Lecture 1

About Functional Analysis

The key objects of study in functional analysis are various kinds of topological vector spaces. The simplest of these are the *Banach* spaces.

Let E be a vector space over the field of real or complex numbers. (In this course we will use $\Bbbk$ to denote the ground field if a result applies equally to $\mathbb{R}$ or $\mathbb{C}$. There is a whole subject of *nonarchimedean functional analysis* over ground fields like the p-adic numbers, but we won't get involved with that. See the book by Peter Schneider, 2001.)

Definition 1.1. A *seminorm* on a vector space E is a function $p\colon X \to \mathbb{R}^+$ that satisfies $p(\lambda x) = |\lambda| p(x)$ for all $\lambda \in \Bbbk$, and $p(x+y) \leqslant p(x) + p(y)$ (the triangle inequality). It is a *norm* if in addition $p(x) = 0$ implies $x = 0$.

A norm, usually written $\|x\|$, defines a metric in the standard way.

Definition 1.2. A *Banach space* is a normed vector space which is *complete* (every Cauchy sequence converges) in the metric arising form the norm.

Standard examples of Banach spaces include function spaces such as $C(X)$, the continuous functions on a compact topological space X, and $L^1(Y, \mu)$, the integrable functions on a measure space (Y, μ). However one is often led to consider spaces of functions that do not admit any natural Banach space topology, such as $C^\infty(S^1)$, the smooth functions on the circle. Moreover, natural notions within the realm of Banach spaces themselves are not described directly by the norm (e.g. pointwise convergence a.e. in $L^1(Y, \mu)$). For these reasons it becomes useful to consider more general *topological vector spaces*.

Definition 1.3. A *topological vector space* is a vector space E equipped with a topology in which the vector space operations (addition and scalar multiplication) are continuous as maps $E \times E \to E$, $\Bbbk \times E \to E$.

Topological vector spaces form a category in which the morphisms are the continuous linear maps. Many authors (e.g. Rudin) restrict attention to *Hausdorff* topological vector spaces, but it is occasionally useful to consider non-Hausdorff examples.

Exercise 1.4. A TVS is Hausdorff if and only if the origin $\{0\}$ is a closed subset.

Exercise 1.5. Show that if E is a topological vector space and F a subspace, then the quotient topology on E/F makes it a TVS. Show that if F is closed in E then E/F is Hausdorff.

As in the case of normed spaces, one of the most important properties of a TVS is *completeness*. Completeness is defined as "Cauchy implies convergent", as for metric spaces. In general however the notion of Cauchy *sequences* is not sufficient and it is necessary instead to consider Cauchy *nets*. Recall that a *net* in a space X is simply a function from a directed set I to X (and is usually denoted $\{x_i\}_{i \in I}$, like a sequence); a net $\{x_i\}$ *converges* to $x \in X$ if, for every open set U containing x, there is an $i_0 \in I$ such that $x_i \in U$ for all $i \geqslant i_0$. (One expresses this last conclusion by saying that the net x_i is "eventually in U".) In this course we will state definitions for nets if that is the appropriate generality; but we shall sometimes only give the proofs in the sequence case (and the reader who is not familiar with nets probably won't lose much by replacing 'net' by 'sequence' all through).

Definition 1.6. A *Cauchy net* in a TVS E is a net $\{x_i\}$ such that, for every open set U containing the origin, there is i_0 such that $x_i - x_j \in U$ whenever $i, j \geqslant i_0$. A TVS is *complete* if every Cauchy net converges.

Proposition 1.7. *A closed subspace of a complete TVS is complete. In a* Hausdorff *TVS, a complete subspace is closed.*

Proof. For example let's prove the second statement. Let $F \leqslant E$, E a Hausdorff TVS, and suppose that F is complete in its own right. If F is not closed, there is $x \in E \setminus F$ such that *every* open neighborhood U of x contains a point x_U of F. The family of such neighborhoods forms a directed set under inclusion, so $\{x_U\}$ can be thought of as a net, which obviously converges to x. Then x_U is a Cauchy net in F, so since F is complete it converges *in* F to some limit. This limit must be x because E is Hausdorff, and this is a contradiction. $\square$

Many examples of topological vector spaces arise from the following construction. Let $\{p_\alpha\}_{\alpha \in \mathcal{A}}$ be a family of seminorms on a vector space E. We can define a topology on E be calling $U \subseteq E$ open if for every $x_0 \in U$ there exist $\alpha_1, \ldots, \alpha_n \in \mathcal{A}$ and real $\epsilon_1, \ldots, \epsilon_n > 0$ such that

$$\{x \in E : p_{\alpha_i}(x - x_0) < \epsilon_i,\ (i = 1, \ldots, n)\} \subseteq U.$$

This topology makes E into a TVS; it is called the *locally convex* topology defined by the seminorms $\{p_\alpha\}$.

Exercise 1.8. Let x_j be a net in a LCTVS whose topology is defined by a family of seminorms $\{p_\alpha\}$. Then $x_j \to x$ if and only if, for each α, $p_\alpha(x_j - x) \to 0$. What condition on the family of seminorms will ensure that E is Hausdorff?

To explain the "locally convex" terminology, suppose that p is a seminorm on the vector space E. Then the set $A = \{x \in E : p(x) \leqslant 1\}$ has the following properties:

- A is *convex*: for any finite subset $\{a_1, \ldots, a_n\}$ of A, and any positive real numbers $\lambda_1, \ldots, \lambda_n$ with $\sum \lambda_i = 1$, the *convex combination* $\sum \lambda_i a_i$ belongs to A.

- A is *balanced*: if $a \in A$ and $\lambda \in \Bbbk$ with $|\lambda| \leqslant 1$, then $\lambda a \in A$ also.

- A is *absorbing*: for any $x \in E$ there is $\lambda > 0$ such that $\lambda x \in A$.

Regarding this terminology, note

Lemma 1.9. *In a TVS every 0-neighborhood is absorbing. Every 0-neighborhood contains a balanced 0-neighborhood; every convex 0-neighborhood contains a convex, balanced 0-neighborhood.*

Proof. These facts follow from the continuity of (scalar) multiplication. For instance, let U be any 0-neighborhood. By continuity, there exist $\epsilon > 0$ and an 0-neighborhood W such that $\lambda W \subseteq U$ for all $|\lambda| < \epsilon$. The union $\bigcup_{|\lambda|<\epsilon} \lambda W$ is then a balanced 0-neighborhood contained in U. ☐

We have seen that each seminorm gives a convex, balanced, absorbing 0-neighborhood. One can also reverse this process:

Lemma 1.10. *Let A be a convex, balanced, absorbing subset of a vector space E. Define $\mu_A \colon E \to \mathbb{R}^+$ (the* Minkowski functional *of A by*

$$\mu_A(x) = \inf\{\lambda \in \mathbb{R}^+ : x \in \lambda A\}.$$

Then μ_A is a seminorm on E.

Proof. If $x \in \lambda A$, $y \in \mu A$ then convexity gives $(x + y) \in (\lambda + \mu)A$, so $\mu_A(x + y) \leqslant \lambda + \mu$. Taking the infimum over allowable λ, μ gives the triangle inequality $\mu_A(x + y) \leqslant \mu_A(x) + \mu_A(y)$. The proof of positive-linearity is similar using the fact that A is balanced. ☐

Thus

Proposition 1.11. *A TVS is locally convex iff there is a basis for the neighborhoods of 0 comprised of convex sets.* ☐

It is natural (and traditional!) to begin a functional analysis course by understanding the *finite-dimensional* examples. It turns out that there aren't very many.

Proposition 1.12. *Let E be a* finite-dimensional *Hausdorff TVS. For any basis $\{v_1, \ldots, v_n\}$ of E the map $\phi\colon \mathbb{k}^n \to E$ defined by $\phi(\lambda_1, \ldots, \lambda_n) = \sum \lambda_i v_i$ is an isomorphism of topological vector spaces.*

Proof. Clearly ϕ is bijective and continuous; we must prove that its inverse is continuous. Equip $\mathbb{k}^n$ with the standard Euclidean norm, and let B be the open unit ball in that norm, S the unit sphere; it suffices to show that $\phi(B)$ contains a neighborhood of 0 (for the given topology on E). Consider the compact set $\phi(S) \subseteq E$; because E is Hausdorff, for each $x \in \phi(S)$ there exist a neighborhood U_x of x and a neighborhood W_x of 0 that do not meet. By compactness, there are finitely many points $x_1, \ldots, x_m$ such that $U_{x_1}, \ldots, U_{x_m}$ cover S.

Let W be a *balanced* 0-neighborhood contained in $W_{x_1} \cap \cdots \cap W_{x_m}$. I claim that $W \subseteq \phi(B)$. Suppose not. Then there exists $v \in E$ with $\|v\| \geqslant 1$ and $\phi(v) \in W$. Then $\phi(v/\|v\|)$ belongs both to W and to $\phi(S)$, which is impossible. This contradiction proves the result. ☐

Proposition 1.13. *A linear functional $\phi\colon E \to \mathbb{k}$ on a topological vector space E is continuous if and only if its kernel is closed.*

Proof. By definition the kernel $K = \ker(\phi) = \phi^{-1}\{0\}$ is the inverse image of a closed set. So, if ϕ is continuous, then K is closed. Conversely suppose that K is closed. Then factor ϕ (algebraically) as

$$\begin{tikzcd} E \arrow[r] & E/K \arrow[r] & \mathbb{k} \end{tikzcd}$$

The first arrow is the canonical quotient map, so is continuous; the second is an isomorphism of finite-dimensional Hausdorff TVS, so it is continuous by proposition 1.12. ☐

Lemma 1.14. *A finite dimensional subspace of a Hausdorff TVS is closed.*

Proof. A finite-dimensional subspace is linearly homeomorphic to $\mathbb{k}^n$; hence it is complete, and therefore closed. ☐

Proposition 1.15. *A Hausdorff TVS is finite-dimensional iff it is locally compact.*

Proof. IF E is finite dimensional it is homeomorphic to $\Bbbk^n$ (by 1.12), hence locally compact. Suppose now that E is locally compact and let W be a 0-neighborhood with compact closure. By compactness, any 0-neighborhood U has $2^n U \supseteq W$ for some n. Thus, $\{2^{-n}W\}$ forms a basis for the 0-neighborhoods in E.

By compactness there is a finite set F such that

$$W \subseteq \overline{W} \subseteq F + \tfrac{1}{2}W \subseteq \langle F \rangle + \tfrac{1}{2}W,$$

where $\langle F \rangle$ denotes the finite-dimensional subspace spanned by F. From this we find

$$W \subseteq \langle F \rangle + \tfrac{1}{2}\big(\langle F \rangle + \tfrac{1}{2}W\big) \subseteq \langle F \rangle + \tfrac{1}{4}W,$$

and iterating the argument gives $W \subseteq \langle F \rangle + 2^{-n}W$ and hence W is contained in the closure of $\langle F \rangle$, which equals $\langle F \rangle$ by lemma 1.14. Since W is absorbing it spans E, so $E = \langle F \rangle$ as required. $\square$

Lecture 2

Examples of Topological Vector Spaces

Let E be a locally convex TVS whose topology is defined by a *countable* family of seminorms $p_1, p_2, \ldots$. The expression

$$d(x,y) = \sum_{n=1}^{\infty} 2^{-n} \left(\frac{p_n(x-y)}{1+p_n(x-y)} \right)$$

is then a translation invariant metric that defines the topology of E. (They key thing to prove is the triangle inequality, which follows from the elementary algebraic fact that

$$a(x_1+x_2) \leqslant a(x_1) + a(x_2) \quad \text{where} \quad a(x) = \frac{x}{1+x}.)$$

Definition 2.1. A complete locally convex TVS whose topology is defined by countably many seminorms is called a *Fréchet space*.

"Completeness" can be taken either in the sense of the metric d, or in the sense of Definition 1.2; they are equivalent here.

An important example of a Fréchet space is the space $C(X)$ of all continuous functions from a σ-compact Hausdorff space X (e.g. the real line) to $\mathbb{k}$. Let $K_1, K_2, \ldots$ be a sequence of compact subsets such that $K_n \subseteq K_{n+1}^\circ$ and $\bigcup K_n = X$, and define $p_n(f) = \sup\{|f(x)| : x \in K_n\}$. Then the p_n are a sequence of seminorms that define a complete vector topology on $C(X)$ — the *topology of uniform convergence on compact sets*.

In a similar way the space $\mathcal{E}(M) = C^\infty(M)$ of *smooth* functions on a σ-compact manifold (e.g. $\mathbb{R}^N$) is a Fréchet space. One now considers (say on $\mathbb{R}^N$ for notational simplicity) the seminorms

$$p_{n,k}(f) = \sup\{|D^\alpha f(x)| : x \in K_n, |\alpha| \leqslant k\}$$

where $\alpha = (\alpha_1, \ldots, \alpha_N)$ is a multi-index, $|\alpha| = \alpha_1 + \cdots + \alpha_N$, and D^α denotes the partial derivative operator

$$D^\alpha = \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_N}}{\partial x_1^{\alpha_N}}.$$

Exercise 2.2. Show that the topologies defined above are independent of the choice of compact exhaustion K_n.

It is sometimes important to know that the topology of a Fréchet space can always be defined by an *increasing* sequence of seminorms ($p_1 \leqslant p_2 \leqslant \cdots$). To see this, just replace the original seminorms by p_1, $p_1 + p_2$, $p_1 + p_2 + p_3$ and so on.

Proposition 2.3. *Let E be a Fréchet space and F a closed subspace. Then F and E/F are Fréchet spaces (in their natural topologies).*

Proof. This is clear for F (the same seminorms define the topology, and closed subspaces of complete spaces are complete). What about E/F? Let $p_1, p_2, \ldots$ be an increasing sequence of seminorms defining the topology of E and define $q_1, q_2, \ldots$ on E/F by

$$q_i(y) = \inf\{p_i(x) \colon \pi(x) = y\},$$

where $\pi \colon E \to E/F$ denotes the canonical projection. Routine arguments show that the q_i are seminorms defining the topology of E/F.

Now to see that E/F is complete, suppose that $\{y_n\}$ is a Cauchy sequence in E/F. By extracting a subsequence we may assume without loss of generality that $\{y_n\}$ is "fast Cauchy" in the sense that $q_n(y_n - y_{n+1}) < 2^{-n}$. We are going to define by induction a sequence $x_n \in E$ with $\pi(x_n) = y_n$ and $p_n(x_n - x_{n+1}) < 2^{-n}$. Start by choosing any x_1 such that $\pi(x_1) = y_1$. Now, assuming $x_1, \ldots, x_n$ have been defined, notice that

$$\inf\{p_n(x_n - x) : \pi(x) = y_{n+1}\} = q_n(y_n - y_{n+1}) < 2^{-n}.$$

Thus it is possible to choose a value of x in this expression for which $p_n(x_n - x) < 2^{-n}$. Call this value x_{n+1}.

If $i \leqslant n < m$ we now have

$$p_i(x_n - x_m) \leqslant p_n(x_n - x_m) \leqslant \sum_{k=n}^{m-1} p_n(x_k - x_{k+1}) < 2 \cdot 2^{-n},$$

so $\{x_n\}$ is a Cauchy sequence in E, hence convergent (say to x); now $\{y_n\}$ converges to $y = \pi(x)$ as required. □

Exercise 2.4. Let (X, μ) be the unit interval with Lebesgue measure, and let E be the vector space of all equivalence classes (modulo equality almost everywhere) of measurable functions $X \to \mathbb{k}$. Recall that a sequence $\{f_n\}$ in E *converges in measure* to $f \in E$ if, for every $\epsilon > 0$, the measure $\mu\{x \in X : |f_n(x) - f(x)| > \epsilon\} \to 0$ as $n \to \infty$. Show that E can be given a vector topology, defined by a complete translation invariant metric, for which convergence is convergence in measure, but that this topology is *not* locally convex.

