Foundations of continuity equations in neural contexts and their operator learning

Andrew Gracyk*

Fall 2025

Abstract

These are reading course notes emphasizing an understanding of gradient flow structure, focusing on the continuity equation, and its theoretical underpinnings that dominate the machine learning optimal transport archetype. We study neural variational PDEs, optimal transport, and connections with these areas to the operator learning archetype of machine learning. In particular, we will study background regarding Wasserstein gradient flows, Wasserstein geometry, and various sub-topics, such as JKO-algorithms.

Contents

1 Deriving the continuity equation

1.1 A more prototypical engineering derivation

Let $V \subseteq \mathbb{R}^d$ be a closed, connected, and bounded volume in Euclidean space and $\rho : \mathcal{V} \times [0, T] \to \mathbb{R}^+$ be a mass density. Let $v : \mathcal{V} \times [0, T] \to \mathbb{R}^d$ be a vector field. We have

$$\text{total mass} = \int_V \rho dV. \tag{1}$$

It is well known via flux theory in calculus that the change of mass through the surface is given by

$$\text{change of mass in the volume} = \int_{\partial V} \rho v \cdot n dS. \tag{2}$$

*Purdue Mathematics; a major thanks to Rongjie Lai for supervising this reading course

The mass is conserved in the volume, and we get

$$\partial_t \int_V \rho dV = -\int_{\partial V} \rho v \cdot n dS. \tag{3}$$

This is generally a well-known identity in engineering (see Leishman (2021)). Here, our surface integral is outward-oriented. We will assume basic regularity properties such as smoothness on ρ, and since our domain is sufficiently nice, we can exchange differentiation and integration. Hence,

$$\int_V \partial_t \rho dV + \int_{\partial V} \rho v \cdot n dS = 0. \tag{4}$$

We cannot combine the integrals into one equation yet due to the first being an integral over Euclidean space and the second a surface integral. Notice the criteria of the divergence theorem are satisfied (we will allow v to be sufficiently smooth), hence

$$\int_V \partial_t \rho dV + \int_V \nabla \cdot (\rho v) dV = \int_V (\partial_t \rho + \nabla \cdot (\rho v)) dV = 0. \tag{5}$$

Just because an integral is zero, that does not imply the integrand is zero. Moreover, ρ is not a (compactly supported) test function.

There are some ways to conclude the proof. For example, we have $\partial_t \rho + \nabla \cdot (\rho v)$ if we permit V to be arbitrary, and if the equation must hold for all V.

Let us justify the negative sign in equation 3 intuitively with a special case. Suppose V is a hypercube and suppose v is outward oriented for each face. Thus, mass is moving outside the volume. But the integral is negative since the inner product is positive (assuming v is not zero) and rho is nonnegative. But mass is leaving the surface, so the left is clearly negative. Thus, the negative sign is needed.

In the next subsection, we will establish an argument in the sense of distributions.

1.2 Continuation equation definition in the weak sense

Let m be a vector-valued measure such that $dm = \rho dx$. Here, ρ is also vector-valued. Let ϕ be a C^∞ compactly supported test function. It is known by integration by parts in the theory of distributions

$$\int \phi d(\nabla \cdot m) = -\int \nabla \phi \cdot dm. \tag{6}$$

There is an equivalence

$$\int \phi (\nabla \cdot \rho) dx = -\int \nabla \phi \cdot \rho dx. \tag{7}$$

By the conservation of mass equation we saw on the previous slide, but in the weak form

$$\partial_t \int \phi \rho_s dx = -\int \phi (\nabla \cdot \rho_s v) dx. \tag{8}$$

Here, $\rho = \rho_s v$ is the vector field multiplied by scalar density.

