Chapter 0
PRELIMINARIES

The purpose of this chapter is to fix some terminology that will be used throughout the book, and to present a few analytical tools which are not included in the prerequisites. It is intended mainly as a reference rather than as a systematic text.

A. Notations and Definitions

Points and sets in Euclidean space

$\mathbb{R}$ will denote the real numbers, $\mathbb{C}$ the complex numbers. We will be working in $\mathbb{R}^n$, and n will always denote the dimension. Points in $\mathbb{R}^n$ will generally be denoted by x, y, ξ, η; the coordinates of x are $(x_1, \ldots, x_n)$. Occasionally $x_1, x_2, \ldots$ will denote a sequence of points in $\mathbb{R}^n$ rather than coordinates, but this will always be clear from the context. Once in a while there will be some confusion as to whether $(x_1, \ldots, x_n)$ denotes a point in $\mathbb{R}^n$ or the n-tuple of coordinate functions on $\mathbb{R}^n$. However, it would be too troublesome to adopt systematically a more precise notation; readers should consider themselves warned that this ambiguity will arise when we consider coordinate systems other than the standard one.

If U is a subset of $\mathbb{R}^n$, $\overline{U}$ will denote its closure and ∂U its boundary. The word **domain** will be used to mean an open set $\Omega \subset \mathbb{R}^n$, not necessarily connected, such that $\partial\Omega = \partial(\mathbb{R}^n \setminus \overline{\Omega})$. (That is, all the boundary points of Ω are "accessible from the outside.")

If x and y are points of $\mathbb{R}^n$ or $\mathbb{C}^n$, we set

$$x \cdot y = \sum_1^n x_j y_j ,$$

so the Euclidean norm of x is given by

$$|x| = (x \cdot \overline{x})^{1/2} \quad (= (x \cdot x)^{1/2} \text{ if } x \text{ is real.})$$

We use the following notation for spheres and (open) balls: if $x \in \mathbb{R}^n$ and $r > 0$,

$$S_r(x) = \{y \in \mathbb{R}^n : |x - y| = r\},$$
$$B_r(x) = \{y \in \mathbb{R}^n : |x - y| < r\}.$$

Measures and integrals

The integral of a function f over a subset Ω of $\mathbb{R}^n$ with respect to Lebesgue measure will be denoted by $\int_\Omega f(x)\,dx$ or simply by $\int_\Omega f$. If no subscript occurs on the integral sign, the region of integration is understood to be $\mathbb{R}^n$. If S is a smooth hypersurface (see the next section), the natural Euclidean surface measure on S will be denoted by $d\sigma$; thus the integral of f over S is $\int_S f(x)\,d\sigma(x)$, or $\int_S f\,d\sigma$, or just $\int_S f$. The meaning of $d\sigma$ thus depends on S, but this will cause no confusion.

If f and g are functions whose product is integrable on $\mathbb{R}^n$, we shall sometimes write

$$\langle f, g\rangle = \int fg, \qquad \langle f \,|\, g\rangle = \int f\overline{g},$$

where $\overline{g}$ is the complex conjugate of g. The Hermitian pairing $\langle f|g\rangle$ will be used only when we are working with the Hilbert space L^2 or a variant of it, whereas the bilinear pairing $\langle f, g\rangle$ will be used more generally.

Multi-indices and derivatives

An n-tuple $\alpha = (\alpha_1, \ldots a_n)$ of nonnegative integers will be called a **multi-index**. We define

$$|\alpha| = \sum_1^n \alpha_j, \qquad \alpha! = \alpha_1!\alpha_2!\cdots\alpha_n!,$$

and for $x \in \mathbb{R}^n$,

$$x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}.$$

We will generally use the shorthand

$$\partial_j = \frac{\partial}{\partial x_j}$$

for derivatives on $\mathbb{R}^n$. Higher-order derivatives are then conveniently expressed by multi-indices:

$$\partial^\alpha = \prod_1^n \left(\frac{\partial}{\partial x_j}\right)^{\alpha_j} = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1}\cdots\partial x_n^{\alpha_n}}.$$

Note in particular that if $\alpha = 0$, ∂^α is the identity operator. With this notation, it would be natural to denote by ∂u the n-tuple of functions $(\partial_1 u, \ldots, \partial_n u)$ when u is a differentiable function; however, we shall use instead the more common notation

$$\nabla u = (\partial_1 u, \ldots, \partial_n u).$$

For our purposes, **a vector field** on a set $\Omega \in \mathbb{R}^n$ is simply an $\mathbb{R}^n$-valued function on Ω. If F is a vector field on an open set Ω, we define the directional derivative ∂_F by

$$\partial_F = F \cdot \nabla,$$

that is, if u is a differentiable function on Ω,

$$\partial_F u(x) = F(x) \cdot \nabla u(x) = \sum_1^n F_j(x)\partial_j u(x).$$

Function spaces

If Ω is a subset of $\mathbb{R}^n$, $C(\Omega)$ will dente the space of continuous complex-valued functions on Ω (with respect to the relative topology on Ω). If Ω is open and k is a positive integer, $C^k(\Omega)$ will denote the space of functions possessing continuous derivatives up to order k on Ω, and $C^k(\overline{\Omega})$ will denote the space of all $u \in C^k(\Omega)$ such that $\partial^\alpha u$ extends continuously to the closure $\overline{\Omega}$ for $0 \le |\alpha| \le k$. Also, we set $C^\infty(\Omega) = \bigcap_1^\infty C^k(\Omega)$ and $C^\infty(\overline{\Omega}) = \bigcap_1^\infty C^k(\overline{\Omega})$.

We next define the Hölder or Lipschitz spaces $C^\alpha(\Omega)$, where Ω is an open set and $0 < \alpha < 1$. (Here α is a real number, not a multi-index; the use of the letter "α" in both these contexts is standard.) $C^\alpha(\Omega)$ is the space of continuous functions on Ω that satisfy a locally uniform Hölder condition with exponent α. That is, $u \in C^\alpha(\Omega)$ if and only if for any compact $V \subset \Omega$ there is a constant $c > 0$ such that for all $y \in \mathbb{R}^n$ sufficiently close to 0,

$$\sup_{x \in V} |u(x + y) - u(x)| \le c|y|^\alpha.$$

(Note that $C^1(\Omega) \subset C^\alpha(\Omega)$ for all $\alpha < 1$, by the mean value theorem.) If k is a positive integer, $C^{k+\alpha}(\Omega)$ will denote the set of all $u \in C^k(\Omega)$ such that $\partial^\beta u \in C^\alpha(\Omega)$ for all muli-indices β with $|\beta| = k$ (or equivalently, with $|\beta| \le k$; the lower-order derivatives are automatically in $C^1(\Omega) \subset C^\alpha(\Omega)$.).

The **support** of a function u, denoted by supp u, is the complement of the largest open set on which $u = 0$. If $\Omega \subset \mathbb{R}^n$, we denote by $C_c^\infty(\Omega)$ the space of all C^∞ functions on $\mathbb{R}^n$ whose support is compact and contained in Ω. (In particular, if Ω is open such functions vanish near $\partial\Omega$.)

The space $C^k(\mathbb{R}^n)$ will be denoted simply by C^k. Likewise for C^∞, $C^{k+\alpha}$, and C_c^∞.

If $\Omega \subset \mathbb{R}^n$ is open, a function $u \in C^\infty(\Omega)$ is said to be **analytic** in Ω if it can be expanded in a power series about every point of Ω. That is, u is analytic on Ω if for each $x \in \Omega$ there exists $r > 0$ such that for all $y \in B_r(x)$,

$$u(y) = \sum_{|\alpha| \ge 0} \frac{\partial^\alpha u(x)}{\alpha!}(y - x)^\alpha,$$

the series being absolutely and uniformly convergent on $B_r(x)$. When referring to complex-analytic functions, we shall always use the word **holomorphic**.

The **Schwartz class** $\mathcal{S} = \mathcal{S}(\mathbb{R}^n)$ is the space of all C^∞ functions on $\mathbb{R}^n$ which, together with all their derivatives, die out faster than any power of x at infinity. That is, $u \in \mathcal{S}$ if and only if $u \in C^\infty$ and for all multi-indices α and β,

$$\sup_{x \in \mathbb{R}^n} |x^\alpha \partial^\beta u(x)| < \infty.$$

Big O and little o

We occasionally employ the big and little o notation for orders of magnitude. Namely, when we are considering the behavior of functions in a neighborhood of a point a (which may be ∞), $O(f(x))$ denotes any function $g(x)$ such that $|g(x)| \le C|f(x)|$ for x near a, and $o(f(x))$ denotes any function $h(x)$ such that $h(x)/f(x) \to 0$ as $x \to a$.

B. Results from Advanced Calculus

A subset S of $\mathbb{R}^n$ is called a **hypersurface of class** C^k $(1 \le k \le \infty)$ if for every $x_0 \in S$ there is an open set $V \subset \mathbb{R}^n$ containing x_0 and a real-valued function $\phi \in C^k(V)$ such that $\nabla\phi$ is nonvanishing on $S \cap V$ and

$$S \cap V = \{x \in V : \phi(x) = 0\}.$$

In this case, by the implicit function theorem we can solve the equation $\phi(x) = 0$ near x_0 for some coordinate x_i — for convenience, say $i = n$ — to obtain

$$x_n = \psi(x_1, \ldots, x_{n-1})$$

for some C^k function ψ. A neighborhood of x_0 in S can then be mapped to a piece of the hyperplane $x_n = 0$ by the C^k transformation

$$x \to (x', x_n - \psi(x')) \qquad (x' = (x_1, \ldots, x_{n-1})).$$

This same neighborhood can also be represented in **parametric form** as the image of an open set in $\mathbb{R}^{n-1}$ (with coordinate x') under the map

$$x' \to (x', \psi(x')).$$

x' may be thought of as giving local coordinates on S near x_0.

Similar considerations apply if "C^k" is replaced by "analytic."

