4 Generalised functions and the Dirac delta distribution

Learning outcomes: *a good understanding of the new conceptual ideas introduced through generalised functions and the ability to prove mathematical statements 'in the sense of distributions', in particular where the Dirac delta distribution is involved. Ability to do basic operations on distributions.*

Relevance to the course: Defining a new type of function in terms of its integrals rather than by its values at points provides a new tool to calculate Green's functions (almost) effortlessly.

4.1 Operator inverses and Green's functions - Unit 4 Lect 1

Learning outcomes: being able to articulate the physical interpretation of Green's functions in the context of differential equations involving a linear operator L, and understand the connection between Green's functions and L^{-1}.

Green's function methods are another tool to solve linear inhomogeneous differential equations of the form

second order linear diff operator → source term.

$$Lu(\boldsymbol{x}) = f(\boldsymbol{x}), \qquad \boldsymbol{x} \in \Omega \subset \mathbb{R}^n, \tag{4.1}$$

where f is a source term and Ω is the domain on which we wish to solve the problem. As before in this course, we consider L to be a second-order differential operator. To find solutions u, one might want to 'invert' such operators to obtain $u(\boldsymbol{x}) = L^{-1}f(\boldsymbol{x})$. This is a very natural idea, but one must make sense of what L^{-1} might be. It turns out that such operators L often have inverses L^{-1} that are *integral* operators. Green's functions appear precisely in the process of identifying inverses of linear operators.

Provided the kernel of L is $\{0\}$, the differential equation $Lu = f$ has a unique solution (for all source terms f in a suitable function space). The inverse operator L^{-1} is then whatever operator that gives the solution $u = L^{-1}f$ for arbitrary but fixed source term f.

We have already seen in Subsection 3.5, equation (3.19) that, when $\mathfrak{L}$ is a Sturm-Liouville operator, the solution to the one-dimensional problem $\mathfrak{L}u(x) = \omega(x)f(x)$ may be written as

Green function

$$u(x) = \sum_{n=0}^{\infty}(-\frac{1}{\lambda_n})\,\langle \hat{u}_n, f\rangle_\omega\, \hat{u}_n(x) = \int_a^b \underbrace{\left\{\sum_{n=0}^{\infty}(-\frac{1}{\lambda_n})\,\hat{u}_n(y)\hat{u}_n(x)\right\}}_{:=G(x,y)} \underbrace{\omega(y)\,f(y)}_{\text{source term}}\,dy, \quad \lambda_n \neq 0\ \forall n.$$

In more general cases than the above, one may thus want a representation of the solution u of the form

$$u(\boldsymbol{x}) = L^{-1}f(\boldsymbol{x}) = \int_\Omega d^n\boldsymbol{\xi}\, G(\boldsymbol{x},\boldsymbol{\xi})\, f(\boldsymbol{\xi}), \quad \boldsymbol{x},\boldsymbol{\xi} \in \Omega \subset \mathbb{R}^n, \tag{4.2}$$

where $G(\boldsymbol{x},\boldsymbol{\xi})$ is called the Green's function or the fundamental solution of the differential equation. In other words, the solution $u(\boldsymbol{x})$ is constructed as a superposition of effects at $\boldsymbol{x}$ from sources at $\boldsymbol{\xi}$ for all $\boldsymbol{\xi}$ in the domain Ω. Therefore we think of the Green's function $G(\boldsymbol{x},\boldsymbol{\xi})$ as being the effect at $\boldsymbol{x}$ from a source f at $\boldsymbol{\xi}$.

We need to figure out what G is mathematically. If G exists, let us see what we can learn about it from applying L to (4.2):

$$Lu(\boldsymbol{x}) = f(\boldsymbol{x}) \overset{(1)}{=} \int_\Omega d^n\boldsymbol{\xi}\, LG(\boldsymbol{x},\boldsymbol{\xi})\, f(\boldsymbol{\xi}), \tag{4.3}$$

where we can bring the operator L under the integral since L does not act on the integration variable $\boldsymbol{\xi}$. Inspection of equality (1) in (4.3) tells us that LG must be sharply peaked around $\boldsymbol{\xi} = \boldsymbol{x}$ so that all of the integral contributions come from $f(\boldsymbol{x})$. In particular, the integral must give an average of $f(\boldsymbol{\xi})$ in an infinitesimal neighbourhood of $\boldsymbol{\xi} = \boldsymbol{x}$.

Let us construct such average in one dimension. For this we use the 'box function' (see Figure 5).

$$\delta_{\Delta x}(x) = \begin{cases} \dfrac{1}{\Delta x} & 0 \le x < \Delta x \\ 0 & \text{otherwise}, \end{cases}$$

$\delta_{\Delta x}(x-s) = \frac{1}{\Delta x} \quad 0 \le x-s \le \Delta x$

which has area 1.

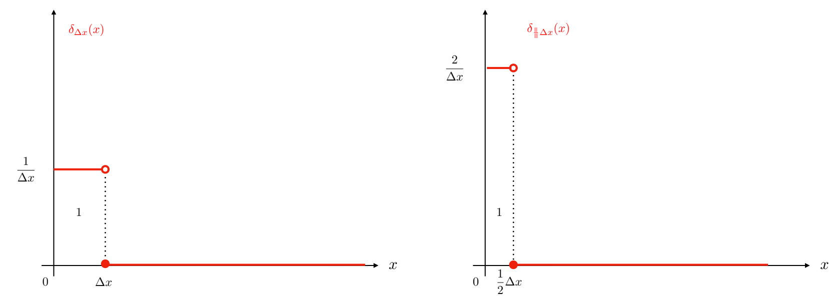
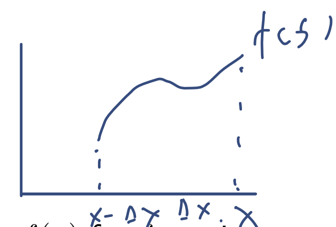


FIGURE 5: *Graphs of two box functions on* $[0, \infty)$ *with the same area of 1. When* $\Delta x \to 0^+$*, the box becomes much narrower and much longer, but the area is again 1. In the limit, we have a mathematical object that spikes at* $x = 0$*: a Diral delta 'distribution'.*



The integral

$$\int d\xi\, \delta_{\Delta x}(x - \xi)\, f(\xi) = \frac{1}{\Delta x} \int_{x-\Delta x}^{x} f(\xi)\, d\xi$$

is the average of f on the interval $[x - \Delta, x]$. As $\Delta x \to 0$, the integral goes to $f(x)$ for f continuous, which is exactly what we want. In other words, if we write the solution as

$$u(x) = \int_a^b G(x, \xi)\, f(\xi) d\xi,$$

we want

$$Lu(x) = f(x) = \int_a^b LG(x, \xi)\, f(\xi) d\xi \text{ with } LG(x, \xi) = \lim_{\Delta x \to 0} \delta_{\Delta x}(x - \xi) := \delta(x - \xi), \tag{4.4}$$

where $\delta(x - \xi)$ is an infinitely high box which is infinitesimally narrow but preserves the area of 1 around $\xi = x$. This mathematical object is not an ordinary function: it requires the introduction of **generalised functions (aka distributions)** to properly handle it. Despite this mathematical glitch, $\delta(x)$ is commonly called the 'Dirac delta *function*'. We shall see next that it should be called the 'Dirac delta *distribution*'.