The construction of a locally convex topology by seminorms is an example of a general process called the construction of *weak* or *induced* topologies. Recall that if $\mathcal{T}_1$ and $\mathcal{T}_2$ are two topologies on the same space, one says that $\mathcal{T}_1$ is *weaker* (or *coarser*)than $\mathcal{T}_2$ if every $\mathcal{T}_1$-open set is also $\mathcal{T}_2$-open. Equivalent terminology is that $\mathcal{T}_2$ is *stronger* or *finer* than $\mathcal{T}_1$.

Let X be a set and let there be given a family of maps $f_\lambda \colon X \to X_\lambda$, $\lambda \in \Lambda$, where the X_λ are topological spaces. The *weak topology* on X induced by these maps is the weakest topology that makes the maps f_λ continuous. In other words, it is the topology defined concretely as follows: a subset U of X is open iff, for every $x \in U$, there exist finitely many $\lambda_1, \ldots, \lambda_n \in \Lambda$ and open subsets $U_j \subseteq X_{\lambda_j}$, $j = 1, \ldots, n$, such that

$$x \in \bigcap_{j=1}^{n} f_{\lambda_j}^{-1}(U_j) \subseteq U.$$

Exercise 2.5. Show that a net $\{x_j\}$ converges to x in the weak topology on X if and only if $f_\lambda(x_j)$ converges to $f_\lambda(x)$ for all $\lambda \in \Lambda$.

Example 2.6. If $X \subseteq Y$, Y a topological space, the *subspace topology* on X is the weak topology induced by the inclusion map.

Example 2.7. The *product topology* (or *Tychonoff topology*) on the Cartesian product $X = \prod_\lambda X_\lambda$ is the weak topology induced by the projection maps $X \to X_\lambda$. We will take for granted the *Tychonoff theorem*: any product of compact spaces is compact. (This theorem is equivalent to the axiom of choice.)

Example 2.8. A locally convex topology on a vector space V generated by seminorms p_α is a weak topology induced by the maps $x \mapsto p_\alpha(x - v)$, for all α and all $v \in V$.

Let E and F be vector spaces. A *nondegenerate pairing* between E and F is a bilinear map $\langle\cdot,\cdot\rangle\colon E\times F\to \Bbbk$ such that

$$[\langle x,y\rangle = 0\forall y]\Rightarrow x=0,\quad [\langle x,y\rangle = 0\forall x]\Rightarrow y=0.$$

(As we will see when we discuss the Hahn-Banach theorem, one very important example of this phenomenon occurs when E is a LCTVS and $F=E^*$ is the space of continuous linear maps $E\to\Bbbk$.) In these circumstances the *$\sigma(E,F)$-topology* (or if the pairing is understood simply the *weak topology*) on E is the weak topology induced by the maps $E\to\Bbbk$ given by pairing with elements of F. Similarly for the *$\sigma(F,E)$-topology* on F.

Exercise 2.9. Show that these topologies make E,F into Hausdorff, locally convex topological vector spaces.

Dually to the construction of weak (induced) topologies, consider a set X and a family $g_\lambda\colon X_\lambda\to X$ of maps from topological spaces to X. The *strong topology coinduced* by these maps is the strongest topology that makes all the maps g_λ continuous. It can be defined explicitly by saying that $U\subseteq X$ is open if and only if $g_\lambda^{-1}(U)$ is open in X_λ for all λ.

Example 2.10. The quotient topology is an example of a strong topology.

Example 2.11. (Inductive limits) Let X be a set and suppose that there is an increasing sequence of subspaces $X_1\subseteq X_2\subseteq\cdots\subseteq X$, whose union is X. Suppose that each X_n is equipped with a topology in such a way that each inclusion $X_n\to X_{n+1}$ is a homeomorphism onto its image. In these circumstances the *inductive limit topology* on X is the topology coinduced by the inclusions $X_n\to X$.

Proposition 2.12. *Let $X=\bigcup X_n$ be equipped with an inductive limit topology, with each X_n Hausdorff. Then X is Hausdorff and, in order that a net x_j in X converge to $x\in X$ it is necessary and sufficient that, for some N and j_0, $x_j\in X_N$ for all $j\geqslant j_0$ and $x_j\to x$ in X_N.*

Proof. We prove the second statement (the proof of the first is similar), and we assume for simplicity that our net is in fact a *sequence*.

Without loss of generality suppose that the limit x belongs to X_1. Assume for a contradiction that the sequence $\{x_j\}$ is not contained in any single X_N; then again, wlog we may assume that $x_j\in X_j\setminus X_{j-1}$. We will define by induction open subsets U_j of X_j such that $x\in U_j$, $U_{j-1}\subseteq U_j$, and none of $x_2,\ldots,x_j$ belong to U_j. The induction starts by choosing any neighborhood U_1 of x in X_1. Now assuming that U_j has been constructed, because X_{j+1} is Hausdorff each point $x \in U_j$ has a neighborhood V_x in X_{j+1} that does not meet $x_2, \ldots, x_{j+1}$; take U_{j+1} to be the union of all these V_x. Finally let $U = \bigcup U_j$; then U is open in X by construction, it contains x and does not contain any x_j; this is impossible since $x_j \to x$ by hypothesis. $\square$

Example 2.13. Consider the space $C_c(X)$ of continuous, compactly supported functions on a locally compact, σ-compact Hausdorff space X. Write X as an increasing union of compact subsets K_n, with $K_n \subseteq K_{n+1}^\circ$. Then we can regard $C_c(X)$ as the inductive limit of the spaces $C(K_n)$ of continuous functions on K_n. The resulting topology on $C_c(X)$ is a complete, Hausdorff, locally convex vector topology; it is not a Fréchet topology (it cannot be defined by countably many seminorms); in particular, it is not the same as the topology induced on $C_c(X)$ by the supremum norm.

Exercise 2.14. Prove the statements made above. Also, show that the topology on $C_c(X)$ can be defined by the *uncountable* family of seminorms

$$p_g(f) = \sup\{|f(x)g(x)| : x \in X\}$$

as g ranges over all continuous functions on X (not necessarily compactly supported).

Example 2.15. Let M be a σ-compact smooth manifold (e.g. $\mathbb{R}^n$). The *test function space* $\mathcal{D}(M) = C_c^\infty(M)$ of compactly supported smooth functions on M can be topologized in a similar way as an inductive limit of the Fréchet spaces $C^\infty(K_n)$ of smooth functions supported within K_n.

Lecture 3

Precompactness and Equicontinuity

For this lecture we will use a few fundamental ideas about nets and convergence. Recall that a *net* in X is a map from a directed set to X. (A *directed set* is a partially ordered set in which any two elements have an upper bound.) Let S be a subset of X. A net $\{x_i\}$ is *eventually in* S if there is i_0 such that $x_i \in S$ for all $i \geqslant i_0$; it is *frequently in* S if it is not eventually in $X \setminus S$. If X is a topological space, then net $\{x_i\}$ *converges* to $x \in X$ if it is eventually in every neighborhood of x.

Definition 3.1. Let D and D' be directed sets. A function $h\colon D' \to D$ is called *final* if, for any $i_0 \in D$, there is $j_0 \in D'$ such that $j \geqslant j_0$ implies $h(j) \geqslant i_0$. Given a net $D \to X$, the result of composing it with a final function h is called a *subnet* or *refinement* of the original one.

Every subnet of a convergent net is convergent (with the same limit); the definition of “final function” is cooked up to make this true.

Definition 3.2. A net $\{x_i\}$ in X is called *universal* if, given any subset S of X whatsoever, either $\{x_i\}$ is eventually in S or else eventually in $X \setminus S$.

The following is a version of the axiom of choice.

Lemma 3.3. *Every net has a universal subnet.* $\square$

A proof is indicated at the end of this lecture, but we probably won’t have time to cover this in class.

The following exercises use these ideas to give an easy proof of the Tychonoff theorem.

Exercise 3.4. Show that the following properties of a topological space X are equivalent: (i) X is compact; (ii) (the finite intersection property) if $\mathcal{F}$ is a family of closed subsets of X, and the intersection of any finite number of members of $\mathcal{F}$ is nonempty, then in fact the intersection of all the members of $\mathcal{F}$ is nonempty; (iii) every universal net in X converges; (iv) every net in X has a convergent subnet.

Exercise 3.5. Prove Tychonoff’s theorem: any product of compact spaces is compact. (Use the characterization (iii) of compactness in the previous exercise.)

Definition 3.6. Let E be a topological vector space. A subset A of E is *precompact* if for every 0-neighborhood $U \subseteq E$ there is a finite subset $F \subseteq X$ with $A \subseteq U + F$.

Obviously, if A is compact, or even if it is contained in a compact set, then it is precompact. We will see that the converse is true in a complete space.

Lemma 3.7. *Let A be a subset of a TVS E. Then the following are equivalent:*

(a) A is precompact.

(b) The closure $\overline{A}$ is precompact.

(c) Every universal net in A is Cauchy.

(d) Every net in A has a Cauchy subnet.

Proof. Suppose A is precompact and let a 0-neighborhood U be given. Choose a balanced 0-neighborhood U' such that $U' + U' \subseteq U$. There is a finite set F such that $F + U'$ covers A. Every point of $\overline{A}$ belongs to $a + U'$ for some $a \in A$. Consequently, $F + U$ covers $\overline{A}$. This proves that (a) and (b) are equivalent.

We prove that (d) implies (a). Suppose that A is *not* precompact. Then there is some neighborhood U of the origin such that no finite union of U-balls covers A. Thus, inductively, we may pick a sequence x_n in A with $x_n \notin \bigcup_{i=1}^{n-1} x_i + U$. If U' is a balanced 0-neighborhood such that $U' + U' \subseteq U$ it follows that the $\{x_n + U'\}$ are all disjoint. Thus $\{x_n\}$ can have no Cauchy subsequence (or subnet).

Lemma 3.3 shows that (c) implies (d).

Finally we show that (a) implies (c). Suppose that x_i is a universal net in the precompact space A and let U be a 0-neighborhood; let U' be a 0-neighborhood with $U' + U' \subseteq U$. There is a finite set F such that $F + U' \supseteq A$, so the net $\{x_i\}$ must be frequently in some $a + U'$, $a \in F$. By universality it must be *eventually* in $a + U'$, so $x_i - x_{i'}$ must be eventually in U. Since this is true for every U, the net is Cauchy. □

We deduce

Proposition 3.8. *Let E be a complete Hausdorff TVS, $A \subseteq E$. Then A is compact if and only if it is precompact and closed.*

Proof. A is compact iff every universal net in A converges in A. □

There is a “Tychonoff theorem” for precompactness just as there is for compactness.

Proposition 3.9. *Let E be a vector space, equipped with a family of linear maps $\phi_\alpha \colon E \to E_\alpha$, where the E_α are topological vector spaces. Suppose that E is equipped with the weak topology induced by these maps. Then a subset $X \subseteq E$ is precompact if and only if $\phi_\alpha(X)$ is precompact in E_α for every α.*

Proof. Consider a universal net $\{x_i\}$ in A. For each α the image $\phi_\alpha(x_i)$ is a universal net in $\phi_\alpha(X)$, so it is Cauchy in E_α. But now $\{x_i\}$ is Cauchy in E by definition of the weak topology. ☐

As we have already remarked, important examples of topological vector spaces are various spaces of functions. To formalize this, let E be a TVS (it could even be the ground field $\Bbbk$) and let X be some topological space. We’ll use the notation $C(X; E)$ for the collection of all continuous maps $X \to E$. Let F be a subspace of $C(X; E)$ and let $\mathfrak{S}$ be any family of subsets of X having the property that $f(S)$ is a bounded subset of F for each $S \in \mathfrak{S}$. A vector topology on F can then be defined by taking as 0-neighborhoods all the sets

$$\{f \in C(X; E) : f(S) \subseteq U\} \quad \text{for } S \in \mathfrak{S}, U \text{ a 0-neighborhood in } E.$$

This is called the topology of *uniform convergence on* $\mathfrak{S}$. Particular examples that we have already met include

- $\mathfrak{S} = \{X\}$ (*uniform* convergence);
- $\mathfrak{S} = \{\{x\} : x \in X\}$ (*pointwise* convergence);
- $\mathfrak{S} = \{K \subseteq X : K \text{ compact}\}$ (*compact* convergence)

Denote by $C_{\mathfrak{S}}(X; E)$ the space $X(X; E)$ with the topology defined above.

Note that the topology of pointwise convergence is simply the weak topology induced by the evaluation maps $f \mapsto f(x)$, $x \in X$.

Proposition 3.10. *In the situation above suppose that E is complete and that every point of X has a neighborhood belonging to $\mathfrak{S}$; then $C_{\mathfrak{S}}(X; E)$ is complete.*

Proof. Let f_i be a Cauchy net in $C_{\mathfrak{S}}(X; E)$. In particular, $f_i(x)$ must be a Cauchy net in E for each $x \in X$; so it converges. Call its limit $f(x)$. This defines a function $f \colon X \to E$.

Is f continuous? Let us check continuity at $x_0 \in X$. Let U be a 0-neighborhood in E and let U' be a balanced 0-neighborhood such that $\overline{U'} + \overline{U'} + \overline{U'} \subseteq U$. By hypothesis, there is a neighborhood S of x_0 belonging to $\mathfrak{S}$. By definition of convergence in $C_{\mathfrak{S}}(X; E)$, there is i_0 such that, for all $i, i' \geqslant i_0$, we have $f_i(x) - f_{i'}(x) \in U'$ for all $x \in S$. Letting i' go to infinity gives $f_i(x) - f(x) \in \overline{U'}$ for all $x \in S$. Since f_{i_0} is continuous at x_0 there is a neighborhood $S' \subseteq S$ of x_0 such that $f_{i_0}(x) - f_{i_0}(x_0) \in U'$ for all $x \in S'$. It follows that $f(x) - f(x_0) \in U$ for all $x \in S'$. Thus f is continuous at x_0.

Does $f_i \to f$ in the topology of $C_{\mathfrak{S}}(X; E)$? Yes. The first part of the previous paragraph shows that for every $S \in \mathfrak{S}$ and every 0-neighborhood U we can find i_0 such that $f_i(x) - f(x) \in U$ for all $x \in S$ and all $i \geqslant i_0$. □

In particular the topology of pointwise convergence is usually not complete (unless we equip X with the discrete topology).

Let X and E be as above and let H be a subset of $C(X; E)$.

Definition 3.11. The subset H is *equicontinuous at* $x_0 \in X$ if for every 0-neighborhood U in E there is a neighborhood V of x_0 in X such that

$$V \subseteq f^{-1}(f(x_0) + U) \quad \forall f \in H.$$

H is *equicontinuous* if it is equicontinuous at every point of X.

Theorem 3.12. *Let X be compact and let H be an equicontinuous subset of $C(X; E)$. Then, when restricted to H, the topology of uniform convergence is the same as the topology of pointwise convergence.*

To put this another way, if you have a sequence of functions that converges pointwise, but does not converge uniformly, then the members of this sequence can't form an equicontinuous set. The classic example is a sequence of functions like

$$f_n(x) = nxe^{-nx}$$

defined on $[0, 1]$.

