Existence of differentiable curves in convex sets and the concept of direction of the flow in mass transportation

Rodolfo Ríos-Zertuche

Abstract

In this paper we consider convex subsets of locally-convex topological vector spaces. Given a fixed point in such a convex subset, we show that there exists a curve completely contained in the convex subset and leaving the point in a given direction if and only if the direction vector is contained in the sequential closure of the tangent cone at that point.

We apply this result to the characterization of the existence of weakly differentiable families of probability measures on a smooth manifold and of the distributions that can arise as their derivatives. This gives us a way to consider the mass transport equation in a very general context, in which the notion of direction turns out to be given by an element of a Colombeau algebra.

Contents

1 Introduction
2 Curves in convex sets
3 Families of positive measures and probabilities
   3.1 Characterization
   3.2 Examples and remarks
   3.3 Mass transportation

1 Introduction

This paper has two parts. The first part is the general theory, presented in Section 2, that characterizes the possible derivatives of differentiable curves contained in convex subsets of locally-convex topological vector spaces. Our main motivation for the development of this general theory is its application to variational and perturbative analysis, and we will do this in upcoming papers.

Like many very general theorems, our main abstract result, Theorem 1, is quite simple. It states that the set of possible derivatives $\gamma'(0)$ of differentiable curves $\gamma: [0, +\infty) \to C$ emanating from a point $p = \gamma(0)$ in a closed, convex subset $C$ of a locally-convex topological vector field, coincides with the sequential closure of the tangent cone of $C$ at $p$; see Section 3.1 for definitions and for
the precise statement. We also give reformulations in Corollaries 2 and 3 that work around the issue of taking the sequential closure, and hence may be more amenable to practical use.

In the second part of the paper, we consider families of measures varying in a weakly differentiable way: we merely ask for the map $t \mapsto \int \phi \, d\mu_t$ to be differentiable at 0 for all $\phi \in C_c^\infty$. Such families arise naturally in many parts of mathematics, including the theory of the heat equation, stochastic dynamics and SPDEs, and the continuity equation in mass transportation, to name a few.

As we show in Examples 8, 15 and 16, the derivative of a family of positive Radon measures is rarely a signed Radon measure, but instead in general it is a distribution, that is, a continuous real-valued functional on the space of $C_c^\infty$ functions. We apply our general theory to characterize in Theorem 4 the distributions that arise as derivatives of families of probabilities on a smooth manifold; we do this in Section 3. Our result states that given a measure $\mu$, the distributions $\eta$ that arise as derivatives of differentiable families $(\mu_t)_{t \geq 0}$ starting at $\mu_0 = \mu$ are precisely those that satisfy that $\langle \eta, f \rangle \geq 0$ for all nonnegative functions $f \in C_c^\infty$ that vanish on the support of $\mu$.

In Section 3.2 we give some important remarks and examples that should aid in understanding the content of the statement of the characterization. We also give a characterization of the existence of curves with two-sided derivatives; this is Corollary 13. Then in Section 3.3 we consider the continuity equation from mass transportation, namely

$$\frac{d\mu_t}{dt} + \text{div}(v_t \mu_t) = 0,$$

in a very weak sense, and we use it to find elements of a Colombeau algebra that provide a meaning to the concept of “direction of the flow,” that is, we find an object that has a role similar to the classical vector field $v_t$, but in a context in which such “movement” can be extremely singular.

An initial version of the results presented in Section 3 was posted previously in the arXiv in [5, 6]. The theory was since completely redeveloped from a different point of view until it attained the form presented here, which is also considerably simpler and more general.

Acknowledgements. The author is very grateful to Patrick Bernard, Jaime Bustillo, and Stefan Suhr for their encouragement, support, suggestions, and numerous discussions.

2 Curves in convex sets

Consider a locally-convex topological vector space $V$ with topological dual space $V^*$. Let $C$ be a closed and convex subset of $V$.

We say that a subset $A$ of $V$ is a cone if $ra$ is contained in $A$ for every $r > 0$ and for every $a$ in $A$. The set $A$ is a closed cone if it is a cone and is closed in the topology of $V$.

Let $p$ be a point in the closed, convex set $C$. The tangent cone of $C$ at $p$ is the set of vectors $v$ such that $p + tv$ is contained in $C$ for some $t > 0$. In other words, the tangent cone is the set $\mathbb{R}_{\geq 0}(C - p)$. This set is a cone, but it is not necessarily closed.
Again with $p$ a point in $C$, the solid tangent cone of $C$ at $p$ is the set of vectors $v$ in $C$ such that $\theta(v) \geq 0$ for every functional $\theta$ in $V^*$ that is nonnegative throughout the translate $C - p$. Since the solid cone is an intersection of sets of the form $\theta^{-1}([0, +\infty))$, which are all closed, the solid tangent cone is a closed cone. It is well known that, conversely, as a consequence of the Hahn-Banach Separation Theorem, the topological closure of the tangent cone is the solid tangent cone.

An intermediate set is the sequential closure of the tangent cone of $C$ at $p$, defined to be the set of vectors $v$ in the solid tangent cone at $p$ that can be approximated by sequences of points $\{v_i\}_{i=1}^\infty$ in the tangent cone at $p$, $v_i \to v$ as $i \to \infty$. Again, the sequential closure of the tangent cone is not necessarily topologically closed, and its closure is the solid tangent cone.

