FUNCTIONAL ANALYSIS AND DISTRIBUTION THEORY: FROM QUANTUM MECHANIC TO MACHINE LEARNING

LEANNE DONG

Contents

1. Representation theorems

We will be using the notion of "linear functionals" throughout. These are simply linear mapping of from a vector space to the real numbers. More precisely, Let X be a vector space. A linear functional on X is a linear mapping

$$\Lambda : X \to \mathbb{R}$$

If $\Lambda(x_n) \to \Lambda(x)$ whenever $x_n \to x$ we say that Λ is a continuous linear functional on X.

Example 1. Integration is a linear functional. To see this, Let $X = C([a, b])$. Then $\Lambda(f) = \int_a^b f(x)\,ds$ is a linear functional on X

Theorem 1.1. *Let F be a bounded linear functional on H. Then there exists an unique element $x_0 \in H$ such that for all $x \in H$, $F(x) = (x, x_0)$, and in fact $\|F\| = \|x_0\|$.*

2. Weak derivatives

The notion of weak or distributional derivatives forms the basis of PDE theory. In essence, the theory allows one to ignore some of the more "problematic points" from classical differential calculus. For instance, take the function $f : [0, 2] \to \mathbb{R}$ defined by

$$f(x) := \begin{cases} x, & \text{if } x \in [0, 1] \\ 1, & \text{if } x \in [1, 2] \end{cases}$$

This has the problem of "cusp" where we fail to define a tangent line. However, if we were able to "ignore" the point at 0, then we might want to say that in some sense

$$f'(x) := \begin{cases} 1, & \text{if } x \in [0, 1] \\ 0, & \text{if } x \in [1, 2] \end{cases}$$

This is how the weak derivative comes into play.

Date: November 17, 2019.

Let build upon our intuition into this subject. Let us start with the 1D case. Let $u,\ v : [0,1] \to \mathbb{R}$ be two differential function. Let $v(0), v(1)$. Then, integration by parts tell us that

$$\int_0^1 u(x)v'(x)\,dx = -\int_0^1 u'(x)v(x)\,dx$$

since the boundary term of $u(x)v(x)$ disappears. Even if u is not differentiable we might be able to make sense of the above formula.

Definition 2.1. Let $u : [0,1] \to \mathbb{R}$ be any real-valued function. We say that $w : [0,1] \to \mathbb{R}$ is the weak derivative of u if for every differentiable function $v : [0,1] \to \mathbb{R}$ with $v(0) = v(1) = 0$, one has that

$$\int_0^1 u(x)v'(x)\,dx = -\int_0^1 w(x)v(x)\,dx$$

To define the n-dimensional weak derivative, we need to do more work.

Suppose Ω is an open and connected region in $\mathbb{R}^n$. That is, for $\varphi \in C_0^\infty(\Omega)$, the support of φ defined by $\text{spt}(\varphi) := \overline{\{x \in \Omega : \varphi(x) \neq 0\}}$ is compact in Ω. We will refer the space as our "test function" space, denoted as $\mathcal{D}(\mathbb{R})$ which as discussed earlier.

Let $\alpha := \{\alpha_1, \cdots, \alpha_n\} \in \mathbb{Z}^n_{\geq 0}$ be multi-index. Then, for any $\varphi \in C^\infty(\mathbb{R}^n)$, one defines the differential operator D^α by

$$D^\alpha := \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_n}}{\partial x_n^{\alpha_n}}$$

We are now can define the multidimensional weak derivative.

Definition 2.2. Suppose $u : \Omega \to \mathbb{R}$ is given. Then, we say that $v : \Omega \to \mathbb{R}$ is the $\alpha-$weak derivative of u for some multi-index α, if for each $\varphi \in C_0^\infty(\Omega)$, the following integration by part formula holds:

$$\int_\Omega u(x) D^\alpha v(x)\,dx = (-1)^{|\alpha|}\int_\Omega v(x)\varphi(x)\,dx$$

where $|\alpha| = |\alpha_1| + \cdots + |\alpha_n|$

3. Distribution theory

The theory of Generalized functions was discovered by Laurent Schwartz in the attempt of seeking convolution operators on the space of test functions, as continuous linear functionals on space $C_0^\infty(\Omega)$. This free differential calculus from certain difficulties due to the existence of nondifferentiable functions. This is accomplished by extending nondifferentiable functions to a class objects (called distributions or generalised functions) which is much larger than the class of differentiable functions to which usual calculus applies.

