DISTRIBUTIONS

TERENCE TAO

1. Distributions

In set theory, a *function* is an object $f : X \to Y$ which assigns to each point x in a domain X precisely one point $f(x)$ in the range Y; thus the fundamental operation available on a function is *evaluation*, $x \mapsto f(x)$. However, this is not necessarily the case when the concept of function is employed in other fields of mathematics. In geometry, for instance, the fundamental property of a function may not necessarily be how it acts on points, but rather how it *pushes forward* or *pulls back* more complicated objects than points (e.g. other functions, bundles and sections, sheaves and schemes, etc.). Similarly, in analysis, a function need not necessarily be defined by how it acts on points, but may instead be defined by how it acts on other objects, such as sets or test functions. The former concept leads to the notion of a *measure*; the latter, to that of a *distribution*.

Of course, all these notions of function and function-like objects are related. It is helpful to think of the various notions of a function in analysis as forming a spectrum[1], ranging from the very "smooth" classes of functions to the very "rough". The smooth classes of functions have more operations available on them, but conversely they are very restrictive in their membership and one cannot necessarily guarantee that one can always work in this category. Conversely, the rough classes of functions are very general and it is easy to ensure that one is working in this category, but the price one pays is that the number of operations available on these functions is often sharply reduced. Nevertheless, the various classes of functions can often be treated in a unified manner, because the smooth classes of functions are often dense (in some suitable topology) in the rough classes of functions, and so any operation defined on smooth classes has a good chance of having a unique extension to the rough classes. Because of this convenient fact, it is often not necessary to care too much exactly what category of function one is working with in analysis (particularly if one has obtained quantitative control on one's functions, which will be stable in passing under limits from smooth to rough functions or vice versa); nevertheless there are subtleties and pitfalls involved if one moves too carelessly between the categories of functions (for instance, multiplying two distributions together is a fairly dangerous thing to attempt without extreme care). The situation is somewhat analogous to that between rational numbers and real numbers; most operations defined on rational numbers (with some exceptions, e.g.

[1]This is an oversimplification; the various function spaces one deals with in analysis do not quite form a totally ordered set. However, this intuitive model will serve as a heuristic first approximation for this discussion.

numerator and denominator) can extend easily enough to the more complicated notion of real number, usually by some sort of limiting argument.

Here is a partial list of some categories of functions one encounters in analysis, in descending order from smoothest to roughest. For simplicity we restrict ourselves to functions (or function-like objects) from the interval $[-1,1]$ to the real line $\mathbf{R}$.

- **The class $C^\omega([-1,1])$ of analytic functions**. These are functions which have a locally convergent Taylor expansion at every point, and include all of the usual algebraic functions (except at their singularities), such as $\exp(x)$, $\sin(x)$, polynomials, etc. These functions are extremely smooth, and also have the very powerful property of extending analytically to some open set in the complex plane, but are also extremely rigid; for example knowing an analytic function on a small open set in fact determines that function everywhere by analytic continuation. As such they are often too restrictive a class to work with in analysis.
- **The class $C^\infty_c([1,1])$ of test functions**. On the interval $[-1,1]$, these are simply the smooth (i.e. infinitely differentiable) functions which vanish on neighbourhoods of the endpoints -1 and 1. They are more numerous than analytic functions and are more tractable for analysis, for instance one can construct smooth cutoff functions available to localize other functions to a small set, whereas such a concept cannot exist in the analytic category (it contradicts unique continuation). Also, all the operations from calculus (differentiation, integration, composition, convolution, evaluation, etc.) are available for these functions.
- **The class $C^0([-1,1])$ of continuous functions**. These functions are regular enough that evaluation $x \mapsto f(x)$ is well-defined for all $x \in [-1,1]$, and one can certainly integrate such functions and perform algebraic operations such as multiplication and composition, but they are not regular enough to perform operations such as differentiation. Still they are usually considered among the smoother examples of functions in analysis.
- **The class $L^2([-1,1])$ of square-integrable functions**. These are measurable functions $f : [-1,1] \to \mathbf{R}$ for which the Lebesgue integral $\int_{-1}^1 |f(x)|^2\, dx$ is finite. Usually one equates any two such functions which agree up to sets of measure zero; this implies in particular that it is usually no longer meaningful to evaluate a square-integrable function $f(x)$ at any specific point x, though one can still talk about the function f on a set of positive measure as being well-defined up to sets of measure zero. In particular one can still Lebesgue integrate these functions even if one cannot evaluate them at individual points. One key point about this class is that it is *self-dual* $L^2([-1,1]) \equiv L^2([-1,1])^*$, in that any two functions in this class can be paired together by the *inner product* $\langle f, g\rangle := \int_{-1}^1 f(x)g(x)\, dx$, and in fact one can reconstruct what a square integrable function f is purely from knowing what its inner products $\langle f, g\rangle$ are with all the other square integrable functions g. Indeed, the continuous linear functionals on $L^2([-1,1])$ are all of the form $g \mapsto \langle f, g\rangle$ for some $f \in L^2([-1,1])$ (this is a special case of one of the *Riesz representation theorems*).

