Rectified Flow

Pramook Khungurn

October 20, 2023

Rectified flow [LGL22] is a new approach to generative modeling that is quite similar to diffusion models [HJA20] but has a totally different formulation.

1 Generative Modeling Through the Optimal Transport Lense

- Generative modeling can be viewed generating data from noise.
  - We have a distribution π_{data} that we want to sample from.
  - We also have a well-known distribution of noise, say $\mathcal{N}(\mathbf{0}, I)$.
  - We want to learn function $\mathbf{f}$ such that, if $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, I)$, then $\mathbf{X} = \mathbf{f}(\mathbf{Z}) \sim \pi_{\mathrm{data}}$.
- This is a special case of transformations between two different probability density π_0 and π_1.
  - Formally, this is the problem of finding a function $\mathbf{f}$ such that $\mathbf{X}_1 = \mathbf{f}(\mathbf{X}_0) \sim \pi_1$ given that $\mathbf{X}_0 \sim \pi_0$.
- A related concept to transformations between densities are coupling of random variables.
  - **Definition 1.** *Let μ_1 and μ_2 be probability measures on the same measurable space $(\Omega, \mathcal{F})$. A* **coupling** *of μ_1 and μ_2 is a probability measure ν on the product space $(\Omega \times \Omega, \mathcal{F} \times \mathcal{F})$ such that the materials of ν coincide with μ_1 and μ_2. In other words,*

    $$\begin{aligned} \nu(A \times \Omega) &= \mu_1(A), \\ \nu(\Omega \times A) &= \mu_2(A) \end{aligned}$$

    for all $A \in \mathcal{F}$.
  - Recall that a random variable is a function from $X : \Omega \to \Omega'$ where $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ are measurable spaces. The **law of** X, denoted by μ_X, is the probability measure on $(\Omega', \mathcal{F}')$ defined by

    $$\mu_X(B) = \mu(X^{-1}(B))$$

    for all $B \in \mathcal{F}'$.
  - **Definition 2.** *For two random variables taking values in $(\Omega, \mathcal{F})$, a* **coupling of** X_1 **and** X_2 *is a joint random variable (Y_1, Y_2) on $(\Omega \times \Omega, \mathcal{F} \times \mathcal{F})$ such that its law as a probability measure is a coupling of the laws of X_1 and X_2.*
  - Recall that the, for a random variable X, its probability density function is the function $\pi_X : \Omega' \to [0, \infty)$ such that, for all $B \in \mathcal{F}'$, we have that

    $$\mu_X(B) = \int_B \pi_X(x)\, \mathrm{d}\mu'(x).$$

- For a joint random variable (Y_1, Y_2), the marginal probability density functions are given by

$$\pi_{Y_1}(y_1) = \int_{y_2 \in \Omega'} \pi_{(Y_1,Y_2)}(y_1, y_2)\, \mathrm{d}\mu'(y_2)$$
$$\pi_{Y_2}(y_2) = \int_{y_1 \in \Omega'} \pi_{(Y_1,Y_2)}(y_1, y_2)\, \mathrm{d}\mu'(y_1).$$

  where $\pi_{(Y_1,Y_2)}(y_1, y_2)$ is the probability density of the joint distribution.

- When (Y_1, Y_2) is a coupling of X_1 and X_2, it follows that $\mu_{Y_1} = \mu_{X_1}$, $\pi_{Y_1} = \pi_{X_1}$, $\mu_{Y_2} = \mu_{X_2}$, and $\pi_{Y_2} = \pi_{X_2}$, which is to say that all the marginal laws and distributions agree.

- Finding transformations between two distributions is a special case of finding a coupling between random variables.
  - Let us say that $\mathbf{X}_1 \in \mathbb{R}^d$ is a random variable with density $\pi_{X_1} = \pi_1$, and $\mathbf{X}_2 \in \mathbb{R}^d$ is another random variable with density $\pi_{X_2} = \pi_2$.
  - Finding a transformation from π_1 to π_2 is equivalent to finding a coupling (Y_1, Y_2) of X_1 and X_2 that is **causal**. In other words, there is a function $\mathbf{f} : \mathbb{R}^d \to \mathbb{R}^d$ such that $Y_2 = \mathbf{f}(Y_1)$.

