Survival Kit on Weak Convergence

Dimitri Cobb

Summer Semester 2024–2025

1 Introduction: Strong Compactness

This short section covers a few notions on weak convergence. The goal is not to provide a thorough presentation of the topic, but merely to give intuition on this notion through a few statements and examples. Most of the proofs will be ignored, and we refer instead to the standard textbook [1], which covers more than what is needed to follow these notes.

As the student will probably have understood from the proofs in these notes, compactness is one of the main tools used to construct solutions of PDEs, and finding compactness properties of a family (f_n) (say of approximate solutions) is the principle challenge of a proof. Said compactness is usually obtained through uniform bounds: assume there is a Banach space X such that

$$\|f_n\|_X \leq R$$

for some constant $R > 0$. Then, if the ball $B(0,R) \subset X$ can be compactly embedded in a Banach (or metric) function space Y, one may extract a limit in Y. There is an extraction $\phi : \mathbb{N} \longrightarrow \mathbb{N}$ and a $f \in Y$ such that

$$f_{\phi(n)} \longrightarrow f \qquad \text{in } Y.$$

For example, take $X = H^1$. Then, the Rellich-Kondrachov theorem states that the embedding $H^1 \subset L^2_{\text{loc}}$ is compact.

Theorem 1 (Rellich-Kondrachov (Theorem 3.16 in [1]))**.** *The embedding $H^1 \subset L^2_{\text{loc}}$ is compact. In other words, any bounded subset $A \subset H^1$ is a relatively compact subset of L^2_{loc}.*

In the above, the space L^2_{loc} is the space of locally L^2 functions on $\mathbb{R}^d$. A measurable function $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ is locally L^2 if the function $\mathbb{1}_K f$ is L^2 for any compact subset $K \subset \mathbb{R}^d$. Note that the space L^2 is *not* a Banach space. Instead, its topology is defined by a metric:

$$d(f,g) := \sum_{k=1}^{\infty} \frac{1}{2^k} \min\big\{1, \|f-g\|_{L^2(B_k)}\big\},$$

where $B_k = B(0,k)$ are balls of increasing radius $k \geq 1$ whose union cover the whole space $\mathbb{R}^d = \bigcup_k B_k$. A sequence of functions (f_n) converges in L^2_{loc} to $f \in L^2_{\text{loc}}$ if and only if the convergence

$$\mathbb{1}_K f_n \longrightarrow \mathbb{1}_K f \qquad \text{in } L^2$$

holds for all compact subset $K \subset \mathbb{R}^d$. Examples of L^2_{loc} functions include (say of $d = 1$) e^x, polynomials, $x^{\epsilon - 1/2}$, *etc.*

Example 2. Consider the sequence of functions $f_n(x) = \exp(-(x-n)^2)$ defined on the real line. It is bounded in H^1_{loc}, and converges locally to zero in L^2_{loc}: for any compact interval $I \subset \mathbb{R}$,

$$\exp(-(x-n)^2) \longrightarrow 0 \qquad \text{in } L^2(I).$$

However, note that (f_n) does not converge to zero in $L^2(\mathbb{R})$, because it is of constant norm $\|f_n\|_{L^2} = C > 0$.

2 Presenting Weak Convergence

We continue to explore the question of extracting a converging sequence from a bounded sequence (f_n) in a Banach space X. Assume again that X is compactly embedded in a (metric) vector space Y, so that we have convergence of an extraction

$$f_{\phi(n)} \longrightarrow f \qquad \text{in } Y.$$

Several questions emerge. Firstly, what can be said about the limit f? Obviously, it is an element of Y, but since Y is a larger space than X, we could hope for more. Especially as the sequence (f_n) remains bounded in X, it seems intuitive that the limit f somehow should also be an element of X. This is generally true.