With S, V, ϕ as above, the vector $\nabla\phi(x)$ is perpendicular to S at x for every $x \in S \cap V$. We shall always suppose that S is **oriented**, that is, that we have made a choice of unit vector $\nu(x)$ for each $x \in S$, varying continuously with x, which is perpendicular to S at x. $\nu(x)$ will be called the **normal** to S at x; clearly on $S \cap V$ we have

$$\nu(x) = \pm\frac{\nabla\phi(x)}{|\nabla\phi(x)|}.$$

Thus ν is a C^{k-1} function on S. If S is the boundary of a domain Ω, we always choose the orientation so that ν points out of Ω.

If u is a differentiable function defined near S, we can then define the **normal derivative** of u on S by

$$\partial_\nu u = \nu \cdot \nabla u.$$

We pause to compute the normal derivative on the sphere $S_r(y)$. Since lines through the center of a sphere are perpendicular to the sphere, we have

$$(0.1) \qquad \nu(x) = \frac{x - y}{r}, \quad \partial_\nu = \frac{1}{r}\sum_1^n (x_j - y_j)\partial_j \qquad \text{on } S_r(y).$$

We will use the following proposition several times in the sequel:

(0.2) Proposition.
Let S be a compact oriented hypersurface of class C^k, $k \geq 2$. There is a neighborhood V of S in $\mathbb{R}^n$ and a number $\epsilon > 0$ such that the map

$$F(x,t) = x + t\nu(x)$$

is a C^{k-1} diffeomorphism of $S \times (-\epsilon, \epsilon)$ onto V.

Proof (sketch): F is clearly C^{k-1}. Moreover, for each $x \in S$ its Jacobian matrix (with respect to local coordinates on $S \times \mathbb{R}$) at $(x,0)$ is nonsingular since ν is normal to S. Hence by the inverse mapping theorem, F can be inverted on a neighborhood W_x of each $(x,0)$ to yield a C^{k-1} map

$$F_x^{-1} : W_x \to (S \cap W_x) \times (-\epsilon_x, \epsilon_x)$$

for some $\epsilon_x > 0$. Since S is compact, we can choose $\{x_j\}_1^N \subset S$ such that the W_{x_j} cover S, and the maps $F_{x_j}^{-1}$ patch together to yield a C^{k-1} inverse of F from a neighborhood V of S to $S \times (-\epsilon, \epsilon)$ where $\epsilon = \min_j \epsilon_{x_j}$. ∎

The neighborhood V in Proposition (0.2) is called a **tubular neighborhood** of S. It will be convenient to extend the definition of the normal derivative to the whole tubular neighborhood. Namely, if u is a differentiable function on V, for $x \in S$ and $-\epsilon < t < \epsilon$ we set

$$\partial_\nu u(x + t\nu(x)) = \nu(x) \cdot \nabla u(x + t\nu(x)). \tag{0.3}$$

If $F = (F_1, \ldots, F_n)$ is a differentiable vector field on a subset of $\mathbb{R}^n$, its **divergence** is the function

$$\nabla \cdot F = \sum_1^n \partial_j F_j.$$

With this terminology, we can state the form of the general Stokes formula that we shall need.

(0.4) The Divergence Theorem.
Let $\Omega \subset \mathbb{R}^n$ be a bounded domain with C^1 boundary $S = \partial\Omega$, and let F be a C^1 vector field on $\overline{\Omega}$. Then

$$\int_S F(y) \cdot \nu(y)\, d\sigma(y) = \int_\Omega \nabla \cdot F(x)\, dx.$$

The proof can be found, for example, in Treves [52, §10].

Every $x \in \mathbb{R}^n \setminus \{0\}$ can be written uniquely as $x = ry$ with $r > 0$ and $y \in S_1(0)$ — namely, $r = |x|$ and $y = x/|x|$. The formula $x = ry$ is called the **polar coordinate** representation of x. Lebesgue measure is given in polar coordinates by

$$dx = r^{n-1}\, dr\, d\sigma(y),$$

where $d\sigma$ is surface measure on $S_1(0)$. (See Folland [14, Theorem (2.49)].)

For example, if $0 < a < b < \infty$ and $\lambda \in \mathbb{R}$, we have

$$\int_{a<|x|<b} |x|^\lambda\, dx = \int_{S_1(0)} \int_a^b r^{n-1+\lambda}\, dr = \begin{cases} \omega_n \frac{b^{n+\lambda} - a^{n+\lambda}}{n+\lambda} & \text{if } \lambda \neq -n, \\ \omega_n \log(b/a) & \text{if } \lambda = -n, \end{cases}$$

where ω_n is the area of $S_1(0)$ (which we shall compute shortly). As an immediate consequence, we have:

(0.5) Proposition.
The function $x \to |x|^\lambda$ is integrable on a neighborhood of 0 if and only if $\lambda > -n$, and it is integrable outside a neighborhood of 0 if and only if $\lambda < -n$.

As another application of polar coordinates, we can compute what is probably the most important definite integral in mathematics:

(0.6) Proposition.
$\int e^{-\pi|x|^2}\, dx = 1.$

Proof: Let $I_n = \int_{\mathbb{R}^n} e^{-\pi|x|^2}\, dx$. Since $e^{-\pi|x|^2} = \prod_1^n e^{-\pi x_j^2}$, Fubini's theorem shows that $I_n = (I_1)^n$, or equivalently that $I_n = (I_2)^{n/2}$. But in polar coordinates,

$$I_2 = \int_0^{2\pi} \int_0^\infty e^{-\pi r^2} r\, dr\, d\theta = 2\pi \int_0^\infty re^{-r^2}\, dr = \pi \int_0^\infty e^{-\pi s}\, ds = 1. \quad ∎$$

This trick works because we know that the measure of $S_1(0)$ in $\mathbb{R}^2$ is 2π. But now we can turn it around to compute the area ω_n of $S_1(0)$ in $\mathbb{R}^n$ for any n. Recall that the **gamma function** $\Gamma(s)$ is defined for $\operatorname{Re} s > 0$ by

$$\Gamma(s) = \int_0^\infty e^{-t} t^{s-1}\, dt.$$

One easily verifies that

$$\Gamma(s+1) = s\Gamma(s), \qquad \Gamma(1) = 1, \qquad \Gamma(\tfrac{1}{2}) = \sqrt{\pi}.$$

(The first formula is obtained by integration by parts, and the last one reduces to (0.6) by a change of variable.) Hence, if k is a positive integer,

$$\Gamma(k) = (k-1)!, \qquad \Gamma(k+\tfrac{1}{2}) = (k-\tfrac{1}{2})(k-\tfrac{3}{2})\cdots(\tfrac{1}{2})\sqrt{\pi}.$$

(0.7) Proposition.
The area of $S_1(0)$ in $\mathbb{R}^n$ is

$$\omega_n = \frac{2\pi^{n/2}}{\Gamma(n/2)}.$$

Proof: We integrate $e^{-\pi|x|^2}$ in polar coordinates and set $s = \pi r^2$:

$$\begin{aligned} 1 &= \int e^{-\pi|x|^2}\,dx = \int_{S_1(0)}\int_0^\infty e^{-\pi r^2} r^{n-1}\,dr\,d\sigma \\ &= \omega_n \int_0^\infty e^{-\pi r^2} r^{n-1}\,dr = \frac{\omega_n}{2\pi^{n/2}}\int_0^\infty e^{-s}s^{(n/2)-1}\,ds \\ &= \frac{\omega_n\Gamma(n/2)}{2\pi^{n/2}}. \end{aligned}$$

∎

Note that, despite appearances, ω_n is always a rational multiple of an integer power of π.

(0.8) Corollary.
The volume of $B_1(0)$ in $\mathbb{R}^n$ is

$$\frac{\omega_n}{n} = \frac{2\pi^{n/2}}{n\Gamma(n/2)}.$$

Proof: $\int_{B_1(0)} dx = \omega_n \int_0^1 r^{n-1}\,dr = \omega_n/n$. ∎

(0.9) Corollary.
For any $x \in \mathbb{R}^n$ and any $r > 0$, the area of $S_r(x)$ is $r^{n-1}\omega_n$ and the volume of $B_r(x)$ is $r^n\omega_n/n$.

C. Convolutions

We begin with a general theorem about integral operators on a measure space (X,μ) which deserves to be more widely known than it is. In our applications, X will be either $\mathbb{R}^n$ or a smooth hypersurface in $\mathbb{R}^n$.

(0.10) Generalized Young's Inequality.
Let (X,μ) be a σ-finite measure space, and let $1 \le p \le \infty$ and $C > 0$. Suppose K is a measurable function on $X \times X$ such that

$$\sup_{x\in X}\int_X |K(x,y)|\,d\mu(y) \le C, \qquad \sup_{y\in X}\int_X |K(x,y)|\,d\mu(x) \le C.$$

If $f \in L^p(X)$, the function Tf defined by

$$Tf(x) = \int_X K(x,y)f(y)\,d\mu(y)$$

is well-defined almost everywhere and is in $L^p(X)$, and $\|Tf\|_p \le C\|f\|_p$.

Proof: Suppose $1 < p < \infty$, and let q be the conjugate exponent $(p^{-1}+q^{-1}=1)$. Then by Hölder's inequality,

$$\begin{aligned} |Tf(x)| &\le \left[\int_X |K(x,y)|\,d\mu(y)\right]^{1/q}\left[\int_X |K(x,y)||f(y)|^p\,d\mu(y)\right]^{1/p} \\ &\le C^{1/q}\left[\int_X |K(x,y)||f(y)|^p\,d\mu(y)\right]^{1/p}. \end{aligned}$$

Raising both sides to the p-th power and integrating, we see by Fubini's theorem that

$$\begin{aligned} \int_X |Tf(x)|^p\,d\mu(x) &\le C^{p/q}\int_X\int_X |K(x,y)||f(y)|^p\,d\mu(y)\,d\mu(x) \\ &\le C^{(p/q)+1}\int_X |f(y)|^p\,d\mu(y), \end{aligned}$$

or, taking pth roots,

$$\|Tf\|_p \le C^{(1/p)+(1/q)}\|f\|_p = C\|f\|_p.$$

These estimates imply, in particular, that the integral defining $Tf(x)$ converges absolutely a.e., so the theorem is proved for the case $1 < p < \infty$. The case $p = 1$ is similar but easier and requires only the hypothesis $\int |K(x,y)|\,d\mu(x) \le C$, and the case $p = \infty$ is trivial and requires only the hypothesis $\int |K(x,y)|\,d\mu(y) \le C$. ∎

In what follows, when we say L^p we shall mean $L^p(\mathbb{R}^n)$ unless another space is specified.