We then have

**Theorem 1.** Let $V$ be a locally-convex topological vector space, and let $C$ be a closed and convex subset of $V$. Given a point $p$ in $C$ and a direction vector $v$ in $V$, there exists a continuous curve $c: [0, +\infty) \to C$ such that $c(0) = p$ and

$$\frac{dc(t)}{dt} \bigg|_{t=0^+} = v$$

if, and only if, $v$ is in the sequential closure of the tangent cone of $C$ at $p$.

**Proof.** To prove the “if” part, let $v_1, v_2, \ldots$ be a sequence of vectors in the tangent cone to $C$ at $p$ that converges to $v$, $v_i \to v$. For $i = 1, 2, \ldots$, let $\varepsilon_i > 0$ be such that $p + \varepsilon_i v_i$ is contained in $C$, and also $\varepsilon_i \to 0$ as $i \to +\infty$. Take a sequence $t_1, t_2, \ldots$ of positive numbers such that $0 < t_{i+1} < t_i \leq \varepsilon_i$. Define $p_j$ to be equal to $p + t_j v_j$. Note that $p_j \to p$ and $t_j \to 0$ as $j \to +\infty$. To define the curve $c$, first set $c(t_j) = p_j$ and then on the intervals $(t_{j+1}, t_j)$ do a linear interpolation between those values. Also set $c(0) = p$; this renders $c$ continuous on $[0, t_1]$. For simplicity, we let $c(t) = p_1$ for $t \geq t_1$. Then $c$ is continuous, and $c(t)$ is in $C$ for all $t > 0$ because of the convexity of $C$ and the fact that $p$ and $p_1$, $i = 1, 2, \ldots$, are all contained in $C$. Also, we have

$$\lim_{i \to +\infty} \frac{p_i - p}{t_i} = \lim_{i \to +\infty} v_j = v.$$

Since $c$ is a linear interpolation between these points,

$$\lim_{t \to 0^+} \frac{c(t) - p}{t} = \lim_{i \to +\infty} \frac{p_i - p}{t_i} = v.$$

Thus the curve $c$ exists.

To prove the converse statement, assume the existence of $c$ as in the statement of the theorem. Let $v_i = i(c(1/i) - p)$ for $i = 1, 2, \ldots$. Since $c(1/i)$ and $p$ are both contained in $C$, $v_i$ is in the tangent cone to $C$ at $p$. Moreover,

$$v = \lim_{t \to 0^+} \frac{c(t) - p}{t} = \lim_{i \to +\infty} \frac{c(1/i) - p}{1/i} = \lim_{i \to +\infty} v_i.$$

So $v$ is in the sequential closure of the tangent cone to $C$ at $p$. \hfill $\Box$

We observe the following reformulation of Theorem 1 that is appropriate for normed spaces.
Corollary 2. Assume that $V$ is a normed space, that $p$ is an element of a closed, convex subset $C$ of $V$, and let $v \in V$. Then there is a continuous curve $c: [0, +\infty) \to C$ such that $c(0) = p$ and $dc(t)/dt|_{t=0^+} = v$ if, and only if, $\theta(v) \geq 0$ for all $\theta \in V^*$ for which $\theta(p) = \inf_{q \in C} \theta(q)$.

This follows easily from Theorem 1 since in this context the topological closure coincides with the sequential closure of any subset of $V$.

Here is the reformulation of the result that we will actually use in the following section.

Corollary 3. Let $V$ be a locally-convex topological vector space, $p$ an element of a convex subset $C$ of $V$, and $v$ be an element of $V$. Assume that $v$ and $C$ are contained in a subset $X$ of $V$ such that there is a norm on $X$ that induces the topology that $X$ inherits from $V$. Assume additionally that $X$ contains the intersection of the tangent cone of $C$ at $p$ with an open subset of $V$ that contains $v$. Then there is a curve $c: [0, +\infty) \to C$ such that $c(0) = p$ and $dc(t)/dt|_{t=0^+} = v$ if, and only if, $\theta(v) \geq 0$ for all $\theta \in V^*$ for which $\theta(p) = \inf_{q \in C} \theta(q)$, that is, for all continuous linear functionals $\theta$ that attain their minimum within $C$ at $p$.

Proof. We construct a sequence $v_1, v_2, \ldots$ in the tangent cone to $C$ at $p$ converging to $v$, $v_i \to v$. For $i \in \mathbb{N}$, we take the balls $B_i$ of radius $1/i$ centered at $v$ with respect to the norm in $X$. These must all intersect that tangent cone because $v$ is in its topological closure. The tangent cone is contained in $X$, so we take $v_i$ to be any point in its intersection with $B_i$. From this point on, the rest of the proof of Theorem 1 works.

3 Families of positive measures and probabilities

3.1 Characterization

The following theorem is the main result of this section and follows from Corollary 2 in a way that will be explained below. Important remarks and some examples of its application can be found in Section 3.2. We explain the natural relation of our result with mass transportation in Section 3.3.

Denote by $C^\infty(P)$ the space of smooth functions on a smooth manifold $P$, and by $C^\infty_c(P)$ the compactly supported ones.

Theorem 4. Let $\mathscr{M}_0$ be the set of positive, Radon measures on a $C^\infty$ manifold $P$. Let $\mu$ be an element of $\mathscr{M}_0$ and let $\eta$ be an element of the space $\mathscr{D}'$ of distributions on $P$. We look for conditions for the existence of a family $(\mu_t)_{t \geq 0}$ in $\mathscr{M}_0$ such that

$$\mu_0 = \mu \quad \text{and} \quad \frac{d}{dt} \int \phi \, d\mu_t \bigg|_{t=0^+} = \langle \eta, \phi \rangle$$

for all $\phi \in C^\infty_c(P)$. Such a family exists if, and only if, $\langle \eta, f \rangle \geq 0$ for all nonnegative functions $f \in C^\infty_c(P)$ that vanish on the support of $\mu$.

