A Proofs of Section 2

A.1 Proof of Proposition 1

Proof. • In the case where $z \sim \hat{p}_{\text{data}}$, conditional probability writes

$$
\begin{aligned}
p(z = x^{(i)}|x, t) &= \frac{p(x, t, z = x^{(i)})}{p(x, t)} && (9)\\
&= \frac{p(x|t, z = x^{(i)})p(t, z = x^{(i)})}{p(x, t)} && (10)\\
&= \frac{p(x|t, z = x^{(i)})p(t, z = x^{(i)})}{\sum_{i'=1}^{n} p(x, t, z = x^{(i')})} && (11)\\
&= \frac{p(x|t, z = x^{(i)})p(t)\overbrace{p(z = x^{(i)})}^{\frac{1}{n}}}{\sum_{i'=1}^{n} p(x|t, z = x^{(i')})p(t)\underbrace{p(z = x^{(i')})}_{\frac{1}{n}}} && (12)\\
&= \frac{p(x|t, z = x^{(i)})}{\sum_{i'=1}^{n} p(x|t, z = x^{(i')})} \quad . && (13)
\end{aligned}
$$

Pluging Equation (13) in Equation (2) yields the closed-formed formula for the velocity field:

$$
\begin{aligned}
u^\star(x, t) &= \sum_{i=1}^{n} u^{\text{cond}}(x, t, z = x^{(i)})p(z = x^{(i)}|x, t) && (14)\\
&= \sum_{i=1}^{n} u^{\text{cond}}(x, t, z = x^{(i)})\frac{p(x|t, z = x^{(i)})}{\sum_{i'=1}^{n} p(x|t, z = x^{(i')})} \quad . && (15)
\end{aligned}
$$

which proves Equation (5); using that $x|t, z = x^{(i)} \sim \mathcal{N}(tx^{(i)}, (1-t)^2 \,\mathrm{Id})$ and $u^{\text{cond}}(x, t, z = x^{(i)}) = \frac{x^{(i)} - x}{1-t}$ yields Equation (6).

• For the case $z \sim p_0 \times \hat{p}_{\text{data}}$,

$$
\begin{aligned}
\hat{u}^\star(x, t) &:= \int_z u^{\text{cond}}(x, t, z)p(z|x, t)\,\mathrm{d}z && (16)\\
&= \int_z u^{\text{cond}}(x, t, z)\frac{p(x, z, t)}{p(x, t)}\,\mathrm{d}z && (17)\\
&= \int_z u^{\text{cond}}(x, t, z)\frac{p(x|z, t)p(z)p(t)}{\int_{z'} p(x|t, z')p(t)p(z')\,\mathrm{d}z'}\,\mathrm{d}z && (18)\\
&= \int_z u^{\text{cond}}(x, t, z)\frac{p(x|z, t)p(z)}{\int_{z'} p(x|t, z')p(z')\,\mathrm{d}z'}\,\mathrm{d}z && (19)
\end{aligned}
$$

Now for the denominator, since $z \sim p_0 \times \hat{p}_{\text{data}}$:

$$
\begin{aligned}
\int_{z'} p(x|t, z')p(z')\,\mathrm{d}z' &= \int_{x_0} \sum_{i=1}^{n} \delta_{(1-t)x_0 + tx^{(i)}}(x)\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}x_0^2)\mathrm{d}x_0 && (20)\\
&= \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2(1-t)^2}(x - tx^{(i)})^2\right) && (21)
\end{aligned}
$$

since $\delta_{(1-t)x_0 + tx^{(i)}}(x)$ imposes $x = (1-t)x_0 + tx^{(i)}$, meaning $x_0 = \frac{x - tx^{(i)}}{1-t}$.

Equivalently, the Dirac imposes $x^{(i)} - x_0 = \frac{x^{(i)}-x}{1-t}$, hence for the numerator:

$$\int_z u^{\text{cond}}(x,t,z)p(x|z,t)p(z)\,\mathrm{d}z \int_{x_0} \sum_{i=1}^{n}(x^{(i)} - x_0)\delta_{(1-t)x_0+tx^{(i)}}(x)\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}\|x_0\|^2)\mathrm{d}x_0 \tag{22}$$

$$= \sum_{i=1}^{n} \frac{x^{(i)} - x}{1-t}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2(1-t)^2}\|x - tx^{(i)})\|^2\right) \tag{23}$$

Taking the ratio of Equations (21) and (23) concludes the proof. □

B Additional details and comments on empirical flow matching

First, recalls on the optimal velocity (Equation (6)) and the empirical flow matching loss (Equations (7) and (8)) are provided in Appendix B.1. The unbiasedness of the estimator is presented in Appendix B.2, and its proof is in Appendix B.3.