- The **optimal transport problem** is the task of finding a coupling (Y_1, Y_2) of $X_1 \sim \pi_1$ and $X_2 \sim \pi_2$ that minimizes

$$E[c(Y_1 - Y_2)]$$

  for some cost function $c : \mathbb{R}^d \to \mathbb{R}$.

- For the generative modeling task, the cost is not of primary concern, but it can give some desirable properties to be worked towards.

2 Rectified Flow

- We return to the distribution transformation problem where we are given two empirical distributions π_0 and π_1.

- An idea that has been popular circa 2023 is create a stochastic process $\{\mathbf{Z}(t) : 0 \leq t \leq 1\}$ that is governed by an ODE

$$\frac{\mathrm{d}\mathbf{Z}}{\mathrm{d}t} = \mathbf{v}(\mathbf{Z}, t) \tag{1}$$

  so that $\mathbf{Z}(0) \sim \pi_0$ and $\mathbf{Z}_1(1) \sim \pi_1$. If we can find this velocity field $\mathbf{v}(\mathbf{Z}, t)$, then we can generate samples according to π_1 by first sampling $\mathbf{Z}(0)$ according to π_0, then compute

$$\mathbf{Z}(1) = \mathbf{Z}(0) + \int_0^1 \mathbf{v}(\mathbf{Z}(t), t)\, \mathrm{d}t,$$

  which can be approximated with any integration scheme.

- The above idea is realized by the probability flow ODE formulation of score-based models [SSDK+21] and DDIM [SME20]. However, π_0 is fixed to the Gaussian distribution $\mathcal{N}(\mathbf{0}, I)$.

- The rectified flow appraoch starts at a stochastic process that has the right boudnary condition but doesn't quite work the way we wants.

- Given two random variables $\mathbf{X}_0 \in \mathbb{R}^d$ and $\mathbf{X}_1 \in \mathbb{R}^d$, their **linear interpolation** is a stochastic process $\{\overline{\mathbf{X}}(t) : 0 \le t \le 1\}$ where $\overline{\mathbf{X}} = (1-t)\mathbf{X}_0 + t\mathbf{X}_1$.

- If we require that $\mathbf{X}_0 \sim \pi_0$ and $\mathbf{X}_1 \sim \pi_1$, then $\overline{\mathbf{X}}(t)$ has the right boundary conditions. However, the problem with $\overline{\mathbf{X}}$ is that it is not causal. Given $\overline{\mathbf{X}}(0)$, we cannot find out which way $\overline{\mathbf{X}}(t)$ is going to evolve without knowing how to precisely sample the destination point from π_1. The velocity field in this case is not deterministic.

- However, suppose we have sampled the destination point $\mathbf{X}_1$. The path traces by $\overline{\mathbf{X}}(t)$ is a straight line in $\mathbb{R}^d$. It's time derivative is a random variable given by

$$\dot{\overline{\mathbf{X}}} = \frac{\mathrm{d}\overline{\mathbf{X}}}{\mathrm{d}t} = \mathbf{X}_1 - \mathbf{X}_0.$$

  We have that $\overline{\mathbf{X}}_t$ is *path-wise continuously differentiable.*

- For a path-wise continously differentiable stochastic process like $\overline{\mathbf{X}}$, it is possible to define the notation of the "expected velocity field."

  Definition 3. *For a path-wise continuously differentiable stochatistic process* $\{\mathbf{X}(t) : 0 \le t \le 1\}$*, its* **expected velocity field** $\mathbf{v}^{\mathbf{X}}$*, is defined as*

$$\mathbf{v}^{\mathbf{X}}(t, \mathbf{x}) = E[\dot{\mathbf{X}}(t) \,|\, \mathbf{X}(t) = \mathbf{x}].$$

  That is $\mathbf{v}^{\mathbf{X}}(t, \mathbf{x})$ *is the expected value of the velocities of the paths that pass point* $\mathbf{x}$ *at time* t*.*

- The idea is the rectified flow approach is this: use the expected velocity field $\mathbf{v}^{\overline{\mathbf{X}}}$ of the linear interpolation $\overline{\mathbf{X}}$ of $\mathbf{X}_0 \sim \pi_0$ and $\mathbf{X}_1 \sim \pi_1$ as the velocity field in Equation (1). If you do this, the marginal distributions would be right.