If $\mu$ is a probability measure, the family $(\mu_t)_{t \geq 0}$ can be realized as a family of probability measures if additionally $\langle \eta, 1 \rangle = 0$.
Let \( g \in E \), \( \| \cdot \| \) given in (1), and let \( \eta \in C_\text{top} \) with respect to \( P \) and \( \mu \) with \( f, \phi \in C_\infty(P) \).

Proof of Theorem 4.

For the proof of Theorem 4 we will need this lemma, which is proved below.

**Lemma 7.** Let \( n \) be a positive integer, \( \gamma > 0 \), and \( U \) a bounded open subset of \( P \), so that its closure \( \overline{U} \) is compact. Consider the subset \( X \) of \( \mathcal{E}' \) consisting of distributions \( \xi \) of degree at most \( n \) with support in \( U \) and such that \( |\xi|_{\mathcal{E}'_n} \leq \gamma \). Then the seminorm \( |\cdot|_{\mathcal{E}'_n} \) is nondegenerate on \( X \), and within \( X \) it acts as a norm that induces the topology that \( X \) inherits from \( \mathcal{E}'_n \), that is, the weak\(^*\) topology with respect to \( C_\infty(P) \). In particular, the sequential closure of subsets of \( X \) coincides with their closure.

**Proof of Theorem 4.** By means of a partition of unity, we may assume at the outset that both \( \mu \) and \( \eta \) are compactly supported.

Let us explain how we will apply Corollary 3. Let \( n > 0 \) be the degree of \( \eta \). Set \( V = \mathcal{E}' \), \( C = \mathcal{M}_0 \), \( p = \mu \), and \( \nu = \eta \), and from the corollary we will get \( c(t) = \mu t \). Fix a bounded, open subset \( U \) of \( P \) that contains the supports of \( \mu \) and \( \eta \), and let \( \gamma = 2|\eta|_{\mathcal{E}'_n} \). Let the set \( X \subset \mathcal{E}'_n \) be the set of distributions \( \xi \) of order at most \( n \) whose supports are contained in \( U \) and such that \( |\xi|_{\mathcal{E}'_n} \leq \gamma \). From Lemma 7 we know that the topology on \( X \) is induced by the seminorm \( |\cdot|_{\mathcal{E}'_n} \), which is nondegenerate within \( X \). We also observe that \( V \) is locally convex. The set \( X \) contains all signed Radon measures \( \nu \) with \( |\nu|_{\mathcal{E}'_n} < 2|\eta|_{\mathcal{E}'_n} \).
so in particular it contains the intersection of the tangent cone of $C$ at $p$ with the set of distributions $\xi$ with $|\xi|_{T_p} < 2|q|_{T_p}$. This set of distributions is an open set that contains $v$. So we can apply Corollary 3, which is then equivalent to the following assertion: A family $(\mu_t)_{t \geq 0}$ as in the statement of Theorem 4 exists if, and only if, $(\eta, f) \geq 0$ for all functions $f \in C^\infty_c(P)$ such that

$$\int f \, d\mu < \int f \, d\nu \quad \text{for all} \quad \nu \in \mathcal{M}_0.$$  \hfill (3)

To finish the proof of the first part of Theorem 4, we just need to understand these functions $f$. Let $f \in C^\infty_c(P)$ satisfy (3). Clearly, the support of $\mu$ must be contained in the set of points $p$ such that $f(p) = \inf_{q \in P} f(q)$, so in order to prove the theorem we just need to argue that this infimum equals zero, or equivalently, that $\int f \, d\mu = 0$. Taking $\nu_a = a\mu \in \mathcal{M}_0$ for $a > 0$, it is easy to see that if $\inf_{q \in P} f(q) \neq 0$, then $\nu_a$ can be taken appropriately so that $\int f \, d\nu_a = a \int f \, d\mu < \int f \, d\mu$, thus contradicting (3).

If $\mu$ is a probability and $(\eta, 1) = 1$, we may divide each measure $\mu_t$ in the family by its total mass to get probability measures $\mu_t/\mu_t(P)$; the condition $(\eta, 1) = 1$ ensures that this procedure does not break the differentiability of the normalized family at $t = 0+$.

\textbf{Proof of Lemma 4.} Let us recall that the strong topology on $\mathcal{E}'$ is the one induced by the seminorms $|\cdot|_{K,m}$ defined in equation (2).

Recall that $\mathcal{E}'$ is a Montel space. A set $A$ in $\mathcal{E}'$ is, by definition, \textit{bounded} if for each $f \in C^\infty_c(P)$ there is a constant $C_f > 0$ such that $\langle \xi, f \rangle \leq C_f$ for all $\xi \in A$. Thus $X$ is bounded. Proposition 34.5 in [7] tells us that, in Montel spaces, the strong and weak* topologies coincide in bounded sets, so this is the case in $X$. It is easy to see that, within $X$, the seminorm $|\cdot|_{T_p}$ dominates all other seminorms of the form $|\cdot|_{K,k}$ with $K$ compact and $k > 0$. Thus, within the set $X$ the seminorm $|\cdot|_{T_p}$ induces the weak* topology. Its nondegeneracy is easy to check as well.