B.1 Recalls

The closed-form formula of the "optimal" velocity field is:

$$\hat{u}^\star(x,t) = \sum_{l=1}^{n} \frac{x^{(l)} - x}{1-t} \cdot \left[\text{softmax}\left(\left(-\frac{\|x - tx^{(k)}\|^2}{2(1-t)^2}\right)_{k=1,\dots,n}\right)\right]_l \quad . \tag{6}$$

The proposed loss uses mini-batches of size M (instead of all n training points) to build an estimator $\hat{u}^\star_M$ of $\hat{u}^\star$:

$$\mathcal{L}_{\text{EFM}}(\theta) = \mathbb{E}_{\substack{t\sim\mathcal{U}([0,1]) \\ x_0\sim p_0 \\ x_1\sim\hat{p}_{\text{data}} \\ x_t=(1-t)x_0+tx_1 \\ b^{(1)}:=x_1\ ;\ b^{(2)},\dots,b^{(M)}\sim\hat{p}_{\text{data}}}} \|u_\theta(x_t,t) - \hat{u}^\star_M(x_t,t)\|^2 \ , \tag{7}$$

with

$$\hat{u}^\star_M(x_t,t) = \sum_{j=1}^{M} \frac{b^{(j)} - x_t}{1-t} \cdot \left[\text{softmax}\left(\left(-\frac{\|x_t - tb^{(k)}\|^2}{2(1-t)^2}\right)_{k=1,\dots,M}\right)\right]_j \quad . \tag{8}$$

Crucially, in Equation (7) the sample $b^{(1)}$ depends on x_t and is reused in the estimate $\hat{u}^\star_M$. This important detail yields an unbiased estimator of $\hat{u}^\star$.

B.2 Theoretical properties of the proposed estimator

Proposition 2. *Let $\hat{p}_{\text{data}}(z|x,t)$ denote $p(z|x,t)$, meaning a probability distribution on $\{x^{(i)}\}_{i=1}^n$, with expression given by Equation (13). With no constraints on the learned velocity field u_θ,*

i) The minimizer of Equation (7) writes, for all (x,t)

$$\mathbb{E}_{b^{(1)}\sim\hat{p}_{\text{data}}(\cdot|x,t)\ ;\ b^{(2)},\dots,b^{(M)}\sim\hat{p}_{\text{data}}}\left[\hat{u}^\star_M(x,t)\right] \ . \tag{24}$$

ii) In addition, for all (x,t), the minimizer of Equation (7) equals the optimal velocity field, i.e.,

$$\mathbb{E}_{b^{(1)}\sim\hat{p}_{\text{data}}(\cdot|x,t)\ ;\ b^{(2)},\dots,b^{(M)}\sim\hat{p}_{\text{data}}}\left[\hat{u}^\star_M(x,t)\right] = \hat{u}^\star(x,t) \ . \tag{25}$$

iii) The conditional variance of the estimator $\hat{u}^\star_M$ is smaller than the usual conditional variance:

$$\text{Var}_{b^{(1)}\sim\hat{p}_{\text{data}}(\cdot|x,t)\ ;\ b^{(2)},\dots,b^{(M)}\sim\hat{p}_{\text{data}}}\left[\hat{u}^\star_M(x,t)\right] \leq \text{Var}_{b^{(1)}\sim\hat{p}_{\text{data}}(\cdot|x,t)}\left[u^{\text{cond}}(x,b^{(1)},t)\right] . \tag{26}$$

The proof of Proposition 2 can be found in Appendix B.3. First, we discuss below the relation between Proposition 2 and the sampling literature.