- Now, let us defined precisely what a rectified flow is.

  Definition 4. *A path-wise continously differentiable stochastic process* $\{\mathbf{X}(t) : 0 \le t \le 1\}$ *is* **rectifiable** *if* $\mathbf{v}^{\mathbf{X}}$ *is locally bounded, and the solution of the integral equation below exists and is unique:*

$$\mathbf{Z}(t) = \mathbf{X}(0) + \int_0^t \mathbf{v}^{\mathbf{X}}(u, \mathbf{Z}(u))\,\mathrm{d}u.$$

  for all $t \in [0,1]$*. In this case, the stochastic process* $\{\mathbf{Z}(t) : 0 \le t \le 1\}$ *is called the* **rectified flow induced from X**.

- Note that the rectified flow $\mathbf{Z}(t)$ is the unique solution to the following initial value problem:

$$\begin{aligned} \mathbf{Z}(0) &\sim \mathbf{X}(0), \\ \frac{\mathrm{d}\mathbf{Z}(t)}{\mathrm{d}t} &= \mathbf{v}^{\mathbf{X}}(t, \mathbf{Z}(t)). \end{aligned} \tag{2}$$

- **Theorem 5.** *Let* $\mathbf{X}$ *be rectifiable, and* $\mathbf{Z}$ *be the rectified flow induced from* $\mathbf{X}$*. We have that the distributions of* $\mathbf{Z}(t)$ *and* $\mathbf{X}(t)$ *agree. In other words,* $\pi_{\mathbf{Z}(t)} = \pi_{\mathbf{X}(t)}$ *for all* $t \in [0,1]$*.*

  Proof. We think of the distributions $\mathbf{X}(t)$ and $\mathbf{Z}(t)$ as functions of signature $\mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$. In other words, $\pi_{\mathbf{X}}(t, \mathbf{x}) := \pi_{\mathbf{X}(t)}(\mathbf{x})$ and $\pi_{\mathbf{Z}}(t, \mathbf{x}) := \pi_{\mathbf{Z}(t)}(\mathbf{x})$. We can show that these functions satisfy the **continuity equation** [Wik23],

$$\frac{\partial \pi(t, \mathbf{x})}{\partial t} + \nabla \cdot (\mathbf{v}^{\mathbf{X}}(t, \mathbf{x}) \pi(t, \mathbf{x})) = 0, \tag{3}$$

which comes from fluid dynamics. Here, the function we want to solve for is π, ahd the divergence operator $\nabla\cdot$ only applies to the spatial dimensions and not the time dimension. The proof of this assertion is available in Lemma 6 in the appendix.

According to Theorem 9 from Ambrosio and Crippa [AC08], the solution to (3) is unique if and only if the solution to the initial value problem (2) is unique. However, the solution to (2) is unique because we assumed that $\mathbf{X}$ is rectifiable. As a result, it follows that $\pi_{\mathbf{X}} = \pi_{\mathbf{Z}}$, which means that $\pi_{\mathbf{X}(t)} = \pi_{\mathbf{Z}(t)}$ for all $t \in [0, 1]$. □

- Suppose that the linear interpolation of $\overline{\mathbf{X}}(t)$ of $\mathbf{X}_0 \sim \pi_0$ and $\mathbf{X}_1 \sim \pi_1$ is rectifiable. Then, its rectified flow $\mathbf{Z}(t)$ has all the right marginal distributions. In particular, $\mathbf{Z}(0) \sim \pi_0$ and $\mathbf{Z}(1) \sim \pi_1$.
- To do generative modeling with rectified flow, we train a network $\mathbf{f}_{\boldsymbol{\theta}}(\cdot,\cdot)$ so that $\mathbf{f}_{\boldsymbol{\theta}}(t,\mathbf{x})$ approximates $\mathbf{v}^{\overline{\mathbf{X}}}(t,\mathbf{x})$. This can be done by minimizing the loss

