Notes on Distribution Theory

Ben Bert

Abstract

In these notes we follow the book *Introduction to the Theory of Distributions* by Friedlander and Joshi [1]. The goal here is to define a functional of the form

$$F_f[\gamma] = \int f(x)\gamma(x) \tag{1}$$

and show that this functional is a distribution, and that every distribution of this form is uniquely determined by the function f. We will then define the Gâteaux derivative to be a distribution of the form of eq. (1) and show that the Gâteaux derivative naturally leads to a definition of the functional derivative.

Contents

1 Notations and basic definitions

Throughout these notes, the notation $\mathbb{R}$ will denote both the field of real numbers and the real line, and the letter $\mathbb{C}$ will stand for both the field of complex numbers and the complex plane. $\mathbb{R}^n$ will be understood to have its usual vector space structure and inner product, so that

$$cx = (cx_1, \ldots, cx_n) \quad \text{if} \quad c \in \mathbb{R} \quad \text{and} \quad x = (x_1, \ldots, x_n) \in \mathbb{R}^n,$$

and

$$x + y = (x_1 + y_1, \ldots, x_n + y_n), \quad x \cdot y = \sum_{j=1}^{n} x_j y_j$$

if $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$. The Euclidean norm $(x \cdot x)^{1/2}$ will be written as $|x|$. We always write $\mathbb{R}$ instead of $\mathbb{R}^1$.

If A is a subset of $\mathbb{R}^n$, then $\bar{A}$ is its closure and ∂A is its boundary. The closure of a set A is the smallest closed set containing A, and is can be written as $\bar{A} = A \cup \partial A$. A is closed if $A = \bar{A}$, *i.e.* if A contains its boundary. It will be recalled that $A \subset \mathbb{R}^n$ is compact if and only if it is both closed and bounded; if $X \subset \mathbb{R}^n$ is an open set then A is a compact subset of X if it is compact and $A \subset X$. If A and B are subsets of $\mathbb{R}^n$, we shall write $A \backslash B$ for the (set-theoretic) difference $\{x : x \in A, x \notin B\}$.

If X and Y are sets, and f is a function on X with range in Y, one writes $f : X \to Y$, and $x \mapsto y$ or $x \mapsto f(x)$ indicates that f maps $x \in X$ to $y = f(x) \in Y$. A map is injective if $f(x') = f(x'')$ implies that $x' = x''$, surjective if every $y \in Y$ is the image of some $x \in X$, and bijective if it is both injective and surjective.

Let $X \subset \mathbb{R}^n$ be an open set, and let k be a nonnegative (>0) integer. The class $C^k(X)$ consists of the complex valued functions on X which have continuous derivatives of order less than, or equal to, k. (The function itself is included, conventionally, as the 'derivative of order zero'.) Likewise, $C^\infty(X)$ consists of the functions which have continuous derivatives of all orders.

The support of a function $f : X \to \mathbb{C}$ is the closure of the set $\mathrm{supp} = \{x \in X : f(x) \neq 0\}$; note that it is a closed subset of X. Functions with compact support play an important part in the theory. We write $C_c^k(X)$ for the subset of $C^k(X)$ consisting of functions with compact support, and $C_c^\infty(X)$ for the subset of $C^\infty(X)$ consisting of functions with compact support. Note that $C^k(X), C^\infty(X), C_c^k(X)$ and $C_c^\infty(X)$ are all vector spaces over $\mathbb{C}$.

Integrals are Lebesgue integrals. When dealing with functions defined on a fixed open subset X of $\mathbb{R}^n$, we omit the domain of integration, so that

$$\int dx\, f(x) = \int_X dx\, f(x)$$

by definition; here, dx is Lebesgue measure.

Let $X \subset \mathbb{R}^n$ be an open set, and $f \in C^\infty(X)$. We shall usually write the derivatives as

$$\partial_j f = \partial f / \partial x_j, \quad j = 1, \ldots, n.$$

Derivatives of higher order can be written concisely by means of the multi-index notation. A multi-index (or, to be precise, an n-multi-index) is an n-tuplet $\alpha = (\alpha_1, \ldots, \alpha_n)$ of nonnegative integers; its length (or order) is $|\alpha| = \alpha_1 + \cdots + \alpha_n$. The sum of two multi-indices α and β is $\alpha + \beta = (\alpha_1 + \beta_1, \ldots, \alpha_n + \beta_n)$. One says that $\beta \leq \alpha$ if $\beta_j \leq \alpha_j$ for $j = 1, \ldots, n$; when $\beta \leq \alpha$ one can also define $\alpha - \beta = (\alpha_1 - \beta_1, \ldots, \alpha_n - \beta_n)$. One now sets