Remark 4.1 **Non-examinable: inverse finite matrices as an inspiration for operator inverses.**
Let us look at a familiar case from a slightly new perspective: the inverse of an $n \times n$ matrix A ($\det A \neq 0$). Suppose

we wish to solve the equation $A\boldsymbol{u} = \boldsymbol{f}$, where $\boldsymbol{f} = \sum_{j=1}^{n} f_j \mathbf{e}_j$, where $\{\mathbf{e}_j\}$ is the canonical orthonormal basis in n-dimensional Euclidean space. Then the solution is of the form,

$$\boldsymbol{u} = A^{-1}\boldsymbol{f} \qquad \text{or} \quad u_i = \sum_{j=1}^{n} (A^{-1})_{ij}\, f_j.$$

We may interpret the matrix elements of A^{-1} as follows:
$(A^{-1})_{ij}$ is the effect (i.e. solution) at i of a source $\boldsymbol{f}$ at j (source localised at $\mathbf{e}_j$). Equally, a fixed column j of A^{-1} is the effect (i.e. solution) **everywhere** from the source at j

Example 4.2 To illustrate the point above, consider the $N \times N$ matrix A emerging from discretising the one-dimensional time-independent Poisson equation on $[0, 1]$ with Dirichlet b.c.:

$$\frac{d^2}{dx^2} u = f, \qquad u(0) = u(1) = 0, \tag{4.5}$$

where $-f$ is a forcing function or source term. The size of the matrix is directly correlated with how we discretise the interval $[0, 1]$. We choose a sequence of N points $x_i, i \in \{1, 2, \ldots N\}$ where $|x_{i+1} - x_i| = h$ for $0 \le i \le N - 1$, and $x_1 = 0$ and $x_N = 1$ coincide with the endpoints of the interval. We then replace $\frac{d^2 u}{dx^2}$ by finite difference approximations,

$$\frac{u_{i-1} - 2u_i + u_{i+1}}{h^2} = f(x_i), \qquad \text{for } 2 \le i \le N - 1, \tag{4.6}$$

where $u_i := u(x_i)$. Together with the b.c. $u_1 = 0$ and $u_N = 0$, the $N - 2$ linear equations (4.6) form a set of N linear equations to be solved. This system can be written as $A\boldsymbol{u} = \boldsymbol{F}$ where $\boldsymbol{u}$ is the N-vector with components $u_i = u(x_i)$ and $\boldsymbol{F}$ is the N-vector with components $(0, f(x_1), \ldots, f(x_{N-1}, 0)$. So we have the solution

$$\boldsymbol{u} = A^{-1}\boldsymbol{F}.$$

The differential equation (4.5) can be viewed as a model of how the quantity u responds to the force f. For instance u could be the displacement of a stretched string under the influence of pressure (force). The graphs in Figure 6 show five columns of the $N \times N$ matrix A^{-1} when $N = 11$, each encoding the effect on the interval $[0, 1]$ of the source (pressure) exerted at the point where each graph is minimum. So solutions (displacements of string) behave in a way you would expect if you would press on the string at the five chosen points.
One can repeat this exercise in 2 dimensions, where now $\boldsymbol{\nabla}^2 u := (\partial^2_{xx} + \partial^2_{yy})u = f$ models a square stretched membrane (think square drum). The two graphs in Figure 7 illustrate two different columns of A^{-1} for a discretised domain of $\Omega = [-1, 1] \times [-1, 1]$, and Dirichlet b.c. Again, the graph matches the intuition of what happens if one presses at one point on the membrane. The location of the minima correspond to the index j of the column, which is the effect of a unit force ($f = 1$) at these very points (and $f = 0$ elsewhere in Ω). □

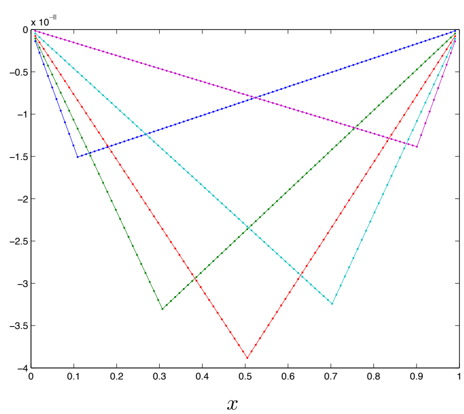


FIGURE 6: *The five graphs are five solutions (columns of A^{-1}) of the discretised time-independent Poisson equation in one dimension with Dirichlet boundary conditions.*

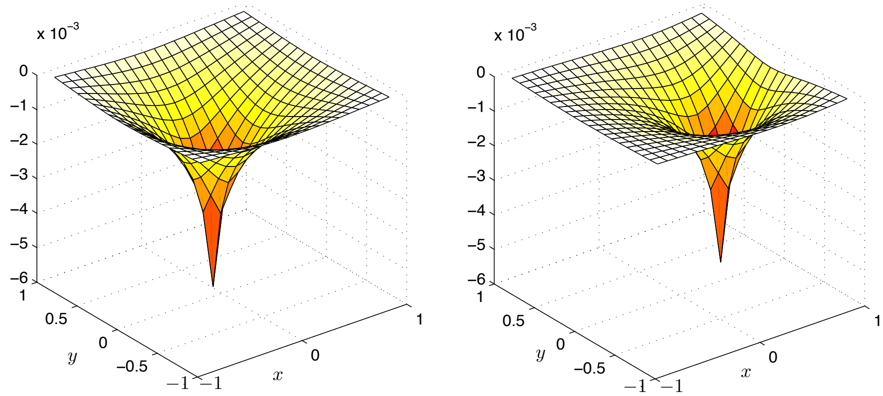


FIGURE 7: *Two solutions of (columns of A^{-1}) of the discretised time-independent Poisson equation in two dimensions with Dirichlet boundary conditions.* ***End of non-examinable remark***

4.2 The space of test functions

The vector space consisting of the class of test functions is the simplest and most commonly used space for defining distributions when physical applications are in mind.