Proof. Let U be a 0-neighborhood in E and let $f \in H$. We must show that the corresponding *uniform* neighborhood of f in H, namely

$$W := \{g \in H : g(x) - f(x) \in U \; \forall x \in X\},$$

contains a *simple* neighborhood of f in H. Choose a balanced 0-neighborhood U' in E with $U' + U' + U' \subseteq U$. By equicontinuity, each $x \in X$ has a neighborhood

V_x with $g(V_x) \subseteq g(x) + U'$ for all $g \in H$. By compactness, X has a finite cover $V_{x_1}, \ldots, V_{x_n}$. Then

$$W' := \{g \in H : g(x_i) - f(x_i) \in U' \; \forall i = 1, \ldots, n\}$$

is a simple neighborhood of f, and $W' \subseteq W$. $\square$

Corollary 3.13. *(Ascoli's theorem) Let X be compact and let H be an equicontinuous subset of $C(X; E)$. If, for each x, the range $\{h(x) : h \in H\}$ is a precompact subset of E, then H is precompact in $C(X; E)$ for the topology of uniform convergence.*

As a special case, any bounded, equicontinuous subset of $C(X)$ is precompact (for the supremum norm).

Proof. H is precompact for pointwise convergence by Proposition 3.9; but pointwise and uniform convergence coincide on H by the previous theorem. $\square$

Appendix to the lecture

Proof. (of Lemma 3.3.) This depends unavoidably on the axiom of choice.

Let $\mathcal{N}\colon D \to X$ be a net in X, parameterized by a directed set D. A collection F of subsets of X will be called an *$\mathcal{N}$-filter* if $\mathcal{N}$ is frequently in every member of F, and if F is closed under finite intersections and the formation of supersets. Such objects exist: for example, $\{X\}$ is an $\mathcal{N}$-filter. The collection of $\mathcal{N}$-filters is partially ordered by inclusion, and every chain in this partially ordered set has an upper bound (the union). Thus Zorn's Lemma provides a maximal $\mathcal{N}$-filter; call it F_0.

Suppose that $S \subseteq X$ has the property that $\mathcal{N}$ is frequently in $A \cap S$ for every $A \in F_0$. Then the union of F_0 with the set of all sets $A \cap S$, $A \in F_0$, is again an $\mathcal{N}$-filter. By maximality we deduce that S itself belongs to F_0.

We will use this property to construct the desired universal refinement. Let D' be the collection of pairs (A, i) with $A \in F_0$, $i \in D$, and $\mathcal{N}(i) \in A$; it is a directed set under the partial order

$$(B, j) \geqslant (A, i) \Leftrightarrow B \subseteq A, \; j \geqslant i.$$

The map $(A, i) \mapsto i$ is final so defines a refinement $\mathcal{N}'$ of the net $\mathcal{N}$. We claim that this refinement is universal.

Let $S \subseteq X$ have the property that $\mathcal{N}'$ is frequently in S. Let $A \in F_0$ and let i be arbitrary. By definition, there exist $B \in F_0$, $B \subseteq A$, and $j \geqslant i$, such that

$\mathcal{N}(j) = \mathcal{N}'(B, j) \in B \cap S \subseteq A \cap S$. We conclude that $\mathcal{N}$ is frequently in $A \cap S$ for every $A \in F_0$ and hence, as observed above, that S itself belongs to F_0.

Now let S be arbitrary. It is not possible that $\mathcal{N}'$ be frequently both in S and in $X \setminus S$, for then (by the above) both S and $X \setminus S$ would belong to F_0, and then their intersection, the empty set, would do so as well, a contradiction. Thus $\mathcal{N}'$ *fails* to be frequently in one of these sets, which is to say that it is eventually in the other one. $\square$

Lecture 4

Convexity and the Hahn-Banach Theorem

A reference for this material is Rudin's Chapter 3.

Let E be a *real* vector space.

Definition 4.1. A map $p\colon E \to \mathbb{R}$ is *sublinear* if $p(x+y) \leqslant p(x) + p(y)$ and $p(\lambda x) = \lambda p(x)$ for $\lambda > 0$.

Obviously, a seminorm is an example of a sublinear map. There are other examples, e.g. the map $f \mapsto \sup f$ on the space of real-valued continuous functions on $[0, 1]$.

Theorem 4.2. *(Hahn-Banach 1) Let E be a real vector space equipped with a sublinear functional p. Let $F \leqslant E$ be a subspace and let $\phi\colon F \to \mathbb{R}$ be a linear functional with $\phi(x) \leqslant p(x)$ for all $x \in F$. Then ϕ can be extended to a linear functional $\psi\colon E \to \mathbb{R}$ with $\psi(x) \leqslant p(x)$ for all $x \in E$.*

We emphasize that no topology is involved in this statement. "Extended" means that the diagram

$$\begin{tikzcd}
E \arrow[r] \arrow[dr, "\phi"'] & F \arrow[d, "\psi"] \\
 & \mathbb{R}
\end{tikzcd}$$

must commute.

Proof. Let $\mathcal{C}$ be the collection of pairs (G, χ), where $F \leqslant G \leqslant E$, and $\chi\colon G \to \mathbb{R}$ is a linear functional extending ϕ and having $\chi \leqslant p$. Order this set by saying that $(G_1, \chi_1) \leqslant (G_2, \chi_2)$ if $G_1 \leqslant G_2$ and χ_2 extends χ_1.

In this partially ordered set, each chain has an upper bound (the union). So by Zorn's Lemma, there is a maximal element (G, χ) in $\mathcal{C}$. We claim that this maximal G is the whole of E.

Suppose not. Choose $x_1 \in E \setminus G$ and try to extend χ to a functional χ_1 on $G_1 = \langle G, x_1 \rangle$. Such an extension is determined by $\alpha = \chi_1(x_1)$; in fact, it is given by the formula

$$\chi_1(x + \lambda x_1) = \chi(x) + \lambda\alpha \quad (x \in G);$$

we need to pick α so that the condition $\chi_1 \leqslant p$ is satisfied.

Taking $\lambda = \pm 1$ it is evident that we must have, for all $x \in G$,

$$\chi(x) + \alpha \leqslant p(x + x_1), \quad \chi(x) - \alpha \leqslant p(x - x_1)$$

and conversely a scaling argument shows that this would be sufficient. Can we choose α to make this condition true? Equivalently, can we choose α so that

$$\chi(y) - p(y - x_1) \leqslant \alpha \leqslant p(x + x_1) - \chi(x) \quad \forall x, y \in G?$$

Such an α can be found if and only if

$$\chi(y) - p(y - x_1) \leqslant p(x + x_1) - \chi(x) \quad \forall x, y \in G,$$

which is equivalent to

$$\chi(x) + \chi(y) \leqslant p(x + x_1) + p(y - x_1).$$

But we have

$$\chi(x) + \chi(y) = \chi(x + y) \leqslant p(x + y) \leqslant p(x + x_1) + p(y - x_1),$$

so the latter condition is true and we have shown that χ can be extended as required. This contradicts the supposed maximality of χ so we conclude that $E = G$ and the theorem is proved. $\square$

Exercise 4.3. A standard application: Let E denote the vector space of bounded sequences $\mathbf{a} = \{a_n\}$ of real numbers. Show that there is a linear map $\ell\colon E \to \mathbb{R}$ such that $\ell(\mathbf{a}) \leqslant \sup\{a_n\}$, such that $\ell(\mathbf{a}) = \lim a_n$ whenever the limit exists, and such that $\ell(\mathbf{a}) = \ell(S\mathbf{a})$ where S is the shift operator $(Sa)_n = a_{n+1}$. (The functional ℓ is called a *Banach limit*. Apply the Hahn-Banach theorem where p is the sublinear functional 'lim sup' and F is the subspace spanned by the constant sequences together with those of the form $\mathbf{a} - S\mathbf{a}$.)

Theorem 4.4. *(Hahn-Banach 2) Let E be a (real or complex) vector space and let p be a seminorm on E. Let $F \leqslant E$ and let $\phi\colon F \to \mathbb{k}$ be a linear functional such that $|\phi(x)| \leqslant p(x)$ for all $x \in F$. Then ϕ can be extended to a linear functional on E satisfying the same bound.*

Proof. (Real case) Apply Theorem 4.2 to obtain an extension ψ with $\psi \leqslant p$. We have $\psi(x) \leqslant p(x)$ and $-\psi(x) = \psi(-x) \leqslant p(-x) = p(x)$, so $|\psi| \leqslant p$ as required.

(Complex case) Consider F_r, E_r, the underlying real vector spaces of F and E; and let $\phi_r = \Re\phi\colon F_r \to \mathbb{R}$. Clearly $|\phi_r| \leqslant p$; so there is an extension $\psi_r\colon E_r \to \mathbb{R}$ with $\psi_r \leqslant p$. We can always find a *complex* linear ψ with $\Re\psi = \psi_r$, namely

$$\psi(x) = \psi_r(x) - i\psi_r(ix)$$

and ψ will extend ϕ. Given $x \in E$ there is θ such that $e^{i\theta}\psi(x) \in \mathbb{R}^+$; then

$$|\psi(x)| = |\psi(e^{i\theta}x)| = \psi_r(e^{i\theta}x) \leqslant p(e^{i\theta}x) = p(x)$$

as required. $\square$

Corollary 4.5. *Let E be a LCTVS, F a subspace of E. Any continuous linear functional on F extends to one on E.* $\square$

The conclusion may be expressed as follows: the ground field $\Bbbk$ is an injective object in the category of LCTVS.

Definition 4.6. The *dual space* E^* of a TVS E is the space of *continuous* linear functionals $E \to \Bbbk$.

Proposition 4.7. *Let S be any subset of a Hausdorff LCTVS E. The closed linear span of S (the closure of the set of finite linear combinations of members of S) is equal to the intersection of the kernels of all $\phi \in E^*$ that annihilate S.*

Proof. Let F be the closed linear span of S. If $\phi \in E^*$ annihilates S, then $\ker \phi$ is a closed subspace containing S, hence containing F. Conversely, suppose that $x_0 \notin F$ and let $F_0 = \langle F, x_0 \rangle$. The quotient space F_0/F is one-dimensional, spanned by the equivalence class of x_0, so (by Proposition 1.12) there is a linear functional

$$F_0 \to F_0/F \to \Bbbk$$

that annihilates F and maps x_0 to 1. Extend this to $\phi \in E^*$ and we have $F \subseteq \ker \phi$, $\phi(x_0) = 1$. This completes the proof. $\square$

Corollary 4.8. *The dual space of a Hausdorff LCTVS E separates points of E; that is, given $x \neq 0$ in E, there exists $\phi \in E^*$ with $\phi(x) \neq 0$.*

Proof. Apply the previous proposition to $S = \{0\}$. $\square$

The next equivalent form of the Hahn-Banach theorem is stated geometrically.

Theorem 4.9. *(Hahn-Banach 3) Let E be a real Hausdorff LCTVS, and let A and B be disjoint closed convex subsets one of which (say A) is compact. Then there is $\phi \in E^*$ such that*

$$\sup\{\phi(a) : a \in A\} < \inf\{\phi(b) : b \in B\}.$$

In other words, A and B are separated by the hyperplane $\{x : \phi(x) = \lambda\}$, for any λ strictly between the sup and the inf above.

Proof. A standard compactness argument shows that there is a 0-neighborhood V such that $A' = A + V$ does not meet B. Since E is locally convex, we may and shall assume that V is convex. Then A' is convex also.

Pick $a_0 \in A$, $b_0 \in B$, let $x_0 = b_0 - a_0$ and let $C = A' - B + x_0$; C is a convex 0-neighborhood. Let p be its Minkowski functional. Convexity implies that p is sublinear. Since $x_0 \notin C$, $p(x_0) \geqslant 1$. Using Theorem 4.2 we find that there is a linear map $\phi\colon E \to \mathbb{R}$ with $\phi \leqslant p$ and $\phi(x_0) = 1$.

Since $p \leqslant 1$ on C, $|\phi(x)| \leqslant 1$ for x in the 0-neighborhood $C \cap (-C)$; hence ϕ is continuous. Moreover, for $a \in A'$ and $b \in B$ we have

$$\phi(a) - \phi(b) + 1 = \phi(a - b + x_0) \leqslant p(a - b + x_0) \leqslant 1,$$

since $a - b + x_0 \in C$. Thus $\phi(a) \leqslant \phi(b)$. Now $\phi(A')$ and $\phi(B)$ are convex subsets of $\mathbb{R}$, that is, intervals; and we have shown that $\phi(A')$ lies to the left of $\phi(B)$. Moreover $\phi(A')$ is *open*. Let $\lambda = \sup \phi(A')$. Then $\inf \phi(B) \geqslant \lambda$ by the inequality above. On the other hand, $\phi(A)$ is a compact interval contained in the open interval $\phi(A')$, so $\sup \phi(A) < \lambda$. This proves the theorem. □

These results lead to an important structure theorem for convex sets. Let S be any subset of a vector space E. The *convex hull* of S is the smallest convex set containing S: it may be defined as the intersection of all convex sets containing S, or more concretely as the collection of all finite convex combinations of members of S. If E is a TVS, the *closed convex hull* of S is the smallest *closed* convex set containing S; it is the closure of the convex hull of S.

Let C be a convex subset of a vector space E. A point $p \in C$ is called an *extreme point* of C if it cannot be written as a nontrivial convex combination of points of C; that is, if $p = \lambda p_0 + (1 - \lambda)p_1$, $\lambda \in (0, 1)$, $p_0, p_1 \in C$, then in fact $p_0 = p_1 = p$.

Theorem 4.10. *(Krein-Milman) Let E be a Hausdorff LCTVS. Any nonempty compact convex subset of E is the closed convex hull of its extreme points. In particular, it has some extreme points.*

An example is the space of Radon probability measures on $[0, 1]$, with its vague topology. This is a compact set (a consequence of the Banach-Alaoglu theorem, which we will prove in the next lecture) and it is clearly convex. The extreme points are the Dirac measures at the various points of $[0, 1]$. The theorem tells us that *any* such measure is the vague limit of a net of convex combinations of Dirac measures.

Proof. Let K be a nonempty compact convex subset of E. Generalizing the notion of extreme *point*, let us call a subset $S \subseteq K$ an extreme *set* if it is closed and nonempty and, for each $p \in S$, if $p = \lambda p_0 + (1 - \lambda)p_1$, $\lambda \in (0, 1)$, $p_0, p_1 \in K$, then in fact $p_0, p_1 \in S$. Extreme sets exist (K is one); they can be partially ordered by inclusion; every decreasing chain has a lower bound (use compactness to show that the intersection of a chain is nonempty). By Zorn's Lemma, there exist minimal extreme sets. We're going to show that they consist of single points.

Let S be any extreme set and $\phi \in E^*$. Let $M = \sup\{\Re\phi(x) : x \in S\}$. It is easy to see that

$$\{x \in S : \Re\phi(x) = M\}$$

is also an extreme set. Thus, if S is *minimal* extreme, every $\phi \in E^*$ is constant on S. By Corollary 4.8, S must consist of a single point only.