Let us see how this works on an example. We look again at the situation from the previous section: let (f_n) be a bounded sequence of functions in H^1. We know that an extraction converges in L^2_{loc},

$$f_{\phi(n)} \longrightarrow f \qquad \text{in } L^2_{\text{loc}}. \tag{1}$$

In particular, for any smooth and compactly supported $\psi \in \mathcal{D}$, we have

$$\int f_{\phi(n)}(x)\psi(x)\,\mathrm{d}x \longrightarrow \int f(x)\psi(x)\,\mathrm{d}x.$$

This implies that the limit f satisfies an inequality of the form

$$\sup_{\|\psi\|_{H^{-1}} \leq 1} \left| \int f(x)\psi(x)\,\mathrm{d}x \right| \leq \varliminf_{n} \|f_n\|_{H^1} < +\infty,$$

where in the above H^{-1} is the dual space of H^1 (see the remark immediately below). Since $\mathcal{D}$ is a dense subspace of H^{-1}, this means that f must be an element of H^1.

Remark 3. The Sobolev space H^1 is a Hilbert space, so the reader may be surprised to see that the dual space $H^{-1} = (H^1)'$ is considered to be a different space than H^1. This is a subtle but important point. The space H^1 is a Hilbert space for the scalar product

$$\langle f, g \rangle_{H^1} := \int \big(fg + \nabla f \cdot \nabla g\big)\,\mathrm{d}x.$$

The Riesz representation theorem implies that any bounded linear map $T \in (H^1)'$ can be represented by a function $g \in H^1$ through the formula $T(f) = \langle g, f \rangle_{H^1}$. However, the bounded linear map T also defines a distribution $h \in \mathcal{D}'$ through the relation

$$\forall \psi \in \mathcal{D}, \qquad T(\psi) := \langle h, \psi \rangle_{\mathcal{D}' \times \mathcal{D}}. \tag{2}$$

If h were a locally integrable function, the bracket above would be equal to $T(\psi) = \int h\psi$. The issue is that h and g are not equal as distributions, since they are not defined by the same formula. In fact, while g is a H^1 function, the distribution T is in general the derivative of a L^2 function. The set of distributions T such that the map (2) is bounded on H^1 is called H^{-1}. Both H^{-1} and H^1 are isometric to the dual space H^1 (and so to each other), but they are not identical as spaces of distributions!

We continue with this example. We have proven that the L^2_{loc} limit f still is a H^1 function. This is more than the simple L^2_{loc} convergence (1) gives by itself, and follows from the fact that (f_n) is a bounded sequence of H^1. The next natural question is then whether the sequence converges to f in some stronger topology than L^2_{loc}. Again, the answer is yes.

Definition 4. *Consider H a Hilbert space and (f_n) a sequence of elements of H. We say that (f_n) converges weakly in H to an element $f \in H$ if*

$$\forall g \in H, \qquad \langle f_n, g\rangle_H \longrightarrow \langle f, g\rangle_H.$$

This is noted $f_n \rightharpoonup g$ (in H).

This convergence is called *weak*, whereas convergence in the norm topology of H is *strong*. To provide intuition, let us give a couple of examples.

Example 5. Consider again the sequence $f_n(x) = \exp(-(x-n)^2)$. Then $f_n \rightharpoonup 0$ in H^1. However, the sequence (f_n) does not converge strongly in H^1.

Example 6. Consider the sequence $f_n(x) = \exp(inx) \in L^2(\mathbb{T})$, where $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$. Then it follows from the Riemann-Lebesgue theorem that $f_n \rightharpoonup 0$ in $L^2(\mathbb{T})$. Again, the sequence (f_n) does not converge strongly in $L^2(\mathbb{T})$.

The weak topology of H is the coarsest topology of H such that all linear maps $T \in H'$ are continuous. The main interest of the weak topology of a Hilbert space H is that is turns H into a locally compact topological space.

Theorem 7. *Let H be a (not necessarily separable) Hilbert space and (f_n) a bounded sequence of elements of H. Then there exists a $f \in H$ and an extraction $\phi : \mathbb{N} \longrightarrow \mathbb{N}$ such that*

$$f_{\phi(n)} \rightharpoonup f \qquad \text{in } H.$$

Moreover, the norm of the weak limit is bounded by

$$\|f\|_H \leq \varliminf_n \|f_n\|_H.$$

Remark 8. The main issue with the notion of weak convergence is that it is poorly suited to non-linear problems: the function product $(f, g) \mapsto fg$ is usually not continuous for the weak topology. For example, we have seen that $e^{\pm inx} \rightharpoonup 0$ in $L^2(\mathbb{T})$. However, $e^{inx}e^{-inx} = 1$ does not tend to zero.