It is a known fact from PDE theory

$$\partial_t \int \phi d\alpha - \int \nabla \phi \cdot v d\alpha = 0 = \partial_t \int \phi d\alpha - \int \phi \nabla \cdot v d\alpha = 0 \implies \partial_t \alpha + \nabla \cdot (v\alpha) = 0. \tag{9}$$

The second equality is due to integration by parts. This is because ϕ is any (smooth compactly supported) function. We will denote $d\alpha = \rho_s dx$. Thus,

$$0 = -\int \phi (\nabla \cdot \rho_s v) dx - \int \nabla \phi \cdot \rho dx \tag{10}$$

$$= \partial_t \int \phi \rho_s dx - \int \nabla \phi \cdot v \rho_s dx \tag{11}$$

$$\implies \partial_t \rho_s + \nabla \cdot (v \rho_s) = 0, \tag{12}$$

and we have the result.

1.3 Theorem deriving the continuity equation

Let $T_t : \mathbb{R}^n \to \mathbb{R}^n$, $0 \leqslant t \leqslant 1$, be a locally Lipschitz family of diffeomorphisms (isomorphism between manifolds) with $T_0 = \text{id}$. Let $v(t, x)$ be the velocity field associated with the trajectories $T_t = X(t, x)$,

$$\partial_t X(t, x) = v(t, X(t, x)). \tag{13}$$

Let α_0 be a probability distribution on $\Omega \subseteq \mathbb{R}^n$ and $\alpha(t, \cdot) = T_t \# \alpha_0$. Then $\alpha(t, \cdot)$ is the unique solution of the following linear transport equation:

$$\begin{cases} \partial_t \alpha + \nabla \cdot (\alpha v) = 0, 0 < t < 1 \\ \alpha(0, \cdot) = \alpha_0. \end{cases} \tag{14}$$

Proof. We previously saw the weak formulation of the continuity equation is such that

$$\partial_t \int \phi \rho_s dx = -\int \phi (\nabla \cdot \rho_s v) dx \tag{15}$$

in the previous section. Note that this formulation is the conservation of mass, which roughly follows from the definition of flux. This conservation of mass equation has heavy flavor from engineering fields, but we will mostly accept it as true. We will adopt $\alpha = \rho_s dx$ to be our measure with respect to Lebesgue measure, so ρ_s is a scalar density. Hence, if we can check for $\varphi \in C_0^\infty(\Omega)$

$$\partial_t \int \phi d\alpha = -\int \phi (\nabla \cdot v) d\alpha, \tag{16}$$

then we have the continuity equation holds. In particular, most of our proof will be attempting to boil down what is given to this form. We will begin with the equivalence Villani (chapter 1)

$$\int_\Omega \varphi(y) dT_t \# \alpha_0 = \int_\Omega \varphi(T_t(x)) d\alpha_0. \tag{17}$$

Note that it is not necessarily the case the domain need be the same in this formulation, but we will operate so that they are equal. Moreover, it is known

$$\frac{\partial}{\partial t}(\varphi \circ T_t) = \nabla \varphi(T_t) \cdot \frac{\partial T_t}{\partial_t} = \nabla \varphi(T_t) \cdot v(t, T_t(x)). \tag{18}$$

The equivalence is by definition. As a reminder, this is a Euclidean inner product. Now, for all $h > 0$, we compute

$$\frac{1}{h}\left(\int_\Omega \varphi d \underbrace{\alpha(t+h, \cdot)}_{= T_{(t+h)} \# \alpha_0} - \int_\Omega \varphi d \underbrace{\alpha(t, \cdot)}_{T_t \# \alpha_0} \right) = \int_\Omega \frac{\varphi(T_{t+h}(x)) - \varphi(T_t(x))}{h} d\alpha_0, \tag{19}$$

which follows from 17. If we let $h \to 0$, then the above formulation is precisely a derivative, hence

$$\lim_{h \searrow 0} \int_\Omega \frac{\varphi(T_{t+h}(x)) - \varphi(T_t(x))}{h} d\alpha_0 = \frac{d}{dt} \int_\Omega \varphi d\alpha(t, \cdot), \tag{20}$$

where we have permitted regularity conditions. Moreover, from the chain rule of 18,

$$\frac{d}{dt} \int_\Omega \varphi d\alpha(t, \cdot) = \int_\Omega \nabla \varphi \cdot v d\alpha. \tag{21}$$

Note that is precisely conservation of mass. So here's what's important: *why doesn't conservation of mass always immediately follow from the chain rule?* This is because it only follows under our specific conditions, i.e. $\alpha_t = T_t \# \alpha_0$. By the result from section 1.2, we have the continuity equation holds.