Definition 4.3 *Let $\Omega \subseteq \mathbb{R}^n$ be an open set, and $\psi : \Omega \to \mathbb{C}$ be a function. Then ψ is a* ***test function*** *if:*

(i) $\psi \in C^\infty(\Omega)$, i.e. ψ has finite derivatives of all orders on Ω (smoothness)

(ii) the support of the function ψ, i.e. the closure of the set of points in which the value of ψ is not zero, is compact. We denote the support of ψ by,

$$\operatorname{supp}\psi = \overline{\{\boldsymbol{x} \in \Omega \,|\, \psi(\boldsymbol{x}) \neq 0\}}.$$

The space of test functions on Ω is denoted $\mathcal{D}(\Omega)$.

Remark 4.4 Since $\operatorname{supp}\psi$ is closed by definition, and since the closed bounded subsets of a Euclidean space are precisely those subsets which are compact, it is enough to ask for the support to be bounded. □

So test functions are smooth functions, which means that they guarantee a trouble-free ride through the theory of distributions, in particular because all their derivatives are also test functions. But do they exist? The answer is yes, however we never need to know much more than the fact they exist: they are a background tool. Here is an example.

Example 4.5 The bump function is the prototype example of a test function in one dimension. It is defined as

$$\Psi(x) := \begin{cases} e^{-1/(1-x^2)} & \text{for } |x| < 1 \\ 0 & \text{else} \end{cases}$$

and is shown in Figure 8. □

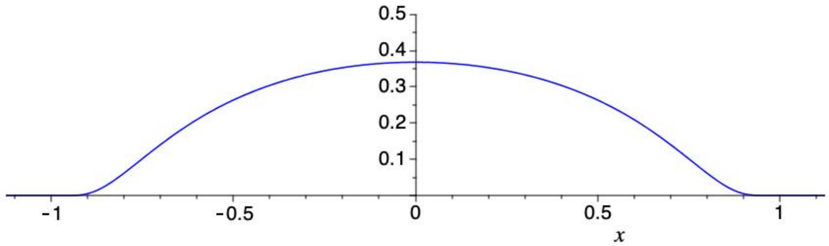


FIGURE 8: *The bump function* $\Psi(x)$

Proposition 4.6 *The space* $\mathcal{D}(\Omega)$ *is a vector space.*

Proof: Let $\psi, \varphi \in \mathcal{D}(\Omega)$ and $a, b \in \mathbb{C}$. If we can show that $a\psi + b\varphi \in \mathcal{D}(\Omega)$, we are done.

- Since $\psi, \varphi \in \mathcal{C}^\infty$, it follows that $a\psi + b\varphi \in \mathcal{D}(\Omega) \in \mathcal{C}^\infty$
- Let $A = \operatorname{supp}\psi$ and $B = \operatorname{supp}\varphi$. Then $a\psi(\boldsymbol{x}) + b\varphi(\boldsymbol{x}) = 0$ for $\boldsymbol{x} \notin A \cup B$. But $A \cup B$ is closed so $\operatorname{supp}(a\psi + b\varphi) \subseteq A \cup B$. Since A and B are bounded, $\operatorname{supp}(a\psi + b\varphi)$ is bounded too.

■

Actually, $\mathcal{D}(\Omega)$ is an uncountably infinite-dimensional vector space, whose 'vectors' are test functions.

It is fairly easy to construct new test functions from known ones. We have

Proposition 4.7 *Let* $\psi \in \mathcal{D}(\mathbb{R}^n)$, $\boldsymbol{\xi} \in \mathbb{R}^n$, $a \in \mathbb{R} - \{0\}$ *and* $g \in \mathcal{C}^\infty(\mathbb{R}^n)$. *Then*

1. $\psi(\boldsymbol{x} + \boldsymbol{\xi})$, $\psi(-\boldsymbol{x}), \psi(a\boldsymbol{x}) \in \mathcal{D}(\mathbb{R}^n)$.
2. $g(\boldsymbol{x})\psi(\boldsymbol{x}) \in \mathcal{D}(\mathbb{R}^n)$.

We need a concept of convergence in $\mathcal{D}(\Omega)$ before introducing distributions. The motivation is that we want distributions to act *continuously* on test functions, and that requirement relies on defining a suitable notion of convergence in $\mathcal{D}(\Omega)$.

Recall: let V be a vector space whose elements are functions, and let V be equipped with a norm $\|\cdot\|$. Then this norm can be used to define a notion of convergence on V as follows: a sequence of functions $(f_m)_{m\in\mathbb{N}}$ converges to a function f in V if $\lim_{m\to\infty} \|f_m - f\| = 0$ as a sequence of numbers. This is called *pointwise convergence*: it requires $f_m(x)$ and $f(x)$ to be close together for all m large enough, but how long one has to wait for $f_m(x)$ and $f(x)$ to get close may depend on which element x one is looking at.

Here though we need a stronger type of convergence, as we want to preserve the defining properties of the test functions (smoothness and compact support) in the convergence process. Here it is.

Definition 4.8 *Let* $\Omega \subseteq \mathbb{R}^n$ *be an open set. A sequence* $(\psi_m)_{m\in\mathbb{N}}$ *whose elements belong to the space* $\mathcal{D}(\Omega)$ *converges to a certain test function* $\psi \in \mathcal{D}(\Omega)$ *when* $m \to \infty$ *if:*

- *there exists a compact subset* K *of* Ω *such that* $\operatorname{supp}\psi_m$ *and* $\operatorname{supp}\psi$ *are subsets of* K *for all* $m \in \mathbb{N}$,

- *the sequence $(\psi_m)_{m\in\mathbb{N}}$ converges* uniformly *to ψ in $\mathcal{D}(\Omega)$ AND the sequence*

$$D^k\psi_m := \psi_m^{(k)} := \frac{\partial^{k_1}}{\partial x_1^{k_1}}\frac{\partial^{k_2}}{\partial x_2^{k_2}}\cdots\frac{\partial^{k_n}}{\partial x_n^{k_n}}\psi_m$$

also converges uniformly to $D^k\psi$ for each multi-index $k=(k_1,\dots,k_n)$ where $k_i\in\mathbb{N}_0$.