$$\partial^\alpha f = \partial^{|\alpha|} f / \partial x_1^{\alpha_1} \ldots \partial x_n^{\alpha_n},$$

so that

$$\partial^\alpha \partial^\beta f = \partial^{\alpha+\beta} f$$

This obviously also applies to $f \in C^k(X)$, provided that $|\alpha + \beta| \leq k$. To complete the multi-index notation, we put

$$\alpha! = \alpha_1! \ldots \alpha_n!$$

for any multi-index α, and if also $x \in \mathbb{R}^n$ we set

$$x^\alpha = x_1^{\alpha_1} \ldots x_n^{\alpha_n}.$$

The formal statement of Taylor's theorem then becomes

$$f(x+h) = \sum_{\alpha \geqslant 0} \frac{x^\alpha}{\alpha!} \partial^\alpha f(h),$$

and the multinomial theorem assumes the concise form

$$(x_1 + \cdots + x_n)^m = \sum_{|\alpha| = m} \frac{m!}{\alpha!} x^\alpha.$$

2 Locally convex topological vector spaces

Definition 2.1: Topological Space

Let X be a set. A *topology* on X is a set T of subsets of X satisfying the following properties:

1. $\varnothing \in T$ and $X \in T$;
2. The union of any subsets of T belongs to T, *i.e.*, if $\{U_\alpha\}_{\alpha \in A} \subseteq T$, then $\bigcup_{\alpha \in A} U_\alpha \in T$;
3. The intersection of any finite number of members of T belongs to T, *i.e.*, if $U_1, U_2, \ldots, U_n \in T$, then $\bigcap_{i=1}^n U_i \in T$.

The pair (X, T) is called a *topological space*, and the members of T are called *open sets*. Often, one omits explicit reference to T and refers simply to the open sets of X.

Definition 2.2: Neighbourhood

Let X be a topological space and let $x \in X$.

- An *open neighbourhood* of x is an open set $U \subseteq X$ such that $x \in U$.
- A *neighbourhood* of x is any set $V \subseteq X$ that contains an open neighbourhood of x, *i.e.*, there exists an open set U with $x \in U \subseteq V$.

Remark 2.1

Suppose $S = \{S_\alpha\}_{\alpha \in A}$ is a family of subsets of X (S has no other special properties). One can always find a topology T on X which contains S as a subset and for which each S_α is open. For example, one can take T to be the set of all subsets of X, *i.e.*, the discrete topology denoted as T_d. T_d can be shown to be a topology. Clearly S is a subset of the discrete topology of X and all S_α are open.

Definition 2.3: Generation of a Topology

Let S be a family of subsets of X. The *topology generated by* S is the smallest topology on X that contains S as a subset.

Lemma 2.1

Let X be a set, and let $S = \{S_\alpha\}$ be a family of subsets of X. Let A be the set of all topologies on X

that contain every $S_\alpha \in S$ as an open set. Then the intersection

$$\tau := \bigcap_{T \in A} T$$

is the smallest topology on X that contains S as a subset *i.e.*, τ is the topology generated by S.

Proof. Set $\tau := \bigcap_{T \in A} T$.

We verify the topology axioms:

1. **Contains $\varnothing$ and X:** Every topology in A contains $\varnothing$ and X by definition. Therefore, $\varnothing, X \in \tau$.
2. **Closed under arbitrary unions:** Let $\{U_\beta\}_{\beta \in B} \subseteq \tau$. Then each $U_\beta \in T$ for all $T \in A$. Since each T is a topology, $\bigcup_{\beta \in B} U_\beta \in T$ for every $T \in A$. Hence,

$$\bigcup_{\beta \in B} U_\beta \in \tau.$$

3. **Closed under finite intersections:** Let $U_1, \ldots, U_n \in \tau$. Each U_i is in every $T \in A$, and each T is closed under finite intersections. Therefore,

$$U_1 \cap \cdots \cap U_n \in T \quad \text{for all } T \in A,$$

which implies

$$U_1 \cap \cdots \cap U_n \in \tau.$$

Thus, τ satisfies all the topology axioms, so it is a topology.

Smallest topology property:

- *τ contains S:* Every $T \in A$ contains each S_α, so τ contains S as well.
- *τ is contained in every topology in A:* By definition, $\tau = \bigcap_{T \in A} T$, so $\tau \subseteq T$ for all $T \in A$. Therefore, no topology in A is smaller than τ.

Hence, τ is the smallest topology containing S as a subset and making each S_α open. □

Definition 2.4: Base of a Topology

If every member of the topology generated by S is a union of members of S, then S is called a base of this topology. Different sets S, or different bases, may generate the same topology.