There is more to prove. We will show uniqueness, i.e. if $\alpha_0 = 0$, then $\alpha(t, \cdot) = 0$ for all $0 < t < 1$. This shows uniqueness because we can take $\alpha_1 - \alpha_2 = 0$, and the result follows from well-posedness of the PDE. We examine the duality method. Assume that we can construct a function $\varphi(t, x)$ satisfying

$$\begin{cases} \frac{\partial \varphi}{\partial t} = -v \cdot \nabla \varphi \\ \varphi(s, \cdot) = \varphi_s, \ \forall s < 1 \text{ fixed}. \end{cases} \tag{22}$$

Here, φ is an arbitrary test function

$$
\begin{aligned}
\frac{d}{dt}\int_\Omega \varphi(t,x)d\alpha(t,\cdot) &\overset{\text{chain rule}}{=} \int_\Omega \frac{\partial\varphi}{\partial t}d\alpha(t,\cdot) + \int_\Omega \varphi(t,x)d\frac{\partial\alpha}{\partial t} && (23)\\
&= \underbrace{\int_\Omega (-v\cdot\nabla\varphi)d\alpha(t,\cdot)}_{\text{follows by assumption}} + \underbrace{\int_\Omega \varphi(t,x)d(-\nabla\cdot(v\alpha))}_{\text{follows from continuity equation}} && (24)\\
&= \int_\Omega (-v\cdot\nabla\varphi)d\alpha + \underbrace{\int_\Omega \nabla\varphi\cdot v d\alpha}_{\text{we refer to 6, which is an integration by parts for divergence of measures}} && (25)\\
&= 0. && (26)
\end{aligned}
$$

From this, we can conclude

$$
\int_{\mathbb{R}^n} \varphi(s,x)d\alpha(s,\cdot) = \int_{\mathbb{R}^n} \varphi(0,x)d\alpha_0 = 0 \text{ since } \alpha_0 \text{ is the trivial measure,} \tag{27}
$$

since the time derivative is zero, thus constant in time. Since φ was arbitrary, we conclude $\alpha(s,\cdot) = 0$, $0 \leqslant s \leqslant 1$.

From 22, we have using this and the chain rule

$$
\begin{aligned}
\partial_t = v\cdot\nabla\varphi = 0 &\implies \frac{d}{dt}\varphi(t,X(t,x)) = 0 && (28)\\
&\implies \varphi(t,T_t(x)) = \varphi(s,T_s(x)) && (29)\\
&\implies \varphi(t,x) = \varphi(s,T_s\circ T_t^{-1}(x)). && (30)
\end{aligned}
$$

□

Remark. If the trajectory has randomness, i.e.

$$
dX(t,x) = v(t,X(t,x))dt + \sqrt{2\epsilon}dW \tag{31}
$$

is obeyed, then the corresponding density follows

$$
\begin{aligned}
\partial_t\rho + \nabla\cdot(v\rho) = \epsilon\Delta\rho && (32)\\
\text{or equivalently} && (33)\\
\partial_t\rho + \sum_i \frac{\partial}{\partial x_i}(v_i\rho) = \epsilon\sum_i \frac{\partial^2}{\partial^2 x_i^2}\rho, && (34)
\end{aligned}
$$

which is a Fokker-Planck equation with nonzero diffusion ($\epsilon \neq 0$).