3.1. **The space $\mathcal{D}$ of test functions.** Distributions are operators map from a certain space of functions to the field of real or complex numbers. To this end, several function spaces can be defined. Here we start with the space of test functions $\mathcal{D}(\Omega)$. Let $\Omega \subset \mathbb{R}^n$.

Definition 3.1. Let $\mathcal{D}(\Omega) = C_0^\infty(\Omega)$ be the class of all infinitely differentiable functions with compact support in Ω: $\text{spt}\varphi \subset \Omega$. The element of $\mathcal{D}(\Omega)$ are called *test* functions.

3.2. **The space of $\mathscr{D}'$ of Distributions.** Let $\mathscr{D}(\Omega)$ be the space of functions on $\mathbb{R}$ taking real values are infinitely differentiable and are 0 outside a bounded set (notation: $C_0^\infty(\Omega)$). One can put seminorm on $\mathscr{D}(\Omega)$.

$$\|\varphi\|_N = \sup_\Omega\{|D^\alpha \varphi(x)| : x \in \Omega, |\alpha| \leq N\} \tag{3.1}$$

A (Schwartz) distribution $T : \mathscr{D}(\Omega) \to \mathbb{R}$ (or $T : \mathscr{D}(\Omega) \to \mathbb{C}$) is a continuous linear functional and there exists $c \geq 0$, $n \geq 0$, $N \geq 1$ so that for all $\varphi \in \mathscr{D}(\Omega)$

$$|T(\varphi)| \leq c\|\varphi\|_N \tag{3.2}$$

Example 2 (Dirac Distribution $T = \delta$). Let φ be a test function in $\mathbb{R}^n$. Each $x \in \Omega$ determines a linear functional δ_x on $\mathscr{D}(\Omega)$ by the formula

$$\delta_x(\phi) = \phi(x)$$

The functional $\delta : \mathscr{D}(\Omega) \to \mathbb{C}$ is given by

$$\delta(\varphi) = \varphi(0) \qquad \varphi \in \mathscr{D}(\Omega) \tag{3.3}$$

is linear on φ.

It is straightforward to check that the dirac delta function is linear, that is,

$$\langle \delta, a\varphi_1 + b\varphi_2\rangle = a\varphi(0) + b\varphi(0) = a\langle \delta, \varphi\rangle + b\langle \delta, \varphi\rangle \tag{3.4}$$

Moreover,

$$|\delta(\varphi)| \leq \sup_{|x|\leq 1} |\varphi(x)| = \|\varphi\| \tag{3.5}$$

One might also think of distribution as coming from integration against a measure (By the famous Riesz representation theorem). The standard notation for the Dirac measure δ_x

Theorem 3.2 (L. Schwartz)**.** *There is a topology on $\mathscr{D}(\Omega)$ which makes it into a locally convex topological vector space, and such that $\mathscr{D}'(\Omega)$ is precisely the set of continuous linear functionals on $\mathscr{D}(\Omega)$. $\mathscr{D}$ is complete in this topology but not metrizable.*

3.3. **Calculus on Distributions.** We have so far presented some examples of distributions. One of the convenient features of distributions is that we can do calculus with them.

Functions and measures as distributions.

Example 3. Let L^1_{loc} be the space of all measurable function on Ω so that for any compact subset $K \subseteq \Omega$ $\int_K |f(t)\, dt < \infty$. Define

$$T_f(\varphi) = \int_\Omega \varphi(t) f(t)\, dt, \qquad \varphi \in \mathscr{D}(\Omega) \tag{3.6}$$

Since

$$\begin{aligned} &|T_f(\varphi)| \\ \leq\ & \int_K |f(t)|\,|\varphi(t)|\, dt \\ \leq\ & \left(\int_K |f|\right) \cdot \|\varphi\|_0 \qquad \varphi \in \mathscr{D}_K \end{aligned}$$

Equation (3.2) shows that $T_f \in \mathscr{D}'(\Omega)$.