Definition 2.5: Continuous Function

Let X and Y be topological spaces, and let $f : X \to Y$ be a function.

- f is *continuous at a point* $x \in X$ if for every open neighbourhood $V \subseteq Y$ of $f(x)$, the inverse image $f^{-1}(V)$ is an open neighbourhood of x in X.
- f is *continuous on* X if it is continuous at every point $x \in X$.

Equivalently, f is continuous on X if and only if for every open set $V \subseteq Y$, the inverse image $f^{-1}(V) \subseteq X$ is open.

Definition 2.6: Convergence of a Sequence

Let X be a topological space, and let $(x_j)_{j=1}^{\infty}$ be a sequence in X. The sequence (x_j) *converges* to a point $x \in X$ (written $x_j \to x$ as $j \to \infty$) if for every neighbourhood U of x, there exists $N \in \mathbb{N}$ such that

$$x_j \in U \quad \text{for all } j \geq N.$$

Equivalently, every neighbourhood of x contains all but finitely many terms of the sequence.

Theorem 2.1

If $f : X \to Y$ is a continuous function and the sequence (x_j) converges to x in X, then the sequence $(f(x_j))$ converges to $f(x)$ in Y.

Proof. We are to prove that the sequence $(f(x_j))$ converges to $f(x)$ in Y. Thus we are to show that for every neighbourhood V of $f(x)$, there exists $N \in \mathbb{N}$ such that $f(x_j) \in V$ for all $j \geq N$.

We shall thus let V be an arbitrary neighbourhood of $f(x)$ in Y. Since f is continuous at x, we know that for the neighbourhood V, $f^{-1}(V) = U$ is a neighbourhood of x in X. Now since (x_j) converges to x in X, we know that for the neighbourhood U of x, there exists $N \in \mathbb{N}$ such that $x_j \in U$ for all $j \geq N$, *i.e.* $x_j \in f^{-1}(V)$ for all $j \geq N$. But this means that $f(x_j) \in V$ for all $j \geq N$, and so the sequence $(f(x_j))$ converges to $f(x)$ in Y. □

Definition 2.7: Hausdorff (Separated) Space

A topological space X is called *Hausdorff* (or *separated*) if it satisfies the following *Hausdorff separation axiom*: For any two distinct points $x, y \in X$, there exist neighbourhoods U of x and V of y such that

$$U \cap V = \varnothing.$$

In a Hausdorff space, the limit of a convergent sequence is unique.

Definition 2.8: Metric Space

A *metric space* is a set X together with a *distance function* (or *metric*)

$$d : X \times X \to \mathbb{R}$$

satisfying, for all $x, y, z \in X$:

1. **Non-negativity and identity of indiscernibles:**

$$d(x, y) \geq 0, \quad d(x, y) = 0 \iff x = y$$

2. **Triangle inequality:**

$$d(x, y) \leq d(x, z) + d(z, y)$$

Remark 2.2: Hausdorff base

In a metric space (X, d), the *open balls*

$$B_r(y) := \{x \in X \mid d(x, y) < r\}, \quad y \in X,\ r > 0,$$

form a base for a Hausdorff topology on X. It is customary to equip a metric space with the Hausdorff topology generated by this base.

Definition 2.9: Cauchy Sequence

Let (X, d) be a metric space. A sequence $(x_j)_{j=1}^{\infty}$ in X is called a *Cauchy sequence* if

$$d(x_j, x_k) \to 0 \quad \text{as } j, k \to \infty,$$

Remark 2.3

Since a metric maps from $X \times X$ to $\mathbb{R}^+$, then $d(x_j, x_k) \to 0$ as $j, k \to \infty$, means that for any neighbourhood of 0 in $\mathbb{R}^+$, there exist $N_1, N_2 \in \mathbb{N}$ such that for all $j \geq N_1$ and $k \geq N_2$, $d(x_j, x_k) \in U$. But any neighbourhood of 0 in $\mathbb{R}^+$ contains an interval of the form $[0, \varepsilon)$ for some $\varepsilon > 0$. Thus, the condition that $d(x_j, x_k) \to 0$ as $j, k \to \infty$ is equivalent to the condition that for every $\varepsilon > 0$, there exist $N_1, N_2 \in \mathbb{N}$ such that for all $j \geq N_1$ and $k \geq N_2$, $d(x_j, x_k) < \varepsilon$. In particular, if we take $N = \max\{N_1, N_2\}$, then we have that for every $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $j, k \geq N$, $d(x_j, x_k) < \varepsilon$.

Definition 2.10: Complete Metric Space

A metric space (X, d) is called *complete* if every Cauchy sequence in X converges to a point in X.