2 The Benamou-Brenier formulation

If the Monge map exists, then the quadratic Wasserstein distance

$$
W_2^2(\alpha_0,\alpha_1) = \min_{T \text{ such that } T\#\alpha_0=\alpha_1} \int_\Omega ||x-T(x)||dx \tag{35}
$$

has the equivalent formulation

$$
A^2(\alpha_0,\alpha_1) = \begin{cases} \min_{\alpha,v} \iint_{[0,1]\times\mathbb{R}^n} ||v||_2^2 d\alpha dt \\ \partial_t\alpha + \nabla\cdot(\alpha v) = 0 \\ \alpha(0,\cdot) = \alpha_0 \\ \alpha(1,\cdot) = \alpha_1. \end{cases} \tag{36}
$$

Proof. Let us consider the class of functions

$$
C(\alpha_0,\alpha_1) = \left\{(\alpha,v)\middle|\partial_t\alpha + \nabla\cdot(\alpha v) = 0, \alpha(0,\cdot) = \alpha_0, \alpha(1,\cdot) = \alpha_1\right\}. \tag{37}
$$

Let $(\alpha, v) \in C(\alpha_0, \alpha_1)$. Consider the problem

$$\begin{cases} \frac{\partial X(t,x)}{\partial t} = v(t, X(t,x)) \\ X(0,x) = x, x \sim \alpha_0 \\ T_t : x \in \mathbb{R}^n \to X(t,x) \in \mathbb{R}^n \\ \alpha(t,\cdot) = T_t \# \alpha_0 \text{ satisfies } \partial_t \alpha + \nabla \cdot (\alpha v) = 0. \end{cases} \tag{38}$$

Thus, we have

$$\iint\limits_{[0,1]\times\mathbb{R}^n} ||v(t,x)||_2^2 d\alpha(t,\cdot) dt \tag{39}$$

$$\overset{\text{pushforward relation}}{=} \iint\limits_{[0,1]\times\mathbb{R}^n} ||v(t,x)||_2^2 dT_t\#\alpha_0 dt \tag{40}$$

$$\overset{\text{Villani}}{=} \iint\limits_{[0,1]\times\mathbb{R}^n} ||v(t,T_t(x))||_2^2 d\alpha_0 dt \tag{41}$$

$$= \iint\limits_{[0,1]\times\mathbb{R}^n} ||v(t,X(t,x))||_2^2 d\alpha_0 dt \tag{42}$$

$$= \iint\limits_{[0,1]\times\mathbb{R}^n} ||v(t,X(t,x))||_2^2 dt d\alpha_0 \tag{43}$$

$$\overset{\text{Jensen's inequality and Fubini's}}{\geqslant} \int_{\mathbb{R}^n} \left|\left| \int_{[0,1]} v(t,X(t,x)) dt \right|\right|_2^2 d\alpha_0 \tag{44}$$

$$= \int_{\mathbb{R}^n} \left|\left| \int_{[0,1]} \frac{\partial X(t,x)}{dt} dt \right|\right|_2^2 d\alpha_0 \tag{45}$$

$$= \int_{\mathbb{R}^n} \left|\left| X(1,x) - X(0,x) dt \right|\right|_2^2 d\alpha_0 \tag{46}$$

$$= \int_{\mathbb{R}^n} \left|\left| T_1(x) - X(0,x) dt \right|\right|_2^2 d\alpha_0 \tag{47}$$

$$\overset{\text{since Wasserstein distance is optimal}}{\geqslant} W_2^2(\alpha_0, \alpha_1) \tag{48}$$

$$\implies A_2^2(\alpha_0, \alpha_1) \geqslant W_2^2(\alpha_0, \alpha_1). \tag{49}$$

Now, let $X(t,x) = x + t(T^*(x) - x), v(t,x) = T^*(x) - x$, then the two inequalities become equality and we have the result.

3 JKO conditions with continuity equation equivalence

We show the JKO conditions as in Lee et al. (2023) are equivalent to a continuity equation.