  $$\mathcal{L}(\boldsymbol{\theta}) = E_{\substack{\mathbf{x}_0\sim\pi_0,\\ \mathbf{x}_1\sim\pi_1,\\ t\sim\mathcal{U}([0,1])}} \Big[\big\| \mathbf{f}_{\boldsymbol{\theta}}\big(t, (1-t)\mathbf{x}_0 + t\mathbf{x}_1\big) - (\mathbf{x}_1 - \mathbf{x}_0) \big\|^2 \Big].$$

  Then, to generate a sample from π_1, we first sample $\mathbf{x}_0 \sim \pi_0$, then we compute

  $$\mathbf{x}_1 = \mathbf{x}_0 + \int_0^1 \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)\,\mathrm{d}t$$

  with an integrator.
- Rectified flow bears a lot of similarity to DDIM, but it is different in multiple ways.
  - Its formulation is a lot simpler.
  - It does not have any hyperparameters related to how the two distributions are interpolated.
    * Linear interpolation is chosen as the first case study. However, other types of interpolation is viable too as long as the paths are differentiable.
  - The starting distribution π_0 for rectified flow does not have to be a Gaussian distribution like in DDIM.

A Proofs

- **Lemma 6.** *Let $\mathbf{X}$ be rectifiable, and $\mathbf{Z}$ be the rectified flow induced from $\mathbf{X}$. The distribution functions $\pi_{\mathbf{X}}$ and $\pi_{\mathbf{Z}}$ satisfy the continuity equation* (3).

  Proof. Let $h : \mathbb{R}^d \to \mathbb{R}$ be a compactly supported differentiable function. We have that

  $$\frac{\mathrm{d}}{\mathrm{d}t} E[h(\mathbf{X}(t))] = E\bigg[\frac{\mathrm{d}h(\mathbf{X}(t))}{\mathrm{d}t}\bigg] = E\bigg[\sum_{i=1}^{d} \frac{\partial h}{\partial x_i}\frac{\mathrm{d}X_i}{\mathrm{d}t}\bigg] = E[\nabla h(\mathbf{X}(t))^T \dot{\mathbf{X}}(t)].$$

  Here, ∇h denotes thes gradient vector $(\partial h/\partial x_1, \partial h/\partial x_2, \dots, \partial h/\partial x_d)$. Using the law of iterated expectation, we can replace $\dot{\mathbf{X}}(t)$ in the above equation with $\mathbf{v}^{\mathbf{X}}(t,\mathbf{X}(t)) = E[\dot{\mathbf{X}} \,|\, \mathbf{X}(t)]$. This gives

  $$\frac{\mathrm{d}}{\mathrm{d}t} E[h(\mathbf{X}(t))] = E[\nabla h(\mathbf{Z}(t))^T \mathbf{v}^{\mathbf{X}}(t, \mathbf{X}(t))].$$

  We also have that $\mathbf{Z}(t)$ satisfies

  $$\frac{\mathrm{d}}{\mathrm{d}t} E[h(\mathbf{Z}(t))] = E[\nabla h(\mathbf{X}(t))^T \mathbf{v}^{\mathbf{X}}(t, \mathbf{Z}(t))].$$

The reasoning is the same as what we did for $\mathbf{X}(t)$ but simpler. This is because $\dot{\mathbf{Z}}(t) = \mathbf{v}^{\mathbf{X}}(t, \mathbf{X}(t))$ by construction.