Remark 2.4: Topology induced by a metric

Let X be a set and let $d : X \times X \to \mathbb{R}$ be a metric. For each $y \in X$ and $r > 0$, define the open ball

$$B_r(y) = \{x \in X : d(x, y) < r\}.$$

The collection of all such balls

$$\mathcal{B} = \{B_r(y) : y \in X,\ r > 0\}$$

forms a base for a topology on X. The topology generated by this base consists of all unions of such balls,

$$\tau_d = \left\{ \bigcup_{i \in I} B_{r_i}(y_i) \right\}.$$

When a topological space (X, τ) satisfies $\tau = \tau_d$ for some metric d, the topology is said to be *derived from the distance function* d, and the space is called *metrizable*.

Definition 2.11: Topological Vector Space

Let X be a vector space over $\mathbf{C}$. A *topological vector space* is a vector space X equipped with a topology such that the vector space operations are continuous. More precisely:

- The *addition map*

  $$+ : X \times X \to X, \qquad (x, y) \mapsto x + y$$

  is continuous.

- The *scalar multiplication map*

  $$\cdot : \mathbf{C} \times X \to X, \qquad (\lambda, x) \mapsto \lambda x$$

  is continuous.

Definition 2.12: Base of Neighbourhoods

Let X be a vector space and let 0 be the zero vector in X. A *base of neighbourhoods* is a family $\mathcal{B} = \{B_i\}_{i \in I}$ of neighbourhoods of 0 such that any neighbourhood of 0 contains at least one member of $\mathcal{B}$, *i.e.*, for every neighbourhood V of 0, there exists $i \in I$ such that $B_i \subseteq V$.

Theorem 2.2: Open Sets in Topological Vector Spaces

A set $G \subset X$ where X is a topological vector space is open if and only if, for every $x \in G$, there is a $U \subset B$ such that $x + U \subset G$, where $x + U = \{y \in X : y - x \in U\}$.

3 Test functions and distributions

Definition 3.1: Test functions

A test function is a function $\phi \in C_c^\infty(\mathbb{R}^n)$.

Definition 3.2: Linear form

If V is a vector space over the field C of complex numbers, then a linear form $\mathscr{L}$ on V is a homomorphism $V \to \mathbb{C}$. In other words, $\mathscr{L} : V \to \mathbb{C}$ is a linear form if

$$\mathscr{L}(\alpha v + \beta w) = \alpha \mathscr{L}(v) + \beta \mathscr{L}(w) \tag{3.1}$$

for all $v, w \in V$ and $\alpha, \beta \in \mathbb{C}$.

Remark 3.1

In what follows, we label our linear forms as $\langle u, \cdot \rangle$, where u is a symbol that serves to identify the linear form. The notation $\langle u, \cdot \rangle$ is meant to suggest that the linear form is a kind of inner product, and so $\langle u, \phi \rangle$ is the value of the linear form $\langle u, \cdot \rangle$ at the vector ϕ. This notation is purely conventional, and does not imply that there is any inner product structure on the vector space on which the linear form is defined.

Definition 3.3: Distribution

Let $X \subset \mathbb{R}^n$ be an open set. A linear form $u : C_c^\infty(X) \to \mathbb{C}$ is called a distribution if, for every compact set $K \subset X$, there is a real number $C \geqslant 0$ and a nonnegative integer N such that

$$|\langle u, \phi \rangle| \leq C \sum_{|\alpha| \leq N} \sup |\partial^\alpha \phi| \,, \tag{3.2}$$

for all $\phi \in C_c^\infty(X)$ with $\operatorname{supp} \phi \subset K$. The vector space of distributions on X is called $\mathscr{D}'(X)$.

Remark 3.2: Boundedness of physical actions and distributions

Distributions and action functionals are different kinds of objects, and the boundedness in the distribution definition is not in tension with unbounded actions — they live at different levels.

Distributions act on test functions, not on fields. A distribution $u \in \mathscr{D}'(X)$ takes a test function $\phi \in C_c^\infty(X)$ as input and returns a number. The boundedness condition is boundedness in ϕ, on each compact set. This is a relatively mild condition — it says u cannot be infinitely sensitive to high-frequency oscillations in ϕ.

Actions act on field configurations. The action $S[\varphi]$ takes an entire field φ as input, a functional on an infinite-dimensional space of fields. There is no reason it should be bounded — indeed $S[\varphi] \to \infty$ as $\|\varphi\| \to \infty$ is typically desirable for the variational problem to be well-posed.