We refer to this type of convergence as $(\psi_m)\underset{\mathcal{D}}{\to}\psi$, where it is understood that $m\in\mathbb{N}$.

In the definition above, the condition of uniform convergence of the sequence of partial derivatives $D^k\psi_m$ is required as derivatives of test functions are again test functions.

One can summarise the notion of convergence given in Definition 4.8 using the 'sup' norm

$$\| f \|_\infty := \sup\{|f(\boldsymbol{x})| : \boldsymbol{x}\in\mathbb{R}^n\}$$

as follows:

Definition 4.9 *The sequence (ψ_m) of test functions in $\mathcal{D}(\Omega)$ converges to $\psi\in\mathcal{D}(\Omega)$ if*

- *there exist a compact set $K\subset\Omega$ such that $\operatorname{supp}\psi_i\subset K$ for all i*
- *for all multi-indices $k=(k_1,\dots k_n)$ and $|k|:=k_1+\ldots+k_n$, $\|D^k\psi_m-D^k\psi\|_\infty$ tends to 0 when $m\to\infty$ as a normal sequence of numbers (this includes the case $|k|=0$).*

We are now ready to enter the world of generalised functions, also called distributions.

End of Unit 4 Lect 1

4.3 Distributions - Definition - Unit 4 Lect 2

Note: Distributions are rules that, given any test function $\psi(\boldsymbol{x})$, return a number. In mathematics, such objects are a particular type of linear maps (also called linear functionals).

Definition 4.10 *Let $\Omega\subseteq\mathbb{R}^n$ be an open set. A **distribution** is a continuous linear map $\mathcal{T}:\mathcal{D}(\Omega)\to\mathbb{R}:\psi\to\mathcal{T}[\psi]$.*

- *$\mathcal{T}$ is linear, i.e. $\forall a,b\in\mathbb{R}$ and for any two test functions ψ and φ in $\mathcal{D}(\Omega)$, $\mathcal{T}[a\psi+b\varphi]=a\mathcal{T}[\psi]+b\mathcal{T}[\varphi]$.*
- *$\mathcal{T}$ is continuous in the following sense:*

$$\forall\psi\in\mathcal{D}(\Omega),\quad \forall(\psi_m)_{m\in\mathbb{N}}\subseteq\mathcal{D}(\Omega)\quad\text{with}\quad(\psi_m)\underset{\mathcal{D}}{\to}\psi,\quad \mathcal{T}[\psi_m]\underset{m\to\infty}{\to}\mathcal{T}[\psi].$$

So this is convergence of a sequence of numbers and one calls $\mathcal{T}$ a sequentially continuous map.

The space of distributions with test functions in $\mathcal{D}(\Omega)$ is denoted $\mathcal{D}'(\Omega)$. It is an infinite-dimensional vector space. We will most often take $\Omega=\mathbb{R}^n$ for some n in this course.

Example 4.11 The Dirac delta 'function' is in fact the distribution

$$\delta:\mathcal{D}(\mathbb{R}^n)\to\mathbb{R}:\psi\mapsto\delta[\psi]:=\psi(\mathbf{0}).$$

Indeed, δ is

- linear: $\delta[a\psi + b\varphi] = a\psi(0) + b\varphi(0) = a\delta[\psi] + b\delta[\varphi]$, for all $a, b \in \mathbb{R}$ and $\psi, \varphi \in \mathcal{D}(\mathbb{R}^n)$, by virtue of $\mathbb{R}$ being a vector space.
- continuous: for all $\psi \in \mathcal{D}(\mathbb{R}^n)$ and for all $(\psi_m)_{m\in\mathbb{N}} \subseteq \mathcal{D}(\mathbb{R}^n)$ with $(\psi_m) \underset{\mathcal{D}}{\to} \psi$, we have in particular that $(\, \psi_m(\boldsymbol{x})\,) \underset{m\to\infty}{\to} \psi(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^n$ (pointwise convergence). Therefore, we have, since 0 is a point of $\mathbb{R}^n$,

  $$\delta[\psi_m] = (\, \psi_m(\mathbf{0})\,) \underset{m\to\infty}{\to} \psi(\mathbf{0}) = \delta[\psi]$$

  as a sequence of points.
  At the moment, we have just *defined* what the Dirac delta distribution is. Our ultimate goal is for this definition to allow an interpretation of its derivative as the unit step function. Here we mean derivative 'in the sense of distributions', which will be defined shortly. □

Example 4.12 Continuous functions are distributions. Let $f \in \mathcal{C}^0(\mathbb{R}^n)$. To treat f as a distribution, we need to define how it acts on test functions. Define

$$\mathcal{T}_f : \mathcal{D}(\mathbb{R}^n) \to \mathbb{R} : \psi \mapsto T_f[\psi] := \int_{\mathbb{R}^n} f(\boldsymbol{x}).\psi(\boldsymbol{x})\, d\boldsymbol{x}.$$

Then $\mathcal{T}_f$ is a distribution. Indeed,

- linear: immediate consequence of integration being a linear operation
- continuous: for all $\psi \in \mathcal{D}(\mathbb{R}^n)$ and for all $(\psi_m)_{m\in\mathbb{N}} \subseteq \mathcal{D}(\mathbb{R}^n)$ with $(\psi_m) \underset{\mathcal{D}}{\to} \psi$, we have uniform convergence in particular (we cannot get away with just pointwise convergence here because that does not allow to take limits through integrals, which we need here). We have

$$\begin{aligned}\lim_{m\to\infty} \mathcal{T}_f[\psi_m] &= \lim_{m\to\infty} \int_{\mathbb{R}^n} f(\boldsymbol{x})\, \psi_m(\boldsymbol{x})\, d\boldsymbol{x} = \int_{\mathbb{R}^n} \lim_{m\to\infty} f(\boldsymbol{x})\, \psi_m(\boldsymbol{x})\, d\boldsymbol{x} \\ &= \int_{\mathbb{R}^n} f(\boldsymbol{x})\, \psi(\boldsymbol{x})\, d\boldsymbol{x} = \mathcal{T}_f[\psi].\end{aligned}$$ □