Consider the density under the flow map $\rho(t, T(t,x))$. By the chain rule,

$$\frac{d}{dt}\rho(t, T(t,x)) = \partial_t \rho + \nabla \rho \cdot \partial_t T(t,x). \tag{50}$$

We are aware that $\partial_t T = v$, so that gives

$$\frac{d}{dt}\rho = \partial_t \rho + \nabla \rho \cdot v. \tag{51}$$

Now, consider the following:

$$\frac{d}{dt}\rho(t,T(t,x)) = \frac{d}{dt}\frac{\rho^k(x)}{\det|J(t,x)|} = -\frac{\rho^k(x)}{(\det|J(t,x)|)^2} \cdot \partial_t \det|J(t,x)| \tag{52}$$

$$= -\frac{\rho^k(x)}{\det|J(t,x)|} \cdot \frac{\partial_t \det|J(t,x)|}{\det|J(t,x)|} = -\frac{\rho^k(x)}{\det|J(t,x)|}\partial_t \log\det|J(t,x)| \tag{53}$$

$$= -\rho(t,T(t,x))\partial_t \log\det|J(t,x)| = -\rho(t,T(t,x))\mathrm{div}(v). \tag{54}$$

Thus,

$$-\rho(t, T(t,x))\text{div}(v) = \partial_t \rho + \nabla \rho \cdot v \implies \partial_t \rho + \nabla \cdot (\rho v) = 0, \tag{55}$$

as desired. The equivalence of $\partial_t \log \det |J(t,x)| = \text{div}(v)$ is due to Jacobi's formula.

4 Derivation of the JKO algorithm

4.1 The Euclidean case

First, we discuss the Euclidean case.

Consider the optimization problem

$$\min_x \left\{ \frac{1}{2\tau} ||x - x_k||_2^2 + \mathcal{E}(x) \right\}. \tag{56}$$

Note that the solution existence is dependent on conditions, such as convexity. A local minima to this problem can be found by taking the gradient

$$\nabla \left[\frac{1}{2\tau} ||x - x_k||_2^2 + \mathcal{E}(x) \right] = \frac{1}{\tau}(x - x_k) + \nabla \mathcal{E}(x) = 0. \tag{57}$$

As a side remark, the first term derivative can be seen more clearly using

$$\nabla \langle x - x_k, x - x_k \rangle = [\partial_j (\sum_i (x_i - x_{k,i})^2)]_j = [2(x_j - x_{k,j})]_j. \tag{58}$$

Here, $[\cdot]_j$ denotes vector concatenation. Rearranging our equation, we see

$$\frac{x - x_k}{\tau} - \nabla \mathcal{E}(x), \tag{59}$$

which is exactly the Euler scheme for the gradient flow equation

$$\partial_t x(t) = -\nabla \mathcal{E}(x(t)). \tag{60}$$

4.2 The Wasserstein case

The Wasserstein gradient flow PDE is

$$\partial_t \rho = \nabla \cdot (\rho \nabla (\mathcal{E}'(\rho))). \tag{61}$$

We have denoted $\mathcal{E}'$ the first variation of the functional. Note that Ansari is a good reference for this section. Similarly as before, consider the optimization problem

$$\min_\rho \left\{ \frac{1}{2\tau} \int |x - T(x)| d\rho + \mathcal{E}(\rho) \right\}. \tag{62}$$

We cannot take the gradient of this optimization problem, as we did in the Euclidean case. Instead, the equivalent is the first variation, so we have a solution to this problem takes the form

$$\delta(\frac{1}{2\tau} \int |x - T(x)| d\rho + \mathcal{E}(\rho)) = \text{constant}. \tag{63}$$

The reason it is constant and not zero is because ρ is constrained, so the constant follows from the method of Lagrange multipliers since the first variation of

$$\delta(\lambda(\int \rho - 1)) = \lambda. \tag{64}$$

It is known $T(x) = x - \nabla \phi(x)$ for some convex ϕ Mokrov et al. (2021)Bao and Zhang (2022) (page 4) Ansari. It is also known the first variation of the Wasserstein distance is Ansari

$$\delta(\frac{1}{2\tau} W_2^2(\rho, \rho_k)) = \frac{1}{\tau} \phi. \tag{65}$$

Thus,

$$\frac{1}{\tau}\phi(x) + \mathcal{E}'(\rho_k)(x) = \text{constant} \implies \frac{1}{\tau}\nabla\phi(x) + \nabla\mathcal{E}'(\rho_k)(x) = 0. \tag{66}$$

Equivalenty,

$$\frac{1}{\tau}(x - T(x)) = -\nabla\mathcal{E}'(\rho_k)(x). \tag{67}$$

In the continuum limit, we get

$$v = -\nabla\mathcal{E}'(\rho_k), \tag{68}$$

which yields our Wasserstein gradient flow when substituted into 61.