3 Weak and Weak-($*$) Convergence

In the analysis of PDEs, it is very frequent to have to deal with Banach spaces that are not Hilbert. In that case, it is convenient to also define a notion of weak convergence.

Definition 9. *Consider a sequence (f_n) of elements of a Banach space X and let $f \in X$. We say that f_n converges weakly to f in X and note*

$$f_n \rightharpoonup f \qquad \text{in } X$$

if for any bounded linear map $g \in X'$, we have

$$\langle g, f_n\rangle_{X'\times X} \longrightarrow \langle g, f\rangle_{X'\times X}.$$

The weak topology of X, sometimes noted $\sigma(X, X')$, is the coarsest topology on X such that all bounded linear maps are continuous. If $X = H$ is a Hilbert space, then weak convergence as defined immediately above is identical to the notion of weak convergence from the previous paragraph.

The weak topology in Banach spaces possesses very nice properties. For instance, a lower semi-continuous convex function on X (for the norm topology) is also lower semi-continuous for the weak topology. This fact plays an important role in the optimization theory of convex functionals. However, it suffers from a fatal flaw: the weak topology is not, for general Banach spaces, locally compact. The typical example of this bad behavior is $L^1(\mathbb{R}^d)$.

Example 10. Consider the sequence of functions $f_n(x) = \mathbb{1}_{[0,1]}(x-n)$, which is bounded in $L^1(\mathbb{R})$. Let $\phi : \mathbb{N} \longrightarrow \mathbb{N}$ be an extraction. Then $(f_{\phi(n)})$ does not converge for the weak topology. To see this, let $g \in L^\infty(\mathbb{R})$ (recall that $(L^1)' = L^\infty$) be defined by

$$g(x) = \begin{cases} 1 & \text{if } x \in [\phi(k), \phi(k)+1[\text{ when } k \text{ is even} \\ -1 & \text{if } x \in [\phi(k), \phi(k)+1[\text{ when } k \text{ is odd.} \end{cases}$$

Then the brackets

$$\langle g, f_{\phi(n)} \rangle_{L^\infty \times L^1} = \int g(x) f_{\phi(n)}(x)\,\mathrm{d}x = (-1)^n$$

never converge. Consequently, no extraction of the sequence (f_n) converges weakly in L^1. Note that the sequence converges to zero in the weak topology of L^2. In other words, the sequence (f_n) has a natural limit $f = 0$, which is an element of L^1, but the weak topology of L^1 is too strong for the sequence to converge to that limit.

Example 11. Let $\chi \geq 0$ be a smooth function on $\mathbb{R}$ that is supported in $[-1,1]$ such that $\int \chi = 1$. We define the sequence (f_n) by

$$f_n(x) = n\chi(nx).$$

Then, for any continuous function $\phi \in C^0(\mathbb{R})$, we have the convergence

$$\int f_n(x)\phi(x)\,\mathrm{d}x \longrightarrow \phi(0). \tag{3}$$

Here the situation is different. The natural limit of the sequence (f_n) is the Dirac delta δ_0, which is *not* an element of L^1. In a sense, L^1 is not "complete" for the weak convergence. This is another example of bounded sequence which does not have a weak accumulation point in L^1 for the weak convergence.

These two examples show the need for another type of convergence. This will be supplied by the notion of weak-$(*)$ convergence.

Definition 12. *Let X be a Banach space that is the (topological) dual a Banach space Y, so that $X = Y'$ (we say that Y is the predual of X). We say that a sequence of elements (f_n) of X converges weakly-$(*)$ to $f \in X$ and note*

$$f_n \overset{*}{\rightharpoonup} f \qquad \text{in } X$$

if for any $g \in Y$ we have

$$\langle f_n, g \rangle_{X \times Y} \longrightarrow \langle f, g \rangle_{X \times Y}.$$

The weak-$(*)$ topology is sometimes noted $\sigma(X, Y)$, and is the topology of bounded linear maps (elements of $Y' = X$) associated to pointwise convergence. Once again, if $X = H$ is a Hilbert space, there is no difference with the notion of weak convergence from the previous paragraph.