\textbf{3.2 Examples and remarks}

\textbf{Example 8.} We show that the derivative of a family of measures $(\mu_t)$, need not itself be a measure. Let $P = \mathbb{R}$ and $\mu_t = \delta_t$, the Dirac delta at $t \in \mathbb{R}$. Then for all $f \in C^\infty_c(P)$ we have

$$\frac{d}{dt} \bigg|_{t=0} \int f \, d\mu_t = \frac{d}{dt} \bigg|_{t=0} f(t) = f'(0),$$

so that the derivative $\eta = d\mu_t/dt|_{t=0}$ is the distribution $\eta = -\partial_x \delta_0$, where $\partial_x$ indicates the derivative in the domain $\mathbb{R} = P$ of the measures, and is taken here in the sense of distributions.

\textbf{Remark 9.} The derivative $d\mu_t/dt|_{t=0+}$ that we are interested in does not coincide with the usual distributional derivative. In order to take the derivative in the sense of distributions, we would need to consider $(\mu_t)_{t \geq 0}$ as a distribution on the set $P \times [0, +\infty)$, and we do not do that.

Let us consider the situation in Example 5 as a purely distributional derivative: we take $\mu_t = \delta_t$ on $P = \mathbb{R}$ and, since it is well defined and smooth for all $t$...
in $\mathbb{R}$, we think of the entire family $\mu_t$ as defining a distribution $\nu$ on $\mathbb{R}^2$, acting on functions $f \in C^\infty(\mathbb{R}^2)$ by integration:

$$\langle \nu, f \rangle = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, t) d\mu(x) dt = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, t) d\nu_t(x) dt = \int_{-\infty}^{+\infty} f(t, t) dt.$$

Then the distribution $\partial_t \nu$ (where the derivative is taken in the sense of distributions) acts like this on $f \in C^\infty_c(\mathbb{R}^2)$:

$$\langle \partial_t \nu, f \rangle = -\langle \nu, \partial_t f \rangle = -\int_{-\infty}^{+\infty} \frac{\partial f}{\partial t}(s, s) ds.$$

Clearly, there is no reason for this to coincide with the result we obtain in Example 8, which for $f \in C^\infty_c(\mathbb{R}^2)$ perhaps amounts to $\partial f/\partial x \big|_{(x, t) = (0, 0)}$. The important point, in any case, is that if we considered the derivative with respect to $t$ as a distributional one, we would need to be taking test functions that depended on that variable, and we do not do this; note that in Example 8 the function $f$ depends only on the real variable corresponding to $P$ and not on $t$.

**Example 10.** Let again $P = \mathbb{R}$, and let us use Theorem 4 to determine the ways in which the Dirac delta at 0, $\delta_0$, can be deformed. The condition in Theorem 4 requires us to look for distributions $\eta$ such that $\langle \eta, f \rangle \geq 0$ for all $f \in C^\infty_c(\mathbb{R})$ such that $f(x) \geq 0$ for all $x \in \mathbb{R}$ and $f(0) = 0$. Since the Taylor expansion of any such function $f$ starts at the quadratic term, which must have a nonnegative coefficient, we conclude that $\eta$ must be of the form

$$\eta = \nu + a \partial_x \delta_0 + b \partial_x^2 \delta_0$$

for a positive measure $\nu$ on $\mathbb{R}$, a real number $a$, and a positive number $b > 0$.

**Example 11.** Going back to the situation of Example 10, we now use the second part of Theorem 4 to determine the ways in which the distribution $\delta_0$ can be deformed within the space of probability measures. In this case, we have the additional requirement that the derivative $\eta$ satisfy $\langle \eta, 1 \rangle = 0$. The reader will immediately recognize that this amounts to the requirement that the measure $\nu$ vanish everywhere or, in other words, that $\eta$ be of the form

$$\eta = a \partial_x \delta_0 + b \partial_x^2 \delta_0$$

for $a \in \mathbb{R}$, $b > 0$.

**Remark 12.** Although the statement of Theorem 4 deals only with one-sided families $(\mu_t)_{t \geq 0}$, it is clear what the requirement is on the distribution $\eta$ for the existence of a two-sided family $(\mu_t)_{t \in \mathbb{R}}$ with derivative $\eta$: $\eta$ and $-\eta$ must satisfy the conditions in Theorem 4. For the reader’s convenience, we state this in full:

**Corollary 13.** Let $\mathcal{M}_0$ be the set of positive, Radon measures on a $C^\infty$ manifold $P$. Let $\mu$ be an element of $\mathcal{M}_0$ and let $\eta$ be an element of the space $\mathcal{D}'$ of distributions on $P$. We look for conditions for the existence of a (two-sided) family $(\mu_t)_{t \in \mathbb{R}}$ in $\mathcal{M}_0$ such that $\mu_0 = \mu$ and $\frac{d}{dt} \int P \phi d\mu_t|_{t=0} = \langle \eta, \phi \rangle$ for all $\phi \in \mathcal{D}'$. For the reader’s convenience, we state this in full:
\(C_c^\infty(P)\).  Such a family exists if, and only if, \(\langle \eta, f \rangle = 0\) for all nonnegative functions \(f \in C_c^\infty(P)\) that vanish in the support of \(\mu\).

If \(\mu\) is a probability measure, the family \((\mu_t)_{t \in \mathbb{R}}\) can be realized as a family of probability measures if additionally \(\langle \eta, 1 \rangle = 0\).