Now, both $\mathbf{X}(t)$ and $\mathbf{Z}(t)$ are examples of random variables $\mathbf{Y}(t)$ that satisfy the equation

$$\frac{\mathrm{d}}{\mathrm{d}t} E[h(\mathbf{Y}(t))] = E[\nabla h(\mathbf{Y}(t))^T \mathbf{v}^{\mathbf{X}}(t, \mathbf{Y}(t))]$$

for any compactly supported differentiable function h. Setting $\mathbf{f}(t, \mathbf{y}) := \mathbf{v}^{\mathbf{X}}(t, \mathbf{y})$ and using Lemma 7, we can conclude that the following equation must hold

$$\frac{\partial \pi_{\mathbf{Y}}(t, \mathbf{y})}{\mathrm{d}t} + \nabla \cdot (\mathbf{v}^{\mathbf{X}}(t, x) \pi_{\mathbf{Y}}(t, \mathbf{y})) = 0,$$

and we can substitute $\mathbf{Y}$ with either $\mathbf{X}$ or $\mathbf{Z}$. We are done. $\square$

- **Lemma 7.** *Let $\mathbf{f} : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$ be any differentiable time-dependent vector field, and $\mathbf{Y}(t)$ be a stochastic process with differentiable paths. Assume that, for any $h : \mathbb{R}^d \to \mathbb{R}$ be any any compactly supported differentiable function, we have that*

  $$\frac{\mathrm{d}}{\mathrm{d}t} E[h(\mathbf{Y}(t))] = E[\nabla h(\mathbf{Y}(t))^T \mathbf{f}(t, \mathbf{Y}(t))].$$

  Then, it holds that

  $$\frac{\partial \pi_{\mathbf{Y}}(t, \mathbf{y})}{\mathrm{d}t} + \nabla \cdot (\mathbf{f}(t, x) \pi_{\mathbf{Y}}(t, \mathbf{y})) = 0 \tag{4}$$

  where $\pi_Y(t, \cdot)$ is defined to be the probability distribution $\pi_{\mathbf{Y}(t)}(\cdot)$.

  Proof. Let $h : \mathbb{R}^d \to \mathbb{R}$ be any any compactly supported differentiable function. We have that $\frac{\mathrm{d}}{\mathrm{d}t} E[h(\mathbf{Y}(t))] = E[\nabla h(\mathbf{Y}(t))^T \mathbf{f}(t, \mathbf{Y}(t))]$. This means that

  $$\frac{\mathrm{d}}{\mathrm{d}t} E[h(\mathbf{Y}(t))] - E[\nabla h(\mathbf{Y}(t))^T \mathbf{f}(t, \mathbf{Y}(t))] = 0.$$

  Our task now is to rewrite the two terms above so that the equation becomes more similar to Equation 4.

  First, using Lemma 8, we have that

  $$\frac{\mathrm{d}}{\mathrm{d}t} E[h(\mathbf{Y}(t))] = \int_{\mathbb{R}^d} h(\mathbf{y}) \frac{\partial \pi_{\mathbf{Y}}(t, \mathbf{y})}{\partial t} \, \mathrm{d}\mathbf{y}.$$

  Next, we have that

  $$-E[\nabla h(\mathbf{Y}(t))^T \mathbf{f}(t, \mathbf{Y}(t))] = -\int_{\mathbb{R}^d} \nabla h(\mathbf{y})^T \mathbf{f}(t, \mathbf{y}) \pi_{\mathbf{Y}}(t, \mathbf{y}) \, \mathrm{d}\mathbf{y}.$$

  Fixing t and setting $\mathbf{F}(t, \mathbf{x}) := \mathbf{f}(t, \mathbf{y}) \pi_{\mathbf{Y}}(t, \mathbf{y})$, we can use Lemma 9 to deduce that

  $$-\int_{\mathbb{R}^d} \nabla h(\mathbf{y})^T \mathbf{f}(t, \mathbf{y}) \pi_{\mathbf{Y}}(t, \mathbf{y}) \, \mathrm{d}\mathbf{y} = \int_{\mathbb{R}^d} h(\mathbf{y}) \nabla \cdot (\mathbf{f}(t, \mathbf{y}) \pi_{\mathbf{Y}}(t, \mathbf{y})) \, \mathrm{d}\mathbf{y}.$$