Let f and g be locally integrable functions on $\mathbb{R}^n$. The **convolution** $f * g$ of f and g is defined by

$$f * g(x) = \int f(x-y)g(y)\,dy = \int f(y)g(x-y)\,dy = g * f(x),$$

provided that the integrals in question exist. (The two integrals are equal by the change of variable $y \to x - y$.) The basic theorem on the existence of convolutions is the following:

(0.11) Young's Inequality.
If $f \in L^1$ and $g \in L^p$ $(1 \le p \le \infty)$, then $f*g \in L^p$ and $\|f*g\|_p \le \|f\|_1\|g\|_p$.

Proof: Apply (0.10) with $X = \mathbb{R}^n$ and $K(x,y) = f(x-y)$. ∎

Remark: It is obvious from Hölder's inequality that if $f \in L^q$ and $g \in L^p$ where $p^{-1} + q^{-1} = 1$ then $f * g \in L^\infty$ and $\|f * g\|_\infty \le \|f\|_q\|g\|_p$. From the Riesz-Thorin interpolation theorem (see Folland [14]) one can then deduce the following generalization of Young's inequality: *Suppose $1 \le p, q, r \le \infty$ and $p^{-1} + q^{-1} = r^{-1} + 1$. If $f \in L^q$ and $g \in L^p$ then $f * g \in L^r$ and $\|f * g\|_r \le \|f\|_q\|g\|_p$.*

The next theorem underlies one of the most important uses of convolutions. Before coming to it, we need a technical lemma. If f is a function on $\mathbb{R}^n$ and $x \in \mathbb{R}^n$, we define the function f_x by

$$f_x(y) = f(x+y).$$

(0.12) Lemma.
If $1 \le p < \infty$ and $f \in L^p$, then $\lim_{x\to 0} \|f_x - f\|_p = 0$.

Proof: If g is continuous with compact support, then g is uniformly continuous, so $g_x \to g$ uniformly as $x \to 0$. Since g_x and g are supported in a common compact set for $|x| \le 1$, it follows also that $\|g_x - g\|_p \to 0$. Now, given $f \in L^p$ and $\epsilon > 0$, choose a continuous g with compact support such that $\|f - g\|_p < \epsilon/3$. Then also $\|f_x - g_x\|_p < \epsilon/3$, so

$$\|f_x - f\|_p \le \|f_x - g_x\|_p + \|g_x - g\|_p + \|g - f\|_p < \|g_x - g\|_p + 2\epsilon/3.$$

But for x sufficiently small, $\|g_x - g\|_p < \epsilon/3$, so $\|f_x - f\|_p < \epsilon$. ∎

Remark: This result is false for $p = \infty$. Indeed, the condition that $\|f_x - f\|_\infty \to 0$ as $x \to 0$ means precisely that f agrees almost everywhere with a uniformly continuous function.

(0.13) Theorem.
Suppose $\phi \in L^1$ and $\int \phi(x)\,dx = a$. For each $\epsilon > 0$, define the function ϕ_ϵ by $\phi_\epsilon(x) = \epsilon^{-n}\phi(\epsilon^{-1}x)$. If $f \in L^p$, $1 \le p < \infty$, then $f * \phi_\epsilon \to af$ in the L^p norm as $\epsilon \to 0$. If $f \in L^\infty$ and f is uniformly continuous on a set V, then $f * \phi_\epsilon \to af$ uniformly on V as $\epsilon \to 0$.

Proof: By the change of variable $x \to \epsilon x$ we see that $\int \phi_\epsilon(x)\,dx = a$ for all $\epsilon > 0$. Hence,

$$f*\phi_\epsilon(x) - af(x) = \int [f(x-y) - f(x)]\phi_\epsilon(y)\,dy = \int [f(x-\epsilon y) - f(x)]\phi(y)\,dy.$$

If $f \in L^p$ and $p < \infty$, we apply the triangle inequality for integrals (Minkowski's inequality; see Folland [14]) to obtain

$$\|f * \phi_\epsilon - af\|_p \le \int \|f_{-\epsilon y} - f\|_p |\phi(y)|\,dy.$$

But $\|f_{-\epsilon y} - f\|_p$ is bounded by $2\|f\|_p$ and tends to zero as $\epsilon \to 0$ for each y, by Lemma (0.12). The desired result therefore follows from the dominated convergence theorem.

On the other hand, suppose $f \in L^\infty$ and f is uniformly continuous on V. Given $\delta > 0$, choose a compact set W so that $\int_{\mathbb{R}^n \setminus W} |\phi| < \delta$. Then

$$\sup_{x \in V} |f * \phi_\epsilon(x) - af(x)| \le \sup_{x\in V,\ y \in W} |f(x - \epsilon y) - f(x)| \int_W |\phi| + 2\|f\|_\infty \delta.$$

The first term on the right tends to zero as $\epsilon \to 0$, and δ is arbitrary, so $f * \phi_\epsilon$ tends uniformly to af on V. ∎

If $\phi \in L^1$ and $\int \phi(x)\,dx = 1$, the family of functions $\{\phi_\epsilon\}_{\epsilon > 0}$ defined in Theorem (0.13) is called an **approximation to the identity**. What makes these useful is that by choosing ϕ appropriately we can get the functions $f * \phi_\epsilon$ to have nice properties. In particular:

(0.14) Theorem.
If $f \in L^p$ $(1 \le p \le \infty)$ and ϕ is in the Schwartz class $\mathcal{S}$, then $f * \phi$ is C^∞ and $\partial^\alpha(f * \phi) = f * \partial^\alpha\phi$ for all multi-indices α.

Proof: If $\phi \in \mathcal{S}$, for every bounded set $V \subset \mathbb{R}^n$ we have

$$\sup_{x \in V} |\partial^\alpha \phi(x-y)| \le C_{\alpha,V}(1+|y|)^{-n-1} \qquad (y \in \mathbb{R}^n).$$

The function $(1+|y|)^{-n-1}$ is in L^q for every q by (0.5), so the integral

$$f * \partial^\alpha \phi(x) = \int f(y) \partial^\alpha \phi(x-y)\, dy$$

converges absolutely and uniformly on bounded subsets of $\mathbb{R}^n$. Differentiation can thus be interchanged with integration, and we conclude that $\partial^\alpha(f * \phi) = f * \partial^\alpha \phi$. ∎

We can get better results by taking $\phi \in C_c^\infty$. In that case we need only assume that f is locally integrable for $f * \phi$ to be well-defined, and the same argument as above shows that $f * \phi \in C^\infty$.

Since the existence of nonzero functions in C_c^∞ is not completely trivial, we pause for a moment to construct some. First, we define the function f on $\mathbb{R}$ by

$$f(t) = \begin{cases} e^{1/(1-t^2)} & (|t| < 1), \\ 0 & (|t| \ge 1). \end{cases}$$

Then $f \in C_c^\infty(\mathbb{R})$, so $\psi(x) = f(|x|^2)$ is a nonnegative C^∞ function on $\mathbb{R}^n$ whose support is $\overline{B_1(0)}$. In particular, $\int \psi > 0$, so $\phi = \psi / \int \psi$ is a function in $C_c^\infty(\mathbb{R}^n)$ with $\int \phi = 1$. It now follows that there are lots of functions in C_c^∞:

(0.15) Lemma.
If f is supported in V and g is supported in W, then $f * g$ is supported in $\{x + y : x \in V,\ y \in W\}$.

Proof: Exercise. ∎

(0.16) Theorem.
C_c^∞ is dense in L^p for $1 \le p < \infty$.

Proof: Choose $\phi \in C_c^\infty$ with $\int \phi = 1$, and define ϕ_ϵ as in Theorem (0.13). If $f \in L^p$ has compact support, it follows from (0.14) and (0.15) that $f * \phi_\epsilon \in C_c^\infty$ and from (0.13) that $f * \phi_\epsilon \to f$ in the L^p norm. But L^p functions with compact support are dense in L^p, so we are done. ∎

Another useful construction is the following:

(0.17) Theorem.
Suppose $V \subset \mathbb{R}^n$ is compact, $\Omega \subset \mathbb{R}^n$ is open, and $V \subset \Omega$. Then there exists $f \in C_c^\infty(\Omega)$ such that $f = 1$ on V and $0 \le f \le 1$ everywhere.

Proof: Let $\delta = \inf\{|x-y| : x \in V,\ y \notin \Omega\}$. (If $\Omega = \mathbb{R}^n$, let $\delta = 1$.) By our assumptions on V and Ω, $\delta > 0$. Let

$$U = \{x : |x-y| < \tfrac{1}{2}\delta \text{ for some } y \in V\}.$$

Then $V \subset U$ and $\overline{U} \subset \Omega$. Let χ be the characteristic function of U, and choose a nonnegative $\phi \in C_c^\infty(B_{\delta/2}(0))$ such that $\int \phi = 1$. Then we can take $f = \chi * \phi$; the simple verification is left to the reader. ∎

We can now prove the existence of "partitions of unity." We state the following results only for compact sets, which is all we need, but they can be generalized.

(0.18) Lemma.
Let $K \subset \mathbb{R}^n$ be compact and let $V_1, \ldots, V_N$ be open sets with $K \subset \bigcup_1^N V_j$. Then there exist open sets $W_1, \ldots, W_N$ with $\overline{W_j} \subset V_j$ and $K \subset \bigcup_1^N W_j$.