**Example 14.** Going back to the situation of Examples 10 and 11 we now determine the possible two-sided deformations of \(\delta_0\) both as two-sided families of positive, Radon measures and as two-sided families of probability distributions on \(P\). From Corollary 13, we see that the condition that needs to be satisfied by \(\eta\) is that \(\langle \eta, f \rangle = 0\) for all \(f \in C_c^\infty(\mathbb{R})\) such that \(f \geq 0\) throughout \(\mathbb{R}\) and \(f(0) = 0\). In this case, the quadratic term of the Taylor expansion of \(f\) impedes \(\eta\) from having a term of order 2, so for the case of arbitrary positive measures we get that \(\eta\) must be of the form

\[
\eta = \nu + a \partial_x \delta_0,
\]

for a positive measure \(\nu\) and a real number \(a\). Similarly, in the case of families of probability measures the additional condition that \(\langle \eta, 1 \rangle = 0\) forces \(\nu\) to vanish and \(\eta\) to be of the form

\[
\eta = a \partial_x \delta_0, \quad a \in \mathbb{R}.
\]

**Example 15.** Distributions of any degree may arise as derivatives of families of measures. Consider the case in which \(\mu\) is the uniform probability measure on a bounded, open set \(U \subseteq P\). Then any distribution \(\eta\) supported in \(U\) satisfies the condition of the first part of either Theorem 4 or Corollary 13 and any distribution \(\eta\) additionally satisfying \(\langle \eta, 1 \rangle = 0\) will be good for the second part of the statements of therein contained.

**Example 16.** We work out Example 13 explicitly in the special case of \(P = \mathbb{R}\), \(U = (-\frac{1}{2}, \frac{1}{2})\), \(\mu = \chi_U dx\) the uniform distribution on \(U\), \(k \in \mathbb{N}\), \(\eta = (-1)^k \partial^k \delta_0\).

Let \(\psi: \mathbb{R} \to [0,1]\) be a compactly-supported, \(C^\infty\) function that is equal to 1 in a neighborhood of 0, satisfies \(\psi(x) = \psi(-x)\), \(\int_{\mathbb{R}} \psi(x) dx = 1\), and is decreasing on \([0, +\infty)\). Pick some \(0 < q < 1\). For \(t \in \mathbb{R}\) with \(|t|\) small enough, we let \(\mu_t\) be the measure that corresponds to the density \(\rho(x, t) dx\) with

\[
\rho(x, t) = \chi_U(x) + (-1)^k (\text{sgn } t) |t|^q \frac{d^k \psi}{dx^k} \left(\frac{|t|}{q+1}, x\right), \quad t \neq 0,
\]

and \(\rho(x, 0) = \chi_U(x)\). We need \(|t|\) to be small enough that \(\rho(x, t)\) will be non-negative for all \(x\); this defines \(\mu_t\) only for small \(|t|\), yet its smooth extension to all of \(t \in \mathbb{R} \setminus \{0\}\) exists. We remark that at \(t = 0\) and \(x = 0\), the function \(\rho\) is not differentiable with respect to \(t\) because the slope is vertical.

Let us check that this family \((\mu_t)_{t \in \mathbb{R}}\) works, that is, that its derivative is indeed equal to \(\eta\). Let \(f \in C^\infty(\mathbb{R})\); then we have

\[
\frac{d}{dt} \int f \, d\mu_t = \frac{d}{dt} \int f(x) \rho(x, t) dx = \frac{d}{dt} \int f(x) \left( \chi_U(x) + (-1)^k \text{sgn } t |t|^q \frac{d^k \psi}{dx^k} \left(\frac{|t|}{q+1}, x\right) \right) dx = \frac{d}{dt} \text{sgn } t |t|^q \frac{d^k \psi}{dx^k} \left(\frac{|t|}{q+1}, x\right) dx
\]
after $k$ integrations by parts. In order to simplify the notation, let $h = d^k f/dx^k$.

We take the derivative and we get

\[ \text{sgn} t \frac{|t|^{q-1+k}}{k+1} \int h(x) \psi(|t|^{q-1+k}x) dx + \frac{|t|^{q-1+k}}{k+1} \int x h(x) \psi'(|t|^{q-1+k}x) dx \]

\[ = \text{sgn} t \frac{|t|^{q-1+k}}{k+1} \int h(x) \psi(|t|^{q-1+k}x) dx - \frac{|t|^{q-1+k}}{k+1} \int (x h(x))' \psi(|t|^{q-1+k}x) dx \]

This is easily seen to tend to $h(0) - \frac{q-1+k}{k+1} 0 h'(0) = h(0) = \langle \eta, f \rangle$ as $t \to 0$, which is what we wanted.

**Remark 17.** The conditions in Corollary [13] force the support of $\eta$ to be contained inside the support of the measure $\mu$. However, take a look at the following example.

**Example 18** (A considerably ugly situation). For the conditions of Theorem [4] or Corollary [13] to be satisfied, it is not true that the support of the distribution $\eta$ must be contained inside the set of points that hold the measure of $\mu$. (Note that the support of $\mu$ is the closure of this set.) Immitating the construction of the Cantor density, it is easy to construct a singular-continuous measure supported on the interval $[0, 1]$ whose measure will be contained in a set of Lebesgue measure zero; for instance, one can recursively subdivide the interval in halves and assign $1/3$ of the measure to the rationals in the lower half and $2/3$ to the rationals in the upper half, and iterating for the halves of each of these intervals, while declaring that the set of irrational numbers in $[0, 1]$ has $\mu$ measure zero. In this case, Theorem [4] and Corollary [13] show that the possible deformations are the same as for the uniform measure on $[0, 1]$; see Example [15].