  These derivation gives

  $$\begin{aligned} \int_{\mathbb{R}^d} h(\mathbf{y}) \frac{\partial \pi_{\mathbf{Y}}(t, \mathbf{y})}{\partial t} \, \mathrm{d}\mathbf{y} + \int_{\mathbb{R}^d} h(\mathbf{y}) \nabla \cdot (\mathbf{f}(t, \mathbf{y}) \pi_{\mathbf{Y}}(t, \mathbf{y})) \, \mathrm{d}\mathbf{y} = 0 \\ \int_{\mathbb{R}^d} h(\mathbf{y}) \bigg(\frac{\partial \pi_{\mathbf{Y}}(t, \mathbf{y})}{\partial t} + \nabla \cdot (\mathbf{f}(t, \mathbf{y}) \pi_{\mathbf{Y}}(t, \mathbf{y})) \bigg) \, \mathrm{d}\mathbf{y} = 0. \end{aligned} \tag{5}$$

Note that Equation 5 is true for all h that is differentiable and compactly supported. This means that we can set h to be an arbitrary narrow peak around any point $\mathbf{y} \in \mathbb{R}^d$ of interest. Thus, if there is any point $\mathbf{y}$ around which $\frac{\partial \pi_{\mathbf{Y}}(t,\mathbf{y})}{\partial t} + \nabla \cdot (\mathbf{f}(t,\mathbf{y})\pi_{\mathbf{Y}}(t,\mathbf{y})) \neq 0$, then Equation 5 would not hold. It follows that

$$\frac{\partial \pi_{\mathbf{Y}}(t,\mathbf{y})}{\partial t} + \nabla \cdot (\mathbf{f}(t,\mathbf{y})\pi_{\mathbf{Y}}(t,\mathbf{y})) = 0$$

as required. □

- **Lemma 8.** *For any compactly supported and continuous function $h : \mathbb{R}^d \to \mathbb{R}$,*

$$\frac{\mathrm{d}}{\mathrm{d}t} E[h(\mathbf{Y}(t))] = \int_{\mathbb{R}^d} h(\mathbf{y}) \frac{\partial \pi_{\mathbf{Y}}(t,\mathbf{y})}{\partial t} \,\mathrm{d}\mathbf{y}.$$

Proof. We have that

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} E[h(\mathbf{Y}(t))] &= \frac{\mathrm{d}}{\mathrm{d}t}\bigg(\int_{\mathbb{R}^d} h(\mathbf{y})\pi_{\mathbf{Y}}(t,\mathbf{y}) \,\mathrm{d}\mathbf{y} \bigg) \\
&= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \bigg(\int_{\mathbb{R}^d} h(\mathbf{y})\pi_{\mathbf{Y}}(t+\Delta t,\mathbf{y}) \,\mathrm{d}\mathbf{y} - \int_{\mathbb{R}^d} h(\mathbf{y})\pi_{\mathbf{Y}}(t,\mathbf{y}) \,\mathrm{d}\mathbf{y} \bigg) \\
&= \int_{\mathbb{R}^d} h(\mathbf{y}) \bigg(\lim_{\Delta t \to 0} \frac{\pi_{\mathbf{Y}}(t+\Delta t,\mathbf{y}) - \pi_{\mathbf{Y}}(t,\mathbf{y})}{\Delta t} \bigg) \,\mathrm{d}\mathbf{y} \\
&= \int_{\mathbb{R}^d} h(\mathbf{y}) \frac{\partial \mu_{\mathbf{Y}}(t,\mathbf{y})}{\partial t} \,\mathrm{d}\mathbf{y}
\end{aligned}$$

as required. □

- **Lemma 9.** *Let $h : \mathbb{R}^d \to \mathbb{R}$ be a compactly supported differentiable function. Let $\mathbf{F} : \mathbb{R}^d \to \mathbb{R}^d$ be a continuously differentiable vector field. We have that*

$$\int_{\mathbb{R}^d} h(\mathbf{y}) \nabla \cdot \mathbf{F}(\mathbf{y}) \,\mathrm{d}\mathbf{y} = -\int_{\mathbb{R}^d} \nabla h(\mathbf{y})^T \mathbf{F}(\mathbf{y}) \,\mathrm{d}\mathbf{y}.$$

Proof. Let H be the support of h. We have that H is compact, and so it is closed and bounded. Expand H to $\widetilde{H}$ such that $\widetilde{H}$ has a piecewise smooth boudnary $\partial \widetilde{H}$ such that $h(\mathbf{y}) = 0$ for all $y \in \partial \widetilde{H}$.