Example 4. Now let $L^p(\mathbb{R})$, $p \geq 1$ be the space of Lebesgue measurable function f on Ω so that $\int_\mathbb{R} |f(t)|^p\, dt < \infty$. If $f \in L^p(\Omega)$

Proof.

$$\|f\|_p = \left(\int_{\mathbb{R}} |f(t)|^p \, dt \right)^{1/p}$$

The mapping $f \to \|f\|_p$ defines a norm on $L^p(\Omega)$ and this space is complete, it is a Banach space. Further as

$$L^p(\Omega) \subset L^1_{\text{loc}}(\Omega)$$

for each $p \geq 1$, T_f is a well defined element of $\mathscr{D}'(\mathbb{R})$. In fact, if $f \in L^p(\mathbb{R})$ and $p > 1$

$$\begin{aligned}
&\int_K |f(t)| \, dt \\
&\leq \left(\int_K |f(t)|^p \right)^{1/p} \left(\int_K 1^q \right)^{1/q} \\
&\leq |K|^{\frac{1}{q}} \|f\|_p < \infty
\end{aligned}$$

where $|K|$ is the Lebesgue measure of K, which is finite and we used

$$\frac{1}{p} + \frac{1}{q} = 1$$

□

Example 5. If f, $g \in L^1(\Omega)$, then $h = f * g \in L^1(\Omega)$.

Proof. We use

$$h(t) = \int_\Omega f(t-s) g(s) \, ds$$

and so

$$|h(t)| \leq \int_{\mathbb{R}} |f(t-s)||g(s)| \, ds$$

and

$$\begin{aligned}
\int_\Omega |h(t)| \, dt \leq & \int_\Omega \left(\int_\Omega |f(t-s)||g(s)| ds \right) dt \\
= & \int_\Omega \left(\int_\Omega |f(t-s)||g(s)| dt \right) ds
\end{aligned}$$

using Fubini theorem with non negative integrands

$$\begin{aligned}
&= \int_\Omega |g(s)| \left(\int_\Omega |f(t')| dt' \right) ds \\
&= \|g\|_1 \|f\|_1
\end{aligned}$$

So we can define

$$\langle f * g, \varphi \rangle = \langle h, \varphi \rangle$$

as $h \in L^1$. (Note, we sometimes write $\langle f, \varphi \rangle \equiv T_f(\varphi)$) In fact, regarding f, g as elements of $\mathscr{D}'(\mathbb{R})$ (f,g are T_f, T_g)

$$\langle f * g, \varphi \rangle = \int_\Omega \int_\Omega f(t) g(s) \varphi(s+t) \, dt \, ds$$

defineds a evaluation map[1]
and it defines a distribution $\in \mathscr{D}(\mathbb{R})$ becomes

$$\begin{aligned} & |\langle f * g, \varphi \rangle| \\ \leq & \int_{\mathbb{R}} \int_{\mathbb{R}} |f(t)||g(s)||\varphi(s+t)| \, dt \, ds \\ \leq & \left(\int_{\mathbb{R}} \int_{\mathbb{R}} |f(t)||g(s)| \, dt \, ds \right) \|\varphi\|_0 \\ = & \|f\|_1 \|g\|_1 \|\varphi\|_0 \end{aligned}$$

So $C = \|f\|_1 \|g\|_1$ holds for any compact set K for which $\operatorname{spt}(\varphi) \subseteq K$.

$$\varphi \to \langle f * g, \varphi \rangle$$

and it defines a distribution $\in \mathscr{D}(\mathbb{R})$ becomes

$$\begin{aligned} & |\langle f * g, \varphi \rangle| \\ \leq & \int_{\mathbb{R}} \int_{\mathbb{R}} |f(t)||g(s)||\varphi(s+t)| \, dt \, ds \\ \leq & \left(\int_{\mathbb{R}} \int_{\mathbb{R}} |f(t)||g(s)| \, dt \, ds \right) \|\varphi\|_0 \\ = & \|f\|_1 \|g\|_1 \|\varphi\|_0 \end{aligned}$$

So $C = \|f\|_1 \|g\|_1$ holds for any compact set K for which $\operatorname{spt}(\varphi) \subseteq K$. □

Similarly, if μ is a Borel measure on Ω, or if μ is a positive measure on Ω with $\mu(K) < \infty$ for every compact set $K \subset \Omega$, the equation

$$\Lambda_\mu(\varphi) = \int_\Omega \varphi \, d\mu \qquad \varphi \in \mathscr{D}(\Omega) \tag{3.7}$$

defines a definition Λ_μ in Ω, which is usually identified with μ.