Many of the distributions thus arising are supported in sets that are disjoint from the set holding the measure; as an example, take $\eta = \partial^k \sqrt{2}$, for any positive integer $k$.

We additionally remark that the existence of ugly examples like these is what forces us to the rather abstract treatment of the theory that we have resorted to in proving the results of Section 3.1 rather than a more constructive and explicit technique.

**Remark 19.** Here, we want to establish the relation between our result and the Carathéodory positivity criterion for holomorphic functions [2]. Carathéodory proved that a necessary and sufficient condition for the signed measure $\mu$ on the interval $[0, 2\pi)$ to be positive is for its Fourier coefficients

\[ a_n = \int_0^{2\pi} e^{-in\theta} d\mu(\theta) \]

to satisfy

\[ \sum_{n} \sum_{m} a_{m-n} \lambda_n \lambda_m \geq 0 \tag{4} \]

for any complex numbers $\lambda_0, \lambda_1, \ldots, \lambda_N$.

Now consider the case in which we have a family of positive measures $(\mu_t)_{t \geq 0}$ with $d\mu_t/dt|_{t=0+} = \eta$, as in Theorem [4]. Then we have Fourier coefficients
\[ a_n(t) = \int_0^{2\pi} e^{-i n \theta} d\mu_t(\theta) \] defined for each \( t \geq 0 \), and the Fourier coefficients of \( \eta \) will be given by \( \langle \eta, e^{-in\theta} \rangle = a'_n(0) \). Since each \( \mu_t \) is positive, the criterion (3) must hold for each \( t \geq 0 \). Thus the condition in Theorem 4 is equivalent to requiring that the Fourier coefficients \( a'_n(0) \) of \( \eta \) satisfy

\[ \sum_m \sum_n a'_{m-n}(0) \lambda_m \overline{\lambda_n} \geq 0 \]

whenever \( \lambda_0, \lambda_1, \ldots, \lambda_N \) are complex numbers such that

\[ \sum_m \sum_n a_{m-n}(0) \lambda_m \overline{\lambda_n} = 0. \]

3.3 Mass transportation

Motivation. For simplicity let \( P = \mathbb{R}^n \), \( n \geq 1 \). Recall that for \( p > 1 \) the Wasserstein metric between two probabilities \( \mu \) and \( \nu \) on \( \mathbb{R}^n \) with finite \( p \)th momenta on \( \mathbb{R}^n \) can be defined as

\[ W_p(\mu, \nu) = \left( \inf_{\gamma \in \Gamma(\mu, \nu)} \int_{\mathbb{R}^n \times \mathbb{R}^n} \text{dist}(x, y)^p \, d\gamma(x, y) \right)^{1/p}, \]

where the infimum is taken over all measures \( \gamma \) on \( \mathbb{R}^n \times \mathbb{R}^n \) whose marginals are \( \mu \) and \( \nu \). When \( (\mu_t)_{t \in \mathbb{R}} \) is a family of smooth probability densities that defines an absolutely continuous curve in the space of probabilities endowed with the metric \( W_p \), it has been shown [1, Chapter 8] that the derivative of \( (\mu_t) \) exists for almost every \( t \), and can be interpreted as the divergence of a vector field, that is, there are vector fields \( v_t : \mathbb{R}^n \rightarrow \mathbb{R}^n \) that satisfy the continuity equation,

\[ \frac{d\mu_t}{dt} + \text{div}(\mu_t v_t) = 0, \quad (5) \]

for almost every \( t \in \mathbb{R} \). The interpretation, which follows from Gauß’s theorem, is that the mass of \( \mu_t \) is being carried or transported by the flow of the vector field \( v_t \). This gives a way to assign a vector field \( v_t \) to the distribution \( d\mu_t/dt \), and this vector field gives a notion of “direction of the movement.” The vector field \( v_t \) is not unique: it is ambiguous up to addition of a vector field \( u_t \) such that \( \text{div}(\mu_t u_t) = 0 \) for all \( t \in \mathbb{R} \). Since the set of possible \( u_t \) is a closed subspace of \( L^p(\mu_t) \), one can choose the vector field \( v_t \) to be the minimizer of the \( L^p(\mu_t) \) norm for each \( t \). It has been shown [1] that the minimizer is in fact a gradient vector field, \( v_t = \nabla \phi_t \) for some functions \( \phi_t : \mathbb{R}^n \rightarrow \mathbb{R} \).

In the theory of \( W_p \)-absolutely continuous families of measures, \( v_t \) exists only for almost every \( t \). Consider the situation of Example 16: the family \( \mu_t = \rho(x, t)dx \) given in the example is \( W_p \)-absolutely continuous, but \( v_0 \) is not defined for \( k \geq 1 \). The results of Section 3.1 indicate that in cases like this, and surely also in the cases of families of measures that are not \( W_p \)-absolutely continuous but are still weakly differentiable, the distributions that can arise as their derivatives can be much more general than those of the form \( \text{div}(\mu_t v_t) \), for some vector field \( v_t \); cf. Example 17. As it turns out, an arbitrary distribution \( \eta \) may arise as the derivative of a family of measures, \( d\mu_t/dt|_{t=t_0} = \eta \), and we can try to use the continuity equation (5) to try to assign an object that will give an idea of direction of the movement determined by \( \eta \).
The assignment of an object conveying the direction of movement and analogous to $v_t$ can be done using Colombeau algebras. These algebras were developed \cite{9} to provide a context in which distributions could be multiplied. The spaces of distributions are subsets of these algebras. Roughly speaking, the solution to the multiplication problem is to record, instead of the distribution itself, all possible smoothings of the distribution, because there is no difficulty in multiplying smooth densities. As we will explain below, an equivalence relation is then introduced on a certain set of families of smooth functions, and its equivalence classes are the elements of the algebra.