Proof: For each $\epsilon > 0$ let V_j^ϵ be the set of points in V_j whose distance from $\mathbb{R}^n \setminus V_j$ is greater than ϵ. Clearly V_j^ϵ is open and $\overline{V_j^\epsilon} \subset V_j$. We claim that $K \subset \bigcup_1^N V_j^\epsilon$ if ϵ is sufficiently small. Otherwise, for each $\epsilon > 0$ there exists $x_\epsilon \in K \setminus \bigcup_1^N V_j^\epsilon$. Since K is compact, the x_ϵ have an accumulation point $x \in K$ as $\epsilon \to 0$. But then $x \in K \setminus \bigcup_1^N V_j$, which is absurd. ∎

(0.19) Theorem.
Let $K \subset \mathbb{R}^n$ be compact and let $V_1, \ldots, V_N$ be bounded open sets such that $K \subset \bigcup_1^N V_j$. Then there exist functions $\zeta_1, \ldots, \zeta_n$ with $\zeta_j \in C_c^\infty(V_j)$ such that $\sum_1^N \zeta_j = 1$ on K.

Proof: Let $W_1, \ldots, W_N$ be as in Lemma (0.18). By Theorem (0.17), we can choose $\phi_j \in C_c^\infty(V_j)$ with $0 \le \phi_j \le 1$ and $\phi_j = 1$ on $\overline{W}_j$. Then $\Phi = \sum_1^N \phi_j \ge 1$ on K, so we can take $\zeta_j = \phi_j / \Phi$, with the understanding that $\zeta_j = 0$ wherever $\phi_j = 0$. ∎

The collection of functions $\{\zeta_j\}_1^N$ is called a **partition of unity on K subordinate to the covering $\{V_j\}_1^N$**.

D. The Fourier Transform

In this section we give a rapid introduction to the theory of the Fourier transform. For a more extensive discussion, see, e.g., Strichartz [47] or Folland [14], [17].

If $f \in L^1(\mathbb{R}^n)$, its **Fourier transform** $\widehat{f}$ is a bounded function on $\mathbb{R}^n$ defined by

$$\widehat{f}(\xi) = \int e^{-2\pi i x\cdot\xi} f(x)\, dx.$$

There is no universal agreement as to where to put the factors of 2π in the definition of $\widehat{f}$, and we apologize if this definition is not the one the reader is used to. It has the advantage of making the Fourier transform both an isometry on L^2 and an algebra homomorphism from L^1 (with convolution) to L^∞ (with pointwise multiplication).

Clearly $\widehat{f}(\xi)$ is well-defined for all ξ and $\|\widehat{f}\|_\infty \le \|f\|_1$. Moreover:

(0.20) Theorem.
If $f, g \in L^1$ then $(f * g)\widehat{\ } = \widehat{f}\widehat{g}$.

Proof: This is a simple application of Fubini's theorem:

$$\begin{aligned}(f*g)\widehat{\ }(\xi) &= \iint e^{-2\pi i x\cdot\xi} f(x-y)g(y)\, dy\, dx \\ &= \iint e^{-2\pi i (x-y)\cdot\xi} f(x-y) e^{-2\pi i y\cdot\xi} g(y)\, dx\, dy \\ &= \widehat{f}(\xi)\int e^{-2\pi i y\cdot\xi} g(y)\, dy = \widehat{f}(\xi)\widehat{g}(\xi).\end{aligned}$$ ∎

The Fourier transform interacts in a simple way with composition by translations and linear maps:

(0.21) Proposition.
Suppose $f \in L^1(\mathbb{R}^n)$.
a. *If $f_a(x) = f(x+a)$ then $(f_a)\widehat{\ }(\xi) = e^{2\pi i a\cdot\xi}\widehat{f}(\xi)$.*
b. *If T is an invertible linear transformation of $\mathbb{R}^n$, then $(f \circ T)\widehat{\ }(\xi) = |\det T|^{-1}\widehat{f}((T^{-1})^*\xi)$.*
c. *If T is a rotation of $\mathbb{R}^n$, then $(f\circ T)\widehat{\ } = \widehat{f}\circ T$.*

Proof: (a) and (b) are easily proved by making the substitutions $y = x + a$ and $y = Tx$ in the integrals defining $(f_a)\widehat{\ }(\xi)$ and $(f\circ T)\widehat{\ }(\xi)$, respectively. (c) follows from (b) since $T^* = T^{-1}$ and $|\det T| = 1$ when T is a rotation. ∎

The easiest way to develop the other basic properties of the Fourier transform is to consider its restriction to the Schwartz class $\mathcal{S}$. In what follows, if α is a multi-index, $x^\alpha f$ denotes the function whose value at x is $x^\alpha f(x)$.

(0.22) Proposition.
Suppose $f \in \mathcal{S}$.
a. *$\widehat{f} \in C^\infty$ and $\partial^\beta \widehat{f} = [(-2\pi i x)^\beta f]\widehat{\ }$.*
b. *$(\partial^\beta f)\widehat{\ } = (2\pi i \xi)^\beta \widehat{f}$.*

Proof: To prove (a), just differentiate under the integral sign. To prove (b), write out the integral for $(\partial^\beta f)\widehat{\ }(\xi)$ and integrate by parts; the boundary terms vanish since f and its derivatives vanish at infinity. ∎

(0.23) Proposition.
If $f \in \mathcal{S}$ then $\widehat{f} \in \mathcal{S}$.

Proof: By Proposition (0.22),

$$\partial^\beta \xi^\alpha \widehat{f} = (-1)^{|\beta|}(2\pi i)^{|\beta|-|\alpha|}[x^\beta \partial^\alpha f]\widehat{\ },$$

so $\partial^\beta\xi^\alpha\widehat{f}$ is bounded for all α, β. It then follows by the product rule for derivatives and induction on β that $\xi^\alpha\partial^\beta\widehat{f}$ is bounded for all α,β, that is, $\widehat{f} \in \mathcal{S}$. ∎

(0.24) The Riemann-Lebesgue Lemma.
If $f \in L^1$ then $\widehat{f}$ is continuous and tends to zero at infinity.

Proof: This is true by Proposition (0.23) if f lies in the dense subspace $\mathcal{S}$ of L^1. But if $\{f_j\} \subset \mathcal{S}$ and $f_j \to f$ in L^1, then $\widehat{f_j} \to \widehat{f}$ uniformly (because $\|\widehat{f_j} - \widehat{f}\|_\infty \le \|f_j - f\|_1$), and the result follows immediately. ∎

(0.25) Theorem.
Let $f(x) = e^{-\pi a|x|^2}$ where $a > 0$. Then

$$\widehat{f}(\xi) = a^{-n/2}e^{-\pi|\xi|^2/a}.$$

Proof: By making the change of variable $x \to a^{-1/2}x$ we may assume $a = 1$. Since the exponential function converts sums into products, by Fubini's theorem we have

$$\widehat{f}(\xi) = \int e^{-2\pi i x\cdot\xi - \pi|x|^2}\, dx = \prod_1^n \int e^{-2\pi i x_j\xi_j - \pi x_j^2}\, dx_j,$$

and it suffices to show that the jth factor in the product is $e^{-\pi\xi_j^2}$, i.e., to prove the theorem for $n = 1$. Now when $n = 1$ we have

$$\int e^{-2\pi i x\xi - \pi x^2}\,dx = e^{-\pi\xi^2}\int e^{-\pi(x+i\xi)^2}\,dx.$$

But $f(z) = e^{-\pi z^2}$ is an entire holomorphic function of $z \in \mathbb{C}$ which dies out rapidly as $|\operatorname{Re} z| \to \infty$ when $|\operatorname{Im} z|$ reamins bounded. Hence by Cauchy's theorem we can shift the contour of integration from $\operatorname{Im} z = 0$ to $\operatorname{Im} z = -\xi$, which together with (0.6) yields

$$e^{-\pi\xi^2}\int e^{-\pi(x+i\xi)^2}\,dx = e^{-\pi\xi^2}\int e^{-\pi x^2}\,dx = e^{-\pi\xi^2}. \qquad \blacksquare$$

(0.26) Theorem.
If $f, g \in \mathcal{S}$ then $\int f\widehat{g} = \int \widehat{f}g$.

Proof: By Fubini's theorem,

$$\int f\widehat{g} = \iint f(x)g(y)e^{-2\pi i x\cdot y}\,dy\,dx = \int \widehat{f}g. \qquad \blacksquare$$

For $f \in L^1$, define the function $f^\vee$ by

$$f^\vee(x) = \int e^{2\pi i x\cdot\xi} f(\xi)\,d\xi = \widehat{f}(-x).$$

(0.27) The Fourier Inversion Theorem.
If $f \in \mathcal{S}$, $(\widehat{f})^\vee = f$.

Proof: Given $\epsilon > 0$ and $x \in \mathbb{R}^n$, set $\phi(\xi) = e^{2\pi i x\cdot\xi - \pi\epsilon^2|\xi|^2}$. Then by Theorem (0.25),

$$\widehat{\phi}(y) = \int e^{-2\pi i(y-x)\cdot\xi}e^{-\pi\epsilon^2|\xi|^2}\,d\xi = \epsilon^{-n}e^{-\pi|x-y|^2/\epsilon^2}.$$

Thus,

$$\widehat{\phi}(y) = \epsilon^{-n}g(\epsilon^{-1}(x-y)) = g_\epsilon(x-y) \text{ where } g(x) = e^{-\pi|x|^2}.$$

By (0.26), then,

$$\int e^{-\pi\epsilon^2|\xi|^2}e^{2\pi i x\cdot\xi}\widehat{f}(\xi)\,d\xi = \int \widehat{f}\phi = \int f\widehat{\phi} = \int f(x)g_\epsilon(x-y)\,dy = f * g_\epsilon(x).$$

By (0.6) and (0.14), $f * g_\epsilon \to f$ uniformly as $\epsilon \to 0$ since functions in $\mathcal{S}$ are uniformly continuous. But clearly, for each x,

$$\int e^{-\pi\epsilon^2|\xi|^2}e^{2\pi i x\cdot\xi}\widehat{f}(\xi)\,d\xi \to \int e^{2\pi i x\cdot\xi}\widehat{f}(\xi)\,d\xi = (\widehat{f})^\vee(x). \qquad \blacksquare$$

(0.28) Corollary.
The Fourier transform is an isomorphism of $\mathcal{S}$ onto itself.