Links with importance sampling. The estimator $\hat{u}^\star$ in Equation (6) can be seen as a form of *importance sampling* (see Robert et al. 1999, Chap. 3 for an in-depth reference). In a nutshell, importance sampling is a way to estimate an expectation when one cannot easily sample from the random variable it depends on. More precisely, in the ideal case $z \sim p_{\text{data}}$ (as opposed to $z \sim \hat{p}_{\text{data}}$), the velocity field formula is the following

$$u^\star(x_t, t) = \mathbb{E}_{z|x_t,t}\left[u^{\text{cond}}(x_t, z, t)\right] \tag{27}$$

$$= \int_z u^{\text{cond}}(x_t, z, t) p(z|x_t, t) \mathrm{d}z \ . \tag{28}$$

When $z \sim p_{\text{data}}$, it is difficult to sample from $z|x_t, t$, but the latter equation can be rewritten as

$$u^\star(x_t, t) = \int_z u^{\text{cond}}(x_t, z, t) \frac{p(z|x_t, t)}{p(z)} p(z) \mathrm{d}z \tag{29}$$

and one can easily sample from $z \sim \hat{p}_{\text{data}}$ using the empirical data distribution $x^{(1)}, \dots, x^{(n)}$

$$u^\star(x_t, t) \approx \frac{1}{n} \sum_{i=1}^{n} u^{\text{cond}}(x_t, x^{(i)}, t) \frac{p(z = x^{(i)}|x_t, t)}{p(x^{(i)})} \tag{30}$$

$$= \sum_{i=1}^{n} u^{\text{cond}}(x_t, x^{(i)}, t) p(z = x^{(i)}|x_t, t) \tag{31}$$

$$:= \hat{u}^\star(x_t, t) \ . \tag{32}$$

Links with self-normalized importance sampling. The estimator $\hat{u}^\star_M$ in Equation (8) draws strong connections with self-normalized importance sampling (see Owen 2013, Chap. 9.2), and Rao-Blackwelled estimators (Casella and Robert, 1996; Cardoso et al., 2022). Very importantly, as discussed in Ryzhakov et al. (2024), self-normalized important sampling estimators are usually biased, in the sense that in general:

$$\mathbb{E}_{b^{(1)}, \dots, b^{(M)} \sim \hat{p}_{\text{data}}} \hat{u}^\star_M(x_t, t) \neq u^\star(x_t, t) \ . \tag{33}$$

A crucial point is that our estimator includes $b^{(1)} \sim \hat{p}_{\text{data}}(\cdot|x_t, t)$, which yields the main result of Proposition 2

$$\mathbb{E}_{b^{(1)} \sim \hat{p}_{\text{data}}(\cdot|x_t,t), b^{(2)}, \dots, b^{(M)} \sim p_{\text{data}}} \left[\hat{u}^\star_M(x_t, t)\right] = u^\star(x_t, t) \ . \tag{34}$$

Finally, this idea of regressing against a more deterministic target derived from the optimal closed-form velocity field has also been empirically explored for diffusion models (Xu et al., 2023).

B.3 Proof of Proposition 2

Proof of Item (i). With no constraints on u_θ, the empirical flow matching loss writes:

$$\mathbb{E}_{\substack{t \sim \mathcal{U}([0,1]) \\ x_1 \sim \hat{p}_{\text{data}} \\ x_t = (1-t)x_0 + t x_1 \\ b^{(1)} := x_1 \ ; \ b^{(2)}, \dots, b^{(M)} \sim \hat{p}_{\text{data}}}} \|u_\theta(x_t, t) - \hat{u}^\star_M(x_t, t)\|^2 \ , \tag{35}$$

$$= \mathbb{E}_{\substack{t \sim \mathcal{U}([0,1]) \\ x_t \sim p_t}} \mathbb{E}_{\substack{b^{(1)} \sim \hat{p}_{\text{data}}(\cdot|x_t,t) \\ b^{(2)}, \dots, b^{(M)}|x_t,t}} \|u_\theta(x_t, t) - \hat{u}^\star_M(x_t, t)\|^2 \ , \tag{36}$$