The main attribute of the weak-$(*)$ topology is that it turns X into a locally compact space (this is the Banach-Alaoglu-Bourbaki theorem, see Theorem 3.16 in [1]). Moreover, if the predual space Y is separable, then X is sequentially compact for the weak-$(*)$ convergence.

Theorem 13 (Banach-Alaoglu, Corollary 3.30 in [1])**.** *Consider X a Banach space which has a separable predual Y. Then any bounded sequence (f_n) of elements of X has an accumulation point for the weak-$(*)$ convergence: there is an extraction $\phi : \mathbb{N} \longrightarrow \mathbb{N}$ and a $f \in X$ such that*

$$f_{\phi(n)} \overset{*}{\rightharpoonup} f \qquad \text{in } X.$$

Moreover, the limit satisfies

$$\|f\|_X \leq \varliminf_n \|f_n\|_X.$$

Example 14. The space $X = L^\infty(\mathbb{R})$ has predual $L^1(\mathbb{R})$, which is a separable Banach space. In particular, any bounded sequence of $L^\infty(\mathbb{R})$ functions has a weak-$(*)$ accumulation point. For example, the sequence (f_n) from Example 10 converges weakly-$(*)$ to $f = 0$.

Example 15. Let us look again at the sequence from Example 11, which is bounded in $L^1(\mathbb{R})$. The issue is that the space $L^1(\mathbb{R})$ is not the dual of any Banach space (this is a consequence of the Krein-Millman and Banach-Alaoglu-Bourbaki theorems), so that the weak-$(*)$ topology cannot be defined on $L^1(\mathbb{R})$. A way to overcome this issue is to embed $L^1(\mathbb{R})$ into the larger space $\mathcal{M}(\mathbb{R})$ of finite measures, which is a Banach space for the total mass norm

$$\forall \mu \in \mathcal{M}(\mathbb{R}), \qquad \|\mu\|_{\mathcal{M}} := \int \mathrm{d}|\mu|(x).$$

In particular, $L^1(\mathbb{R})$ can be seen as the (closed) subspace of $\mathcal{M}(\mathbb{R})$ of finite measures which are absolutely continuous with respect to the Lebesgue measure. The space $\mathcal{M}(\mathbb{R})$ is the dual of the space $C_0(\mathbb{R})$ of continuous functions f that have limit $f(x) \longrightarrow 0$ at $x \to \pm\infty$. Note that the space $C_0(\mathbb{R})$ is separable, so that any bounded sequence of finite measures should have a weak accumulation point for the weak-$(*)$ convergence. In the case of the sequence (f_n) from Example 11, the convergence (3) shows that

$$f_n \overset{*}{\rightharpoonup} \delta_0 \qquad \text{in } \mathcal{M}(\mathbb{R}).$$

Example 16. In the case of Lebesgue spaces $L^p(\mathbb{R})$ for $1 < p < +\infty$, things are much more easier. Indeed, these spaces are reflexive, which means that they can be canonically[1] identified with their bi-dual $(L^p)'' = L^p$. For reflexive Banach spaces, the weak and weak-$(*)$ topologies are the same. Unfortunately, the spaces $\mathcal{M}(\mathbb{R})$ and $L^\infty(\mathbb{R})$ are not reflexive, which makes the notion of weak-$(*)$ topology indispensable.

Finally, as a conclusion to this discussion on weak convergence, we emphasize that the separability assumption for the predual in the Banach-Alaoglu Theorem 13 is absolutely necessary. Otherwise the the weak-$(*)$ topology is locally compact, but not sequentially locally compact. In order to stress this point, we give another series of examples.