Let K' be the closed convex hull of the extreme points of K. Clearly, $K' \subseteq K$, and thus K' is compact. Suppose that $x_0 \in K \setminus K'$. Then by Theorem 4.9, there is $\phi \in E^*$ with $\Re\phi(x) < \Re\phi(x_0)$ for all $x \in K'$. Reprising a previous argument, let $M = \sup\{\Re\phi(x) : x \in K\}$. Then $\{x \in K : \Re\phi(x) = M\}$ is an extreme set and it does not meet K'. Since every extreme set contains a minimal extreme set, i.e. an extreme point, this is a contradiction. $\square$

Lecture 5

Baire Category Arguments

Begin with a few remarks about continuity and boundedness for linear maps.

Definition 5.1. Let E be a TVS. A subset $A \subseteq E$ is *bounded* if for every 0-neighborhood U there exists $r > 0$ with $A \subseteq rU$.

Remarks 5.2. A *sequence* that tends to 0 is bounded. A set A is bounded iff every sequence that it contains is bounded. In a LCTVS with topology defined by seminorms p_α, a set A is bounded iff each p_α is bounded on A. All of these are easy exercises.

Definition 5.3. A linear map $T\colon E \to F$ between topological vector spaces is said to be *bounded* if it takes bounded sets to bounded sets. A collection H of such maps is *uniformly bounded* or *equibounded* if, for each bounded subset B of E, the union $\bigcup_{T \in H} T(B)$ is bounded in F.

Proposition 5.4. *Let E, F be TVS and let H be a set of linear maps from E to F. If H is equicontinuous, then it is equibounded. The converse holds provided that E is metrizable.*

The most important case is when H consists of a single map.

Proof. We'll take it for granted that (equi)continuity for linear maps need only be checked at 0.

Suppose that H is equicontinuous and that B is bounded in E. Let V be a 0-neighborhood in F. By equicontinuity there is an 0-neighborhood U in E with $T(U) \subseteq V$ for all $T \in H$. Then there is $r > 0$ such that $B \subseteq rU$. It follows that $\bigcup_{T \in H} T(B) \subseteq rV$. Thus $\bigcup_{T \in H} T(B)$ is bounded.

Suppose that H is equibounded and that E is metrizable, with metric d. Assume for a contradiction that H *fails* to be equicontinuous. Then there exists a 0-neighborhood V in F with the following property: for each $n = 1, 2, \ldots$ there exist $x_n \in E$ and $T_n \in H$ with $d(x_n, 0) < 1/n$ while $T_n(x_n) \notin V$.

Define the integer ℓ_n by $\ell_n = \min\{n, \lfloor d(x_n, 0)^{-1/2} \rfloor\}$. Then $\ell_n \to \infty$. On the other hand, repeated application of the triangle inequality and the translation-invariance of the metric gives $d(\ell_n x_n, 0) \leqslant \ell_n d(x_n, 0) \to 0$, so $\ell_n x_n \to 0$, and hence $\{\ell_n x_n\}$ is a bounded set. By equiboundedness the sequence $\ell_n T_n(x_n)$ is also bounded, and it follows that $\ell_n T_n(x_n) \in rV$ for some $r > 0$. But then $T_n(x_n) \in V$ as soon as $\ell_n > r$, and this is a contradiction. $\square$

Exercise 5.5. Consider the space of all continuous functions on $[0,1]$. Let E be this space equipped with the topology of pointwise convergence, and let F be the same space equipped with the topology of convergence in measure. Show that the identity map $E \to F$ is bounded, and sequentially continuous, but nevertheless not continuous. Compare with the result above.

Theorem 5.6. *(The Baire Category Theorem) In a complete metric space, the intersection of countably many dense open sets is dense.*

Proof. It's enough to prove that such a countable intersection is *nonempty*, since this assertion can then be applied to any closed ball in the original space.

Let $U_1, U_2, \ldots$ be dense open sets. By induction define a sequence of open balls $B(x_i;\delta_i)$ with $\overline{B}(x_1;\delta_1) \subseteq U_1$ and

$$\overline{B}(x_n;\delta_n) \subseteq B(x_{n-1};\delta_{n-1}) \cap U_n, \quad \delta_n < 1/n.$$

The induction proceeds because the intersection on the right hand side must be *nonempty* and open, since U_n is dense.

By construction the centers $\{x_n\}$ form a Cauchy sequence, so they converge to a limit x which is in each $\overline{B}(x_n;\delta_n)$ and therefore is in each U_n. □

There is also a version of the theorem (with almost the same proof) for locally compact Hausdorff spaces, but we won't need that. An equivalent statement is that a complete metric space cannot be written as the union of countably many closed sets with empty interior.

Exercise 5.7. Let X be the Banach space $C[0,1]$ and let X_n be the set of functions f such that, for some $t \in [0,1]$, the inequality $|f(s) - f(t)| \leqslant n|s-t|$ holds for all $s \in [0,1]$. Show that each nonempty open set in X contains a nonempty open subset that does not meet X_n. Use the Baire category theorem to deduce the existence of a continuous, nowhere differentiable function.

Let E be a topological vector space.

Definition 5.8. A *barrel* in E is a closed, balanced, convex, absorbing set. The space E is *barreled* if every barrel contains a neighborhood of 0.

Proposition 5.9. *Every Fréchet space is barreled.*

Proof. Let B be a barrel in a Fréchet space E. Then $\bigcup_{k=1}^{\infty} kB = E$. Since E has a complete metric the Baire category theorem applies to show that some kB has nonempty interior. Then $2kB \supseteq kB - kB$ contains a neighborhood of the origin, so by rescaling B itself contains a neighborhood of the origin. $\square$

Theorem 5.10. *(Banach-Steinhaus) Let E be a barreled space and let F be any LCTVS. Let H be a set of continuous linear maps from E to F. Suppose that for each $x \in E$, the set*

$$H(x) := \{h(x) : h \in H\}$$

is bounded in F. Then H is equicontinuous.

Proof. Let W be a closed convex balanced 0-neighborhood in F and let $B = \bigcap\{h^{-1}(W) : h \in F\}$. For every $x \in E$, the set $H(x)$ is bounded so it is contained in some rW, which implies that x itself is contained in rB. Thus, B is absorbing, and it is clear that B is also closed, convex and balanced. Hence, B is a barrel, and contains a neighborhood of the origin. Since $H(B) \subseteq W$, this proves equicontinuity. $\square$

If E is a Fréchet space, so that equicontinuity and equiboundedness are the same thing, the result can be restated: a family of continuous linear maps defined on a Fréchet space is *uniformly bounded* if and only if it is *pointwise bounded*. This is the *uniform boundedness principle*.

Corollary 5.11. *With notation as above, suppose that $\{T_n\}$ is a sequence of continuous linear maps from a barreled space E to a locally convex space F, and that $T_n(x)$ converges, say to $T(x)$, for each $x \in E$. Then the linear map T is also continuous.* $\square$

Exercise 5.12. In the above result it is *not* asserted that $T_n \to T$ in any sense stronger than pointwise convergence. Underline this by constructing an example where E and F are Banach spaces, and the conditions of the corollary are satisfied, but T_n does not converge in norm to T.

We will use the following corollary in our discussion of tensor products and the Schwartz kernels theorem.

Proposition 5.13. *Let E, F be Fréchet spaces and G a LCTVS. Let $L\colon E \times F \to G$ be a bilinear map which is separately continuous (i.e., $L(x, y)$ is a continuous function of x for each fixed y and vice versa). Then L is continuous.*

Proof. Let B and C be bounded subsets of E and F respectively. Consider the collection $H = \{L_x : x \in B\}$ of linear maps from F to G, where $L_x(y) = L(x, y)$. For each $y \in F$ the collection $H(y) = \{L_x(y)\}$ is bounded (by continuity of L in the first variable). Thus H is pointwise bounded, hence uniformly bounded by the Banach-Steinhaus theorem. In particular, $\bigcup_{y \in C} H(y)$ is bounded in G. But this is $L(B \times C)$, so L is a bounded map, hence continuous (Proposition 5.4). $\square$

A continuous mapping between topological spaces is said to be *open* if it carries open sets to open sets. A quotient mapping is a standard example.

We are going to prove that any surjective continuous linear map between Fréchet spaces is open. This is the *open mapping theorem*. First, a weaker result.

Lemma 5.14. *Let E be a LCTVS, F a barreled space, and $T\colon E \to F$ a continuous linear surjection. Then for each 0-neighborhood U in E, the* closure $\overline{T(U)}$ *is an 0-neighborhood in F.*

Proof. We may assume that U is balanced and convex. Being a 0-neighborhood, it is also absorbing. Then $\overline{T(U)}$ is closed, balanced, convex, and absorbing—i.e., a barrel. The result follows. $\square$

To remove the closure here requires some additional argument that uses a complete metric on E.

Theorem 5.15. *Let E and F be Fréchet spaces and let $T\colon E \to F$ be a continuous linear surjection. Then T is an open map.*

Proof. Fix complete metrics defining the topologies of E and F. For each $r > 0$ let $U = B(0; r) \subseteq E$. It suffices to show that $T(U)$ is an 0-neighborhood in F.

Let $U_k = B(0; 2^{-k}r) \subseteq E$, $k = 2, 3, \ldots$. Let $y \in \overline{T(U_2)}$, which is a 0-neighborhood by Lemma 5.14. It will suffice to show that $y \in T(U)$.

Inductively define sequences $\{x_k\}$ in E and $\{y_k\}$ in F as follows. The induction starts with $y = y_1 \in \overline{T(U_2)}$. Suppose for induction that $y_k \in \overline{T(U_{k+1})}$. By definition, one can find $x_{k+1} \in U_{k+2}$ such that $y_k - T(x_{k+1})$ belongs to the 0-neighborhood $\overline{T(U_{k+2})} \cap B_F(0; 1/k)$. Define $y_{k+1} = y_k - T(x_{k+1})$ to complete the induction.

By construction $y_k \to 0$ and the partial sums of the series $\sum x_k$ form a fast Cauchy sequence. Hence $\sum x_k$ converges, say to x, with $d(x, 0) \leqslant \sum_2^\infty r2^{-k} < r$, so $x \in U$. We have

$$T(x) = \lim_{n\to\infty} \sum_{k=2}^{n} T(x_k) = \lim_{n\to\infty} \sum_{k=2}^{n} (y_k - y_{k+1}) = y.$$

This completes the proof ☐

We can restate this result in language which is a bit more category-theoretic.

Definition 5.16. We say that an *isomorphism* of TVS is a linear homeomorphism (a continuous linear map with continuous inverse). A *homomorphism* will be a linear map $T\colon E \to F$ with the property that the canonical map

$$E/\ker(T) \to \mathrm{Im}(T)$$

is an isomorphism (in the sense defined above).

Corollary 5.17. *A continuous linear map between Fréchet spaces is an isomorphism if and only if it is bijective. It is a homomorphism if and only if its range is closed.*

Proof. Let $T\colon E \to F$ be such a map. If it is bijective, then it is open (by 5.15) so its inverse is continuous.

Now suppose T is a homomorphism. Then $\mathrm{Im}(T)$ is isomorphic to the Fréchet space $E/\ker(T)$, so it is complete, hence closed. Conversely, if $\mathrm{Im}(T)$ is closed, then the canonical map $E/\ker(T) \to \mathrm{Im}(T)$ is a continuous linear bijection between Fréchet spaces, hence an isomorphism. ☐

Let $T\colon E \to F$ be a map. The *graph* of T is the subset $\{(x, y) : y = Tx\}$ of $E \times F$.

Theorem 5.18. *A linear map between Fréchet spaces is continuous if and only if its graph is a closed subset of $E \times F$.*

The condition that the graph is closed can be stated in the following way: let $\{x_n\}$ be a sequence in E and suppose there exist points $x \in E$, $y \in F$, such that $x_n \to x$ and $Tx_n \to y$. Then $Tx = y$. This formulation makes it clear why the closed graph condition is a priori weaker than continuity: in the statement above the convergence of Tx_n is part of the *hypothesis*, whereas in the definition of continuity it is part of the *conclusion*.

Proof. It is obvious that if T is continuous, its graph is closed. To prove the converse let G be the graph, considered as a subspace of the Fréchet space $E \times F$. Since G is closed, it is itself a Fréchet space. There are projection maps $p\colon G \to E$ and $q\colon G \to F$. The projection p is a continuous bijection, hence a homeomorphism by Theorem 5.15; and $T = q \circ p^{-1}$. ☐

Remark 5.19. Suppose that $T\colon E \to F$ is a linear map and that $J\colon F \to G$ is a continuous, linear, injective map to a Hausdorff TVS G (for instance, G might be F with some topology weaker than its original topology). If $J \circ T$ is continuous, then the graph of T is closed (the proof is easy). Thus, if E and F are Fréchet spaces, T is continuous into the stronger topology of F. This is often how the CGT is applied in practise.

Exercise 5.20. One cannot dispense with completeness in these theorems. Let E be the Hilbert space ℓ^2; let F be the same vector space, but equipped with the ℓ^∞ norm instead. Show that the identity map $i\colon E \to F$ is continuous, and has closed graph, but is not an isomorphism.

Exercise 5.21. Now let F be any Banach space. Using the axiom of choice, show that F has a Hamel basis $\{f_\alpha\}$ consisting of unit vectors and having the property that $\inf\{\|f_\alpha - f_\beta\|\} = 0$. Let E be the same vector space as F but equipped with the norm

$$\left\| \sum_{j=1}^{n} \lambda_j e_{\alpha_j} \right\|_E = \sum_{j=1}^{n} |\lambda_j|.$$

Show that the identity map $i\colon E \to F$ is continuous, and has closed graph, but is not an isomorphism.

Lecture 6

Duality and Weak Topologies

Let E be a Banach space. Recall that the *dual space* E^* is the space of continuous linear maps $E \to \Bbbk$. It is a Banach space also, with the norm $\|\phi\| = \sup\{|\phi(x)| : \|x\| \leqslant 1\}$. The process of forming duals is functorial: a continuous linear map $T\colon E \to F$ induces a continuous linear map $T^*\colon F^* \to E^*$. Moreover, the process can be iterated. The double dual E^{**} is a Banach space and there is a canonical isometric embedding $E \to E^{**}$. If this embedding is an isomorphism (of Banach spaces) then E is called *reflexive*.

In the next couple of lectures we shall generalize this familiar material to the context of locally convex TVS.

Definition 6.1. Let E be a LCTVS. The *dual space* E^* is the vector space of continuous linear maps $E \to \Bbbk$.

The Hahn-Banach theorem shows that the pairing between E and E^* is nondegenerate in the sense discussed in lecture 2. This pairing gives rise to natural topologies.

Definition 6.2. In the above circumstances the $\sigma(E, E^*)$-topology on E is called its *weak topology*. The $\sigma(E^*, E)$-topology on E^* is called its *weak-$*$ topology*.

The weak topology on E is weaker (usually strictly weaker) than its original topology. Similarly the weak-$*$ topology on E^* is the weakest of the natural topologies that can be put on this space.