**Construction.** To define the relevant Colombeau algebra, we follow \cite{4, Section 8.5}. Let $E(\mathbb{R}^n)$ be the set of families $(f_\varepsilon)_{0 < \varepsilon < 1}$ of functions $f_\varepsilon \in C^\infty(\mathbb{R}^n)$ indexed by $0 < \varepsilon < 1$, such that for each compact set $K \subset \mathbb{R}^n$ and every multi-index $I$ there are $N \in \mathbb{N}$, $\gamma > 0$, and $c > 0$ such that

$$\sup_{x \in K} |\partial^I f_\varepsilon(x)| \leq \frac{c}{\varepsilon^N} \text{ if } 0 < \varepsilon < \gamma.$$ 

We define the ideal $\mathcal{N}(\mathbb{R}^n)$ of $E(\mathbb{R}^n)$ to be the set of families $(f_\varepsilon)_{0 < \varepsilon < 1}$ such that for all compact sets $K$, for all multi-indices $I$, and for all $q \in \mathbb{N}$ there exist $c > 0$ and $\gamma > 0$ such that

$$\sup_{x \in K} |\partial^I f_\varepsilon(x)| \leq ce^q \text{ if } 0 < \varepsilon < \gamma.$$ 

This means that the elements of $\mathcal{N}(\mathbb{R}^n)$ have a fast decay (faster than any power of $\varepsilon$) as $\varepsilon \searrow 0$. The Colombeau algebra is the quotient

$$G(\mathbb{R}^n) = E(\mathbb{R}^n)/\mathcal{N}(\mathbb{R}^n).$$ 

All distributions $\eta$ are represented in $G(\mathbb{R}^n)$ because the families of smoothings by convolution with a mollifier are contained there. We denote by $[\eta]$ the set of elements of $G(\mathbb{R}^n)$ that are associated to the distribution $\eta$. That is, $(f_\varepsilon)_{\varepsilon}$ belongs to the subset $[\eta]$ of $G(\mathbb{R}^n)$ if, for all $\phi \in C^\infty_c(\mathbb{R}^n)$,

$$\lim_{\varepsilon \to 0^+} \int \phi(x) f_\varepsilon(x) \, dx = \langle \eta, \phi \rangle.$$ 

We now proceed to use the Colombeau algebra $G(\mathbb{R}^n)$ to define the object corresponding to the direction of the motion. Let $\mu$ be a probability measure on $\mathbb{R}^n$ and let $\eta$ be a distribution satisfying the conditions in Corollary \cite{14} including the last part for families of probability measures, so that there exists a family of probabilities $(\mu_t)_{t \in \mathbb{R}}$ with derivative $\eta$ at 0 and $\mu_0 = \mu$.

Let $[\mu]$ be the subset of $G(\mathbb{R}^n)$ associated to $\mu$. Note that $[\mu]$ always contains representatives $(f_\varepsilon)_{\varepsilon} \in [\mu]$ satisfying

$$\int_{\mathbb{R}^n} f_\varepsilon(x) \, dx = 1 \text{ and } f_\varepsilon \geq 0 \text{ for } 0 < \varepsilon < 1$$ 

because these can be obtained by convolution of $\mu$ with a mollifier of mass 1. Let $V(\mu)$ be the set of families $(v_\varepsilon)_{0 < \varepsilon < 1}$ of smooth vector fields such that, for all representatives satisfying $f_\varepsilon v_\varepsilon$, we have

$$(\text{div}(f_\varepsilon v_\varepsilon))_{0 < \varepsilon < 1} \in E(\mathbb{R}^n).$$
Note that the set $\mathcal{V}(\mu)$ is a vector space.

Choose a representative $(g_\varepsilon)_{0<\varepsilon<1}$ in the subset $[\eta]$ that additionally satisfies that the support of $g_\varepsilon$ is contained in the support of $f_\varepsilon$; by Remark 17 such a representative can always be constructed by convolution.

Consider the continuity equation
\[
g_\varepsilon + \text{div}(f_\varepsilon v_\varepsilon) = 0. \tag{7}
\]
The space $\mathcal{V}(\mu)$ always contains a solution $(v_\varepsilon)_{0<\varepsilon<1}$ to this equation; we know this because for each $\varepsilon$ this is just the classical smooth case, treated in [1, Chapter 8]. Accordingly, we may choose $v_\varepsilon$ to be minimal with respect to the $L^2$ norm, and it will correspond to a gradient vector field $\nabla \phi_\varepsilon$ for some functions $\phi_\varepsilon: \mathbb{R}^n \to \mathbb{R}$, $0<\varepsilon<1$.

The (non-unique) family $(v_\varepsilon)_{0<\varepsilon<1} \in \mathcal{V}(\mu)$ is the object that we have been pursuing, as it gives precise meaning to the notion of “direction of the movement” of $\mu$ when the change of the distribution of mass is given by $\eta$.

Remarks and examples. In [1], the authors make the choice of taking the tangent space to the set of probabilities to be the set of vector fields that arise as $v_\varepsilon$ in the continuity equation. What tangent space to take is a question that depends a bit on the intended application and a bit on personal taste. Our description above suggests an alternative in the form of a subset of the corresponding Colombeau algebra, and Theorem 4 suggests yet another alternative.