Let T_f be defined as (3.6), where f is locally integrable. Let α be a multi-index. We define the α-th derivative of T_f by

$$(D^\alpha T_f)(\varphi) = (-1)^\alpha T_f(D^\alpha \varphi) \qquad \varphi \in \mathscr{D}(\Omega) \tag{3.8}$$

defines a linear functional $D^\alpha T$ on $\mathscr{D}(\Omega)$. If

$$|T\varphi| \leq C\|\varphi\|_N$$

for all $\varphi \in \mathscr{D}_K$, then

$$|(D^\alpha T_f)\varphi| \leq |T_f D^\alpha \varphi| \leq C\|D^\alpha \varphi\|_0 C\|\varphi\|_{|\alpha|}$$

Why (3.8) is the natural way to define $D^\alpha T_f$. Roughly speaking it follows from the integration by parts formula. Nevertheless, let us give some intuition. Suppose f and g are two differentiable functions, then

$$\int_K f'g = -\int_K fg' \tag{3.9}$$

[1] In functional analysis, this is $e_x f = f(x)$

assuming g has compact support contained in K. Perhaps it is easier to think of the one dimensional case. We know

$$\int_a^b f'(x)g(x)\,dx = [f(x)g(x)]_a^b - \int_a^b f(x)g'(x)\,dx \tag{3.10}$$

However, if f and g both hit zero at $x = a$ and $x = b$, then we obtain

$$\int_a^b f'(x)g(x)\,dx = -\int_a^b f(x)g'(x)\,dx$$

This is the same as (3.9). The idea here if that, if f and g are zero on the boundary of the region of integration, which is exactly what happens in this case where they have compact support, then the first term on the right hand side of (3.10) will equal zero. This explains (3.9).

In general if we integrate by parts $|\alpha|$ times, we will have

$$\int_K (D^\alpha f)g = (-1)^\alpha \int_K f D^\alpha g$$

Why does this matter? Notice that the function f which defines our distribution T_f is simply any locally integrable function. We never claimed that it was differentiable. But rather say that it is $|\alpha|$ times differentiable.

What we want to do is to define the derivative of the distribution T_f. Since $T_f(\varphi) = \int f\varphi$ we ought to define the derivative of T_f, that is $D^\alpha T_f$ based on the rule

$$D^\alpha T_f \varphi = \int (D^\alpha f)\varphi$$

The problem is that $D^\alpha f$ may not exist, since we have not assumed that f is differentiable. But, $D^\alpha \varphi$ always exists, since the test function is infinite differentiable ($\varphi \in C^\infty$). The integration by parts formula tells us that if $D^\alpha f$ exists and φ has compact support, then

$$\int_K (D^\alpha f)\varphi(x)\,dx = (-1)^{|\alpha|} \int_K f D^\alpha \varphi\,dx \tag{3.11}$$

Hence it is logical to define the derivative of a distribution T_f according to the rule

$$(D^\alpha T_f)(\varphi) = (-1)^{|\alpha|} T_f(D^\alpha \varphi)$$

Now define L^1_{loc} to be the space of locally integrable functions in Ω so that for any compact subset $K \subseteq \Omega$, $\int_K |f(t)\,dt < \infty$.

We know that if $f \in L^1_{\text{loc}}(\Omega)$, then T_f is a distribution, and its derivative is defined by (3.8). It is natural to extend the definition to a more general distribution. For any general distribution $T \in \mathcal{D}'$ one defines

$$(D^\alpha T)(\varphi) = (-1)^{|\alpha|} T(D^\alpha \varphi)$$

3.4. **Applications.** For concreteness, Let $\alpha = 1$. There are three types of Distributions one may concern with. Namely,

(i) $\mathcal{D}'(\mathbb{R})$ test space $\mathcal{D}(\mathbb{R}) = C_0^\infty(\mathbb{R})$.
(ii) $\mathcal{S}'(\mathbb{R})$ test space $\mathcal{S}(\mathbb{R})$
(iii) $\mathcal{E}'(\mathbb{R})$ test space $\mathcal{E}(\mathbb{R}) = C^\infty(\mathbb{R})$.