Example 17. Consider the space $Y = \ell^\infty(\mathbb{N})$ of bounded sequences, which is not separable, and note $X = \ell^\infty(\mathbb{N})'$ its dual[2]. We define a sequence of bounded linear maps $T_n : \ell^\infty(\mathbb{N}) \longrightarrow \mathbb{R}$ by setting

$$\forall u \in \ell^\infty(\mathbb{N}), \qquad T_n(u) := u(n).$$

Then, the sequence (T_n) converges to $T \in \ell^\infty(\mathbb{N})'$ weakly-$(*)$ if and only if the limit

$$T_n(u) = u(n) \longrightarrow T(u)$$

holds for all $u \in \ell^\infty(\mathbb{N})$. In other words, the convergence happens if and only if any bounded sequence $(u(n))$ has a limit at $n \to +\infty$, which is absurd. This shows that the sequence (T_n) does not have an accumulation point for the weak-$(*)$ convergence, although the weak-$(*)$ topology $\sigma\big(\ell^\infty(\mathbb{N})', \ell^\infty(\mathbb{N})\big)$ is (topologically) locally compact.

Example 18. We consider the sequence (T_n) from the previous example, but from a different point of view. Let $c_\ell(\mathbb{N})$ be the space of sequences u that are convergent: for any $u \in c_\ell(\mathbb{N})$, the limit $\lim_n u(n)$ exists. The space $c_\ell(\mathbb{N})$, equipped with the uniform norm $\|\,.\,\|_\infty$ is a separable

[1]The adjective canonically means that the canonical embedding $L^p \longrightarrow (L^p)''$ is onto. There are non-reflexive Banach spaces X that are isomorphic to their bi-dual X'', but such that the canonical embedding is not onto.

[2]The dual space $\ell^\infty(\mathbb{N})$ is a very complicated and-non explicit Banach space. It contains the space $\ell^1(\mathbb{N})$ of summable sequences, by the canonical embedding, but the construction of elements $T \in \ell^\infty(\mathbb{N})'$ which are not in $\ell^1(\mathbb{N})$ requires the axiom of choice. It can be shown that $\ell^\infty(\mathbb{N})'$ is the space of finite measures on the Stone-Cech compactification $\beta\mathbb{N}$ of the natural numbers.

Banach space, and the linear maps $T_n : c_\ell(\mathbb{N}) \longrightarrow \mathbb{R}$ are bounded. The sequence (T_n) converges weakly-$(*)$, and the limit T is given by

$$\forall u \in c_\ell(\mathbb{N}), \qquad T(u) := \lim_n u(n).$$

Example 19. Finally, we look at the space $\ell^1(\mathbb{N})$ of summable sequences. It is the dual of the space $c_0(\mathbb{N})$ of sequences u which converge to zero $u(n) \longrightarrow 0$ (equipped with the uniform norm). Since the space $c_0(\mathbb{N})$ is separable, $\ell^1(\mathbb{N})$ is locally sequentially compact for the weak-$(*)$ convergence. This is in sharp contrast with the space $L^1(\mathbb{R})$, although both spaces are Lebesgue spaces of integrable functions: the first one for functions $\mathbb{N} \longrightarrow \mathbb{R}$ with the counting measure, and the second one for functions $\mathbb{R} \longrightarrow \mathbb{R}$ with the Lebesgue measure.

4 Distributions and Distributional Convergence

In the paragraphs above, we have defined and studied increasingly weaker notions of convergence. Here, we study one of the weakest topologies: that of distributions. As we will see, the notion of distribution is not absolutely necessary to study PDEs: in most practical examples, one can find a suitable weak topology (usually H^{-k} for some k) to work with, and not bother with distributions. However, distributions are very convenient as a unifying *language*. One could compare the theory of distributions to that of categories: in itself, knowing that a given family of objects has a categorical structure is useless, all the interesting properties are obtained through hard work on the concrete objects. However, it provides a language in which to express theorems. The modern formulation of the theory is largely due to Laurent Schwartz.[3]

The main idea of the theory of distributions is that of duality. As we know, for $p > 1$, the Lebesgue space L^p can be identified to the dual $L^{p'}$, since every function $f \in L^p$ corresponds to exactly one bounded linear functional

$$I_f : \begin{array}{c} L^{p'} \longrightarrow \mathbb{R} \\ g \longmapsto \int fg. \end{array}$$

We may generalize this. Let $\mathcal{D} := C_c^\infty$ be the set of smooth and compactly supported functions. Then every function $f \in L^1_{\mathrm{loc}}$ corresponds to exactly one linear functional

$$I_f : \begin{array}{c} \mathcal{D} \longrightarrow \mathbb{R} \\ \phi \longmapsto \int f\phi. \end{array} \tag{4}$$

However, the continuity properties of this functional are not clear, as we have not defined a topology on $\mathcal{D}$.

