form, from the toolkit sheet.
(b) In $d = 1$: train a small network on samples (y, x) with the square loss. Plot it against your closed form.
(c) Replace the target X by $Z = X + \eta$ with $\mathbb{E}[\eta \mid Y] = 0$. Retrain. The **losses** differ; the **learned function** does not.

This exercise has a secret identity, to be revealed in U3.

References

Lipman, Yaron, Marton Havasi, Peter Holderrieth, Neta Shaul, Matt Le, Brian Karrer, Ricky T. Q. Chen, David Lopez-Paz, Heli Ben-Hamu, and Itai Gat. 2024. “Flow Matching Guide and Code.” https://arxiv.org/abs/2412.06264.