Here we stick with (i) in the list. For those who are interested in working with Fourier Transform, use (ii). For Laplace transform, use (iii) but replace $\mathbb{R}$ with $\mathbb{R}^+$.

From our earlier discussion we know that $\Lambda \in \mathcal{D}'(\mathbb{R})$ if

- Λ is linear on $\mathcal{D}(\mathbb{R})$

- For any compact subset K of $\mathbb{R}$, there is a constant $c > 0$ and integer $N \geq 0$ so that when $\varphi \in \mathcal{D}(\mathbb{R})$ with $\mathrm{spt}(\varphi) \subset K$,

$$|\Lambda(\varphi)| \leq C\|\varphi\|_N \tag{3.12}$$

Now we need some definitions.

- $L^1(\mathbb{R})$. We say that $f \in L^1(\mathbb{R})$ if f is Lebesque measurable and

$$\|f\|_1 = \int_{\mathbb{R}} |f(x)|\, dx < \infty$$

- For $p \geq 1$, $L^p(\mathbb{R})$. We say that $f \in L^p(\mathbb{R})$ if f is Lebesgue measurable on $\mathbb{R}$ and

$$\|f\|_p = \left(\int_{\mathbb{R}} |f(x)|^p\, dx\right)^{1/p} < \infty$$

- If $p = \infty$, $L^\infty(\mathbb{R})$ are the essentially bounded functions on $\mathbb{R}$. We say that $f \in L^\infty(\mathbb{R})$ if there is $M > 0$ so that the Lebesgue measure of the set

$$\{x \in \mathbb{R} | f(x) > M\}$$

  is zero.
- $L^1_{\mathrm{loc}}(\mathbb{R})$. We say that $f \in L^1_{\mathrm{loc}}(\mathbb{R})$ if f is Lebesgue measurable and for each compact subset K of $\mathbb{R}$

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

Remark. For $1 \leq p \leq \infty$, $L^p(\mathbb{R}) \subset L^1_{\mathrm{loc}}(\mathbb{R})$

For each $f \in L^1_{\mathrm{loc}}(\mathbb{R})$ one can associate a distribution $\Lambda = T_f$ defined by

$$\Lambda(\varphi) = T_f(\varphi) = \int_{\mathbb{R}} f(t)\varphi(t)\, dt \tag{3.13}$$

for $\varphi \in \mathcal{D}(\mathbb{R})$. Now, (3.13) defines a distribution if $K \subseteq \mathbb{R}$ is compact and $\mathrm{spt}(\varphi) \subseteq K$ then

$$|T_f(\varphi)| \leq \left(\int_K |f(t)|\, dt\right) \sup_{x \in K} |\varphi(x)|$$

So $C = \int_K |f(t)|\, dt < \infty$, $N = 0$.

Remark. 1. If f is continuous differentiable then $f' \in L^1_{\mathrm{loc}}(\mathbb{R})$ and

$$\begin{aligned} T_{f'}(\varphi) &= \int_{\mathbb{R}} f'(t)\varphi(t)\, dt, \;\; \varphi \in \mathcal{D}(\mathbb{R}) \\ &= -\int_{\mathbb{R}} f(t)\varphi'(t)\, dt \end{aligned}$$

In fact, using our earlier argument, in fact, if $\mathrm{spt}(\varphi) \subset [a, b]$ then

$$\begin{aligned} T_{f'}(\varphi) &= \int_a^b f'(t)\varphi(t)\, dt \\ &= [f(t)\varphi(t)]_a^b - \int_a^b f(t)\varphi'(t)\, dt = 0 - \int_a^b f(t)\varphi'(t)\, dt = -\int_{\mathbb{R}} f(t)\varphi'(t)\, dt \end{aligned}$$

This motivates

$$D\Lambda(\varphi) = -\Lambda(\varphi'), \qquad \varphi \in \mathcal{D}(\mathbb{R})$$

as the derivative of a distribution. If Λ satisfies (3.12), then

$$|D\Lambda(\varphi)| = |\Lambda(\varphi')| \leq C\|\varphi\|_{N+1}$$

for φ with $\operatorname{spt}(\varphi) \subseteq K$. So $D\Lambda \in \mathscr{D}(\mathbb{R})$ where $\Lambda \in \mathscr{D}'(\mathbb{R})$

Here are some more examples.