Lemma 6.3. *Let E and F be nondegenerately paired vector spaces. Let E_w denote E considered as a topological vector space with the $\sigma(E, F)$-topology. Then the natural map $F \to (E_w)^*$ is an isomorphism.*

Proof. Nondegeneracy says that the natural map is injective; we must show that it is surjective, or in other words that every continuous linear functional on E_w comes from an element of F. Let ϕ be such a linear functional. By continuity and the definition of the weak topology, there are finitely many $f_1, \ldots, f_m \in F$ such that

$$|\phi(x)| \leqslant C(|f_1(x)| + \cdots + |f_m(x)|)$$

for all $x \in X$. But then $\bigcap_{j=1}^{m} \operatorname{Ker} f_j \subseteq \operatorname{Ker} \phi$, so by standard linear algebra[1] ϕ is a linear combination of the f_j, and is therefore in F. $\square$

Corollary 6.4. *A linear functional on a LCTVS E is originally continuous (that is, continuous w.r.t. the original topology on E) iff it is weakly continuous.*

The weak topology has many fewer closed sets than the original topology. Nevertheless, it has the same *convex* closed sets:

Proposition 6.5. *A* convex *subset A of a LCTVS E is originally closed iff it is weakly closed.*

Proof. It is clear that weakly closed implies originally closed. Suppose that A is originally closed and let $x \notin A$. By the third version of the Hahn-Banach theorem (4.9) there are a linear functional $\phi_x \in E^*$ and a constant C_x such that $\Re\phi_x(x) > C_x$ while $\Re\phi_x(a) \leqslant C_x$ for all $a \in A$. Let $H_x = \{y : \Re\phi_x(y) \leqslant C_x\}$. Then H_x is weakly closed and

$$A = \bigcap \{H_x : x \notin A\}.$$

Thus A is weakly closed as required. $\square$

Proposition 6.6. *A subset A of a LCTVS E is originally bounded iff it is weakly bounded.*

Proof. It is clear that originally bounded implies weakly bounded. Suppose that A is weakly bounded and let p be one of the seminorms defining the original topology of E. We make use of a "localization" construction which can also be applied elsewhere.

Note that $\operatorname{Ker}(p) := \{x \in E : p(x) = 0\}$ is a subspace of E, and that p gives rise to a norm on the quotient space $E/\operatorname{Ker}(p)$. Let $(E_p, \|\cdot\|_p)$ denote the Banach space completion of this normed space. By construction there is a continuous map $\pi_p\colon E \to E_p$, and $\|\pi_p(x)\|_p = p(x)$. It suffices therefore to show that $B = \pi_p(A)$ is bounded for every p. We know that B is weakly bounded. (In other words, at this point we have reduced the problem to the case of a Banach space.)

---

[1] Here is an explanation of the standard linear algebra. Clearly there is no loss of generality in assuming the f_j to be linearly independent. Let $K = \bigcap_{j=1}^{m} \operatorname{Ker} f_j$. The list $f = (f_1, \ldots, f_n)$ describes a surjection $\pi\colon E \to \Bbbk^n$ with kernel K. Since $K \subseteq \ker\phi$, the linear map ϕ can be factored through this surjection: there is $\psi\colon \Bbbk^n \to \Bbbk$ such that $\phi = \psi \circ \pi$. Writing ψ in terms of the dual basis on $(\Bbbk^n)^*$ amounts to writing ϕ as a linear combination of the $f_1, \ldots, f_n$.

Consider the natural isometric embedding $\iota\colon E_p \to (E_p)^{**}$, and consider $H = \iota(B) \subseteq (E_p)^{**}$ as a collection of linear functionals on $(E_p)^*$. The weak boundedness of B tells us that $H(\phi)$ is bounded for each $\phi \in (E_p)^*$, i.e., H is point bounded. By the Banach-Steinhaus Theorem (5.10), H is uniformly bounded—that is, bounded in norm. But since ι is an isometry that implies that the original B is bounded in norm, that is, $p(A)$ is bounded. The proof is complete. ☐

Exercise 6.7. Let K be a compact metrizable space. Show that a sequence $\{f_n\}$ in the Banach space $C(K)$ converges weakly to f if and only if the norms $\|f_n\|$ are uniformly bounded and $f_n(x) \to f(x)$ for each $x \in K$.

Exercise 6.8. The weak topology of E is usually not complete. In fact, show that the completion of E_w is the *algebraic* dual space of E^*, provided with its E^*-weak topology.

Definition 6.9. Let E be a LCTVS. Let A be a subset of E and let B be a subset of E^*. The *polar* of A is the subset

$$A^\circ = \{\phi \in E^* : |\phi(x)| \leqslant 1 \ \forall x \in A\}$$

of E^*. Similarly the *prepolar* of B is the subset

$$^\circ B = \{x \in E : |\phi(x)| \leqslant 1 \ \forall \phi \in B\}$$

of E.

Note that if A or B are subspaces, these reduce to the usual definitions of annihilator and preannihilator. If A is the unit ball of a Banach space E, then A° is the unit ball of its dual.

The polar is weak-$*$ closed; the prepolar is weakly closed (by Proposition 6.5 this is equivalent to saying that it is closed in the original topology of E). It is clear that both are convex, balanced sets.

Exercise 6.10. Show that the polar of a bounded subset of E is absorbing, and that the prepolar of a weak-$*$ bounded subset of E^* is absorbing.

Proposition 6.11. *(Bipolar theorem) Let E be a LCTVS and E^* its dual. For any closed balanced convex subset $A \subseteq E$, one has $^\circ(A^\circ) = A$; and for any weak-$*$ closed balanced convex subset $B \subseteq E^*$, one has $(^\circ B)^\circ = B$.*

It follows that for any $A \subseteq E$ at all, the *bipolar* $^\circ(A^\circ)$ is the *closed balanced convex hull* of A (the smallest closed balanced convex set containing A). There is a similar statement for subsets of E^*.

Proof. Clearly $^\circ(A^\circ)$ is closed, balanced, convex, and contains A. Suppose now that $x \in E \setminus A$. By the Hahn-Banach theorem (4.9) there is a linear functional $\phi \in E^*$ such that $\Re\phi(a) \leqslant 1$ for all $a \in A$, while $\Re\phi(x) > 1$. Since A is balanced we can promote the first inequality to say that $|\phi(a)| \leqslant 1$ for all $a \in A$. We see that $\phi \in A^\circ$ and thus $x \notin {}^\circ(A^\circ)$.

For the other case, apply the first case to E^* (with its weak-$*$ topology) and its dual space which is E (by Lemma 6.3). □

The next result is specific to the weak-$*$ topology.

Theorem 6.12. *(Alaoglu) Let E be any TVS. The polar of any 0-neighborhood in E is weak-$*$ compact.*

Proof. Let V be an 0-neighborhood in E. Because V is absorbing, for every $x \in E$ there is $r(x) > 0$ such that $x \in r(x)V$.

For each $x \in E$ let D_x be the closed disc in $\Bbbk$ with center the origin and radius $r(x)$. Let X be the product $\prod_{x \in E} D_x$, equipped with the product topology. A point of x is therefore a function (not necessarily linear or continuous) $f\colon E \to \Bbbk$, with $|f(x)| \leqslant r(x)$ for all $x \in E$. We give X the Tychonoff (product) topology, which makes it a compact space.

Map V° to X by sending any $\phi \in V^\circ$ to itself, considered as a map of the kind described above. (The condition $\phi \in V^\circ$ exactly says that $|\phi(x)| \leqslant r(x)$ for all x.) Obviously this map $j\colon V^\circ \to X$ is injective. Moreover, the weak-$*$ topology on V° is the weak topology coming from the evaluation maps $\phi \mapsto \phi(x)$, $x \in V$; in other words, it is the restriction to $j(V^\circ)$ of the product topology on X. It suffices then to show that $j(V^\circ)$ is closed in X. But, in fact, $j(V^\circ)$ is comprised exactly of the *continuous linear* maps in X. The condition of linearity can be expressed by saying that $j(V^\circ)$ is the intersection of the sets

$$F_{x,y,\lambda,\mu} = \{f \in X : f(\lambda x + \mu y) - \lambda f(x) - \mu f(y) = 0\}$$

each of which manifestly is closed in the product topology; thus, the linear elements of X form a closed set. Moreover, if $f \in X$ is linear, then the inverse image of the unit disc (under f) contains the 0-neighborhood V; thus f is automatically continuous. We conclude that $j(V^\circ)$ is closed in X, and this completes the proof. □

Corollary 6.13. *A weak-$*$ closed, equicontinuous subset of E^* is compact.* ☐

Exercise 6.14. Does the Alaoglu Theorem remain true if we merely assume that V is an absorbing subset of E? Why or why not?

Remark 6.15. Suppose E is a *separable* space and let A be a weak-$*$ compact subset of E^*. Then there are countably many continuous functions on A (given by evaluation on a countable dense subset of E) which (collectively) separate the points of A. This implies that A is metrizable; therefore, it is sequentially compact (not just covering compact!). Warning: E itself need *not* be metrizable.

Lecture 7

Duality and Linear Mappings

Let E be a LCTVS and F a closed subspace. There are two natural maps relating the embedding $F \to E$ with the formation of polars (=annihilators).

(a) Let $\phi \in F^\circ$ be a continuous linear functional annihilating F. Then ϕ defines a continuous map $E/F \to \Bbbk$, by the homomorphism theorem. This process gives rise to a continuous linear map $F^\circ \to (E/F)^*$.

(b) Let ψ be a continuous linear functional on F. It can be extended, via the Hahn-Banach theorem, to a continuous linear functional ϕ on E. Two such choices of extension differ by an element of F°. Therefore we obtain a well-defined continuous linear map $F^* \to E^*/F^\circ$.

Proposition 7.1. *The maps defined in (a) and (b) above are isomorphisms of topological vector spaces (in their weak-$*$ topologies). Moreover, the quotient topology on E/F induced by the weak topology of E is the $\sigma(E/F, F^\circ)$-topology.*

Proof. (a) Let $\pi\colon E \to E/F$ be the canonical quotient map. Composition with π defines a linear map $\pi^*\colon (E/F)^* \to E^*$, whose image is contained in F°. This map is a two-sided inverse to the map $\Psi_1\colon F^\circ \to (E/F)^*$ specified in (a) above. Hence, Ψ_1 is an isomorphism of vector spaces.

To see that Ψ_1 is a *topological* isomorphism, note that the weak-$*$ topology on $(E/F)^*$ is, by definition, the weakest topology that makes all the evaluation functionals at points $\pi(x)$, $x \in E$, continuous. On the other hand the subspace topology of F° (induced by the weak-$*$ topology on E^*) is the weakest topology that makes all the evaluation functionals at points $x \in E$ continuous. Clearly, these topologies agree under π^*.

Consider the quotient topology on E/F induced by the weak topology on E. It is defined by the seminorms $x \mapsto \inf\{|\phi(x+v)| : v \in F\}$, as ϕ ranges over E^*. Now, if $\phi \notin F^\circ$, then the set of values $\phi(v)$, $v \in F$, covers the entire field $\Bbbk$, and therefore the infimum here is 0. The remaining seminorms correspond to those functionals $\phi \in F^\circ$, and so they define the $\sigma(E/F, F^\circ)$-topology.

(b) Let $j\colon F \to E$ denote the inclusion map and $j^*\colon E^* \to F^*$ the corresponding restriction map on dual spaces. Clearly $\mathrm{Ker}(j^*) = F^\circ$ so j^* induces a linear map $E^*/F^\circ \to F^*$ which is inverse to the map $\Psi_2\colon F^* \to E^*/F^\circ$ defined in (b) above. Hence , Ψ_2 is an isomorphism of vector spaces.

To see that Ψ_2 is a *topological* isomorphism, note (as in the last paragraph of (a) above) that the quotient topology on E^*/F° is defined by seminorms $\phi \mapsto \inf\{|\phi(x)+\theta(x)| : \theta \in F^\circ\}$, as x ranges over E. If $x \notin F$ then the set of values $\{\theta(x) : \theta \in F^\circ\}$ is all of $\Bbbk$ (by the Hahn-Banach theorem), so the corresponding seminorm is 0. The remaining seminorms correspond to the elements of F, so they define the weak-$*$ topology on F^*. □

Let E be a LCTVS. In addition to its weak-$*$ topology the dual space E^* can be provided with several other topologies.

Definition 7.2. The *strong* topology on E^* is the topology of uniform convergence on bounded subsets of E.

An equivalent definition is to say that a basis of 0-neighborhoods for the strong topology is the collection of *polars* B° of bounded subsets $B \subseteq E$. For example, when E is a Banach space, this is the usual (norm) topology on the dual space. The strong topology is often a more natural topology to consider on E^* than the weak-$*$ topology. We will not go far into its theory here, however.

The following observation is a version of the Banach-Steinhaus theorem.

Proposition 7.3. *Let E be a barreled LCTVS (e.g. a Fréchet space) and let H be a subset of E^*. The following properties are equivalent:*

(a) H is weak-$*$ bounded;

(b) H is strongly bounded;

(c) H is equicontinuous.

Proof. It is clear that (c) implies (b) implies (a). But (a) implies (c) by Theorem 5.10. □

Remark 7.4. A corollary of this statement is that for a barreled LCTVS E, its original topology is the same as the strong topology coming from its pairing with E^* (that is, the topology of uniform convergence on bounded subsets of E^*.) For the prepolar of every bounded (hence equicontinuous) subset of E^* is clearly a 0-neighborhood in E, and every closed balanced convex 0-neighborhood in E is the prepolar of its polar (which is a bounded subset of E^*).

Let E and F be locally convex topological vector spaces and let $T\colon E \to F$ be a continuous linear map. For any $\phi \in F^*$, the composite $\phi \circ T$ is a continuous linear map from E to $\Bbbk$. We denote this map by $T^*\phi$ and in this way we have obtained a mapping $T^*\colon F^* \to E^*$.

Definition 7.5. The map T^* defined above is called the *dual* of T. (Rudin calls it the *adjoint* but I prefer to keep this terminology for the Hilbert space case.)

Proposition 7.6. *In the above circumstances, $T^*\colon F^* \to E^*$ is a linear map, and it is continuous both for the weak-$*$ and for the strong topologies on the dual spaces (the same one on each side, of course!).*

Proof. The proof that T^* is linear is routine. To see the continuity, in either topology, let S be a subset of E. One clearly has

$$(T^*)^{-1}(S^\circ) = T(S)^\circ$$

since a functional $\phi \in F^*$ belongs to either side iff $|\phi(Tx)| \leqslant 1\ \forall x \in S$. Recall now that a basis for the 0-neighborhoods in the weak-$*$ topology is given by the polars of finite sets; in the strong topology, by the polars of bounded sets. Since T takes finite sets to finite sets (obviously) and bounded sets to bounded sets (5.4) we see that T^* is continuous in either of the topologies. ☐

Proposition 7.7. *Let E and F be LCTVS and let $T\colon E \to F$ be a linear map. If T is originally continuous (that is, continuous relative to the originally given topologies on E and F) then it is weakly continuous. The converse holds if E and F are barreled spaces.*

Proof. This is like the previous proof. Let S be a subset of F^*. We have

$$T^{-1}({}^\circ S) = {}^\circ T^*(S),$$

and T^* is weak-$*$ continuous (even if T is only weakly continuous). Now a basis for the 0-neighborhoods in the weak topology is given by the polars of finite sets; in the barreled case a basis for the 0-neighborhoods in the original topology is given by the polars of weak-$*$ bounded sets (Remark 7.4). Since T^* takes finite sets to finite sets (obviously) and bounded sets to bounded sets (5.4) we see that T is continuous in either of the topologies. ☐

Proposition 7.8. *Let $T\colon E \to F$ be a continuous linear map between LCTVS. We have the following identities*

$$\ker(T) = {}^\circ\operatorname{Im}(T^*), \quad \ker(T^*) = \operatorname{Im}(T)^\circ$$

(where we note that the polars, applied to subspaces, are really annihilators.)