Theorem 3.1

Let $X \subset \mathbb{R}^n$ be an open set, and let $f \in C^0(X)$. Then

$$\langle f, \phi \rangle_{L^2} \equiv \int_X dx\, f\phi, \quad \phi \in C_c^\infty(X) \tag{3.3}$$

is a distribution. Furthermore, if the second member (*i.e.* RHS) of eq. (3.3) vanishes for all $\phi \in C_c^\infty(X)$, then $f = 0$ on X.

Proof. Let $K \subset X$ be a compact set and suppose support of the test function ϕ is contained in K, then $\phi(x) = 0$ for all x in $X \backslash K$, and so

$$\begin{aligned}
|\langle f, \phi \rangle_{L^2}| &= |\int_X dx\, f\phi| && (3.4a)\\
&\leq \int_X dx\, |f(x)\phi(x)| && (3.4b)\\
&= \int_K dx\, |f(x)\phi(x)| && (3.4c)\\
&= \int_K dx\, |f(x)||\phi(x)| && (3.4d)\\
&\leq \sup_K |\phi| \int_K dx\, |f(x)|, && (3.4e)
\end{aligned}$$

Now both $\phi(x)$ and $f(x)$ are continuous functions, and K is a compact set, so $\sup_K |\phi|$ and $\int_K |f(x)| dx$ are finite, and so we can write

$$|\langle f, \phi \rangle_{L^2}| \leq C \sup_K |\phi|, \quad \phi \in C_c^\infty(X), \tag{3.5}$$

where C is a real number that depends on f and K. Thus there exists an integer $N > 0$ (namely $N = 1$) such that

$$|\langle f, \phi \rangle_{L^2}| \leq C \sum_{|\alpha| \leq N} \sup_K |\partial^\alpha \phi|, \quad \phi \in C_c^\infty(X), \tag{3.6}$$

and so $\langle f, \cdot \rangle_{L^2}$ is a distribution. Now, by way of contradiction, suppose that the second member of eq. (3.3) vanishes for all $\phi \in C_c^\infty(X)$, but that $f(y) \neq 0$ at some point $y \in X$. Since f is continuous and because $f(y) \neq 0$, then the function $h(x) = \mathrm{Re}\,(f(x)/f(y))$ is also continuous and $h(y) = 1$. By the continuity of h, for all $\epsilon' > 0$ there exists a $\delta' > 0$ such that $|h(x) - h(y)| < \epsilon'$ if $|x - y| < \delta'$. In particular, if we choose $\epsilon' = \frac{1}{2}$, then there exists a $\delta' > 0$ such that $|h(x) - 1| < \frac{1}{2}$ if $|x - y| < \delta'$, which is to say $-\frac{1}{2} < h(x) - 1 < \frac{1}{2}$ which then implies $h(x) > \frac{1}{2}$. Thus there always exists a $\delta' > 0$ such that $\mathrm{Re}(f(x)/f(y)) > \frac{1}{2}$ provided that $|x - y| < \delta'$. Now, choose $\delta \leq \delta'$ such that $\{x : |x - y| \leq \delta\} \subset X$. Now, let $\rho \in C_c^\infty(\mathbb{R}^n)$ be such that

$$\rho \geqslant 0, \quad \mathrm{supp}\, \rho \subset \{|x| \leq 1\}, \quad \int dx\, \rho = 1, \tag{3.7}$$

Define $\phi(x) = \rho\left(\frac{x-y}{\delta}\right)/f(y)$, which is a test function. Then by the assumption that the second member of eq. (3.3) vanishes for all $\phi \in C_c^\infty(X)$, we have

$$0 = \int_X dx\, f(x)\phi(x) \tag{3.8a}$$

$$= \int_X dx\, f(x)\frac{\rho\left(\frac{x-y}{\delta}\right)}{f(y)} \tag{3.8b}$$

$$\implies 0 = \int_X dx\, \mathrm{Re}\left(\frac{f(x)}{f(y)}\right)\rho\left(\frac{x-y}{\delta}\right) \tag{3.8c}$$

Since $\rho(x)$ only has support for $|x| \leq 1$, then $\rho\left(\frac{x-y}{\delta}\right)$ only has support for $|x-y| \leq \delta$, and so

$$0 = \int_{|x-y|\leq\delta} dx\, \mathrm{Re}\left(\frac{f(x)}{f(y)}\right)\rho\left(\frac{x-y}{\delta}\right) \tag{3.9a}$$

$$> \int_{|x-y|\leq\delta} dx\, \frac{1}{2}\rho\left(\frac{x-y}{\delta}\right) \tag{3.9b}$$

$$= \frac{1}{2}\int_{|x-y|\leq\delta} dx\, \rho\left(\frac{x-y}{\delta}\right) \tag{3.9c}$$

$$= \frac{1}{2}\delta^n, \tag{3.9d}$$

which is a contradiction. So the theorem is proved. □

Theorem 3.2: Uniqueness of the distribution $\langle f, \cdot\rangle_{L^2}$

Every distribution of the form eq. (3.3) is uniquely determined by the function f.