$$= \mathbb{E}_{\substack{t \sim \mathcal{U}([0,1]) \\ x_t \sim p_t}} \mathbb{E}_{\substack{b^{(1)} := \hat{p}_{\text{data}}(\cdot|x_t,t) \\ b^{(2)}, \dots, b^{(M)} \sim \hat{p}_{\text{data}}}} \|u_\theta(x_t, t) - \hat{u}^\star_M(x_t, t)\|^2 \ \text{because } b^{(2)}, \dots, b^{(M)} \perp\!\!\!\perp x_t, t \ , \tag{37}$$

which is minimized when for all x_t, t

$$u_\theta(x_t, t) = \mathbb{E}_{\substack{b^{(1)} \sim \hat{p}_{\text{data}}(\cdot|x_t,t) \\ b^{(2)}, \dots, b^{(M)} \sim \hat{p}_{\text{data}}}} \left[\hat{u}^\star_M(x_t, t)\right] \ . \tag{38}$$

□

Proof of Item (ii). The minimizer for a given (x_t, t), removing these elements from the notation for conciseness and abstraction, is a weighted mean:

$$\hat{u}^\star(x_t, t) = \hat{u}^\star = \sum_{l=1}^{n} w^{(l)} u^{(l)} \ \ , \text{ with} \tag{39}$$

$$w^{(l)} = \hat{p}_{\text{data}}(z = x^{(l)} | t, x_t) \ \ , \ \sum_{l=1}^{n} w^{(l)} = 1 \tag{40}$$

$$u^{(l)} = u^{\text{cond}}(x_t, x^{(l)}, t) \tag{41}$$

We express a mini-batch as an M-valued vector of indices, $\boldsymbol{i} \in [\![1, n]\!]^M$. The mini-batch estimate from Equation (7), considering the definition of the $\mathrm{softmax}$, can be expressed as a mini-batch weighted-mean:

$$\hat{u}^\star_M(\mathbf{i}) = \frac{\sum_{j=1}^{M} w^{(\mathbf{i_j})} u^{(\mathbf{i_j})}}{\sum_{j=1}^{M} w^{(\mathbf{i_j})}} \tag{42}$$

The categorical distribution over $[\![1, n]\!]$ with probabilities following the weights w in (40) is denoted $\mathrm{Cat}(w)$ and the uniform distribution, *i.e.,* $\mathrm{Cat}(\mathbf{1}/n)$), is denoted Unif.

The main result of the following is that, in expectation over the biased-mini-batches, **where the first point is drawn according to** w and the $M-1$ other points are drawn uniformly, the mini-batch weighted-mean is an unbiased estimate of the w-weighted-mean $\hat{u}^\star$.

$$\mathbb{E}\left[\hat{u}^\star_M(\mathbf{i})\right] := \mathbb{E}_{\mathbf{i_1} \sim \mathrm{Cat}(w)} \mathbb{E}_{\boldsymbol{i}_2, \dots, \boldsymbol{i}_M \sim \mathrm{Unif}} \left[\hat{u}^\star_M(\mathbf{i})\right] \tag{43}$$

$$= \sum_{\boldsymbol{i}_1 = 1}^{n} w^{(\boldsymbol{i}_1)} \mathbb{E}_{\boldsymbol{i}_2, \dots, \boldsymbol{i}_M \sim \mathrm{Unif}} \left[\hat{u}^\star_M(\mathbf{i})\right] \tag{44}$$

$$= \sum_{\boldsymbol{i}_1 = 1}^{n} \mathbb{E}_{\boldsymbol{i}_2, \dots, \boldsymbol{i}_M \sim \mathrm{Unif}} \left[w^{(\boldsymbol{i}_1)} \hat{u}^\star_M(\mathbf{i})\right] \tag{45}$$

$$= n \sum_{\boldsymbol{i}_1 = 1}^{n} \frac{1}{n} \mathbb{E}_{\boldsymbol{i}_2, \dots, \boldsymbol{i}_M \sim \mathrm{Unif}} \left[w^{(\boldsymbol{i}_1)} \hat{u}^\star_M(\mathbf{i})\right] \tag{46}$$

$$= n \ \mathbb{E}_{\boldsymbol{i}_1 \sim \mathrm{Unif}} \mathbb{E}_{\boldsymbol{i}_2, \dots, \boldsymbol{i}_M \sim \mathrm{Unif}} \left[w^{(\boldsymbol{i}_1)} \hat{u}^\star_M(\mathbf{i})\right] \tag{47}$$

$$= n \ \mathbb{E}_{\boldsymbol{i}_1, \dots, \boldsymbol{i}_M \sim \mathrm{Unif}} \left[w^{(\boldsymbol{i}_1)} \hat{u}^\star_M(\mathbf{i})\right] \tag{48}$$