Example 4.13 Consider the Gaussian functions $f_m(x) := \frac{1}{\sqrt{\pi}} m e^{-m^2x^2}$, $m \in \mathbb{N}$. They are elements of $\mathcal{C}^0(\mathbb{R})$, and therefore are distributions $\mathcal{T}_{f_m} : \mathcal{D}(\mathbb{R}) \to \mathbb{R}$ acting on test functions as

$$\mathcal{T}_{f_m}[\psi] := \int_{\mathbb{R}} f_m(x)\psi(x)\, dx = \int_{\mathbb{R}} \frac{1}{\sqrt{\pi}} m e^{-m^2x^2}\, \psi(x)\, dx.$$

Looking at Figure 9, we see that when the Gaussian $f_m(x)$ is very narrow, $\psi(x)$ is virtually constant (and equal to $\psi(0)$) over the width of the Gaussian $f_m(x)$. Then the integral becomes nearly $\psi(0)\int_{\mathbb{R}} f_m(x)dx = \psi(0)$ since by design, $\int_{\mathbb{R}} f_m(x)dx = 1$. So the sequence of numbers $\mathcal{T}_{f_m}[\psi]$ converges to $\psi(0) =: \delta[\psi]$ as $m \to \infty$. It is in that sense that one says that the Gaussians $f_m(x)$ approximate the delta distribution δ.

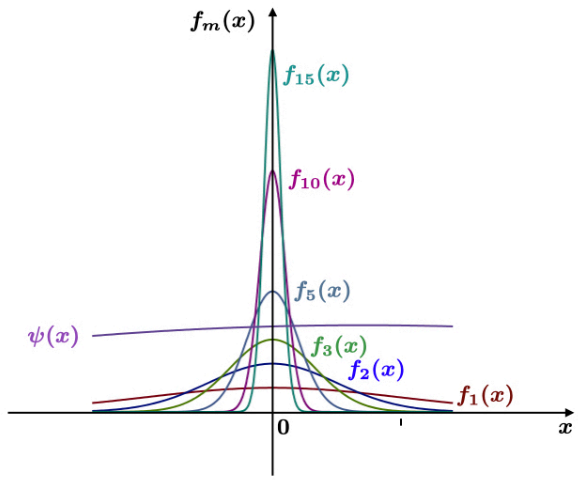


FIGURE 9: *Averaging of a sequence of Gaussian functions* $f_m(x) = \frac{1}{\sqrt{\pi}} m e^{-m^2 x^2}$ *using a test function* $\psi(x)$.

□

4.4 Regular and singular distributions

Definition 4.14 *A function* $f : \mathbb{R}^n \to \mathbb{R}$ *is called* ***locally integrable*** *if for all compact* $K \subseteq \mathbb{R}^n$*,*

$$\int_K |f(\boldsymbol{x})| \, d\boldsymbol{x} < \infty.$$

The set of locally integrable functions on $\mathbb{R}^n$ *is denoted* $L^1_{\text{loc}}(\mathbb{R}^n)$*.*

All integrable functions are locally integrable, but the reverse is not true. For example, $f : \mathbb{R} \to \mathbb{R} : x \mapsto x^2$ is not integrable since $\int_{\mathbb{R}} x^2 \, dx$ is not finite, but it is locally integrable as it is continuous and its integral over any compact subset of $\mathbb{R}$ is therefore finite.

The set of locally integrable functions is much larger than the set of continuous functions though.

Definition 4.15 $\mathcal{T} \in \mathcal{D}'(\mathbb{R}^n)$ *is called a* ***regular distribution*** *if there exists a locally integrable function* f *such that*

$$\mathcal{T}[\psi] = \mathcal{T}_f[\psi] := \int_{\mathbb{R}^n} f(\boldsymbol{x}) \psi(\boldsymbol{x}) \, d\boldsymbol{x}, \quad \psi \in \mathcal{D}(\mathbb{R}^n). \tag{4.7}$$

Note that the integral is over a compact set (the support of ψ) and hence it is well-defined.

The existence of regular distributions is fantastic news, although there is more excitement to come as not all distributions are regular. Regular distributions provide interesting insights, and in particular lift the problems caused in classical analysis when functions have special characteristics at a few isolated points. Let us show in an example. Consider the locally integrable function

$$f : \mathbb{R} \to \mathbb{R} : x \mapsto \begin{cases} 1 & x = 0 \\ 0 & \text{otherwise} \end{cases},$$

which defines a regular distribution $\mathcal{T}_f[\psi] = \int_{-\infty}^{\infty} f(x)\psi(x)\,dx = 0$. But no matter what ψ is, the integral must give zero because $f(x)$ is only nonzero (and finite) at a single point, *with zero area*. So *in the sense of distributions*, one can say that $f = 0$ even though $f \neq 0$ in the ordinary function case. We have the following result:

In general, any two ordinary locally integrable functions that only differ by finite amounts at isolated points define the same regular distribution.

Remark 4.16 We will make use later on of the Heaviside (unit step) function $\Theta(x - x_0)$ defined as (see Figure 10)

$$\Theta(x - x_0) = \begin{cases} 0 & x \leq x_0 \\ 1 & x > x_0. \end{cases} \tag{4.8}$$

In the sense of distributions, the value taken at $x = x_0$ by the Heaviside function is not important (provided it is finite), and is a matter of choice. Several choices are present in the literature, depending on what must be achieved. □

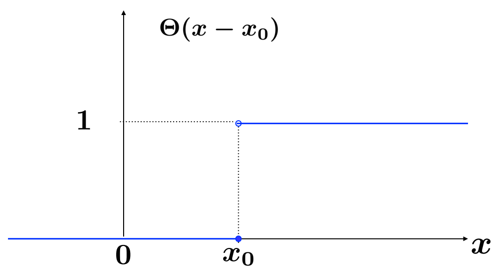


FIGURE 10: *The Heaviside (unit step function with step at $x = x_0$.*

Definition 4.17 *If there is no $f \in L^1_{\rm loc}(\mathbb{R}^n)$ such that a distribution $\mathcal{T}$ can be written as $\mathcal{T}_f$, the distribution is called **singular**.*

In analogy with the definition of regular distributions, it is customary to use the integral form (4.7) **in a symbolic way**, writing

$$\mathcal{T}[\psi] := \int_{\mathbb{R}^n} \mathcal{T}(\boldsymbol{x})\psi(\boldsymbol{x})\,d\boldsymbol{x} =: \langle \mathcal{T}, \psi \rangle. \tag{4.9}$$