Proof. Let f and g be functions in $C^0(X)$ such that

$$\langle f, \phi\rangle_{L^2} = \langle g, \phi\rangle_{L^2}, \quad \phi \in C_c^\infty(X), \tag{3.10}$$

then

$$\int dx\, (f-g)\phi = 0, \quad \phi \in C_c^\infty(X), \tag{3.11}$$

and so $f - g = 0$ on X by theorem 3.1, which is to say that $f = g$ on X. □

Example 3.1: Examples of distributions

From theorem 3.1, any continuous function defines a unique distribution. In particular, the following are distributions:

$$\langle 1, \phi \rangle_{L^2} = \int_{\mathbb{R}} dx\ \phi(x) \tag{3.12a}$$

$$\langle x, \phi \rangle_{L^2} = \int_{\mathbb{R}} dx\ x\phi(x) \tag{3.12b}$$

$$\langle e^x, \phi \rangle_{L^2} = \int_{\mathbb{R}} dx\ e^x\phi(x) \tag{3.12c}$$

$$\langle 1/x, \phi \rangle_{L^2} = \int_{\mathbb{R}\setminus\{0\}} dx\ \frac{1}{x}\phi(x) \tag{3.12d}$$

A particularly useful example of a distribution is the Dirac delta distribution δ, which is defined by

$$\langle \delta, \phi \rangle \equiv \phi(0), \quad \phi \in C_c^\infty(\mathbb{R}). \tag{3.13}$$

That $\langle \delta, \cdot \rangle$ defines a distribution does not follow from theorem 3.1, but it can be shown by direct verification of eq. (3.2).

Remark 3.3: Local integrability as a weaker condition for distributions

It follows immediately from eq. (3.4e) that the expression

$$\langle f, \phi \rangle = \int_X f(x)\phi(x)\, dx$$

defines a distribution on X whenever f is *locally integrable*, that is, when f is measurable and

$$\int_K |f(x)|\, dx < \infty$$

for every compact set $K \subset X$. Such a distribution does not determine f uniquely. Indeed, the above pairing is unchanged if f is replaced by a function g satisfying $f = g$ almost everywhere. This means that the set

$$\{x \in X : f(x) - g(x) \neq 0\}$$

has measure zero. A set is said to have measure zero if, for every $\varepsilon > 0$, it can be covered by a countable collection of rectangles whose total volume is less than ε. Nevertheless, one can show that if f is locally integrable and

$$\int_X f(x)\phi(x)\, dx = 0$$

for all $\phi \in C_c^\infty(X)$, then $f = 0$ almost everywhere on X.

4 Functional derivatives

Definition 4.1: Functional

A functional is a map $F : C^0(X) \to \mathbb{R}$ where $X \subset \mathbb{R}^n$ is an open set.

Definition 4.2: Gâteaux Derivative

Let $F : C^0(X) \to \mathbb{R}$ be a functional, and let $f \in C^0(X)$. Now, suppose $D_G F : C_c^\infty(X) \to \mathbb{R}$ is a distribution in the form of eq. (3.3) defined by

$$D_G F[\phi] \equiv \int_X dx \, \frac{\partial F}{\partial f(x)} \phi(x) \tag{4.1}$$

where $\partial F / \partial f(x)$ is a function in $C^0(X)$. If for all test functions $\phi \in C_c^\infty(X)$, the following limit exists and is equal to zero,

$$\lim_{\epsilon \to 0} \frac{F[f + \epsilon\phi] - F[f] - \epsilon D_G F[\phi]}{\epsilon} = 0, \tag{4.2}$$

then $D_G F$ is called the Gâteaux derivative of F with respect to f in the direction of ϕ.

Remark 4.1

We call $\partial F / \partial f(x)$ the functional derivative of F with respect to f at the point x. By theorem 3.1, eq. (4.1) defines the Gâteaux derivative as a distribution, and by theorem 3.2, the Gâteaux derivative is uniquely determined by the functional derivative $\partial F / \partial f(x)$.