5 Wasserstein geometry

In this section, we discuss the Riemannian metric of Wasserstein space. Recall the Riemannian metric is defined as

$$g|_p : \mathrm{Tan}_p M \times \mathrm{Tan}_p M \to \mathbb{R}^*, g_{ij} = g(\partial_{x^i}|_p, \partial_{x^j}|_p). \tag{69}$$

We have used nonstandard notation here, and we will denote $\mathbb{R}^* = \mathbb{R}^+$ corresponding to $i = j$ and $\mathbb{R}^* = \mathbb{R}$ corresponding to $i \neq j$.

Consider the set of Borel measures on M such that Takatsu

$$\mathcal{B}_2(M) = \left\{\mu : \text{Borel sets on } M \to \mathbb{R}^+ \cup \{0\} : \int_M d(x, x_0)^2 d\mu(x) < \infty, x_0 \in M, \int_M d\mu(x) = 1\right\}. \tag{70}$$

We remark this notation is also a bit nonstandard, but is very clear, so we will use it. As a remark, recall that $M = \mathbb{R}^d$ is indeed a valid Riemannian manifold. Our continuity equation stems from this set.

The Riemannian metric on Wasserstein space is Ay (2024)

$$g_W(\nabla\phi, \nabla\psi) = \int_M \langle\nabla\phi, \nabla\psi\rangle d\mu(x). \tag{71}$$

Here, $\mu = \mu_0 \in \mathcal{B}_2(M)$ is the initial measure corresponding to a continuity equation

$$\partial_t\mu_t + \nabla\cdot(\mu_t\nabla\Psi) = 0, \tag{72}$$

which holds in the sense of distributions Ambrosio et al. (2008)

$$\iint\limits_{[0,1]\times\mathbb{R}^d} (\partial_t\varphi + \langle v_t, \nabla\varphi\rangle) d\mu_t dt = 0, \quad \varphi \text{ is a suitable compactly supported test function}, \tag{73}$$

where the divergence is defined in the sense of distributions (or we can define the gradient of the measure using the Radon-Nikodym derivative, which is the density), and

$$\nabla\phi, \nabla\psi \in \left\{\nabla\Psi : \Psi \text{ is a plausible potential}\right\} = \Gamma, \Gamma^* = \overline{\left\{\nabla\Psi : \Psi \in \mathcal{C}(M)\right\}}^{L^p(\mu;X)} : \tag{74}$$

are two velocity fields in the continuity equations corresponding to two distinct tangent elements. Γ is a generalized set of admissible potentials, and Γ^* is a more prototypical example. Here, the closure is taken to ensure the set is compact. $\mathcal{C}$ is a set of admissible smooth functions. We refer to *Ambrosioet al.* (2008), chapter 8 for discussion of this. In particular, it is known Ambrosio et al. (2008)

$$v_t \in \overline{\left\{j_q(\nabla\Psi) : \Psi \in \mathrm{Cyl}(X)\right\}}^{L_p(\mu_t;X)} \tag{75}$$

and j_q is the map $j_q(v) = ||v||_2^{q-2}v$. Here Cyl denotes the cylindrical class of functions

6 JKO variational equivalence

Consider the OT problem Lee et al. (2023)

$$\begin{cases} (\rho, v) = \arg\min \iint_{[0,1]\times\mathbb{R}^d} \rho \|v\|_2^2 dxdt + 2\Delta t \mathcal{E}(\rho(1,\cdot)) \\ \text{subject to } \partial_t \rho + \nabla(\rho v) = 0, \rho(0,x)\rho_0. \end{cases} \tag{76}$$

Equivalenty, this can be written

$$\begin{cases} \min_v \iint_{[0,1]\times\mathbb{R}^d} \rho^n(x) \|v(T(t,x))\|_2^2 dxdt + 2\Delta t \mathcal{E}(T(1,\cdot)\#\rho^n) \\ \text{subject to } \frac{d}{dt}T = v, T(0,x) = x. \end{cases} \tag{77}$$

We discuss why these are equivalent. We already proved in section 3 why the Lagrangian flow map condition $\frac{d}{dt}T = v$ is equivalent to a continuity equation.