Definition 20. *Consider a sequence (ϕ_n) of $\mathcal{D}$ functions. We say that (ϕ_n) converges to a function $\phi \in \mathcal{D}$ if and only if, for every compact $K \subset \mathbb{R}^d$ and every $k \in \mathbb{N}$,*

$$\|\nabla^k \phi_n - \nabla^k \phi\|_{L^\infty(K)} \underset{n\to\infty}{\longrightarrow} 0.$$

This notion of convergence equips $\mathcal{D}$ a locally convex (although non-metric) topology.

A *distribution* is simply a continuous linear functional on $\mathcal{D}$.

Definition 21. *We call a distribution any linear functional $I : \mathcal{D} \longrightarrow \mathbb{R}$ such that, for any compact $K \subset \mathbb{R}^d$, there is a $k_0 \in \mathbb{N}$ and a constant $C(K, k_0) \geq 0$ for which the inequality*

$$\forall \phi \in \mathrm{div}, \forall k \leq k_0, \qquad \mathrm{supp}(\phi) \subset K \quad \Rightarrow \quad |I(\phi)| \leq C(K, k_0) \left(\|\phi\|_{L^\infty(K)} + \|\nabla^{k_0}\phi\|_{L^\infty(K)} \right).$$

[3] We refer to the excellent lecture notes [2] of F. Golse (in French) for a solid intriduction to the theory of distributions.

The number k_0 is called the order of I on K. In particular, any function $f \in L^1_{\text{loc}}$ uniquely defines a distribution I_f of order zero (on any compact K) through (4). The set of all distributions is noted $\mathcal{D}'$.

The previous definition shows that there is a natural embedding $L^1_{\text{loc}} \subset \mathcal{D}'$ of locally integrable functions in the space of distributions. Consequently, it is customary to use a slight abuse of notation[4] and write $f \in \mathcal{D}'$ when $f \in L^1_{\text{loc}}$. However, distributions can be much more general that L^1_{loc} functions. We give a few examples.

Example 22. Consider the distribution $\delta_0 : \phi \in \mathcal{D} \longmapsto \phi(0) \in \mathbb{R}$. Then δ_0 is the Dirac delta, and is an order zero distribution. More generally, any finite measure $\mu \in \mathcal{M}$ defines a unique order zero distribution through the formula $I_\mu(\phi) = \int \phi \, \mathrm{d}\mu$. We have the embedding $\mathcal{M} \subset \mathcal{D}'$.

Example 23. Consider a non-negative distribution, that is a distribution $I \in \mathcal{D}'$ such that if $\phi \in \mathcal{D}$ satisfies $\phi \geq 0$, then $I(\phi) \geq 0$. In particular, we have $I(\|\phi\|_{L^\infty} - \phi) \geq 0$ and so

$$\forall \phi \in \mathcal{D}, \qquad \big|I(\phi)\big| \leq \|\phi\|_{\mathrm{L}^\infty} |I(1)|.$$

This implies that I defines a non-negative Radon measure on $\mathbb{R}^d$, and, thanks to the embedding $\mathcal{M} \subset \mathcal{D}'$ of the previous example, we can identify I with a non-negative measure $I \in \mathcal{M}$.

Example 24. Consider the (tensor-valued) distribution $\nabla^m \delta_0 : \phi \in \mathcal{D} \longmapsto \nabla^m \phi(0) \in \mathbb{R}^d$. Then $\nabla^m \delta_0 \in \mathcal{D}'$ is an order m distribution. In particular, it cannot be identified to any locally integrable function, or to any (locally finite) measure.