Example 4.1: Electrodynamics, Functional Derivatives and Gâteaux Derivatives

Consider the Lagrangian of the electromagnetic field in vacuum:

$$L[A^\mu, \partial_\nu A^\mu] = \int_{\mathbb{R}^3} d^3x \;\; \mathscr{L} = -\frac{1}{4} \sum_{\mu,\nu} \int_{\mathbb{R}^3} d^3x \;\; F_{\mu\nu} F^{\mu\nu}, \tag{4.3}$$

with $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$. Let us suppose that the electromagnetic field we are studying corresponds to static point particle. In this case, the field configuration $A^\mu(x)$ is given by the Coulomb potential,

$$A^0(x) = \frac{1}{4\pi |\vec{x}|}, \quad A^i(x) = 0. \tag{4.4}$$

The Lagrangian of the electromagnetic field from eq. (4.3) evaluated on the Coulomb potential is then given by

$$
\begin{aligned}
L[A^\mu, \partial_\nu A^\mu] &= -\frac{1}{4}\sum_{\mu,\nu}\int_{\mathbb{R}^3} d^3x \;\; F_{\mu\nu}F^{\mu\nu} && \text{(4.5a)}\\
&= -\frac{1}{4}\sum_{i}\int_{\mathbb{R}^3} d^3x \;\; 2\partial_i A_0 \partial^i A^0 && \text{(4.5b)}\\
&= -\frac{1}{2}\sum_{i}\int_{\mathbb{R}^3} d^3x \;\; \partial_i A_0 \partial^i A^0 && \text{(4.5c)}\\
&= \frac{1}{2}\int_{\mathbb{R}^3} d^3x \;\; \nabla A_0 \cdot \nabla A_0 && \text{(4.5d)}\\
&= \frac{1}{2}\int_{\mathbb{R}^3} d^3x \;\; \nabla A_0 \cdot \nabla A_0 && \text{(4.5e)}\\
&= \frac{1}{2}\int_{\mathbb{R}^3} d^3x \;\; \frac{1}{16\pi^2|\vec{x}|^4} && \text{(4.5f)}\\
&= \frac{1}{32\pi^2}\int_0^{2\pi} d\phi \int_0^{\pi} d\theta \sin\theta \int_0^{\infty} dr\; r^2 \frac{1}{r^4} && \text{(4.5g)}\\
&= \frac{1}{8\pi}\int_0^{\infty} dr\; \frac{1}{r^2}. && \text{(4.5h)}
\end{aligned}
$$

We see that the Lagrangian of the electromagnetic field evaluated on the Coulomb potential is ultraviolet divergent.

We now compute the Gâteaux derivative of the Lagrangian of the electromagnetic field with respect to A^μ at the Coulomb potential in the direction of a test function ϕ^μ. We have

$$
\begin{aligned}
D_G L[\phi^\mu] &= \left.\frac{\partial L[A^\mu + \epsilon\phi^\mu, \partial_\nu A^\mu]}{\partial\epsilon}\right|_{\epsilon=0} && \text{(4.6a)}\\
&= 0 && \text{(4.6b)}
\end{aligned}
$$

Next we compute the Gâteaux derivative of the Lagrangian of the electromagnetic field with respect to $\partial_\nu A^\mu$ at the Coulomb potential in the direction of a test function $\phi^{\mu\nu}$. We have

$$D_G L[\phi^{\mu\nu}] = \frac{\partial L[A^\mu, \partial^\nu A^\mu + \epsilon\phi^{\mu\nu}]}{\partial\epsilon}\bigg|_{\epsilon=0} \tag{4.7a}$$

$$= -\frac{1}{4}\sum_{\rho,\sigma}\int_{\mathbb{R}^3} d^3x \;\frac{\partial}{\partial\epsilon}\bigg|_{\epsilon=0} F_{\epsilon,\rho\sigma}F_\epsilon^{\rho\sigma} \tag{4.7b}$$

$$= -\frac{1}{2}\sum_{\rho,\sigma}\int_{\mathbb{R}^3} d^3x \;F_{\rho\sigma}\frac{\partial}{\partial\epsilon}\bigg|_{\epsilon=0} F_\epsilon^{\rho\sigma} \tag{4.7c}$$

$$= -\frac{1}{2}\sum_{\rho,\sigma}\int_{\mathbb{R}^3} d^3x \;F_{\rho\sigma}(\frac{\partial}{\partial\epsilon}\bigg|_{\epsilon=0}\partial^\rho A_\epsilon^\sigma - \frac{\partial}{\partial\epsilon}\bigg|_{\epsilon=0}\partial^\sigma A_\epsilon^\rho) \tag{4.7d}$$

$$= -\frac{1}{2}\sum_{\rho,\sigma}\int_{\mathbb{R}^3} d^3x \;F_{\rho\sigma}(\delta^\rho_\mu\delta^\sigma_\nu - \delta^\rho_\nu\delta^\sigma_\mu)\phi^{\mu\nu} \tag{4.7e}$$

$$= -\frac{1}{2}\int_{\mathbb{R}^3} d^3x \,(F_{\mu\nu} - F_{\nu\mu})\phi^{\mu\nu} \tag{4.7f}$$

$$= \int_{\mathbb{R}^3} d^3x \,F_{\nu\mu}\phi^{\mu\nu} \tag{4.7g}$$

From eq. (4.7g), we can read off the functional derivative of the Lagrangian of the electromagnetic field with respect to $\partial^\nu A^\mu$ as

$$\frac{\partial L}{\partial(\partial^\nu A^\mu(x))} = F_{\nu\mu}(x). \tag{4.8}$$