The expression in Equation (48) is invariant with respect to order of the indices $\boldsymbol{i}_1, \dots, \boldsymbol{i}_M$: the indices in expectation in Equation (48) can be exchanged, and one thus has

$$\forall k \in [\![1, M]\!], \ \mathbb{E}\left[\hat{u}^\star_M(\mathbf{i})\right] = n \ \mathbb{E}_{\boldsymbol{i}_1, \dots, \boldsymbol{i}_M \sim \mathrm{Unif}} \left[w^{(\boldsymbol{i}_k)} \hat{u}^\star_M(\mathbf{i})\right] \ \ . \tag{49}$$

Averaging Equation (49) over the indices $k \in [\![1, M]\!]$ yields the desired result

$$\frac{1}{M}\sum_{k=1}^{M}\mathbb{E}\hat{u}^{\star}_{M}(\mathbf{i}) = \frac{1}{M}\sum_{k=1}^{M} n\ \mathbb{E}_{\boldsymbol{i}_1,\ldots,\boldsymbol{i}_M\sim\text{Unif}}\left[w^{(\boldsymbol{i}_k)}\hat{u}^{\star}_{M}(\mathbf{i})\right] \tag{50}$$

$$\mathbb{E}\hat{u}^{\star}_{M}(\mathbf{i}) = \frac{1}{M}n\ \mathbb{E}_{\boldsymbol{i}_1,\ldots,\boldsymbol{i}_M\sim\text{Unif}}\left[\sum_{k=1}^{M} w^{(\boldsymbol{i}_k)}\hat{u}^{\star}_{M}(\mathbf{i})\right] \tag{51}$$

$$= \frac{1}{M}n\ \mathbb{E}_{\boldsymbol{i}_1,\ldots,\boldsymbol{i}_M\sim\text{Unif}}\left[\sum_{k=1}^{M} w^{(\boldsymbol{i}_k)}\frac{\sum_{j=1}^{M} w^{(\mathbf{i_j})}u^{(\mathbf{i_j})}}{\sum_{j=1}^{M} w^{(\mathbf{i_j})}}\right] \tag{52}$$

$$= \frac{1}{M}n\ \mathbb{E}_{\boldsymbol{i}_1,\ldots,\boldsymbol{i}_M\sim\text{Unif}}\left[\left(\sum_{k=1}^{M} w^{(\boldsymbol{i}_k)}\right)\frac{\sum_{j=1}^{M} w^{(\mathbf{i_j})}u^{(\mathbf{i_j})}}{\left(\sum_{j=1}^{M} w^{(\mathbf{i_j})}\right)}\right] \tag{53}$$

$$= \frac{1}{M}n\ \mathbb{E}_{\boldsymbol{i}_1,\ldots,\boldsymbol{i}_M\sim\text{Unif}}\left[\sum_{j=1}^{M} w^{(\mathbf{i_j})}u^{(\mathbf{i_j})}\right] \tag{54}$$

$$= \frac{1}{M}n\ \sum_{j=1}^{M}\mathbb{E}_{\boldsymbol{i}_1,\ldots,\boldsymbol{i}_M\sim\text{Unif}}\left[w^{(\mathbf{i_j})}u^{(\mathbf{i_j})}\right] \tag{55}$$

$$= \frac{1}{M}n\ \sum_{j=1}^{M}\mathbb{E}_{\boldsymbol{i}_j\sim\text{Unif}}\left[w^{(\mathbf{i_j})}u^{(\mathbf{i_j})}\right] \tag{56}$$

$$= \frac{1}{M}n\ M\ \mathbb{E}_{l\sim\text{Unif}}\left[w^{(l)}u^{(l)}\right] \tag{57}$$