Proof. Routine. For instance, let's check the first statement. We have $x \in \ker T$ if and only if $\phi(Tx) = 0$ for all $\phi \in F^*$ (since F^* separates points on F by Hahn-Banach). This is equivalent to the identity $(T^*\phi)(x) = 0$ for all $\phi \in F^*$, which says that x is in the (pre)annihilator of every element of $\mathrm{Im}(T^*)$. ☐

Corollary 7.9. *In the above circumstances, if* $\ker(T^*) = 0$, *then* $\mathrm{Im}(T)$ *is (originally) dense in* F*; if* $\ker(T) = 0$, *then* $\mathrm{Im}(T^*)$ *is weak-$*$ dense in* E^*.

For applications, e.g. to the solvability of PDE, one would often like to have a criterion in terms of the dual for when T is actually *surjective* (not simply having dense range). This motivates the study of *closed range theorems*.

Proposition 7.10. *Let E, F be LCTVS and $T\colon E \to F$ a continuous linear map.*

(a) In order that $\mathrm{Im}(T)$ be weakly closed (= originally closed by 6.5) in F it is necessary and sufficient that $T^*\colon F^* \to E^*$ be a homomorphism for the weak-$*$ topology.

(b) In order that $\mathrm{Im}(T^*)$ be weak-$*$ closed in E^* it is necessary and sufficient that $T\colon E \to F$ be a homomorphism for the weak topology.

(The definition of a homomorphism was given in 5.16.)

Proof. The proofs are identical (after interchanging spaces and their duals). Let us prove (b). Let $K \leqslant E$ be the kernel of T, which is the preannihilator of $\mathrm{Im}(T^*)$; so $K^\circ = \overline{\mathrm{Im}(T^*)}$. For T to be a (weak) homomorphism it is necessary and sufficient that the linear bijection u in the factorization

$$\begin{tikzcd}
E \arrow[r, two heads] \arrow[rrr, bend left, "T"] & E/K \arrow[r, "u"] & \mathrm{Im}(T) \arrow[r, tail] & F
\end{tikzcd}$$

be a homeomorphism, where E and F are equipped with their weak topologies. Let us compare the quotient topology on E/K with the topology pulled back via u from the subspace topology on $\mathrm{Im}(T)$.

- The quotient topology on E/K is the weak topology coming form its pairing with its dual space K° (see 7.1), that is, with $\overline{\mathrm{Im}(T^*)}$. If $\phi \in \overline{\mathrm{Im}(T^*)} \subseteq E^*$ then ϕ gives a well-defined linear functional on E/K and the quotient topology is the weakest making all these functionals continuous.

- On the other hand, the subspace topology of $\mathrm{Im}(T)$ is the weakest topology making all $\psi \in F^*$ continuous. Bearing in mind the definition of u, this induces on E/K the weakest topology making all the functionals $T^*\psi$ continuous, that is the weak topology coming from the pairing with $\mathrm{Im}(T)$ (not the closure).

Let $V = E/K$, $V^* = K^\circ = \overline{\mathrm{Im}(T^*)}$ its dual space, and let $W = \mathrm{Im}(T) \subseteq V^*$. For T to be a homomorphism it is necessary and sufficient that the $\sigma(V, W)$-topology on V be the same as the $\sigma(V, V^*)$-topology. But since the dual of the TVS $(V, \sigma(V, W))$ is W (Lemma 6.3) this happens iff $W = V^*$, that is, iff $\mathrm{Im}(T^*)$ is closed. ☐

Theorem 7.11. *Let E and F be Fréchet spaces. The following are equivalent for a continuous linear map $T\colon E \to F$.*

(a) $T\colon E \to F$ is a homomorphism for the original topology.

(b) $T\colon E \to F$ is a homomorphism for the weak topology.

(c) The subspace $\mathrm{Im}(T)$ is originally (= weakly) closed in F.

(d) $T^*\colon F^* \to E^*$ is a homomorphism for the weak-$*$ topology.

(e) The subspace $\mathrm{Im}(T^*)$ is weak-$*$-closed in E^*.

Proof. The previous proposition shows that (c) is equivalent to (d), and that (b) is equivalent to (e). It follows from the open mapping theorem that (a) is equivalent to (c) (Corollary 5.17). Finally we check that (a) is equivalent to (b). It is enough to prove that a map between Fréchet spaces is an *isomorphism* for the weak topology iff it is an isomorphism for the original topology; but this follows from Proposition 7.7. ☐

It is (apparently) unknown whether the "strong" versions of (d) and (e) can be added to the list in general. However, this *is* true if E, F are Banach spaces:

Proposition 7.12. *In theorem 7.11 suppose that E and F are in fact Banach spaces. Then the items (a)–(e) listed in the theorem are also equivalent to*

(f) The subspace $\mathrm{Im}(T^*)$ is closed for the norm topology on E^*,

(g) $T^*\colon F^* \to E^*$ is a homomorphism for the norm topologies on the dual spaces.

Proof. Clearly (e) implies (f), and (f) and (g) are equivalent by the open mapping theorem (5.17) again. Suppose (g). Apply Theorem 7.11 and the first part of the present proposition to T^* to deduce that $T^{**}\colon E^{**} \to F^{**}$ is a homomorphism for the norm topologies (this is the implication (a) ⇒ (g), applied to T^*). Now E is a complete subspace of E^{**}, and the image under a homomorphism of a complete space is complete. Thus $T(E)$ is complete, and it is a subspace of F which is closed in F^{**}, so $T(E)$ is closed in F. □

Lecture 8

The Fourier Transform

The next section of the course is devoted to *distribution theory*. This theory, which was given its present form by Laurent Schwarz in the 1950s (though the work of Sobolev in the 1930s is also closely related) arises from the desire to extend the domain of standard operations of linear partial differential equations—such as differentiation, convolution, or the Fourier transform—to encompass objects such as Dirac's delta "function" $\delta(x)$, which is supposed to have the properties that $\delta(x) = 0$ for all $x \neq 0$, while $\int \delta(x)dx = 1$. Of course no actual function has these properties, but distribution theory allows us to make sense of them as well as such apparently paradoxical statements as

$$H'(x) = \delta(x)$$

where $H(x)$ is the step function equal to 0 for $x \leqslant 0$ and 1 for $x > 0$, and

$$\int_{-\infty}^{\infty} e^{itx}\, dt = 2\pi\delta(x)$$

which turns out to be a version of the Fourier inversion formula.

Before we embark on generalizing these operations to the context of distributions, we shall review them in their classical context. Let f be a smooth function on $\mathbb{R}^n$. A *multi-index* α is an n-tuple $(\alpha_1, \ldots, \alpha_n)$ of nonnegative integers, and the notation D^α is short for the differential operator

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

I take it for granted that "mixed partials commute" (we will actually prove this later in a more general context). Similarly, if $x = (x_1, \ldots, x_n)$, the notation x^α will refer to the monomial

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

The notation X^α will refer to the *operator* (on a suitable vector space of functions) defined by multiplication by the function x^α. If P is a polynomial in n variables the notations $P(D)$ and $P(X)$ will be used in the obvious way. Finally, we define $|\alpha| = \alpha_1 + \cdots + \alpha_n$.

If $x = (x_1, \ldots, x_n)$ we will let $dx = dx_1 \ldots dx_n$ denote Lebesgue measure on $\mathbb{R}^n$. We won't introduce extraneous factors of 2π into the measure (this is a different convention from that used in Rudin's Chapter 7).

Definition 8.1. Let G be a locally compact abelian topological group. The *Pontrjagin dual* of G is the group

$$\widehat{G} = \hom(G, \mathbb{T}),$$

where $\mathbb{T}$ is the group of complex numbers of modulus 1.

The group operation is pointwise multiplication (in $\mathbb{T}$); the topology is the topology of uniform convergence on compact subsets. Elements of the Pontrjagin dual are called *characters* of G.

Exercise 8.2. Show that every character of $\mathbb{R}^n$ is of the form $x \mapsto e^{it\cdot x}$, where t is a fixed element of $\mathbb{R}^n$ and the dot denotes the usual inner product. Thus the dual group of $\mathbb{R}^n$ is another copy of $\mathbb{R}^n$. In similar terms identify the dual groups of $\mathbb{Z}^n$ and $\mathbb{T}^n$.

The *Pontrjagin duality theorem* says that the natural map $G \to \widehat{\widehat{G}}$ is an isomorphism for all locally compact abelian G. We won't discuss the general case, but simply note that the above exercise proves the theorem for the most important examples $\mathbb{R}^n$, $\mathbb{Z}^n$ and $\mathbb{T}^n$. In the rest of the discussion we will restrict attention to these cases, but bear in mind that many of the basic results of Fourier analysis generalize with little effort to any G.

Definition 8.3. Let f be an integrable function on $\mathbb{R}^n$. The *Fourier transform* of f is the function $\mathscr{F}(f) = \hat{f}$ on $\widehat{\mathbb{R}}^n$ defined by

$$\hat{f}(t) = (2\pi)^{-n/2} \int e^{-it\cdot x} f(x) dx.$$

By the Dominated Convergence Theorem , $\hat{f}$ is continuous (though it might not be integrable). Warning: The multiples of 2π are not standard. The theory has to have some 2π's in it somewhere, but different authors push them into different places.

As is well known, the Fourier transform "converts differentiation into multiplication". To formulate this statement accurately one needs some topological vector spaces in which differentiation and multiplication by polynomials appear on an equal footing. These are the *Schwarz spaces*.

Definition 8.4. A function f on $\mathbb{R}^n$ belongs to the *Schwarz class* if it is smooth and, for all multi-indices α and β,

$$p_{\alpha\beta}(f) := \sup\{|x^\alpha||D^\beta f(x)| : x \in \mathbb{R}^n\} < \infty.$$

The collection of all such functions forms a locally convex topological vector space with seminorms $p_{\alpha\beta}$. This is called the *Schwarz space* $\mathscr{S}(\mathbb{R}^n)$.

Exercise 8.5. Show that $\mathscr{S}(\mathbb{R}^n)$ is complete (so it is a Fréchet space).

Exercise 8.6. Show that for any polynomial P, the operators $P(D)$ and $P(X)$ are continuous $\mathscr{S}(\mathbb{R}^n) \to \mathscr{S}(\mathbb{R}^n)$. (Using the closed graph theorem saves some work here, but one can also do without it.)

Lemma 8.7. *The smooth function with compact support are dense in the Schwarz space $\mathscr{S}(\mathbb{R}^n)$.*

Proof. We need the existence of smooth bump functions: there exists a smooth function ϕ on $\mathbb{R}^n$ with $\phi(x) = 1$ for $|x| \leqslant 1$ and $\phi(x) = 0$ for $|x| \geqslant 2$. Granted this, let $f \in \mathscr{S}(\mathbb{R}^n)$ and let

$$f_n(x) = f(x)\phi(x/n).$$

The functions f_n are smooth and compactly supported. Let us estimate $p_{\alpha\beta}(f_n - f)$. We have

$$x^\alpha D^\beta(f_n - f) = x^\alpha \sum_{\gamma \leqslant \beta} c_{\beta\gamma}(D^{\beta-\gamma} f)(x) n^{-|\gamma|} D^\gamma[\phi - 1](x/n),$$

where the $c_{\beta\gamma}$ are certain constants arising from Liebniz' rule. The term $D^\gamma[\phi - 1](x/n)$ vanishes when $|x| < n$. But since $f \in \mathscr{S}$, the terms $x^\alpha(D^{\beta-\gamma} f)(x)$ tend to zero at infinity; thus the sum tends to zero (uniformly on $\mathbb{R}^n$) and we have shown that $p_{\alpha\beta}(f_n - f) \to 0$ as required. □

Proposition 8.8. *Let $f \in \mathscr{S}(\mathbb{R}^n)$. Then the Fourier transform $\hat{f}$ belongs to $\mathscr{S}(\widehat{\mathbb{R}}^n)$. Moreover, the identities*

$$\mathscr{F}(P(D)f) = P(i\widehat{X})\mathscr{F}(f), \quad \mathscr{F}(P(X)f) = P(i\widehat{D})\mathscr{F}(f)$$

are satisfied for any polynomial P. The Fourier transformation is a continuous linear map $\mathscr{F} \colon \mathscr{S}(\mathbb{R}^n) \to \mathscr{S}(\widehat{\mathbb{R}}^n)$.

Proof. Let $C_b(\widehat{\mathbb{R}}^n)$ denote the Banach space of bounded, continuous functions on $\widehat{\mathbb{R}}^n$. Certainly $\mathscr{F} \colon \mathscr{S}(\mathbb{R}^n) \to C_b(\widehat{\mathbb{R}}^n)$ is continuous. Suppose that $f \in \mathscr{S}(\mathbb{R}^n)$ and $g = \partial f/\partial x_k$. By integration by parts,

$$\hat{g}(t) = \int \frac{\partial f}{\partial x_k}(x)e^{-it\cdot x}dx = -\int f(x)(-it_k e^{it\cdot x})dx = it_k\hat{f}(t);$$

the identity $\mathscr{F}(P(D)f) = P(i\widehat{X})\mathscr{F}(f)$ follows by induction. Similarly,

$$\frac{\partial}{\partial t_k}\hat{f}(t) = \int f(x)(-ix_k e^{-it\cdot x})dx$$

by differentiating under the integral sign (justified by the dominated convergence theorem, since the derivative is uniformly bounded by an integrable function). In particular, $\hat{f}$ is differentiable. The identity $\mathscr{F}(P(D)f) = P(i\widehat{X})\mathscr{F}(f)$ now similarly follows by induction. Since we already know that the operators X^α and D^β are continuous on $\mathscr{S}(\mathbb{R}^n)$ (Exercise 8.6), we find that the maps

$$f \mapsto \widehat{X}^\alpha \widehat{D}^\beta \mathscr{F}(f) = i^{-|\alpha|-|\beta|}\mathscr{F}(D^\alpha X^\beta f)$$

are continuous $\mathscr{S}(\mathbb{R}^n) \to C_b(\widehat{\mathbb{R}^n})$. But this just says that $\mathscr{F}$ is continuous $\mathscr{S}(\mathbb{R}^n) \to \mathscr{S}(\widehat{\mathbb{R}^n})$. $\square$

From these results we can compute a useful Fourier transform.

Lemma 8.9. *Let $f \in \mathscr{S}(\mathbb{R}^n)$ be the Gaussian $f(x) = e^{-|x|^2/2}$. Then its Fourier transform is $\hat{f}(t) = e^{-|t|^2/2}$.*

Proof. It suffices to prove the 1-dimensional case (the n-dimensional version follows using Fubini's theorem). So, let $f(x) = e^{-x^2/2}$ be a Gaussian on $\mathbb{R}$. It satisfies the differential equation

$$df/dx + xf = 0.$$

Taking the Fourier transform and using the previous proposition, we see that $\hat{f}$ satisfies the differential equation

$$it\hat{f} + id\hat{f}/dt = 0$$

which is the same! The equation has general solution $Ce^{-t^2/2}$, and the constant C may be computed by evaluating $\hat{f}(0) = (2\pi)^{-1/2}\int e^{-x^2/2}dx = 1$ which is a standard integral. $\square$

Exercise 8.10. Shoe that $\mathscr{S}(\mathbb{R}^n)$ is dense in $L^1(\mathbb{R}^n)$. Use this to prove the *Riemann-Lebesgue Lemma*, which states that if $f \in L^1$, then $\hat{f}(t) \to 0$ as $|t| \to \infty$.