Define $\mathbf{G}(\mathbf{y}) := h(\mathbf{y})\mathbf{F}(y)$. We have that $\mathbf{G}$ is a continuously differentiable vector field such that $\mathbf{G}(\mathbf{y}) = \mathbf{0}$ for all $\mathbf{y} \in (\mathbb{R}^d - \widetilde{H}) \cup \partial \widetilde{H}$. We have that

$$\int_{\mathbb{R}^d} \nabla \cdot \mathbf{G}(\mathbf{y}) \,\mathrm{d}\mathbf{y} = \int_{\widetilde{H}} \nabla \cdot \mathbf{G}(\mathbf{y}) \,\mathrm{d}\mathbf{y} = \int_{\partial \widetilde{H}} \mathbf{G}(\mathbf{y}) \cdot \hat{\mathbf{n}}(\mathbf{y}) \,\mathrm{d}\mathbf{y} = \int_{\partial \widetilde{H}} \mathbf{0} \cdot \hat{\mathbf{n}}(\mathbf{y}) \,\mathrm{d}\mathbf{y} = 0.$$

Here, $\hat{n}(\mathbf{y})$ denotes the unit normal vector to the boundary at $\mathbf{y}$. Note the equation

$$\int_{\widetilde{H}} \nabla \cdot \mathbf{G}(\mathbf{y}) \,\mathrm{d}\mathbf{y} = \int_{\partial \widetilde{H}} \mathbf{G}(\mathbf{y}) \cdot \hat{\mathbf{n}}(\mathbf{y}) \,\mathrm{d}\mathbf{y}$$

is the divergence theorem.

So, we have that

$$
\begin{aligned}
0 &= \int_{\mathbb{R}^d} \nabla \cdot (h(\mathbf{y})\mathbf{F}(\mathbf{y}))\, \mathrm{d}\mathbf{y} \\
&= \int_{\mathbb{R}^d} \sum_{i=1}^{d} \frac{\partial (hF_i)}{\partial y_i}\, \mathrm{d}\mathbf{y} \\
&= \int_{\mathbb{R}^d} \sum_{i=1}^{d} \left(\frac{\partial h}{\partial y_i} F_i + h \frac{\partial F_i}{\partial i} \right) \mathrm{d}\mathbf{y} \\
&= \int_{\mathbb{R}^d} \sum_{i=1}^{d} \frac{\partial h}{\partial y_i} F_i \, \mathrm{d}\mathbf{y} + \int_{\mathbb{R}^d} h \frac{\partial F_i}{\partial i}\, \mathrm{d}\mathbf{y} \\
&= \int_{\mathbb{R}^d} \nabla h(\mathbf{y})^T \mathbf{F}(\mathbf{y})\, \mathrm{d}y + \int_{\mathbb{R}^d} \mathbf{h}(\mathbf{y}) \nabla \cdot \mathbf{F}(\mathbf{y})\, \mathrm{d}\mathbf{y}.
\end{aligned}
$$

It follows that $\int_{\mathbb{R}^d} h(\mathbf{y}) \nabla \cdot \mathbf{F}(\mathbf{y})\, \mathrm{d}\mathbf{y} = -\int_{\mathbb{R}^d} \nabla h(\mathbf{y})^T \mathbf{F}(\mathbf{y})\, \mathrm{d}\mathbf{y}$. □

References

[AC08] Luigi Ambrosio and Gianluca Crippa. Existence, uniqueness, stability and differentiability properties of the flow associated to weakly differentiable vector fields. *In: Transport Equations and Multi-D Hyperbolic Conservation Laws, Lecture Notes of the Unione Matematica Italiana*, 5, 2008. cvgmt preprint.

[HJA20] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *CoRR*, abs/2006.11239, 2020.

[LGL22] Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow, 2022.

[SME20] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models, 2020.

[SSDK+21] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.

[Wik23] Wikipedia. Continuity equation. `https://en.wikipedia.org/wiki/Continuity_equation`, 2023. Accessed: 2023-10-20.