However, $\mathcal{T}(\boldsymbol{x})$ is NOT a function and to specify which distribution is under consideration, one must give a separate definition of what value $\mathcal{T}[\psi]$ actually takes. The action of a distribution $\mathcal{T} \in \mathcal{D}'(\mathbb{R}^n)$ on a test function $\psi \in \mathcal{D}(\mathbb{R}^n)$ may be represented by a *duality pairing*, as it resembles an inner product

$$\langle \cdot, \cdot \rangle : \mathcal{D}'(\mathbb{R}^n) \times \mathcal{D}(\mathbb{R}^n) \to \mathbb{R},$$

which is why we have adopted the symbolic notation $\langle \mathcal{T}, \psi \rangle$. This has some advantages when performing calculations as we will amply see later, but do not be fooled: singular distributions are NOT functions, and it does not make sense to think of evaluating them at a particular point $\boldsymbol{x} \in \mathbb{R}^n$, which is why the notation $\mathcal{T}(\boldsymbol{x})$ is symbolic.

Example 4.18 The delta distribution $\boldsymbol{\delta}$ is not a regular distribution, i.e. there is no locally integrable function $f : \mathbb{R}^n \to \mathbb{R}$ such that $\delta[\psi] = \mathcal{T}_f[\psi]$ for all $\psi \in \mathcal{D}(\mathbb{R}^n)$. In other words, there is no function $f \in L^1_{\rm loc}$ such that

$$\psi(\boldsymbol{0}) = \int_{\mathbb{R}^n} f(\boldsymbol{x})\psi(\boldsymbol{x})d\boldsymbol{x}.$$

The δ distribution is often written in the following symbolic integral form,

$$\delta[\psi] = \langle \delta, \psi \rangle = \int_{\mathbb{R}^n} \delta(\boldsymbol{x})\psi(\boldsymbol{x})d\boldsymbol{x},$$

but $\delta(\boldsymbol{x})$ is not a function: it is the symbol for what is called historically (and abusively) the Dirac delta *function*. The separate definition specifying how the δ distribution acts on test functions was already given in Example 4.11, namely

$$\delta[\psi] := \psi(\mathbf{0}).$$

Putting everything together, we have

$$\delta[\psi] = \langle \delta, \psi \rangle = \int_{\mathbb{R}^n} \delta(\boldsymbol{x})\psi(\boldsymbol{x})d\boldsymbol{x} \overset{(1)}{=} \begin{cases} \psi(\mathbf{0}) & \text{for } \boldsymbol{x} = \mathbf{0} \\ 0 & \text{otherwise,} \end{cases} \qquad (\textbf{sifting property of } \boldsymbol{\delta} \textbf{ on } \mathbb{R}^{n}). \tag{4.10}$$

In (4.10), the relation (1) which *defines* $\delta(\boldsymbol{x})$ is called the *sifting property* of the delta distribution. More generally for $\Omega \subset \mathbb{R}^n$,

$$\delta[\psi] = \langle \delta, \psi \rangle = \int_{\Omega} \delta(\boldsymbol{x})\psi(\boldsymbol{x})d\boldsymbol{x} \overset{(1)}{=} \begin{cases} \psi(\mathbf{0}) & \text{if } \mathbf{0} \in \Omega \\ 0 & \text{otherwise,} \end{cases} \qquad (\textbf{sifting property of } \boldsymbol{\delta} \textbf{ on } \boldsymbol{\Omega} \subset \mathbb{R}^{\boldsymbol{n}}). \tag{4.11}$$

A shifted version of the sifting property of the δ distribution is presented in Example 4.19.

Notations and conventions: if $n = 1$, i.e. if considering $\mathbb{R}$, one writes

$$\delta(\boldsymbol{x}) = \delta(x),$$

and analogously for $n = 2$ and $n = 3$, one writes $\delta(\boldsymbol{x}) = \delta(x)\delta(y)$ and $\delta(\boldsymbol{x}) = \delta(x)\delta(y)\delta(z)$ for $\boldsymbol{x} = (x, y) \in \mathbb{R}^2$ and $\boldsymbol{x} = (x, y, z) \in \mathbb{R}^3$. For instance, consider the test function $\psi(\boldsymbol{x}) = \exp(-1/(1-\boldsymbol{x}^2)) \in \mathcal{D}(\mathbb{R}^2)$. Then

$$\int_{\mathbb{R}^2} \delta(\boldsymbol{x})\psi(\boldsymbol{x})\, d\boldsymbol{x} = \int_{\mathbb{R}} \delta(y) \left(\int_{\mathbb{R}} \delta(x) e^{-1/(1-x^2-y^2)}\, dx \right) dy = \int_{\mathbb{R}} \delta(y)\, e^{-1/(1-y^2)}\, dy = e^{-1}.$$

So we use the same symbol δ irrespectively of the dimension of its argument. The context is usually clear enough to avoid any ambiguity. □

End of Unit 4 Lect 2

4.5 Operations on distributions - Unit 4 Lect 3

Some operations on distributions affect the argument of the test function ψ appearing in the integral defining the action of the distribution $\mathcal{T}$. It is therefore necessary in some occasions to write $\mathcal{T}[\psi(\cdot)]$ instead of just $\mathcal{T}[\psi]$, where $\cdot$ is the argument of ψ within the defining integral.

One has, for all $\psi \in \mathcal{D}(\mathbb{R}^n)$, and $\mathcal{T}_1, \mathcal{T}_2, \mathcal{T} \in \mathcal{D}'(\mathbb{R}^n)$,

1. **addition of distributions** $\mathcal{T}_1 + \mathcal{T}_1$: $(\mathcal{T}_1 + \mathcal{T}_2)[\psi] := \mathcal{T}_1[\psi] + \mathcal{T}_2[\psi]$
2. **multiplication of a distribution by a constant** $c\mathcal{T}$: $(c\mathcal{T})[\psi] := c\mathcal{T}[\psi]$, $c \in \mathbb{R}$
3. **shifting of a distribution** $\mathcal{T}_{\boldsymbol{\xi}}$**,** $\boldsymbol{\xi} \in \mathbb{R}^n$:

$$\mathcal{T}_{\boldsymbol{\xi}}[\psi(\boldsymbol{x})] := \int_{\mathbb{R}^n} \mathcal{T}(\boldsymbol{x} - \boldsymbol{\xi})\psi(\boldsymbol{x})\, d\boldsymbol{x} \overset{(\mathbf{y} = \boldsymbol{x} - \boldsymbol{\xi})}{=} \int_{\mathbb{R}^n} \mathcal{T}(\mathbf{y})\psi(\mathbf{y} + \boldsymbol{\xi})\, d\mathbf{y} =: \mathcal{T}[\psi(\boldsymbol{x} + \boldsymbol{\xi})].$$