Exercise 8.11. If $f, g \in L^1(\mathbb{R}^n)$ their *convolution* $f * g$ is defined by

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

Show that the Fourier transform 'converts convolution into multiplication': the Fourier transform $\mathscr{F} f * g$ is equal to a constant times $\mathscr{F}(f)\mathscr{F}(g)$.

Definition 8.12. A sequence g_k in $\mathscr{S}(\mathbb{R}^n)$ is called an *approximate identity* if $g_n \geqslant 0$, $\int g_k = 1$, and for every $\delta > 0$ one has

$$\int_{|x|>\delta} g_k(x)\, dx \to 0$$

as $n \to \infty$.

Note that the sequence of functions $g_k(x) = (k/2\pi)^{n/2} e^{-kx^2/2}$ ("Gaussian bumps") forms an approximate identity. We will use these below. Later, we will recognize an approximate identity as a sequence in $\mathscr{S}(\mathbb{R}^n)$ that converges weakly to the Dirac delta distribution. In fact, that is the content of the next lemma.

Lemma 8.13. *Let $\{g_k\}$ be an approximate identity and suppose that $f \in \mathscr{S}(\mathbb{R}^n)$ (in fact, it is enough to suppose that f is continuous and bounded). Then*

$$\int g_k(x)f(x)dx \to f(0)$$

as $k \to \infty$.

Proof. Let a_k denote the left side of the display above. Fix x and let $\epsilon > 0$ be given. There is $\delta > 0$ such that $|f(x) - f(0)| < \epsilon$ if $|x| < \delta$, Then there is K such that $\int_{|u|>\delta} g_k(x)dx < \epsilon$ for $k > K$. Now write

$$a_k - f(0) = \int (f(x) - f(0))g_k(x)dx.$$

Thus

$$\begin{aligned}|a_k - f(0)| &\leqslant \int |f(x) - f(0)| g_k(x) dx \\ &= \int_{|x|<\delta} |f(x) - f(0)| g_k(x) dx + \int_{|x|>\delta} |f(x) - f(0)| g_k(x) dx \\ &\qquad\qquad\qquad\qquad\qquad\qquad \leqslant \epsilon + 2\epsilon \sup |f|.\end{aligned}$$

for $k > K$. It follows that $a_k \to f(0)$ as required. ☐

Exercise 8.14. Let g_k be an approximate identity. Show that for every $f \in \mathscr{S}$, $f * g_k \to f$ in $\mathscr{S}$ as $k \to \infty$.

Theorem 8.15. *(Fourier inversion theorem) Let $f \in \mathscr{S}(\mathbb{R}^n)$. Then*

$$f(x) = (2\pi)^{-n/2} \int e^{itx} \hat{f}(t) dt.$$

Consequently, the Fourier transformation $\mathscr{F} : \mathscr{S}(\mathbb{R}^n) \to \mathscr{S}(\widehat{\mathbb{R}}^n)$ is an isomorphism of topological vector spaces.

Proof. We want to evaluate

$$g(x) = (2\pi)^{-n/2} \int e^{it\cdot x} \hat{f}(t) dt.$$

Because $\hat{f}$ is integrable, we can use the dominated convergence theorem to write

$$g(x) = \lim_{k\to\infty} (2\pi)^{-n/2} \int e^{it\cdot x} e^{-t^2/2k} \hat{f}(t) dt.$$

Expand the definition of the Fourier transform in the integral appearing on the right hand side here to get

$$(2\pi)^{-n} \iint e^{it\cdot(x-y)-t^2/2k} f(y) dy dt.$$

The integrand belongs to $L^1(\mathbb{R}^{2n})$ (why?), so we may apply Fubini's theorem to rearrange it as

$$(2\pi)^{-n} \int \left(\int e^{it\cdot(x-y)-t^2/2k} dt \right) f(y) dy.$$

From lemma 8.9 the inner integral (the one in parentheses) is equal to $(2\pi k)^{n/2}e^{-k(x-y)^2/2}$, so the whole integral is

$$\int (k/2\pi)^{n/2}e^{-k(x-y)^2/2}f(y)dy.$$

But the sequence of functions $u \mapsto (k/2\pi)^{n/2}e^{-ku^2/2}$ is an approximate identity, so this tends to $f(x)$ as $n \to \infty$ (by Lemma 8.13). This completes the proof of the inversion formula.

As for the last statement of the theorem, we have already shown that $\mathscr{F}: \mathscr{S}(\mathbb{R}^n) \to \mathscr{S}(\widehat{R}^n)$ is a continuous linear map. Now we see that it has an inverse given by the formula

$$\widehat{\mathscr{F}}h(x) = (2\pi)^{-n/2}\int h(t)e^{it\cdot x}dt$$

which defines a continuous linear map $\mathscr{S}(\widehat{\mathbb{R}}^n) \to \mathscr{S}(\mathbb{R}^n)$ for exactly the same reasons. Consequently, both $\mathscr{F}$ and $\widehat{\mathscr{F}}$ are homeomorphisms. □

Theorem 8.16. *(Parseval-Plancherel) The Fourier transformation $\mathscr{F}$ extends to an* isometry *of the Hilbert space $L^2(\mathbb{R}^n)$ onto $L^2(\widehat{\mathbb{R}}^n$.*

Proof. It is easy to check that the Schwarz space $\mathscr{S}$ is dense in L^2. Thus, it suffices to prove that $\mathscr{F}: \mathscr{S}(\mathbb{R}^n) \to \mathscr{S}(\widehat{R}^n)$ is an isometry for the L^2-norm.

Let $f \in \mathscr{S}$ and write

$$\int |f(x)|^2dx = \int \bar{f}(x)\left((2\pi)^{-n/2}\int e^{it\cdot x}\hat{f}(t)dt\right)dx$$

by the Fourier inversion formula. Rearranging the double integral using Fubini's theorem gives

$$\int \hat{f}(t)\left((2\pi)^{-n/2}\int e^{it\cdot x}\bar{f}(x)dx\right)dt = \int |\hat{f}(t)|^2dt$$

since the inner integral on the left hand side is the complex conjugate of $\hat{f}(t)$. □

Lecture 9

Distributions

Throughout the present discussion, Ω will denote a fixed open subset of the Euclidean space $\mathbb{R}^n$. (With a little extra technicality, Ω could be any σ-compact n-dimensional manifold; we may say more about this later.) λ will denote Lebesgue measure on Ω (in the manifold case, λ denotes any measure, fixed once and for all, belonging to the smooth measure class.)

Recall that $\mathscr{D}(\Omega)$ denotes the space of smooth functions $\Omega \to \mathbb{C}$ that have compact support. This space is naturally topologized as an inductive limit of Fréchet spaces (a so-called *LF space*); let $K_1 \subseteq K_2 \subseteq \ldots \subseteq \Omega$ be an increasing sequence of compact subsets whose union in Ω, and regard $\mathscr{D}(\Omega)$ as the inductive limit (2.11) of the sequence

$$\mathscr{D}_{K_1}(\Omega) \to \mathscr{D}_{K_2}(\Omega) \to \cdots$$

where $\mathscr{D}_K(\Omega)$ is the Fréchet space of smooth functions on Ω whose support lies in K. Recall that the inductive limit topology has the following properties (see Proposition 2.12):

(a) The subspace topology that it induces on any $\mathscr{D}_K(\Omega)$ is the same as the original Fréchet topology of $\mathscr{D}_K(\Omega)$. All such subspaces are closed.

(b) A net f_i converges in $\mathscr{D}(\Omega)$ if and only if it belongs to some $\mathscr{D}_K(\Omega)$ and converges there.

(c) A subset A of $\mathscr{D}(\Omega)$ is closed if and only if $A \cap \mathscr{D}_{K_n}(\Omega)$ is closed in $\mathscr{D}_{K_n}(\Omega)$ for all n.

(d) $\mathscr{D}(\Omega)$ is a barreled space.

Exercise 9.1. Prove that the only bounded sets in $\mathscr{D}(\Omega)$ are the bounded subsets of some $\mathscr{D}_K(\Omega)$, where K is a compact set.

From the definition of the topology on $\mathscr{D}(\Omega)$, we see that a linear map of $\mathscr{D}(\Omega)$ to another locally convex space F is continuous if and only if its restriction to each $\mathscr{D}_K(\Omega)$ is continuous. From the previous exercise, we also see that this takes place if and only if the linear map is bounded.

In particular, the differential operators D^α, together with the operators of multiplication by smooth functions on Ω, are continuous as maps $\mathscr{D}(\Omega) \to \mathscr{D}(\Omega)$. It follows that *every* linear differential operator is continuous as a map on $\mathscr{D}(\Omega)$. (This latter statement makes sense even if Ω is a manifold.)

Definition 9.2. The dual space $\mathscr{D}'(\Omega)$ is called the space of *distributions* on Ω. It is a locally convex space (which we usually equip with its strong topology).

Exercise 9.3. Show that closed, bounded subsets of $\mathscr{D}(\Omega)$ are compact. Thus, the strong topology on $\mathscr{D}'(\Omega)$ is the same as the topology of uniform convergence on compact sets.

Explicitly, then, a distribution is a linear functional Λ, defined on compactly supported smooth functions on Ω, that has the following property: for every compact subset K of Ω there exist a constant C_K and an integer N_K such that

$$|\Lambda(f)| \leqslant C_K \sup\{|D^\alpha f(x)| : x \in K, |\alpha| \leqslant N_K\}$$

for all smooth f having support within K. If the integers N_K are bounded, the distribution Λ is said to be of *finite order* and the least upper bound of the N_K is the *order* of the distribution.

We can regard functions as a special kind of distributions. Consider the Hilbert space $H = L^2(\Omega, \lambda)$. The inclusion map $j\colon \mathscr{D}(\Omega) \to H$ is continuous and injective, with dense range. By Proposition 7.8 it follows that the dual map $j^*\colon H^* \to \mathscr{D}'(\Omega)$ is also injective with dense range. But the map sending $f \in H$ to the linear functional $\phi_f(g) = \int fg d\lambda$ is an isomorphism (the Riesz representation theorem for Hilbert spaces). Combining these facts we obtain a continuous injection with weak-$*$ dense range

$$\mathscr{D}(\Omega) \to H \to H^* \to \mathscr{D}'(\Omega).$$

Let's state this as a proposition:

Proposition 9.4. *Each function $f \in \mathscr{D}(\Omega)$ may be identified with the distribution*

$$\Lambda_f(g) = \int f(x)g(x)\, d\lambda(x).$$

This identification makes $\mathscr{D}(\Omega)$ a weak-$*$ dense subspace of $\mathscr{D}'(\Omega)$.

The density tells us that every distribution is a limit of smooth functions (in the weak-$*$ topology of $\mathscr{D}'(\Omega)$, which is of course a very weak one). This statement is in fact true in the strong topology (a more powerful density result), but to prove that seems to require a more explicit calculation.

Remark 9.5. The same formula as above makes sense when f is merely a *locally integrable* function on Ω, and it defines a distribution.

Example 9.6. The *Dirac delta function* at a point $p \in \Omega$ is the distribution $\delta_p(f) = f(p)$.

Exercise 9.7. Verify that δ_p is a distribution. Use an approximate identity (Lemma 8.13) to construct an explicit sequence of elements of $\mathscr{D}(\Omega)$ that tends to δ_p.

Example 9.8. The *dipole* at a point $p \in \mathbb{R}$ is the distribution $\delta'(p) \in \mathscr{D}'(\mathbb{R})$ defined by $\delta'_p(f) = -f'(p)$.

Exercise 9.9. Construct an explicit sequence of elements of $\mathscr{D}(\Omega)$ converging to δ'_p.

Example 9.10. (Cauchy principal values) Consider the "integral"

$$\int_{-\infty}^{\infty} \frac{f(x)}{x}\,dx$$

where $f \in \mathscr{D}(\mathbb{R})$. This integral does not make sense in the conventional fashion because $f(x)/x$ does not belong to L^1 (unless $f(0) = 0$). Cauchy introduced the idea of considering the "principal value"

$$PV \int \frac{f(x)}{x}\,dx = \lim_{\epsilon \to 0} \left(\int_{-\infty}^{-\epsilon} + \int_{\epsilon}^{\infty} \frac{f(x)}{x}\,dx \right)$$

Integration by parts gives

$$\int_{\epsilon}^{\infty} \frac{f(x)}{x}\,dx = [\log(x) f(x)]_{\epsilon}^{\infty} - \int_{\epsilon}^{\infty} \log(x) f'(x) dx.$$

When we add this formula to the corresponding one for the interval $[-\infty, -\epsilon]$, the "square bracket" terms cancel in the limit and we get

$$PV \int \frac{f(x)}{x}\,dx = -\int \log|x| f'(x) dx.$$

The right hand side here is a well defined distribution (of order 1), since $x \mapsto \log|x|$ is a locally integrable function. Thus the Cauchy principal value ”regularizes” the divergent integral as a distribution.

Example 9.11. Any Borel measure μ on Ω can be regarded as a distribution: the functional $f \mapsto \int f d\mu$.

We’re now going to extend the operations of multiplication (by a smooth function) and differentiation from the dense subspace $\mathscr{D}(\Omega)$ to the whole space of distributions $\mathscr{D}'(\Omega)$. Note that since $\mathscr{D}(\Omega)$ is dense in $\mathscr{D}'(\Omega)$, there can be only one such extension (as a continuous linear map).

Proposition 9.12. *For any smooth function ϕ on Ω, and for any multi-index α, there exist (strongly) continuous linear maps $M_\phi\colon \mathscr{D}'(\Omega) \to \mathscr{D}'(\Omega)$ and $D^\alpha\colon \mathscr{D}'(\Omega) \to \mathscr{D}'(\Omega)$ that extend the operations of multiplication by ϕ and differentiation with multi-index α defined on the dense subspace $\mathscr{D}(\Omega)$ of $\mathscr{D}'(\Omega)$.*

Proof. Define these operations in terms of the duals of the corresponding operations on spaces of functions. Thus, for example, D^α on distributions is defined to be $(-1)^{|\alpha|}$ times the dual of the operation D^α defined on $\mathscr{D}(\Omega)$. The identity

$$\Lambda_{D^\alpha f}(g) = \int (D^\alpha f) g d\lambda = (-1)^{|\alpha|} \int f (D^\alpha g) d\lambda = (-1)^{|\alpha|}((D^\alpha)^* \Lambda)(g),$$

which comes from integrating by parts $|\alpha|$ times, shows that this definition of D^α on distributions agrees with the existing definition of D^α on functions. Being defined as a dual, D^α is strongly continuous.

The construction for M_ϕ is similar, but easier: we define the operation M_ϕ on distributions simply to be the dual of the corresponding operation on functions. □

Remark 9.13. The consistency statement in this proposition applies to smooth functions only. Beware that if we assume only that f is locally integrable, and that it is differentiable a.e. with derivative that is also locally integrable, then the distribution defined by the a.e. derivative of f may not equal the distributional derivative of the distribution defined by f. A simple example of this phenomenon is the Heaviside (unit step) function $H(x)$ on $\mathbb{R}$: this is differentiable a.e., with derivative 0, but the *distributional* derivative of H is the delta function.

At least in one variable it is true that if $f \in L^1$ is differentiable *everywhere* with L^1 derivative, then its distributional and ordinary derivatives agree: but to prove the result in this generality requires some tricky measure theory (basically the Banach-Zarecki theorem: a function satisfying the listed conditions has to be absolutely continuous).