It would certainly be interesting to know which parts of the theory of mass transport still hold in the context of weakly-differentiable families of probabilities, a question that we leave for later research.

The following examples are intended to clarify the construction above and its interpretation.

Example 20. We go back to the situation of Example 8 and we compute a representative $(v_\varepsilon)_{\varepsilon}$ of the direction of movement object. Let $P = \mathbb{R}$, $\mu_t = \delta_0$, $\mu = \mu_0 = \delta_0$, $\eta = d\mu_t/dt|_{t=0} = -\partial_x \delta_0$. Let $\psi: \mathbb{R} \to [0,1]$ be a compactly-supported, $C^\infty$ function that is equal to 1 in a neighborhood of 0, and satisfies $\psi(x) = \psi(-x)$ and $\int_\mathbb{R} \psi(x)dx = 1$. Also let $\psi_\varepsilon(x) = \frac{1}{\varepsilon} \psi(x/\varepsilon)$, $f_\varepsilon = \psi_\varepsilon * \mu = \psi_\varepsilon$, and $g_\varepsilon = \psi_\varepsilon * \eta = -\psi_\varepsilon'$, where the last equality is true because for all test functions $\phi \in C_c^\infty(\mathbb{R})$, we have
\[
(g_\varepsilon, \phi) = (\eta, \psi_\varepsilon * \phi) = (-\partial_x \delta_0, \psi_\varepsilon * \phi) = (\delta_0, \partial_\varepsilon(\psi_\varepsilon * \phi)) = (\delta_0, (\partial_x \psi_\varepsilon) * \phi) = (\partial_x \psi_\varepsilon) * \phi(0) = (-\partial_x \psi_\varepsilon, \phi). \tag{8}
\]
Also, in dimension 1 the divergence is simply the derivative. Thus in this case the smoothed version (7) of the continuity equation (5) becomes
\[
(-\psi_\varepsilon' + (\psi_\varepsilon v_\varepsilon)' = 0.
\]
Taking $v_\varepsilon \equiv 1$ solves this equation. It also makes sense: this can be interpreted as movement to the right, which is exactly what the mass of the family $(\delta_t)_{t \in \mathbb{R}}$ is doing.

Example 21. We work out explicitly the same situation as discussed in Example 16, namely, $P = \mathbb{R}$, $U = (-\frac{1}{2}, \frac{1}{2})$, $\mu = \chi_U dx$ the uniform distribution on
$U, k \in \mathbb{N}, \eta = (-1)^k \partial^k \delta_0$. We will produce a representative $(v^\varepsilon)_\varepsilon$ of the subset of $G(\mathbb{R})$ that gives the notion of "direction of the movement" of the mass of $\mu$ when it is deformed in direction $\eta$.

Let $\psi_\varepsilon$ be as in Example 20 and $f_\varepsilon = \psi_\varepsilon \ast \mu$, the convolution. Note that for $0 < \varepsilon \ll 1$ small enough, we have that $f_\varepsilon \equiv 1$ in a neighborhood of 0 (in fact this is true in most of $U$) and $\psi_\varepsilon \equiv 0$ outside a small neighborhood of 0.

We also let $g_\varepsilon = \psi_\varepsilon \ast \eta$. For reasons analogous to those outlined in (5), we then have that $g_\varepsilon = (-1)^k \psi_\varepsilon^{(k)}$.

In this context, equation (7) takes the form

\[
(-1)^k \psi_\varepsilon^{(k)} = f_\varepsilon' v^\varepsilon + f_\varepsilon (v^\varepsilon)'.
\]

(9)

Close to 0, since $f_\varepsilon \equiv 1$, this becomes

\[
(-1)^k \psi_\varepsilon^{(k)} = (v^\varepsilon)'',
\]

so the solution in this region is obviously $v_\varepsilon = (-1)^k \psi_\varepsilon^{(k-1)}$. It is easy to see that this solution works globally since $\psi_\varepsilon \equiv 0$ slightly further away from zero. So $(v^\varepsilon)_\varepsilon = ((-1)^k \psi_\varepsilon^{(k-1)})_{0 < \varepsilon < 1}$ is the representative we were looking for.

References

[1] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. Gradient flows in metric spaces and in the space of probability measures. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[2] C. Carathéodory. Über den Variabilitätsbereich der Koeffizienten von Potenzreihen, die gegebene Werte nicht annehmen. Mathematische Annalen, 64(1):95–115, March 1907.

[3] Jean-François Colombeau. New generalized functions and multiplication of distributions, volume 84 of North-Holland Mathematics Studies. North-Holland Publishing Co., Amsterdam, 1984. Notas de Matemática [Mathematical Notes], 90.

[4] Jean-François Colombeau. Multiplication of distributions, volume 1532 of Lecture Notes in Mathematics. Springer-Verlag, Berlin, 1992. A tool in mathematics, numerical engineering and theoretical physics.

[5] Rodolfo Ríos-Zertuche. Distributions that arise as derivatives of families of measures. Preprint. arXiv:1408.3683 [math.PR].

[6] Rodolfo Ríos-Zertuche. The variational structure of the space of holonomic measures. Preprint. arXiv:1408.5785 [math.OC].

[7] François Trèves. Topological vector spaces, distributions and kernels. Dover Publications, Inc., Mineola, NY, 2006. Unabridged republication of the 1967 original.