Now, using the Coulomb potential, we have

$$D_G L[\phi^{\mu\nu}] = \int_{\mathbb{R}^3} d^3x \,(\partial_\nu A_\mu - \partial_\mu A_\nu)\phi^{\mu\nu} \tag{4.9a}$$

$$= \int_{\mathbb{R}^3} d^3x \,\partial_\nu A_\mu\phi^{\mu\nu} - \partial_\mu A_\nu\phi^{\mu\nu} \tag{4.9b}$$

$$\tag{4.9c}$$

The only non-zero terms are from ϕ^{i0} and ϕ^{0i}, computing $D_G L[\phi^{i0}]$ we get

$$D_G L[\phi^{i0}] = \int_{\mathbb{R}^3} d^3x \,\partial_0 A_i\phi^{i0} - \partial_i A_0\phi^{i0} \tag{4.10a}$$

$$= -\int_{\mathbb{R}^3} d^3x \,\partial_i A_0\phi^{i0} \tag{4.10b}$$

$$= \int_{\mathbb{R}^3} d^3x \;\frac{x_i}{|\vec{x}|^3}\phi^{i0} \tag{4.10c}$$

$$\sim \int_0^\epsilon dr\; r^2\frac{1}{r^2} \qquad \text{(for small } r\text{)} \tag{4.10d}$$

$$< \infty \tag{4.10e}$$

We see that the Gâteaux derivative of the Lagrangian of the electromagnetic field with respect to $\partial_\nu A^\mu$ at the Coulomb potential in the direction of a test function $\phi^{\mu\nu}$ can be written as

$$D_G L[\phi^{i0}] = \langle f, \phi^{i0}\rangle_{L^2} = \int_{\mathbb{R}^3} d^3x \;\frac{x_i}{|\vec{x}|^3}\phi^{i0}, \tag{4.11}$$

with $f(\vec{x}) = x_i/|\vec{x}|^3$, which is locally integrable, and thus the Gâteaux derivative of the Lagrangian of the electromagnetic field with respect to $\partial_\nu A^\mu$ at the Coulomb potential in the direction of a test function $\phi^{\mu\nu}$ is a distribution.

If we had chosen $A^0 \sim 1/|\vec{x}|^\alpha$, we would have obtained

$$D_G L[\phi^{i0}] \sim \int_{\mathbb{R}^3} d^3x \, \frac{1}{|\vec{x}|^{2+\alpha}} \phi^{i0}, \tag{4.12}$$

which would not define a distribution for $\alpha > 0$. Equation (4.12) is not a distribution because the function $1/|\vec{x}|^{2+\alpha}$ is not locally integrable for $\alpha > 0$.

This example shows that even if a functional is unbounded, its Gâteaux derivative can still be a distribution, provided that the Gâteaux is evaluated at a suitable function.

Note that we can make the same conclusions for the Gâteaux derivative of the action of the electromagnetic field, since the action is

$$S[A^\mu, \partial_\nu A^\mu] = \int dt \; L[A^\mu, \partial_\nu A^\mu], \tag{4.13}$$

and thus

$$D_G S[\phi^{i0}] = \int dt \; D_G L[\phi^{i0}] \tag{4.14a}$$

$$= -\int_{\infty}^{\infty} dt \int_{\mathbb{R}^3} d^3x \; \partial_i A_0(\vec{x}) \phi^{i0}(\vec{x}, t). \tag{4.14b}$$

Naively, one might worry that the integral over time might diverge. However, if $\partial_i A_0$ is locally integrable over $\mathbb{R}^3$ then the integral over space is finite. We thus obtain

$$D_G S[\phi^{i0}] = \int_{\infty}^{\infty} dt \; g\phi^{i0}(t) \tag{4.15a}$$

where g is a finite number and is independent of time. Since the integral of a constant over a time domain with compact support is finite, we see that g is locally integrable over the time domain, and thus $D_G S[\phi^{i0}]$ is a distribution.

References

[1] F. G. Friedlander and M. S. Joshi. *Introduction to the Theory of Distributions*. Cambridge University Press, Cambridge, 2 edition, 1998.