It is also known

$$\rho(\tau,\cdot) = T(\tau,\cdot)\#\rho^n. \tag{78}$$

There is a known equivalence Villani (chapter 1)

$$\int_Y \varphi(y) d\nu(y) = \int_X \varphi(T(x)) d\mu(x) \tag{79}$$

when $T\#\mu = \nu$. This implies $\rho(1,\cdot) = T(1,\cdot)\#\rho^n$. The result follows from these facts. In particular, we can take $\varphi = \|v\|_2^2$ and we see

$$\iint\limits_{[0,1]\times\mathbb{R}^d} \rho(t,x) \|v(t,x)\|_2^2 dxdt = \iint\limits_{[0,1]\times\mathbb{R}^d} \|v(t,x)\|_2^2 d\rho dt \tag{80}$$

$$= \iint\limits_{[0,1]\times\mathbb{R}^d} \|v(t,T(t,x))\|_2^2 d\rho^n dt = \iint\limits_{[0,1]\times\mathbb{R}^d} \rho^n(x) \|v(t,T(t,x))\|_2^2 dxdt. \tag{81}$$

Here, we have assumed our measures are absolutely continuous with respect to Lebesgue measure. In particular, it can be noted that Lebesgue measure coincides with Riemann integration if we are allowed the assumption of Riemann integrability, which we will indeed. As a reminder, we say a measure Γ is absolutely continuous with respect to Lebesgue measure λ^*, or $\Gamma \ll \lambda^*$, if there exists a density ρ such that

$$\Gamma(A) = \int_A f d\lambda^*, \tag{82}$$

where A is any measurable set on the Lebesgue σ-algebra.

7 Normalizing flows

I think this section is important because this is an area of research that we are more inclined to make contributions towards. I have some preliminary ideas in this area.

References

Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient Flows*. Lectures in Mathematics. ETH Zürich. Birkhäuser Basel, 2 edition, 2008. ISBN 978-3-7643-8722-8. doi: 10.1007/978-3-7643-8722-8. URL https://doi.org/10.1007/978-3-7643-8722-8.

Abdul Fatir Ansari. Introduction to gradient flows in the 2-wasserstein space. URL https://abdulfatir.com/blog/2020/Gradient-Flows/.

Nihat Ay. Information geometry of the otto metric. *Information Geometry*, 2024. doi: 10.1007/s41884-024-00149-w. URL https://doi.org/10.1007/s41884-024-00149-w.

Gang Bao and Yixuan Zhang. Optimal transportation for electrical impedance tomography, 2022. URL https://arxiv.org/abs/2210.16082.

Wonjun Lee, Li Wang, and Wuchen Li. Deep jko: time-implicit particle methods for general nonlinear gradient flows, 2023. URL https://arxiv.org/abs/2311.06700.

J. Gordon Leishman. *Conservation of Mass: Continuity Equation*. 2021. URL https://eaglepubs.erau.edu/introductiontoaerospaceflightvehicles/chapter/conservation-of-mass-continuity-equation/.

Petr Mokrov, Alexander Korotin, Lingxiao Li, Aude Genevay, Justin Solomon, and Evgeny Burnaev. Large-scale wasserstein gradient flows, 2021. URL https://arxiv.org/abs/2106.00736.

Asuka Takatsu. Wasserstein geometry of gaussian measures. URL https://www2.sonycsl.co.jp/person/nielsen/infogeo/Seminar/gauss-Takatsu.pdf.

Cédric Villani. Optimal transport, old and new. URL https://www.ceremade.dauphine.fr/~mischler/articles/VBook-O&N.pdf.