(0.29) The Plancherel Theorem.
The Fourier transform on $\mathcal{S}$ extends uniquely to a unitary isomorphism of L^2 onto itself.

Proof: Since $\mathcal{S}$ is dense in L^2 (Theorem (0.16)), by Corollary (0.28) it suffices to show that $\|\widehat{f}\|_2 = \|f\|_2$ for $f \in \mathcal{S}$. If $f \in \mathcal{S}$, set $g(x) = \overline{f(-x)}$. One easily checks that $\widehat{g} = \overline{\widehat{f}}$. Hence by Theorems (0.20) and (0.27),

$$\begin{aligned}\|f\|_2^2 &= \int f(x)\overline{f(x)}\,dx = f * g(0) = \int (f*g)\widehat{\ }(\xi)\,d\xi = \int \widehat{f}(\xi)\overline{\widehat{f}(\xi)}\,d\xi \\ &= \|\widehat{f}\|_2^2. \qquad \blacksquare\end{aligned}$$

The results (0.20)–(0.29) are the fundamental properties of the Fourier transform which we shall use repeatedly. We shall also sometimes need the Fourier transform as an operator on tempered distributions, to be discussed in the next section, and the following result.

(0.30) Proposition.
If $f \in L^1$ has compact support, then $\widehat{f}$ extends to an entire holomorphic function on $\mathbb{C}^n$. If $f \in C_c^\infty$, then $\widehat{f}(\xi)$ is rapidly decaying as $|\operatorname{Re}\xi| \to \infty$ when $|\operatorname{Im}\xi|$ remains bounded.

Proof: The integral $\widehat{f}(\xi) = \int e^{-2\pi i x\cdot\xi}f(x)\,dx$ converges for every $\xi \in \mathbb{C}^n$, and $e^{-2\pi i x\cdot\xi}$ is an entire function of $\xi \in \mathbb{C}^n$. Hence one can take complex derivatives of $\widehat{f}$ simply by differentiating under the integral. Moreover, if $f \in C_c^\infty$ and f is supported in $\{x : |x| \le K\}$, for any multi-index α we have

$$|(2\pi i\xi)^\alpha\widehat{f}(\xi)| = \left|\int e^{-2\pi i x\cdot\xi}\partial^\alpha f(x)\,dx\right| \le e^{K|\operatorname{Im}\xi|}\|\partial^\alpha f\|_1,$$

which yields the second assertion. $\blacksquare$

E. Distributions

We now outline the elements of the theory of distributions. The material sketched here is covered in more detail in Folland [14] and Rudin [41], and a more extensive treatment at an elementary level can be found in Strichartz [47]. See also Treves [49] and Hörmander [27, vol. I] for a deeper study of distributions.

Let Ω be an open set in $\mathbb{R}^n$. We begin by defining a notion of sequential convergence in $C_c^\infty(\Omega)$. Namely, we say that $\phi_j \to \phi$ in $C_c^\infty(\Omega)$ if the ϕ_j's are all supported in a common compact subset of Ω and $\partial^\alpha \phi_j \to \partial^\alpha \phi$ uniformly for every multi-index α. (This notion of convergence comes from a locally convex topology on $C_c^\infty(\Omega)$, whose precise description we shall not need. See Rudin [41] or Treves [49].)

If u is a linear functional on the space $C_c^\infty(\Omega)$, we denote the number obtained by applying u to $\phi \in C_c^\infty(\Omega)$ by $\langle u, \phi\rangle$ (or sometimes by $\langle \phi, u\rangle$: it is convenient to maintain this flexibility). A **distribution** on Ω is a linear functional u on $C_c^\infty(\Omega)$ that is continuous in the sense that if $\phi_j \to \phi$ in $C_c^\infty(\Omega)$ then $\langle u, \phi_j\rangle \to \langle u, \phi\rangle$. A bit of functional analysis (cf. Folland [14, Prop. (5.15)]) shows that this notion of continuity is equivalent to the following condition: for every compact set $K \subset \Omega$ there is a constant C_K and an integer N_K such that for all $\phi \in C_c^\infty(K)$,

$$(0.31) \qquad |\langle u, \phi\rangle| \le C_K \sum_{|\alpha| \le N_K} \|\partial^\alpha \phi\|_\infty .$$

The space of distributions on Ω is denoted by $\mathcal{D}'(\Omega)$, and we set $\mathcal{D}' = \mathcal{D}'(\mathbb{R}^n)$. We put the weak topology on $\mathcal{D}'(\Omega)$; that is, $u_j \to u$ in $\mathcal{D}'(\Omega)$ if and only if $\langle u_j, \phi\rangle \to \langle u, \phi\rangle$ for every $\phi \in C_c^\infty(\Omega)$.

Every locally integrable function u on Ω can be regarded as a distribution by the formula $\langle u, \phi\rangle = \int u\phi$, which accords with the notation introduced earlier. (The continuity follows from the Lebesgue dominated convergence theorem.) This correspondence is one-to-one if we regard two functions as the same if they are equal almost everywhere. Thus distributions can be regarded as "generalized functions." Indeed, we shall often pretend that distributions are functions and write $\langle u, \phi\rangle$ as $\int u(x)\phi(x)\,dx$; this is a useful fiction that makes certain operations involving distributions more transparent.

Every locally finite measure μ on Ω defines a distribution by the formula $\langle \mu, \phi\rangle = \int \phi\, d\mu$. In particular, if we take μ to be the point mass at 0, we obtain the granddaddy of all distributions, the **Dirac δ-function** $\delta \in \mathcal{D}'$ defined by $\langle \delta, \phi\rangle = \phi(0)$. Theorem (0.13) implies that if $u \in L^1$, $\int u = a$, and $u_\epsilon(x) = \epsilon^{-n} u(\epsilon^{-1}x)$, then $u_\epsilon \to a\delta$ in $\mathcal{D}'$ when $\epsilon \to 0$.

If $u, v \in \mathcal{D}'(\Omega)$, we say that $u = v$ on an open set $V \subset \Omega$ if $\langle u, \phi\rangle = \langle v, \phi\rangle$ for all $\phi \in C_c^\infty(V)$. The **support** of a distribution u is the complement of the largest open set on which $u = 0$. (To see that this is well-defined, one needs to know that if $\{V_\alpha\}_{\alpha\in A}$ is a collection of open sets and $u = 0$ on each V_α, then $u = 0$ on $\bigcup V_\alpha$. But if $\phi \in C_c^\infty(\bigcup V_\alpha)$, $\operatorname{supp}\phi$ is covered by finitely many V_α's. By means of a partition of unity on $\operatorname{supp}\phi$ subordinate to this covering, one can write $\phi = \sum_1^N \phi_j$ where each ϕ_j is supported in some V_α. It follows that $\langle u, \phi\rangle = \sum \langle u, \phi_j\rangle = 0$, as desired.)

The space of distributions on $\mathbb{R}^n$ whose support is a compact subset of the open set Ω is denoted by $\mathcal{E}'(\Omega)$, and we set $\mathcal{E}' = \mathcal{E}'(\mathbb{R}^n)$.

Suppose $u \in \mathcal{E}'$. Let Ω be a bounded open set such that $\operatorname{supp} u \subset \Omega$, and choose $\psi \in C_c^\infty(\Omega)$ with $\psi = 1$ on a neighborhood of $\operatorname{supp} u$ (by Theorem (0.17)). Then for any $\phi \in C_c^\infty$ we have

$$\langle u, \phi\rangle = \langle u, \psi\phi\rangle.$$

This has two consequences. First, u is of "finite order": indeed, by (0.31) with $K = \overline{\Omega}$,

$$|\langle u, \phi\rangle| \le C_{\overline{\Omega}} \sum_{|\alpha| \le N_{\overline{\Omega}}} \|\partial^\alpha(\psi\phi)\|_\infty .$$

Expanding $\partial^\alpha(\psi\phi)$ by the product rule, we see that

$$(0.32) \qquad |\langle u, \phi\rangle| \le C \sum_{|\alpha| \le N} \sup_{x\in\Omega} |\partial^\alpha \phi(x)|,$$

where $N = N_{\overline{\Omega}}$ and C depends only on $C_{\overline{\Omega}}$ and the constants $\|\partial^\beta \psi\|_\infty$, $|\beta| \le N$. Second, $\langle u, \psi\phi\rangle$ makes sense for all $\phi \in C^\infty$, compactly supported or not, so if we define $\langle u, \phi\rangle$ to be $\langle u, \psi\phi\rangle$ for all $\phi \in C^\infty$, we have an extension of u to a linear functional on C^∞. This extension is clearly independent of the choice of ψ, and it is unique subject to the condition that $\langle u, \phi\rangle = 0$ whenever $\operatorname{supp}\phi$ and $\operatorname{supp} u$ are disjoint. Thus distributions with compact support can be regarded as linear functionals on C^∞ that satisfy estimates of the form (0.32). Conversely, the restriction to C_c^∞ of any linear functional on C^∞ satisfying (0.32) is clearly a distribution supported in $\overline{\Omega}$.

The general philosophy for extending operations from functions to distributions is the following. Let T be a linear operator on $C_c^\infty(\Omega)$ that is continuous in the sense that if $\phi_j \to \phi$ in $C_c^\infty(\Omega)$ then $T\phi_j \to T\phi$ in $C_c^\infty(\Omega)$. Suppose there is another such operator T' such that $\int (T\phi)\psi = \int \phi(T'\psi)$ for all $\phi, \psi \in C_c^\infty(\Omega)$. (We call T' the **dual** or **transpose** of T.) We can then extend T to act on distributions by the formula

$$\langle Tu, \phi\rangle = \langle u, T'\phi\rangle.$$

The linear functional Tu on $C_c^\infty(\Omega)$ defined in this way is continuous on $C_c^\infty(\Omega)$ since T' is assumed continuous. The most important examples are the following; in all of them the verification of continuity is left as a simple exercise.