- **The class $C^0([-1,1])^*$ of finite Borel measures**. A measure μ does not necessarily have a value (or more precisely, a density $\frac{d\mu}{dx}$) at any given point, but it can still assign a number to a measurable set, or to a measurable function (if the latter is absolutely integrable). A finite Borel measure μ, in particular, assigns a number $\mu(E)$ to every open set E, and also assigns a number $\langle \mu, g\rangle := \int_{-1}^1 g\ d\mu$ to every continuous function $g \in C^0([-1,1])$. For instance, every square-integrable function $f(x)$ is associated to a finite Borel measure $f(x)\ dx$, which explains the repeated use of the inner product notation $\langle , \rangle$. Indeed, one can define a finite Borel measure μ as a continuous linear functional $g \mapsto \langle \mu, g\rangle$ on the space of continuous functions (this is another of the *Riesz representation theorems*).
- **The class $C^\infty([-1,1])^*$ of distributions**. Just as measures can be viewed as continuous linear functionals on $C^0([-1,1])$, a distribution μ is a continuous linear functional on $C_c^\infty([-1,1])$ (endowed with the smooth topology), thus a distribution can be viewed as a "virtual function" which cannot itself be directly evaluated, but which can still be paired with any test function $g \in C_c^\infty([-1,1])$, producing a number $\langle \mu, g\rangle$. A famous example is the *Dirac distribution* δ_0, defined as the functional which when paired with any test function g returns the evaluation $g(0)$ of g at zero: $\langle \delta_0, g\rangle := g(0)$. Similarly we have the derivative Dirac distribution $-\delta_0'$, which when paired with any test function g returns the derivative $g'(0)$ of g at zero: $\langle -\delta_0', g\rangle := g'(0)$. (The reason for the minus sign will be explained later). Since test functions have so many operations available to them, the class of distributions is quite large. While they cannot be evaluated or integrated on open sets, there are still many operations available to them; we discuss this later.
- **The class $C^\omega([-1,1])^*$ of hyperfunctions**. There are classes of functions more general still than distributions; for instance there are hyperfunctions, which roughly speaking one can think of as linear functionals that can only be tested against analytic functions $g \in C^\omega([-1,1])$ rather than test functions $g \in C^\infty([-1,1])$. However as the class of analytic functions is so sparse, hyperfunctions tend not to be as useful as distributions in analysis.

At first glance, the concept of a distribution has limited utility, as all a distribution μ is empowered to do is to be tested against test functions g to produce inner products $\langle \mu, g\rangle$. However, using this inner product, one can often take operations which are initially only defined on test functions, and extend them to distributions by *duality*. A typical example is with differentiation. Suppose one wants to know how to define the derivative μ' of a distribution, or in other words how to define $\langle \mu', g\rangle$ for any test function g and distribution μ. If μ was itself a test function $\mu = f$, then we could evaluate this using integration by parts (recalling that test functions vanish on the boundary $-1, 1$) we have

$$\langle f', g\rangle = \int_{-1}^1 f'(x)g(x)\ dx = -\int_{-1}^1 f(x)g'(x)\ dx = -\langle f, g'\rangle.$$

Note that if g is a test function then so is g'. Thus we can generalize this formula to arbitrary distributions by defining

$$\langle \mu', g\rangle := -\langle \mu, g'\rangle.$$

Thus for instance $\langle \delta_0', g\rangle = -\langle \delta_0, g'\rangle = -g'(0)$. More formally, what we have done here is computed the adjoint of the differentiation operation (as defined on the dense space of test functions), and then taken adjoints again to define the differentiation operation for general distributions. This procedure is well-defined and works for many other concepts also, thus one can add two distributions, multiply a distribution by a smooth function, convolve two distributions, and compose distributions on both left and right with suitably smooth functions. One can even take Fourier transforms of distributions; for instance, the Fourier transform of the Dirac delta δ_0 is the constant function 1, and conversely (this is essentially the Fourier inversion formula), while the distribution $\sum_{n\in\mathbf{Z}} \delta_0(x-n)$ is its own Fourier transform (this is essentially the Poisson summation formula). Thus the space of distributions is quite a good space to work in, in that it contains a large class of functions (e.g. all measures and integrable functions), and is also closed under a large number of common operations in analysis. Because the test functions are dense in the space of distributions, the operations as defined on distributions are usually compatible with those on test functions; for instance, if f and g are test functions and $f' = g$ in the sense of distributions, then $f' = g$ will also be true in the classical sense. This often allows one to manipulate distributions as if they were test functions without fear of confusion or inaccuracy. The main operations one has to be careful about are evaluation $x \mapsto \mu(x)$ and pointwise multiplication $\mu_1, \mu_2 \mapsto \mu_1\mu_2$ of distributions, both of which are usually not well defined (e.g. the square of the Dirac delta distribution is not well defined as a distribution).

Another way to view distributions is as the *weak limit* of test functions. A sequence of functions f_n is said to *converge weakly* to a distribution μ is if $\langle f_n, g\rangle \to \langle \mu, g\rangle$ for all test functions g. For instance, if φ is a test function with total integral $\int_{-1}^{1} \varphi = 1$, then the test functions $f_n(x) := n\varphi(nx)$ can be shown to converge weakly to the Dirac delta distribution δ_0, while the functions $f_n' = n^2\varphi'(nx)$ converge weakly to the derivative δ_0' of the Dirac delta. On the other hand, the functions $g_n(x) := \cos(nx)\varphi(x)$ converge weakly to zero (this is a variant of the *Riemann-Lebesgue lemma*). Thus weak convergence has some unusual features not present in stronger notions of convergence, in that severe oscillations can sometimes "disappear" in the limit. One advantage of working with distributions instead of smoother functions is that one often has some compactness in the space of distributions under weak limits (e.g. by the Banach-Alaoglu theorem). Thus distributions can be thought of as asymptotic extremes of behavior of smoother functions, just as real numbers can be thought of as limits of rational numbers.

The theory of distributions is particularly useful in the theory of linear partial differential equations. For instance, to solve a PDE such as $Lu = f$ where L is a constant-coefficient differential operator, and f is a given test function, one can often use distributions to obtain a (smooth) solution of the form $u = f * K$, where K is a distribution known as the *fundamental solution* of L. In particular, distributions can be useful even for questions which only involve classical functions.

Department of Mathematics, UCLA, Los Angeles CA 90095-1555

E-mail address: tao@math.ucla.edu