Example 6. Let $f(t) = 0$ for $t < 0$ and $f(t) = t$ for $t \geq 0$. Then f is continuous on $\mathbb{R}$, $f \in L^1_{\text{loc}}(\mathbb{R})$ but $f \notin L^p(\mathbb{R})$ for any $1 \leq p \leq \infty$. We can define $\Lambda = T_f$ and

$$\Lambda(\varphi) = \int_0^\infty t\varphi(t)\,dt \qquad \varphi \in \mathscr{D}(\mathbb{R})$$

then

$$\begin{aligned} D\Lambda(\varphi) &= -\Lambda(\varphi') \\ &= -\int_0^\infty t\varphi'(t)\,dt \end{aligned}$$

and if $\varphi(t) = 0$ for $t > b$, then

$$\begin{aligned} D\Lambda(\varphi) &= -\int_0^b t\varphi'(t)\,dt \\ &= [-t\varphi(t)]_0^b + \int_0^b \varphi(t)\,dt \\ &= \int_0^\infty \varphi(t)\,dt \\ &= T_H(\varphi) \end{aligned}$$

where

$$H(t) = \begin{cases} 0, & \text{if } t < 0 \\ 1, & \text{if } t \geq 0 \end{cases}$$

the heavyside function. Of course $H \in L^1_{\text{loc}}(\mathbb{R})$ but $H \in L^\infty(\mathbb{R})$ but not $H \in L^p(\mathbb{R})$ for $1 \leq p < \infty$.

$$\begin{aligned} D^2\Lambda(\varphi) &= -D\Lambda(\varphi') \\ &= -\int_0^\infty \varphi'(t)\,dt \end{aligned}$$

assuming $\varphi(t) = 0$ for $t > b$

$$\begin{aligned} &= -\int_0^b \varphi'(t)\,dt \\ &= [-\varphi(t)]_0^b = \varphi(0) = \delta_0(\varphi) \end{aligned}$$

where δ_0 is the dirac delta function, that is $D^2T_f = \delta_0$ in $\mathscr{D}'(\mathbb{R})$. We can also define $D\delta_0 = D^3T_f$

$$D\delta_0(\varphi) = -\varphi'(0)$$

Example 7. All $\Lambda \in \mathscr{D}(\mathbb{R})$ roughly arise in the same way. Recall Theorem 6.26 (Rudin) that, for any compact set $K \subseteq \mathbb{R}$, there exists a continuous function f on $\mathbb{R}$ and some integer $k \geq 0$ so that

$$\Lambda(\varphi) = (-1)^k \int_{\mathbb{R}} f(t)\varphi^{(k)}(t)\,dt \tag{3.14}$$

for all $\varphi \in \mathscr{D}(\mathbb{R})$ and $\mathrm{spt}(\varphi) \subseteq K$. It follows from this that (See theorem 6.28 Rudin) If $\Lambda \in \mathscr{D}'(\mathbb{R})$, there exists continuous functions $q_1, q_2, \cdots$ and nonnegative integers $k_1, k_2, \cdots$ so that

$$\Lambda = \sum_{i=1}^{\infty} D^{k_i} T_{g_i} \tag{3.15}$$

and for any compact set $K \subset \mathbb{R}$, only a finite number of g_i are non-zero on K.

Moreover, if Λ has finite order if the choice of N in (3.12) does not depend on K. In that case a finite number of $g_1, g_2, \cdots$ are needed in (3.15)

Our earlier example shows

$$\delta_0 = D^2 T_f$$

$$f(t) = \begin{cases} 0, & \text{if } t < 0 \\ t, & \text{if } t < 0 \end{cases}$$

illustrates this.