1. Let T be multiplication by the function $f \in C^\infty(\Omega)$. Then $T' = T$, so we can multiply any distribution u by $f \in C^\infty(\Omega)$ by the formula $\langle fu, \phi\rangle = \langle u, f\phi\rangle$.
2. Let $T = \partial^\alpha$. By integration by parts, $T' = (-1)^{|\alpha|}\partial^\alpha$. Hence we can differentiate any distribution as often as we please to obtain other distributions by the formula $\langle \partial^\alpha u, \phi\rangle = (-1)^{|\alpha|}\langle u, \partial^\alpha \phi\rangle$.
3. We can combine (1) and (2). Let $T = \sum_{|\alpha|\le k} a_\alpha \partial^\alpha$ be a differential operator of order k with C^∞ coefficients a_α. Integration by parts shows that the dual operator T' is given by $T'\phi = \sum_{|\alpha|\le k}(-1)^{|\alpha|}\partial^\alpha(a_\alpha\phi)$. For any distribution u, then, we define Tu by $\langle Tu, \phi\rangle = \langle u, T'\phi\rangle$.

Clearly, if $u \in C^k(\Omega)$, the distribution derivatives of u of order $\le k$ are just the pointwise derivatives. The converse is also true:

(0.33) Proposition.
If $u \in C(\Omega)$ and the distribution derivatives $\partial^\alpha u$ are in $C(\Omega)$ for $|\alpha| \le k$ then $u \in C^k(\Omega)$.

Proof: By induction it suffices to assume that $k = 1$. Since the conclusion is of a local nature, moreover, it suffices to assume that Ω is a cube, say $\Omega = \{x : \max|x_j - y_j| \le r\}$ for some $y \in \mathbb{R}^n$. For $x \in \Omega$, set

$$v(x) = \int_{y_1}^{x_1} \partial_1 u(t, x_2, \ldots, x_n)\, dt + u(y_1, x_2, \ldots, x_n).$$

It is easily checked that v and u agree as distributions on Ω, hence $v = u$ as functions on Ω. But $\partial_1 u$ is clearly a pointwise derivative of v. Likewise for $\partial_2 u, \ldots, \partial_n u$; thus $u \in C^1(\Omega)$. ∎

We now continue our list of operations on distributions. In all of the following, we take $\Omega = \mathbb{R}^n$.

4. Given $x \in \mathbb{R}^n$, let $T\phi = \phi_x$, where $\phi_x(y) = \phi(x+y)$. Then $T'\phi = \phi_{-x}$. Thus for any distribution u, we define its translate u_x by $\langle u_x, \phi\rangle = \langle u, \phi_{-x}\rangle$.
5. Let $T\phi = \widetilde{\phi}$, where $\widetilde{\phi}(x) = \phi(-x)$. Then $T' = T$, so for any distribution u we define its reflection in the origin $\widetilde{u}$ by $\langle \widetilde{u}, \phi\rangle = \langle u, \widetilde{\phi}\rangle$.

6. Given $\psi \in C_c^\infty$, define $T\phi = \phi * \psi$, which is in C_c^∞ by (0.14) and (0.15). It is easy to check that $T'\phi = \phi * \widetilde{\psi}$, where $\widetilde{\psi}$ is defined as in (5). Thus, if u is a distribution, we can define the distribution $u * \psi$ by $\langle u * \psi, \phi\rangle = \langle u, \phi * \widetilde{\psi}\rangle$. On the other hand, notice that $\phi * \psi(x) = \langle \phi, (\psi_x)^\sim\rangle$, so we can also define $u * \psi$ pointwise as a continuous function by $u * \psi(x) = \langle u, (\psi_x)^\sim\rangle$.
   In fact, these two definitions agree. To see this, let $\phi \in C_c^\infty$, let K be a compact set containing $\operatorname{supp}(\psi_x)^\sim$ for all $x \in \operatorname{supp}\phi$, and let N_K be as in (0.31). From the relation $\phi * \widetilde{\psi}(y) = \int \phi(x)(\psi_x)^\sim(y)\, dx$ it is is not hard to see that there is a sequence of Riemann sums $\sum \phi(x_j)(\psi_{x_j})^\sim \Delta x_j$ that converge uniformly to $\phi * \widetilde{\psi}$ together with their derivatives of order $\le N_K$. But then (0.31) implies that if $u * \psi$ is defined as a continuous function, we have
   $$\begin{aligned}\langle u * \psi, \phi\rangle &= \lim \sum u * \psi(x_j)\phi(x_j)\, \Delta x_j \\ &= \lim \sum \langle u, (\psi_{x_j})^\sim\rangle \phi(x_j)\, \Delta x_j = \langle u, \phi * \widetilde{\psi}\rangle,\end{aligned}$$
   which is the action of the distribution $u * \psi$ on ϕ.
   Moreover, by (2) and integration by parts, we see that the distribution $\partial^\alpha(u * \psi)$ is given by
   $$\langle \partial^\alpha(u*\psi), \phi\rangle = (-1)^{|\alpha|}\langle u, (\partial^\alpha\phi) * \widetilde{\psi}\rangle = \langle u, \phi * (\partial^\alpha\psi)^\sim\rangle = \langle u * \partial^\alpha\psi, \phi\rangle,$$
   so $\partial^\alpha(u*\psi) = u * \partial^\alpha\psi$ is a continuous function. Hence $u * \psi$ is actually a C^∞ function.
7. The same considerations apply when $u \in \mathcal{E}'$ and $\psi \in C^\infty$. That is, we can define $u * \psi$ either as a distribution by $\langle u * \psi, \phi\rangle = \langle u, \phi * \widetilde{\psi}\rangle$, or as a C^∞ function by $u * \psi(x) = \langle u, (\psi_x)^\sim\rangle$.
8. If $u \in \mathcal{E}'$ and $\psi \in C_c^\infty$, as in (0.15) we see that $u * \psi \in C_c^\infty$. Hence we can consider the operator $T\psi = u * \psi$ on C_c^∞, whose dual is clearly $T'\psi = \widetilde{u} * \psi$. It follows that if $u \in \mathcal{E}'$ and $v \in \mathcal{D}'$, $u * v$ can be defined as a distribution by the formula $\langle u * v, \psi\rangle = \langle v, \widetilde{u} * \psi\rangle$. We leave it as an exercise to verify that for any multi-index α we have $\partial^\alpha(u * v) = (\partial^\alpha u) * v = u * (\partial^\alpha v)$.

We shall also need to consider the class of "tempered distributions." We endow the Schwartz class $\mathcal{S}$ with the Fréchet space topology defined by the family of norms $\|\phi\|_{(\alpha,\beta)} = \|x^\alpha\partial^\beta\phi\|_\infty$. That is, $\phi_j \to \phi$ in $\mathcal{S}$ if and only if

$$\sup_x |x^\alpha\partial^\beta(\phi_j - \phi)(x)| \to 0 \text{ for all } \alpha, \beta.$$

A **tempered distribution** is a continuous linear functional on $\mathcal{S}$; the space of tempered distributions is denoted by $\mathcal{S}'$. Since C_c^∞ is a dense subspace of $\mathcal{S}$ in the topology of $\mathcal{S}$, and the topology on C_c^∞ is stronger than the topology on $\mathcal{S}$, the restriction of every tempered distribution to C_c^∞ is a distribution, and this restriction map is one-to-one. Hence, every tempered distribution "is" a distribution. On the other hand, Proposition (0.32) shows that every distribution with compact support is tempered. Roughly speaking, the tempered distributions are those which "grow at most polynomially at infinity." For example, every polynomial is a tempered distribution, but $u(x) = e^{|x|}$ is not. (Exercise: prove this.)

One can define operations on tempered distributions as above, simply by replacing C_c^∞ by $\mathcal{S}$. For example, if $u \in \mathcal{S}'$, then:

1. $\partial^\alpha u$ is a tempered distribution for all multi-indices α;
2. fu is a tempered distribution for all $f \in C^\infty$ such that $\partial^\alpha f$ grows at most polynomially at infinity for all α;
3. $u * \phi$ is a tempered distribution, and also a C^∞ function, for any $\phi \in \mathcal{S}$.

The importance of tempered distributions lies in the fact that they have Fourier transforms. Indeed, since the Fourier transform maps $\mathcal{S}$ continuously onto itself and is self-dual by Proposition (0.26), for any $u \in \mathcal{S}'$ we can define $\widehat{u} \in \mathcal{S}'$ by $\langle \widehat{u}, \phi \rangle = \langle u, \widehat{\phi} \rangle$ ($\phi \in \mathcal{S}$), which is consistent with the definition for functions. It is easy to see that Propositions (0.21) and (0.22) are still valid when $f \in \mathcal{S}'$, as is the Fourier inversion theorem (0.27), provided $f^\vee$ is defined by $\langle f^\vee, \phi \rangle = \langle f, \phi^\vee \rangle$. Also, if $u \in \mathcal{S}'$ and $\phi \in \mathcal{S}$, we have $(u * \phi)\widehat{\ } = \widehat{\phi}\widehat{u}$; the proof is left as an easy exercise.

For example, the Fourier transform of the Dirac δ-function is given by $\langle \widehat{\delta}, \phi \rangle = \langle \delta, \widehat{\phi} \rangle = \widehat{\phi}(0) = \int \phi(x)\,dx = \langle 1, \phi \rangle$, so $\widehat{\delta}$ is the constant function 1. It then follows from the inversion theorem that $\widehat{1} = \delta$, and from Proposition (0.22) that $(\partial^\alpha \delta)\widehat{\ } = (2\pi i)^{|\alpha|}\xi^\alpha$ and that $(x^\alpha)\widehat{\ } = (i/2\pi)^{|\alpha|}\partial^\alpha \delta$.

F. Compact Operators

Let $\mathcal{X}$ be a Banach space and let T be a bounded linear operator on $\mathcal{X}$. We denote the nullspace and range of T by $\mathcal{N}(T)$ and $\mathcal{R}(T)$. T is called **compact** if whenever $\{x_j\}$ is a bounded sequence in $\mathcal{X}$, the sequence $\{Tx_j\}$ has a convergent subsequence. Equivalently, T is compact if it maps bounded sets into sets with compact closure. T is said to be of **finite rank** if $\mathcal{R}(T)$ is finite-dimensional. Clearly every bounded operator of finite rank is compact.