Example 25. Consider the map $I : \mathcal{D}(\mathbb{R}_+^*) \longrightarrow \mathbb{R}$ defined by

$$I(\phi) := \sum_{n=0}^{\infty} \phi^{(n)} \left(\frac{1}{n} \right).$$

This map is well defined because every $\phi \in D(\mathbb{R}_+^*)$ is supported away from $x = 0$. However, I cannot be extended to a distribution (a map on $\mathcal{D}'(\mathbb{R})$), because it cannot have a finite order on a compact neighborhood of $x = 0$.

Example 26. The 1D measurable function $f(x) = 1/|x|$ is not locally integrable, and therefore does not define a distribution.

Due to the linearity of the definition of $\mathcal{D}$, all linear operations can be performed on distributions. In addition, it is also possible to define the derivative of a distribution.

Definition 27. *Consider a distribution $I \in \mathcal{D}'$. We define the derivative $\nabla I \in \mathcal{D}'$ through the formula*

$$\forall \phi \in \mathcal{D}, \qquad \nabla I(\phi) := -I(\nabla \phi).$$

In particular, for any Sobolev function $f \in W^{1,1}_{\text{loc}}$, the distributional derivative ∇I_f and the weak derivative (in the sense of Sobolev spaces) $\nabla f \equiv I_{\nabla f}$ are the same, thanks to integration by parts. In other words,

$$\forall \phi \in \mathcal{D}, \qquad \nabla I_f(\phi) = -\int \nabla \phi f \, \mathrm{d}x = \int \phi \nabla f \, \mathrm{d}x = I_{\nabla f}(\phi).$$

Therefore, the notion of distributional derivative extends that of weak derivative for Sobolev space functions. Therefore, for any distribution $f \in \mathcal{D}'$, we will also note ∇f it's (distributional) derivative.

[4] This is comparable to the abuse of notation that allows us to identify a locally measurable function $f : \mathbb{R} \longrightarrow \mathbb{R}$ with its class up to measure zero sets $f \in L^2$, and with an element of the dual space $f \in (L^2)'$.

This has all kinds of applications. One is that it is always possible to consider the derivative of a locally integrable function as a distribution, although may possibly not be identified to an element of L^1_{loc}. Another one, is that it is possible to consider distributional solutions of PDEs. For example, the function $f(t,x) = \mathbb{1}_{x \le t}$ solves

$$\partial_t f + \partial_x f = 0 \qquad \text{in the sense of } \mathcal{D}'(\mathbb{R} \times \mathbb{R}).$$

Likewise, the distribution $\nabla^m \delta_0$ defined earlier is indeed the m-th derivative of the Dirac delta δ_0.

Remark 28. A word of warning: the distributional derivative is distinct of the notion of almost everywhere derivative, for almost everywhere differentiable functions. For example, the Cantor "staircase" function is almost everywhere differentiable, and the almost everywhere derivative is zero, while the distributional derivative is a non-zero and non-negative measure.

As the (topological) dual space of $\mathcal{D}$, the space of distributions also has a notion of convergence, which we state here.

Definition 29. *Consider a sequence of distributions (I_n). We say that (I_n) converges to a distribution $I \in \mathcal{D}'$ if and only if*

$$\forall \phi \in \mathcal{D}, \qquad I_n(\phi) \longrightarrow I(\phi) \qquad \text{as } n \to +\infty.$$

In particular, the addition of distributions and the distributional derivative are continuous operations with respect to distributional convergence.

Distributional convergence has the usual properties of limits. For example, the distributional limit is unique. In particular, this means that the distributional limit extends all the notions of weak limits we have seen above: for example, any sequence of bounded functions (f_n) that converges weakly-($*$) in L^∞ to some $f \in L^\infty$ also converges in the sense of distributions. Essentially, distributional convergence is the weakest type of convergence that one can reasonable study in the immense majority of PDE problems.[5]

Example 30. Consider $f_n(x) = e^{inx}$. Then $f_n \rightharpoonup 0$ in $L^2(\mathbb{T})$. In particular, we deduce that $f_n \longrightarrow 0$ in $\mathcal{D}'(\mathbb{T})$. The continuity of the distributional derivative implies that we also have, for every $k \in \mathbb{N}$, the convergence $(in)^k e^{inx} \longrightarrow 0$ in $\mathcal{D}'(\mathbb{R})$.