4. **transposition of a distribution $\mathcal{T}^t$**:

$$\mathcal{T}^t[\psi(\boldsymbol{x})] := \int_{\mathbb{R}^n} \mathcal{T}(-\boldsymbol{x})\psi(\boldsymbol{x})\,d\boldsymbol{x} \overset{(\mathbf{y}=-\boldsymbol{x})}{=} \int_{\mathbb{R}^n} \mathcal{T}(\mathbf{y})\psi(-\mathbf{y})\,d\mathbf{y} =: \mathcal{T}[\psi(-\boldsymbol{x})].$$

5. **dilation of a distribution $\mathcal{T}_{(\alpha)}$**, $\alpha \in \mathbb{R}$:

$$\mathcal{T}_{(\alpha)}[\psi(\boldsymbol{x})] := \int_{\mathbb{R}^n} \mathcal{T}(\alpha\boldsymbol{x})\psi(\boldsymbol{x})\,d\boldsymbol{x} \overset{(\mathbf{y}=\alpha\boldsymbol{x})}{=} \frac{1}{|\alpha|^n}\int_{\mathbb{R}^n} \mathcal{T}(\mathbf{y})\psi(\frac{\mathbf{y}}{\alpha})\,d\mathbf{y} =: \frac{1}{|\alpha|^n}\mathcal{T}[\psi(\frac{\boldsymbol{x}}{\alpha})]. \quad (4.12)$$

6. **multiplication of a distribution by a smooth function $\phi\mathcal{T}$**:

$$(\phi\mathcal{T})[\psi] := \mathcal{T}[\phi\psi] \quad \text{for all} \quad \phi \in C^\infty(\mathbb{R}^n). \quad (4.13)$$

These operations arise naturally in the case of regular distributions. The idea is then to extend the definition to singular distributions, using the integral $\langle \mathcal{T}, \psi\rangle$ of (4.9) in a symbolic way.

Example 4.19 Let us look at the δ distribution in the case $\delta \in \mathcal{D}'(\mathbb{R})$.

- First shift by $\xi \in \mathbb{R}$ according to operation 3 above to obtain

$$\delta_\xi[\psi(x)] = \int_{\mathbb{R}} \delta(x-\xi)\psi(x)dx \overset{(y=x-\xi)}{=} \int_{\mathbb{R}} \delta(y)\psi(y+\xi)\,dy \overset{(y=x)}{=} \int_{\mathbb{R}} \delta(x)\psi(x+\xi)\,dx \overset{(1)}{=} \psi(\xi)$$
$$\overset{(2)}{=} \delta[\psi(x+\xi)], \quad (4.14)$$

where equality (1) uses the sifting property of the δ distribution, while equality (2) uses the definition of a distribution. We call the equation

$$\int_{\mathbb{R}} \delta(x-\xi)\psi(x)dx = \psi(\xi) \quad (4.15)$$

the **(shifted) sifting property of the delta distribution in one dimension**. More generally, if $\Omega \subset \mathbb{R}$ is an open set, and $\psi \in \mathcal{D}(\Omega)$, the shifted sifting property is given by,

$$\delta_\xi[\psi(x)] = \int_{\Omega} \delta(x-\xi)\psi(x)dx = \begin{cases} \psi(\xi) & \text{if } \xi \in \Omega \\ 0 & \text{otherwise.} \end{cases} \quad (4.16)$$

- Second, as an example of (4.13), consider the Dirac δ_ξ distribution multiplied by a smooth function $\phi(x)$. Using the symbolic integral notation, we have

$$(\phi\delta_\xi)[\psi] = \langle \phi\delta_\xi, \psi\rangle = \int_{\mathbb{R}} \phi(x)\delta(x-\xi)\psi(x)dx = \int_{\mathbb{R}} \delta(x-\xi)\phi(x)\psi(x)dx = \phi(\xi)\psi(\xi), \quad \psi \in \mathcal{D}(\Omega).$$

But

$$\phi(\xi)\psi(\xi) = \phi(\xi)\int_{\mathbb{R}} \delta(x-\xi)\psi(x)\,dx = \int_{\mathbb{R}} \phi(\xi)\delta(x-\xi)\psi(x)\,dx,$$

so, symbolically,

$$\phi(x)\delta(x-\xi) = \phi(\xi)\delta(x-\xi), \quad (4.17)$$

which we shall call the **sifting of a smooth factor of the delta distribution**.

Remark 4.20 Let $\psi \in \mathcal{D}(\mathbb{R})$. The formula

$$\int_{\mathbb{R}} \delta(x-q_1)\,\delta(x-q_2)\,\psi(x)\,dx = \delta(q_1-q_2)\,\psi(q_2) \qquad \text{`in the sense of distributions'}$$

is often used in physics. It only makes sense if q_1 and q_2 are *variables* (which are integrated further in the problem to solve). The proof of this identity proceeds by integrating the LHS and the RHS against a test function $\phi(x) \in \mathcal{D}(\mathbb{R})$ and show that both sides produce the same output:

$$\begin{aligned} LHS &: \int_{\mathbb{R}} \int_{\mathbb{R}} \delta(x-q_1)\,\delta(x-q_2)\,\psi(x)\,\phi(q_1)\,dx\,dq_1 = \int_{\mathbb{R}} \delta(x-q_2)\,\psi(x)\,\phi(x)\,dx \\ &= \psi(q_2)\,\phi(q_2) \\ RHS &: \int_{\mathbb{R}} \delta(q_1-q_2)\,\psi(q_2)\,\phi(q_1)\,dq_1 = \psi(q_2)\,\phi(q_2). \end{aligned}$$

One may even replace the test function ψ by the constant function 1 (which does not have compact support) because the delta distribution has compact support (a fact we did not prove in this course). So one may see the identity written as

$$\int_{\mathbb{R}} \delta(x-q_1)\,\delta(x-q_2)\,dx = \delta(q_1-q_2),$$

but q_1 and q_2 are not constants.