(0.34) Theorem.
The set of compact operators on $\mathcal{X}$ is a closed two-sided ideal in the algebra of bounded operators on $\mathcal{X}$ with the norm topology.

Proof: Suppose T_1 and T_2 are compact, and $\{x_j\} \subset \mathcal{X}$ is bounded. We can choose a subsequence $\{y_j\}$ of $\{x_j\}$ such that $\{T_1 y_j\}$ converges, and then choose a subsequence $\{z_j\}$ of $\{y_j\}$ such that $\{T_2 z_j\}$ converges. It follows that $a_1 T_1 + a_2 T_2$ is compact for all $a_1, a_2 \in \mathbb{C}$. Also, it is clear that if T is compact and S is bounded, then TS and ST are compact. Thus the set of compact operators is a two-sided ideal.

Suppose $\{T_m\}$ is a sequence of compact operators converging to a limit T in the norm topology. Given a seqence $\{x_j\} \subset \mathcal{X}$ with $\|x_j\| \le C$ for all j, choose a subsequence $\{x_{1j}\}$ such that $\{T_1 x_{1j}\}$ converges. Proceeding inductively, for $m = 2, 3, 4, \ldots$, choose a subsequence $\{x_{mj}\}$ of $\{x_{(m-1)j}\}$ such that $\{T_m x_{mj}\}$ converges. Setting $y_j = x_{jj}$, one easily sees that $\{T_m y_j\}$ converges for all m. But then

$$\begin{aligned}\|Ty_j - Ty_k\| &\le \|(T - T_m)y_j\| + \|T_m(y_j - y_k)\| + \|(T_m - T)y_k\| \\ &\le 2C\|T - T_m\| + \|T_m y_j - T_m y_k\|.\end{aligned}$$

Given $\epsilon > 0$, we can choose m so large that $\|T - T_m\| \le \epsilon/4C$, and then with this choice of m we have $\|T_m y_j - T_m y_k\| < \epsilon/2$ when j and k are sufficiently large. Thus $\{Ty_j\}$ is convergent, so T is compact. ∎

(0.35) Corollary.
If T is a bounded operator on $\mathcal{X}$ and there is a sequence $\{T_m\}$ of operators of finite rank such that $\|T_m - T\| \to 0$, then T is compact.

In case $\mathcal{X}$ is a Hilbert space, this corollary has a converse.

(0.36) Theorem.
If T is a compact operator on a Hilbert space $\mathcal{X}$, then T is the norm limit of operators of finite rank.

Proof: Suppose $\epsilon > 0$, and let B be the unit ball in $\mathcal{X}$. Since $T(B)$ has compact closure, it is totally bounded: there is a finite set $y_1, \ldots, y_n$ of elements of $T(B)$ such that every $y \in T(B)$ satisfies $\|y - y_j\| < \epsilon$ for some j. Let P_ϵ be the orthogonal projection onto the space spanned by $y_1, \ldots, y_n$, and set $T_\epsilon = P_\epsilon T$. Then T_ϵ is of finite rank. Also, since $T_\epsilon x$ is the element closest to Tx in $\mathcal{R}(P_\epsilon)$, for $x \in B$ we have

$$\|Tx - T_\epsilon x\| \le \min_{1 \le j \le n} \|Tx - y_j\| < \epsilon.$$

In other words, $\|T - T_\epsilon\| < \epsilon$, so $T_\epsilon \to T$ as $\epsilon \to 0$. ∎

Remark: For many years it was an open question whether Theorem (0.36) were true for general Banach spaces. The answer is negative even for some separable, reflexive Banach spaces; see Enflo [12].

(0.37) Theorem.
The operator T on the Banach space $\mathcal{X}$ is compact if and only if the dual operator T^* on the dual space $\mathcal{X}^*$ is compact.

Proof: Let B and B^* be the unit balls in $\mathcal{X}$ and $\mathcal{X}^*$. Suppose T is compact, and let $\{f_j\}$ be a bounded sequence in $\mathcal{X}^*$. Multiplying the f_j's by a small constant, we may assume $\{f_j\} \subset B^*$. The functions $f_j : \mathcal{X} \to \mathbb{C}$ are equicontinuous and uniformly bounded on bounded sets, so by the Arzelà-Ascoli theorem there is a subsequence (still denoted by $\{f_j\}$) which converges uniformly on the compact set $\overline{T(B)}$. Thus $T^* f_j(x) = f_j(Tx)$ converges uniformly for $x \in B$, so $\{T^* f_j\}$ is Cauchy in the norm topology of $\mathcal{X}^*$. Hence T^* is compact.

Likewise, if T^* is compact then T^{**} is compact on $\mathcal{X}^{**}$. But $\mathcal{X}$ is isometrically embedded in $\mathcal{X}^{**}$, and T is the restriction of T^{**} to $\mathcal{X}$, so T is compact. ∎

We now present the main structure theorem for compact operators. This theorem was first proved by I. Fredholm (by different methods) for certain integral operators on L^2 spaces. In the abstract Hilbert space setting it is due to F. Riesz, and it was later extended to arbitrary Banach spaces by J. Schauder. For this reason it is sometimes called the *Riesz-Schauder theory.* We shall restrict attention to the Hilbert space case, which is all we shall need, and for which the proof is easier; see Rudin [41] for the general case.

(0.38) Fredholm's Theorem.
Let T be a compact operator on a Hilbert space $\mathcal{X}$ with inner product $\langle \cdot | \cdot \rangle$. For each $\lambda \in \mathbb{C}$, let

$$\mathcal{V}_\lambda = \{x \in \mathcal{X} : Tx = \lambda x\}, \qquad \mathcal{W}_\lambda = \{x \in \mathcal{X} : T^* x = \lambda x\}.$$

Then:

a. *The set of $\lambda \in \mathbb{C}$ for which $\mathcal{V}_\lambda \neq \{0\}$ is finite or countable, and in the latter case its only accumulation point is 0. Moreover, $\dim(\mathcal{V}_\lambda) < \infty$ for all $\lambda \neq 0$.*
b. *If $\lambda \neq 0$, $\dim(\mathcal{V}_\lambda) = \dim(\mathcal{W}_{\overline{\lambda}})$.*
c. *If $\lambda \neq 0$, $\mathcal{R}(\lambda I - T)$ is closed.*

Proof: (a) is equivalent to the following statement: For any $\epsilon > 0$, the linear span of the spaces $\mathcal{V}_\lambda$ with $|\lambda| \geq \epsilon$ is finite-dimensional. Suppose to the contrary that there exist $\epsilon > 0$ and an infinite sequence $\{x_j\} \subset \mathcal{X}$ of linearly independent elements such that $Tx_j = \lambda_j x_j$ with $|\lambda_j| \geq \epsilon$ for all j. Since $|\lambda_j| \leq \|T\|$, by passing to a subsequence we can assume that $\{\lambda_j\}$ is a Cauchy sequence. Let $\mathcal{X}_m$ be the linear span of $x_1, \ldots, x_m$. For each m, choose $y_m \in \mathcal{X}_m$ with $\|y_m\| = 1$ and $y_m \perp \mathcal{X}_{m-1}$. Then $y_m = \sum_1^m c_{mj} x_j$ for some scalars c_{mj}, so

$$\begin{aligned}\lambda_m^{-1} T y_m &= c_{mm} x_m + \sum_1^{m-1} c_{mj} \lambda_j \lambda_m^{-1} x_j = y_m + \sum_1^{m-1} c_{mj} (\lambda_j \lambda_m^{-1} - 1) x_j \\ &= y_m \pmod{\mathcal{X}_{m-1}}.\end{aligned}$$

If $n < m$, then,

$$\lambda_m^{-1} T y_m - \lambda_n^{-1} T y_n = y_m \pmod{\mathcal{X}_{m-1}}.$$

Therefore, since $y_m \perp \mathcal{X}_{m-1}$, the Pythagorean theorem yields

$$\|\lambda_m^{-1} T y_m - \lambda_n^{-1} T y_n\| \geq 1.$$

But then

$$1 \leq |\lambda_m^{-1}| \|T y_m - T y_n\| + |\lambda_m^{-1} - \lambda_n^{-1}| \|T y_n\|,$$

or

$$\|T y_n - T y_m\| \geq |\lambda_m| - |1 - \lambda_m \lambda_n^{-1}| \|T y_n\|.$$

As $m, n \to \infty$ the second term on the right tends to zero since $\|T y_n\| \leq \|T\|$ and $\lambda_m \lambda_n^{-1} \to 1$, and the first term is bounded below by ϵ. Thus $\{T y_m\}$ has no convergent subsequence, contradicting compactness.

Now consider (b). Given $\lambda \neq 0$, by Theorem (0.36) we can write $T = T_0 + T_1$ where T_0 has finite rank and $\|T_1\| < |\lambda|$. The operator $\lambda I - T_1 = \lambda(I - \lambda^{-1} T_1)$ is invertible (the inverse being given by the convergent geometric series $\sum_0^\infty \lambda^{-k-1} T_1^k$), and we have

$$(0.39) \quad (\lambda I - T_1)^{-1}(\lambda I - T) = (\lambda I - T_1)^{-1}(\lambda I - T_1 - T_0) = I - (\lambda I - T_1)^{-1} T_0.$$

Set $T_2 = (\lambda I - T_1)^{-1} T_0$. Then clearly $x \in \mathcal{V}_\lambda$ if and only if $x - T_2 x = 0$. On the other hand, taking the adjoint of both sides of (0.39), we have

$$(\overline{\lambda} I - T^*)(\overline{\lambda} I - T_1^*)^{-1} = I - T_2^*,$$

so $y = (\overline{\lambda} I - T_1^*)^{-1}x$ is in $\mathcal{W}_{\overline{\lambda}}$ if and only if $x - T_2^* x = 0$. We must therefore show that the equations $x - T_2 x = 0$ and $x - T_2^* x = 0$ have the same number of independent solutions.