Example 31. Note that the pointwise product of functions is *not*, in general, continuous (or even well defined) in distributional topology. A classical example is, again $f_n(x) = e^{inx}$, which converges to zero as $n \to \infty$. However, $f_n \overline{f_n} = 1$ does not converge to zero in the sense of distributions.

Example 32. Another example of bad behavior of a product is the Dirac mass δ_0 defined above. If $\delta_{1/n}$ is the distribution defined by $\delta_{1/n}(\phi) = \phi(1/n)$, then any reasonable definition of the product should have $\delta_0 \delta_{1/n} = 0$, although continuity would imply that the product tends to δ_0.

One of the very convenient properties of distributional convergence is that a limit of distributions, whenever it exists, always is a distribution. This is essentially a variant of the Banach-Steinhaus theorem.

Proposition 33. *Consider a sequence of distributions (I_n) such that the limit $I(\phi) := \lim_n I_n(\phi)$ exists for every $\phi \in \mathcal{D}$. Then this limit is a distribution $I \in \mathcal{D}'$.*

[5]By choosing an even smaller space of test functions (*i.e.* a subspace $X \subset \mathcal{D}$ such as analytic functions on a neighborhood of some real interval), it is possible to define even larger spaces than $\mathcal{D}'$. This is the idea behind the theory of hyperfunctions, although the usefulness of such spaces is rather limited for the study of non-linear PDEs, which usually involve some regularity.

As a concluding remark, one should note that, although the language of distributions is a considerable generalization of all usual Sobolev functions (for example elements of H^{-k}), provide a global framework to most every type of weak convergence used in PDEs, and hence might seem highly appealing to the student, we must nuance our enthusiasm by two sobering remarks.

Firstly, most non-linear PDEs do not have well-defined solutions unless some level of regularity is granted. For example, it only makes sense to study the Burgers equation

$$\partial_t u + \frac{1}{2}\partial_x(u^2) = 0$$

for solutions u that have a well-defined square u^2. In particular, the equation has no clear meaning when $u \notin L^2_{\mathrm{loc}}$. Therefore, the full power of distribution theory cannot be applied here. In fact, the full generality of distributions (or hyperfunctions) is mostly useful for studying *linear* equations with *smooth* coefficients, with techniques that border on algebra, and are more distant to the methods of non-linear analysis.

Secondly, most non-linear PDEs or linear PDEs with non-smooth coefficients have non-unique distributional solutions, but unique smooth (enough) solutions. In other words, any solution that has interesting properties, usually also has some given regularity, while general distributional solutions (when they make sense) may have a pathological behavior. For example, the Burgers equation above has unique smooth solutions for a given initial datum, but an infinite set of weak irregular solutions (say solutions $u \in L^\infty(\mathbb{R} \times \mathbb{R})$ in the sense of distributions). But this is a topic in itself.

5 Summary

Let us summarize the different ideas and types of convergence we have discussed so far.

- **In a Hilbert space:** any bounded sequence has a weak accumulation point. The weak topology is very nice. If only all spaces were Hilbert spaces.
- **The weak convergence** has nice properties, but the weak topology is usually not locally compact. This creates issues in spaces like $L^1(\mathbb{R})$.
- **Weak-$(*)$ convergence:** the topology is locally compact, and even locally sequentially compact if the predual is separable. The typical spaces in which weak-$(*)$ convergence is used are $L^\infty(\mathbb{R})$ and $\mathcal{M}(\mathbb{R})$. Bad things happen when the predual is not separable.
- **For reflexive spaces** the weak and weak-$(*)$ convergences are the same. For example, in $L^p(\mathbb{R})$ when $1 < p < +\infty$. Outside of Hilbert spaces, this is the most agreeable case.

References

[1] H. Brezis: *Functional analysis, Sobolev spaces and partial differential equations.* Universitext Springer, New York, 2011. *xiv* + 599 pp.

[2] F. Golse *Distributions, analyse de Fourier, équations aux dérivées partielles.* Lecture notes (in French), Octobre 2012.