$$= n\ \mathbb{E}_{l\sim\text{Unif}}\left[w^{(l)}u^{(l)}\right] \tag{58}$$

$$= n\ \sum_{l=1}^{n}\frac{1}{n}\left[w^{(l)}u^{(l)}\right] \tag{59}$$

$$= \sum_{l=1}^{n}\left[w^{(l)}u^{(l)}\right] \tag{60}$$

$$= \hat{u}^{\star} \tag{61}$$

$\square$

Proof of Item (iii). Using the same ideas as for Item (ii), one has

$$\mathbb{E}_{x^{(1)}\sim\hat{p}_{\mathrm{data}}(\cdot|x_t,t)\,;\,b^{(2)},\ldots,b^{(M)}\sim\hat{p}_{\mathrm{data}}}\left[\hat{u}^{\star}_M(x_t,t)^2\right] \tag{62}$$

$$= n\mathbb{E}_{\boldsymbol{i}_1,\ldots,\boldsymbol{i}_M\sim\mathrm{Unif}}\left[w^{(\boldsymbol{i}_1)}\hat{u}^{\star}_M(\boldsymbol{i})^2\right] \tag{63}$$

$$= n\mathbb{E}_{\boldsymbol{i}_1,\ldots,\boldsymbol{i}_M\sim\mathrm{Unif}}\left[w^{(\boldsymbol{i}_k)}\hat{u}^{\star}_M(\boldsymbol{i})^2\right], \forall k \in [\![1, M]\!] \tag{64}$$

$$= n\frac{1}{M}\mathbb{E}_{\boldsymbol{i}_1,\ldots,\boldsymbol{i}_M\sim\mathrm{Unif}}\left[\sum_{k=1}^{M} w^{(\boldsymbol{i}_k)}\hat{u}^{\star}_M(\boldsymbol{i})^2\right] \tag{65}$$

$$= n\frac{1}{M}\mathbb{E}_{\boldsymbol{i}_1,\ldots,\boldsymbol{i}_M\sim\mathrm{Unif}}\left[\sum_{k=1}^{M} w^{(\boldsymbol{i}_k)}\left(\frac{\sum_{j=1}^{M} w^{(\boldsymbol{i}_j)}u^{(\boldsymbol{i}_j)}}{\sum_{j=1}^{M} w^{(\boldsymbol{i}_j)}}\right)^2\right] \tag{66}$$

$$\leq n\frac{1}{M}\mathbb{E}_{\boldsymbol{i}_1,\ldots,\boldsymbol{i}_M\sim\mathrm{Unif}}\left[\sum_{k=1}^{M} w^{(\boldsymbol{i}_k)}\frac{\sum_{j=1}^{M} w^{(\boldsymbol{i}_j)}(u^{(\boldsymbol{i}_j)})^2}{\sum_{j=1}^{M} w^{(\boldsymbol{i}_j)}}\right] \text{ by convexity of } x \mapsto x^2 \tag{67}$$

$$= n\frac{1}{M}\mathbb{E}_{\boldsymbol{i}_1,\ldots,\boldsymbol{i}_M\sim\mathrm{Unif}}\left[\left(\sum_{k=1}^{M} w^{(\boldsymbol{i}_k)}\right)\frac{\sum_{j=1}^{M} w^{(\boldsymbol{i}_j)}(u^{(\boldsymbol{i}_j)})^2}{\sum_{j=1}^{M} w^{(\boldsymbol{i}_j)}}\right] \tag{68}$$

$$= n\frac{1}{M}\mathbb{E}_{\boldsymbol{i}_1,\ldots,\boldsymbol{i}_M\sim\mathrm{Unif}}\left[\sum_{j=1}^{M} w^{(\boldsymbol{i}_j)}(u^{(\boldsymbol{i}_j)})^2\right] \tag{69}$$

$$= \mathbb{E}_{\boldsymbol{i}_1\sim\mathrm{Unif}}\left[w^{(\boldsymbol{i}_1)}(u^{(\boldsymbol{i}_1)})^2\right] \tag{70}$$

$$= \mathbb{E}_{l\sim\mathrm{Unif}}\left[w^{(l)}(u^{(l)})^2\right] \quad . \tag{71}$$