Since T_0 has finite rank, so does T_2. Let $u_1, \ldots, u_N$ be an orthonormal basis for $\mathcal{R}(T_2)$. Then for any $x \in \mathcal{X}$ we have $T_2 x = \sum_1^N f_j(x) u_j$ where $\sum_1^N |f_j(x)|^2 = \|T_2 x\|^2$. It follows that $x \to f_j(x)$ is a bounded linear functional on $\mathcal{X}$, so there exist $v_1, \ldots, v_N$ such that

$$T_2 x = \sum_1^N \langle x \mid v_j \rangle u_j \qquad (x \in \mathcal{X}).$$

Set $\beta_{jk} = \langle u_j \mid v_k \rangle$, and given $x \in \mathcal{X}$, set $\alpha_j = \langle x \mid v_j \rangle$. If $x - T_2 x = 0$, then $x = \sum_1^n \alpha_j u_j$, and we see by taking the scalar prodict with v_k that

$$\alpha_k - \sum_j \beta_{jk} \alpha_j = 0, \qquad k = 1, \ldots, N. \tag{0.40}$$

Conversely, if $\alpha_1, \ldots, \alpha_n$ satisfy (0.40), then $x = \sum \alpha_j u_j$ satisfies $x - T_2 x = 0$. On the other hand, one easily verifies that $T_2^* x = \sum \langle x \mid u_j \rangle v_j$, so by the same reasoning, $x - T_2^* x = 0$ if and only if $x = \sum_1^N \alpha_j v_j$, where

$$\alpha_k - \sum_j \overline{\beta}_{kj} \alpha_j = 0, \qquad k = 1, \ldots, N. \tag{0.41}$$

But the matrices $(\delta_{jk} - \beta_{jk})$ and $(\delta_{jk} - \overline{\beta}_{kj})$ are adjoints of each other and so have the same rank. Thus (0.40) and (0.41) have the same number of independent solutions.

Finally, we prove (c). Suppose we have a sequence $\{y_j\} \subset \mathcal{R}(\lambda I - T)$ which converges to an element $y \in \mathcal{X}$. We can write $y_j = (\lambda I - T)x_j$ for some $x_j \in \mathcal{X}$; if we set $x_j = u_j + v_j$ where $u_j \in \mathcal{V}_\lambda$ and $v_j \perp \mathcal{V}_\lambda$, we have $y_j = (\lambda I - T)v_j$. We claim that $\{v_j\}$ is a bounded sequence. Otherwise, by passing to a subsequence we may assume $\|v_j\| \to \infty$. Set $w_j = v_j / \|v_j\|$; then by passing to another subsequence we may assume that $\{Tw_j\}$ converges to a limit z. Since the y_j's are bounded and $\|v_j\| \to \infty$,

$$\lambda w_j = T w_j + \frac{y_j}{\|v_j\|} \to z, \qquad (j \to \infty).$$

Thus $z \perp \mathcal{V}_\lambda$ and $\|z\| = |\lambda|$, but also

$$(\lambda I - T)z = \lim(\lambda I - T)\lambda w_j = \lim \frac{\lambda y_j}{\|v_j\|} = 0,$$

so $z \in \mathcal{V}_\lambda$. This is a contradiction since we assume $\lambda \neq 0$.

Now, since $\{v_j\}$ is a bounded sequence, by passing to a subsequence we may assume that $\{Tv_j\}$ converges to a limit x. But then

$$v_j = \lambda^{-1}(y_j + Tv_j) \to \lambda^{-1}(y + x),$$

so

$$y = \lim(\lambda I - T)v_j = (\lambda I - T)\lambda^{-1}(y + x).$$

Thus $y \in \mathcal{R}(\lambda I - T)$, and the proof is complete. ∎

(0.42) Corollary.
Suppose $\lambda \neq 0$. Then:
i. The equation $(\lambda I - T)x = y$ has a solution if and only if $y \perp \mathcal{W}_{\overline{\lambda}}$.
ii. $\lambda I - T$ is surjective if and only if it is injective.

Proof: (i) follows from part (c) of the theorem and the fact that $\overline{\mathcal{R}(S)} = \mathcal{N}(S^*)^\perp$ for any bounded operator S. (ii) then follows from (i) and part (b) of the theorem. ∎

In general it may happen that the spaces $\mathcal{V}_\lambda$ are all trivial. (It is easy to construct an example from a weighted shift operator on l^2.) However, if T is self-adjoint, there are lots of eigenvectors.

(0.43) Lemma.
If T is a compact self-adjoint operator on a Hilbert space $\mathcal{X}$, then either $\|T\|$ or $-\|T\|$ is an eigenvalue for T.

Proof: Clearly we may assume $T \neq 0$. Let $c = \|T\|$ (so $c > 0$), and consider the operator $A = c^2 I - T^2$. For all $x \in \mathcal{X}$ we have

$$\langle Ax \mid x \rangle = c^2 \|x\|^2 - \|Tx\|^2 \geq 0.$$

Choose a sequence $\{x_j\} \subset \mathcal{X}$ with $\|x_j\| = 1$ and $\|Tx_j\| \to c$. Then $\langle Ax_j \mid x_j \rangle \to 0$, so applying the Schwarz inequality to the nonnegative Hermitian form $(u, v) \to \langle Au \mid v \rangle$, we see that

$$\begin{aligned} \|Ax_j\|^2 = \langle Ax_j \mid Ax_j \rangle &\leq \langle Ax_j \mid x_j \rangle^{1/2} \langle A^2 x_j \mid Ax_j \rangle^{1/2} \\ &\leq \langle Ax_j \mid x_j \rangle^{1/2} \|A^2 x_j\|^{1/2} \|Ax_j\|^{1/2} \leq \|A\|^{3/2} \langle Ax_j \mid x_j \rangle^{1/2} \to 0, \end{aligned}$$

so $Ax_j \to 0$. By passing to a subsequence we may assume that $\{Tx_j\}$ converges to a limit y, which satisfies

$$\|y\| = \lim \|Tx_j\| = c > 0, \qquad Ay = \lim ATx_j = \lim TAx_j = 0.$$

In other words,

$$y \neq 0 \quad \text{and} \quad Ay = (cI + T)(cI - T)y = 0.$$

Thus either $Ty = cy$ or $cy - Ty = z \neq 0$ and $Tz = -cz$. ∎

(0.44) The Spectral Theorem.
If T is a compact self-adjoint operator on a Hilbert space $\mathcal{X}$, then $\mathcal{X}$ has an orthonormal basis consisting of eigenvectors for T.

Proof: It is a simple consequence of the self-adjointness of T that (i) eigenvectors for different eigenvalues are orthogonal to each other, and (ii) if $\mathcal{Y}$ is a subspace of $\mathcal{X}$ such that $T(\mathcal{Y}) \subset \mathcal{Y}$, then also $T(\mathcal{Y}^\perp) \subset \mathcal{Y}^\perp$. In particular, let $\mathcal{Y}$ be the closed linear span of all the eigenvectors of T. If we pick an orthonormal basis for each eigenspace of T and take their union, by (i) we obtain an orthonormal basis for $\mathcal{Y}$. By (ii), $T|\mathcal{Y}^\perp$ is a compact operator on $\mathcal{Y}^\perp$, and it has no eigenvectors since all the eigenvectors of T belong to $\mathcal{Y}$. But this is impossible by Lemma (0.43) unless $\mathcal{Y}^\perp = \{0\}$, so $\mathcal{Y} = \mathcal{X}$. ∎

We conclude by constructing a useful class of compact operators on $L^2(\mu)$, where μ is a σ-finite measure on a space S. To simplify the argument a bit, we shall make the (inessential) assumption that $L^2(\mu)$ is separable. If K is a measurable function on $S \times S$, we formally define the operator T_K on functions on S by

$$T_K f(x) = \int K(x, y) f(y)\, d\mu(y).$$

If $K \in L^2(\mu \times \mu)$, K is called a **Hilbert-Schmidt kernel**.

(0.45) Theorem.
Let K be a Hilbert-Schmidt kernel. Then T_K is a compact operator on $L^2(\mu)$, and $\|T_K\| \leq \|K\|_2$.

Proof: First we show that T_K is well-defined on $L^2(\mu)$ and bounded by $\|K\|_2$. By the Schwarz inequality,

$$|T_K f(x)| \leq \int |K(x,y)||f(y)|\, d\mu(y) \leq \left[\int |K(x,y)|^2\, d\mu(y)\right]^{1/2} \|f\|_2.$$

This shows that $T_K f$ is finite almost everywhere, and moreover

$$\|T_K f\|_2^2 = \int |T_K f(x)|^2\, d\mu(x) \leq \|f\|_2^2 \iint |K(x,y)|^2\, d\mu(y)\, d\mu(x) = \|K\|_2^2 \|f\|_2^2,$$

so $\|T_K\| \leq \|K\|_2$.

Now let $\{\phi_j\}_1^\infty$ be an orthonormal basis for $L^2(\mu)$. It is an easy consequence of Fubini's theorem that if $\psi_{ij}(x,y) = \phi_i(x)\phi_j(y)$, then $\{\psi_{ij}\}_{i,j=1}^\infty$ is an orthonormal basis for $L^2(\mu \times \mu)$. Hence we can write $K = \sum a_{ij}\psi_{ij}$. For $N = 1, 2, \ldots$, let

$$K_N(x,y) = \sum_{i+j \leq N} a_{ij}\psi_{ij}(x,y) = \sum_{i+j \leq N} a_{ij}\phi_i(x)\phi_j(y).$$

Then $\mathcal{R}(T_{K_N})$ lies in the span of $\phi_1, \ldots, \phi_N$, so T_{K_N} has finite rank. On the other hand,

$$\|K - K_N\|_2^2 = \sum_{i+j > N} |a_{ij}|^2 \to 0 \text{ as } N \to \infty,$$

so by the previous remarks,

$$\|T_K - T_{K_N}\| \leq \|K - K_N\|_2 \to 0 \text{ as } N \to \infty.$$

By Corollary (0.35), then, T_K is compact. ∎