- Third, we illustrate the dilation operation (4.12). In one dimension, and for $\alpha \in \mathbb{R}$ we get

$$\delta_{(\alpha)}[\psi] := \int_{\mathbb{R}} \delta(\alpha x)\psi(x)dx \overset{(y=\alpha x)}{=} \frac{1}{|\alpha|}\int_{\mathbb{R}} \delta(y)\psi(\frac{y}{\alpha})dy \overset{(1)}{=} \frac{1}{|\alpha|}\psi(0) \overset{(2)}{=} \frac{1}{|\alpha|}\delta[\psi(\frac{x}{\alpha})]$$

where equality (1) uses the sifting property of the δ distribution while equality (2) uses the definition of distribution. Furthermore,

$$\frac{1}{|\alpha|}\psi(0) = \frac{1}{|\alpha|}\int_{\mathbb{R}} \delta(x)\,\psi(x)\,dx = \frac{1}{|\alpha|}\,\delta[\psi],$$

so that

$$\delta_{(\alpha)} = \frac{1}{|\alpha|}\delta, \quad \text{or symbolically,} \quad \delta(\alpha x) = \frac{1}{|\alpha|}\delta(x).$$

Note that this covers the case $\alpha = 0$, as one defines $\delta(0) = \infty$. As a special case, consider $\alpha = -1$, which leads to $\delta(-x) = \delta(x)$ (some people talk of 'even' distribution in this case).

- Fourth, combining shift and dilation in one dimension, we find

$$\int_{\mathbb{R}} \delta(ax+b)\,\psi(x)\,dx \overset{(y=ax+b)}{=} \int_{\mathbb{R}} \delta(y)\,\psi(\frac{y-b}{a})\,\frac{1}{|a|}\,dy = \frac{1}{|a|}\psi(-\frac{b}{a}).$$

- Fifth, in n dimensions, the most general linear transformation A on $\boldsymbol{x}$ combined with a shift $\boldsymbol{b}$ yields (for A non singular, i.e. such that $\det A \neq 0$),

$$\int_{\mathbb{R}^n} \delta(A\boldsymbol{x}-\boldsymbol{b})\,\psi(\boldsymbol{x})\,d\boldsymbol{x} \overset{(\boldsymbol{y}=A\boldsymbol{x}-\boldsymbol{b})}{=} \int_{\mathbb{R}^n} \delta(\boldsymbol{y})\psi(A^{-1}(\boldsymbol{y}+\boldsymbol{b}))\,\frac{1}{|\det A|}\,d\boldsymbol{y} = \frac{1}{|\det A|}\,\psi(A^{-1}\boldsymbol{b}). \quad (4.18)$$

□

More generally, for any function $f : \Omega \to \mathbb{R}$ with $f \in \mathcal{C}^1(\Omega)$, we have that $\delta(f(x))$ is zero everywhere except at the zeros of f, so that an integral involving $\delta(f(x))$ times a test function receives no contribution outside an arbitrarily small neighbourhood of the zeros of f. In particular if the only zeros of f are simple zeros at points $\{x_1, \ldots, x_n\} \in \Omega$ then

$$\int_\Omega \delta(f(x))\psi(x)dx = \sum_{i=1}^{n} \int_{x_i-\epsilon}^{x_i+\epsilon} \delta(f(x))\psi(x)dx.$$

Using the fact that $f(x) \approx (x - x_i)f'(x_i)$ when $x \to x_i$, we have

$$\int_\Omega \delta(f(x))\psi(x)dx = \sum_{i=1}^{n}\left[\int_{x_i-\epsilon}^{x_i+\epsilon} \delta(\,(x-x_i)\,f'(x_i)\,)\,\psi(x)dx\right] \overset{(1)}{=} \sum_{i=1}^{n}\left[\frac{1}{|f'(x_i)|}\int_{x_i-\epsilon}^{x_i+\epsilon} \delta(x-x_i)\psi(x)\,dx\right]$$
$$\overset{(2)}{=} \sum_{i=1:f(x_i)=0}^{n} \frac{\psi(x_i)}{|f'(x_i)|}, \tag{4.19}$$

where the equality (1) uses the scaling property $\delta(\alpha x) = \frac{1}{|\alpha|}\delta(x)$, while equality (2) uses the sifting property of the delta function. □

Example 4.21 Let $f : \mathbb{R} \to \mathbb{R} : x \mapsto f(x) = x^2 - b^2$ with $b \in \mathbb{R}, b > 0$. Since $f(x)$ has two simple zeros at $x = \pm b$, one finds

$$\int_{\mathbb{R}} \delta(x^2 - b^2)\psi(x)dx = \frac{1}{2|b|}\{\psi(b) + \psi(-b)\}.$$

So the important point is that the integral is localized to an arbitrarily small neighbourhood of the zeros of the argument of the δ-function. □

Remark 4.22 In general there is no way to define the product of two distributions. This is a drawback of the theory of distributions. □

We end this survey of operations on distributions with a criterion allowing to decide if two distributions are equal.

Definition 4.23 *Two distributions $\mathcal{T}_1$ and $\mathcal{T}_2$ are said to be equal if*

$$\int_\Omega \mathcal{T}_1(x)\psi(x)dx = \int_\Omega \mathcal{T}_2(x)\psi(x)dx \text{ for every test function } \psi \in \mathcal{D}(\Omega), \tag{4.20}$$

or formally, if $\langle \mathcal{T}_1, \psi\rangle = \langle \mathcal{T}_2, \psi\rangle$ $\forall \psi \in \mathcal{D}(\Omega)$.

This criterion is handy in many concrete calculations on distributions.

Example 4.24 For which value $A \in \mathbb{R}$ are the following distributions equal?

$$T_1(x) = x^2\delta(x^3) \quad \text{and} \quad T_2(x) = A\delta(x), \qquad T_1, T_2 \in \mathcal{D}'(\mathbb{R})?$$

Integrating each distribution against a test function $\psi \in \mathcal{D}(\mathbb{R})$, one finds

$$\int_{\mathbb{R}} x^2\delta(x^3)\,\psi(x)\,dx \overset{(y=x^3)}{=} \frac{1}{3}\int_{\mathbb{R}} \delta(y)\psi(y^{1/3})\,dy = \frac{1}{3}\psi(0)$$
$$\int_{\mathbb{R}} A\delta(x)\psi(x)\,dx = A\psi(0).$$

Hence the two distributions are equal if $A = 1/3$. □

One may define further useful operations on distributions besides the basic operations reviewed above. One of them is differentiation. **End of Unit 4 Lect 3**