Hence

$$\mathbb{E}_{x^{(1)}\sim\hat{p}_{\mathrm{data}}(\cdot|x_t,t)\,;\,b^{(2)},\ldots,b^{(M)}\sim\hat{p}_{\mathrm{data}}}\left[\hat{u}^{\star}_M(x_t,t)^2\right] - (\hat{u}^{\star})^2 \leq \mathbb{E}_{l\sim\mathrm{Unif}}\left[w^{(l)}(u^{(l)})^2\right] - (\hat{u}^{\star})^2 \quad , \tag{72}$$

which is exactly

$$\mathrm{Var}_{x^{(1)}\sim\hat{p}_{\mathrm{data}}(\cdot|x_t,t)\,;\,b^{(2)},\ldots,b^{(M)}\sim\hat{p}_{\mathrm{data}}}\left[\hat{u}^{\star}_M(x_t,t)\right] \leq \mathrm{Var}_{x^{(1)}\sim\hat{p}_{\mathrm{data}}(\cdot|x_t,t)}\left[u^{\mathrm{cond}}(x_t,x^{(1)},t)\right] \quad . \tag{73}$$

□

C Experiments details

For all the experiment we used all the same learning hyperparameters, the default ones form Tong et al. (2024). The hyperparameter values are summarized in Table 1. The details specific to each figure are described in Appendices C.1 to C.4

# Channels	Batch Size	Learning Rate	EMA Decay	Gradient Clipping
128	128	0.0002	0.9999	1

Table 1: Learning hyperparameters for all the CIFAR-10 and CelebA 64 experiments.

C.1 Figures 1b and 2

For Figure 1b no deep learning is involved: the datasets 2-moons and CIFAR-10 are loaded. Then, 256 points from $p_0 \times \hat{p}_{\mathrm{data}}$ are drawn, and one computes the mean of the cosine similarities between $\hat{u}^{\star}((1-t)x_0 + tx_1, t)$ and $u^{\mathrm{cond}}((1-t)x_0 + tx_1, z = x_1, t) = x_1 - x_0$, for each value of $t \in \{0, 1/100, 2/100, \ldots, 99/100\}$.

No deep learning either is involved in Figure 2: the Imagenette dataset is loaded and spatially subsampled to resolution $\dim = 8$, $\dim = 16$, $\dots$, $\dim = 256$, *i.e.,* with $d = " \cdot 8^2$, $d = 3 \cdot 16^2, \dots, d = 3 \cdot 256^2$. Then, as for Figure 1b, batches of 256 points from p_0 and p_{data} are drawn, and one computes the percentage of cosine similarities between $\hat{u}^\star((1-t)x_0 + tx_1, t)$ and $u^{\text{cond}}((1-t)x_0 + tx_1, z = x_1, t) = x_1 - x_0$, that are larger than 0.9, for multiple time values t.

C.2 Figure 3

In Figure 3, networks are trained with a vanilla conditional flow matching, with the standard 34 million parameters U-Net for diffusion by Nichol and Dhariwal (2021), with default settings from the `torchfm` codebase [3] (Tong et al., 2024). Training uses the CFM loss. For this specific experiment, **we removed the usual random flip transform**, for $\hat{u}^\star$ to be simpler and easier to estimate by u_θ. For each "data" subsampling of the dataset, we trained the model for $5 \cdot 10^4$ iterations, with a batch size of 128, *i.e.,* we trained the models for 128 epochs.

C.3 Figure 4

In Figure 4, for each dataset (CIFAR-10 and CelebA 64), one network is trained using a vanilla conditional flow matching with the default parameters of Tong et al. (2024) (the most important ones are recalled in Table 1). Then images are generated first following the closed-form formula of the optimal velocity field $\hat{u}^\star$ from 0 to τ. And then following the velocity field learned with a usual conditional flow matching u_θ from τ to 1.

C.4 Figure 5

For experiments involving training on CIFAR-10 (Figures 3 and 4), we rely on the standard 34 million parameters U-Net for diffusion by Nichol and Dhariwal (2021), with default settings from the `torchfm` codebase (Tong et al., 2024). For each algorithm, the networks are trained for 500k iterations with batch size 128, *i.e.,* 1280 epochs.

For CelebA 64×64 (Figure 4), we rely on the training script of `pnpflow` library[4] (Martin et al., 2025), which uses a U-Net from Huang et al. (2021); Ho et al. (2020b).

[3] https://github.com/atong01/conditional-flow-matching
[4] https://github.com/annegnx/PnP-Flow