Let $\mathscr{S}(\mathbb{R}^n)$ denote the Schwarz space defined in the preceding lecture. Clearly, $\mathscr{D}(\mathbb{R}^n)$ may be regarded as a dense subspace of $\mathscr{S}(\mathbb{R}^n)$. By duality, $\mathscr{S}'(\mathbb{R}^n)$, the dual space of $\mathscr{S}(\mathbb{R}^n)$, may be regarded as a subspace of $\mathscr{D}'(\mathbb{R}^n)$. Notice that $\mathscr{S}(\mathbb{R}^n)$ is dense in $\mathscr{S}'(\mathbb{R}^n)$.

Definition 9.14. $\mathscr{S}'(\mathbb{R}^n)$ is called the space of *temperate* (or *tempered*, depending on which book you read) distributions on $\mathbb{R}^n$.

Proposition 9.15. *There exists a continuous linear map* $\mathscr{F} : \mathscr{S}'(\mathbb{R}^n) \to \mathscr{S}'(\widehat{\mathbb{R}}^n)$ *which extends the Fourier transform* $\mathscr{S}(\mathbb{R}^n) \to \mathscr{S}(\widehat{\mathbb{R}}^n)$.

Proof. This is done by duality again. Temporarily identify $\mathbb{R}^n$ with $\widehat{\mathbb{R}}^n$. One way to state the Plancherel-Parseval theorem is then: for $f, g \in \mathscr{S}(\mathbb{R}^n)$,

$$\int \mathscr{F}(f)(x)g(x)d\lambda(x) = \int f(y)\mathscr{F}(g)(y)d\lambda(y).$$

This shows that if we define the operation $\mathscr{F}$ on temperate distributions simply to be the dual of the operation $\mathscr{F}$ on $\mathscr{S}(\mathbb{R}^n)$, the operation so defined will be (strongly) continuous and will extend the definition of $\mathscr{F}$ on $\mathscr{S}(\mathbb{R}^n)$. Such an extension is necessarily unique, so our choice of identification of $\mathbb{R}^n$ with $\widehat{\mathbb{R}}^n$ is irrelevant. ☐

Example 9.16. The locally integrable function $x \mapsto e^{isx}$ may be considered as a temperate distribution on $\mathbb{R}$. What is its Fourier transform? By definition, its Fourier transform is the distribution Λ which has the property that

$$\Lambda(f) = \int e^{isx}(\mathscr{F}f)(x)dx = (2\pi)^{1/2}f(s)$$

using the Fourier inversion theorem. Thus, $\Lambda = (2\pi)^{1/2}\delta_s$. Taking $s = 0$ and writing out the definition of the Fourier transform as an integral gives the mysterious identity

$$\int_{-\infty}^{\infty} e^{-itx}dx = 2\pi\delta(t)$$

which we mentioned at the beginning.

Returning for a moment to the general case, let $\Lambda \in \mathscr{D}'(\Omega)$ be a distribution and let $U \subseteq \Omega$ be open. If $\Lambda(f) = 0$ for all f supported within U, we say that Λ *vanishes on* U.

Proposition 9.17. *Let Ω be as above, let $\mathscr{U} = \{U_\alpha\}_{\alpha \in \mathscr{A}}$ be any family of open subsets of Ω and suppose that $\Lambda \in \mathscr{D}'(\Omega)$ vanishes on each of the sets U_α. Then Λ vanishes on their union $U := \bigcup_{\alpha \in \mathscr{A}} U_\alpha$.*

Proof. We need partitions of unity. A *locally finite partition of unity* subordinated to the cover $\mathscr{U}$ is a collection $\{\phi_\beta\}$ of smooth, compactly supported functions $\Omega \to [0,1]$ having the following properties:

(i) Each set $K_\beta := \mathrm{Support}(\phi_\beta)$ is a subset of some member of $\mathscr{U}$.

(ii) The sets $\{K_\beta\}$ form a *locally finite system*, which means that each $x \in \Omega$ has a neighborhood meeting only finitely many of the K_β.

(iii) The sum $\sum_\beta \phi_\beta(x)$ (which is well defined by (b)) is identically equal to 1 on U.

It is known that partitions of unity exist subordinated to any given open cover $\mathscr{U}$. Granted this construction the proof is easy: Let $f \in \mathscr{D}(\Omega)$. By local finiteness, the (compact) support of f meets the support of only finitely many of the ϕ_β. Thus we may write $f = \sum_\beta f\phi_\beta$, where only finitely many terms are nonzero, and then $\Lambda(f) = \sum_\beta \Lambda(f\phi_\beta)$. Because each term $f\phi_\beta$ is supported in some U_α, all the terms here are zero. □

Let Λ be a distribution. From the above proposition we can deduce that there is a *largest* open set on which Λ vanishes — namely, the union of all the open sets that have this property.

Definition 9.18. The complement of the largest open set on which Λ vanishes is called the *support* of Λ.

Exercise 9.19. Let $\mathscr{U}$ be a family of open sets in $\mathbb{R}^n$, with union U. Let $B_j = \overline{B}(p_j; r_j)$ be an enumeration of all the closed balls with rational centers and radii that lie in some member of $\mathscr{U}$, and let $V_j = B(p_j; r_j/2)$ be open balls of the same centers and half the radii. For each j let $\psi_j \colon \Omega \to [0,1]$ be a smooth function equal to 1 on V_j and vanishing outside B_j. Define

$$\phi_1 = \psi_1,\ \phi_2 = (1-\psi_1)\psi_2,\ \phi_3 = (1-\psi_1)(1-\psi_2)\psi_3, \ldots$$

and so on. Prove that the ϕ_j are a locally finite partition of unity subordinated to $\mathscr{U}$.

Lecture 10

Distributions and Supports (continued)

At the end of the last lecture we defined the *support* of a distribution: the support of Λ is the complement of the largest open set on which Λ vanishes. Notice that a weak limit of distributions vanishing on U is also a distribution vanishing on U. Thus, the collection of distributions supported within a given closed set K is a *closed* subspace of $\mathscr{D}'(\mathbb{R}^n)$.

A basic result about supports is the following.

Proposition 10.1. *A distribution whose support consists of a single point is a linear combination of derivatives of the Dirac δ distribution at that point.*

Proof. Let Λ be a distribution supported at the origin in $\mathbb{R}^n$, say. Since Λ has compact support, it has finite order, say N. We are going to show that if $f \in \mathscr{D}(\mathbb{R}^n)$ with $D^\alpha f(0) = 0$ for all $|\alpha| \leqslant N$, then $\Lambda(f) = 0$. This will show that $\operatorname{Ker}(\Lambda) \supseteq \bigcap_{|\alpha|\leqslant N} \operatorname{Ker} D^\alpha\delta$ and therefore (by "standard linear algebra") that Λ is a linear combination of the functionals $D^\alpha\delta$ for $|\alpha| \leqslant n$.

If the derivatives of f vanish up to order N, then from Taylor's theorem we have $|f(x)| = O(|x|^{N+1}$ and, more generally, $|D^\beta f(x)| = O(|x|^{N+1-|\beta|})$ for $|\beta| \leqslant N$. Let ϕ be a smooth bump function, equal to 1 in a neighborhood of 0 and supported in the unit ball, and let $\phi_n(x) = \phi(nx)$. Since Λ has support $\{0\}$, we have $\Lambda(f) = \Lambda(\phi_n f)$ for all n. On the other hand, for $|\alpha| \leqslant N$ we have

$$D^\alpha(\phi_n f)(x) = \sum_{\beta+\gamma=\alpha} c_{\beta\gamma} D^\beta f(x) n^{|\gamma|}(D^\gamma\phi)(nx)$$

where the $c_{\beta\gamma}$ are certain constants. Consider the term under the summation sign. It is supported in the ball of radius $1/n$, and thus it is bounded by a multiple of

$$n^{-N-1+|\beta|} \cdot n^{|\gamma|}.$$

Since $|\beta| + |\gamma| = |\alpha| \leqslant N$, this tends to zero as $n \to \infty$. We conclude (since Λ has order N) that $\Lambda(\phi_n f) \to 0$ as $n \to \infty$. Hence, $\Lambda(f) = 0$. $\square$

Example 10.2. (Borel's theorem) A theorem of Borel states that every formal power series in n variables (without regard for any questions of convergence) is the Taylor series expansion at the origin of some function in $\mathscr{D}(\mathbb{R}^n)$. Let's prove this (in one variable for simplicity). We will make the space $F = \mathbb{C}[[z]]$ of formal power series into a Fréchet space by taking as seminorms the functionals

$$p_n\left(\sum_{k=0}^{\infty} a_k z^k\right) = |a_n|.$$

It is easy to see that the dual F^* is the space $\mathbb{C}[w]$ of polynomials, which is paired with F by

$$\langle w^m, z^n \rangle = \delta_{mn}.$$

Consider the map $T\colon \mathscr{D}_{[-1,1]} \to F$ of Fréchet spaces that sends a smooth function to its Taylor expansion at the origin (considered as a formal power series). Each polynomial function is its own Taylor expansion, and the polynomials are dense in F, so the image of T is clearly dense in F. It suffices then to prove that the image of T is *closed* in F. By Theorem 7.11, this will be true if and only if the image of T^* is weakly-$*$ closed in $\mathscr{D}^*$ (that is, if it is a weakly closed space of distributions). Now Taylor's theorem tells us that the coefficient of z^n in the Taylor series of f is $f^{(n)}(0)/n!$; this translates to say

$$T^*(w^n) = (n!)^{-1}\delta^{(n)}$$

where δ denotes the δ-distribution at the origin. Thus, $\mathrm{Im}(T^*)$ is the space of distributions spanned by δ and its derivatives. By Proposition 10.1 this coincides with the space of distributions having support $\{0\}$. Thus, it is weakly closed as required.

Example 10.3. (Characterization of the sine function) Let f_n, $n \in \mathbb{Z}$, be smooth functions on $\mathbb{R}$, each of which is the derivative of the preceding one. Suppose that the $|f_n(x)|$ are all bounded by a fixed constant C (independent of n and x) or even by a fixed polynomial in $|x|$ (independent of n). Then $f_0(x) = a\sin(x + c)$ for some constants a and c.

To prove this notice that the condition implies that the $\{f_n\}$, considered as temperate distributions, form an *equicontinuous* set. Consequently, their Fourier transforms also form an equicontinuous set of temperate distributions. It suffices therefore to prove the following statement: if $\{\phi_n\}$ is an equicontinuous set of temperate distributions on $\mathbb{R}$, having $\phi_{n+1} = X\phi_n$ for all n, then each of the ϕ_n has support in $\{-1, 1\}$. (For if the Fourier transform of f has support in $\{-1, 1\}$, then it is a linear combination of derivatives of the δ-functions at these two points. It follows that $f(x) = p_1(x)e^{ix} + p_{-1}(x)e^{-ix} = q(x)\sin x + r(x)\cos x$, where

p, q, r are polynomials. If f is bounded, then both q and r have degree zero, so $f(x) = q \sin x + r \cos x = (q^2 + r^2)^{1/2} \sin(x + c)$, where $\tan c = q/r$.)

Indeed, $U = (a, b)$ with $b > a > 1$. If $f \in \mathscr{D}(U)$ then it is easy to see that $X^{-n} f \to 0$ in $\mathscr{D}(U)$ (hence in $\mathscr{S}(U)$) as $n \to \infty$. Consequently

$$\phi_0(f) = \phi_n(X^{-n} f) \to 0$$

by equicontinuity. We deduce that $\phi_0(f) = 0$, so $\operatorname{Support}(\phi_0)$ does not meet (a, b). Exactly the same argument shows that $\operatorname{Support}(\phi_0)$ does not meet $(-b, -a)$ with $-a < -1$. Finally, to show that $\operatorname{Support}(\phi_0)$ does not meet (a, b) with $-1 < a < b < 1$, we can consider

$$\phi_0(f) = \phi_{-n}(X^n f)$$

and use the fact that in this case $X^n f \to 0$ in $\mathscr{S}(\mathbb{R})$ as $n \to \infty$.

Exercise 10.4. Let Λ be a distribution on $\mathbb{R}$ and suppose that the distributional derivative Λ' is equal to 0 (as a distribution). Prove that there is a constant C such that $\Lambda(f) = C \int f(x) dx$. In other words, Λ "is" the constant function C.

We will now discuss *convolution* of distributions. First, recall the corresponding operation for functions. If $f, g \in L^1(\mathbb{R}^n)$ then the integral

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

exists almost everywhere and defines an integrable function of x, called the *convolution* of f and g. The operation of convolution is commutative and associative, and $\|f * g\|_1 \leqslant \|f\|_1 \|g\|_1$. If both of f, g are smooth and compactly supported, so is $f * g$.

Now if f is a *distribution* and $g \in \mathscr{D}(\mathbb{R}^n)$, the integral above still makes sense when considered as the result of applying the *distribution* f to the smooth *function* $y \mapsto g(x - y)$. We call the result the *convolution* of f and g and note that it is a *function* of x.

Proposition 10.5. *Let $f \in \mathscr{D}'(\mathbb{R}^n)$, $g \in \mathscr{D}(\mathbb{R}^n)$. Then the convolution $f * g$ is a smooth function on $\mathbb{R}^n$, and it is compactly supported if the distribution f is compactly supported. Moreover, if also $h \in \mathscr{D}(\mathbb{R}^n)$, we have associativity: $(f * g) * h = f * (g * h)$.*

Proof. Let e be a unit vector in $\mathbb{R}^n$ and let

$$g_t(z) = \frac{g(z+te) - g(z)}{t}, \quad t \in \mathbb{R}^+.$$

By Taylor's theorem with remainder

$$|g_t(z) - D_e g(z)| \leqslant M|t|$$

where M is a constant (half the supremum of the second derivatives). Thus $g_t \to D_e g$ uniformly as $t \to 0$. Similar arguments show that all derivatives of g_t tend uniformly to the corresponding derivatives of $D_e g$ as $t \to 0$, so that in fact $g_t \to D_e g$ in $\mathscr{D}$. Now it follows that $(f*g(x+te) - f*g(x))/t = f*g_t(x) \to f*D_e g(x)$ as $t \to 0$, so $f*g$ is differentiable in the e-direction with derivative $f*D_e$. Iterating this argument shows that $f*g$ is smooth.

The statement about supports is clear: in fact, $\mathrm{Support}(f*g) \subseteq \mathrm{Support}(f) + \mathrm{Support}(g)$.

For the proof of associativity, we use a similar idea of "finite approximations". To keep notation simple work in the case $n=1$. Then the key lemma is

Lemma 10.6. *Let $h, k \in \mathscr{D}(\mathbb{R})$. For $p \in \mathbb{N}$ define*

$$\Phi_p^{h,k}(x) = \sum_{q=-\infty}^{\infty} p h(x - q/p) k(q/p).$$

(The sum is finite.) Then $\Phi_p^{h,k} \to h*k$ in $\mathscr{D}(\mathbb{R})$ as $p \to \infty$.

Assume this result, and assume also that f is compactly supported. Then we have

$$\begin{aligned}
f*(g*h)(x) &= \lim_{p\to\infty} \int f(y)\Phi_p^{g,h}(x-y)dy \\
&= \lim_{p\to\infty} \sum_{q=-\infty}^{\infty} h(q/p) \int f(y) g(x-y-q/p)dy \\
&= \lim_{p\to\infty} \Phi_p^{f*g,h}(x) = (f*g)*h(x)
\end{aligned}$$

This gives the result for compactly supported f. For general f, choose a smooth compactly supported function ϕ equal to 1 on the unit ball, and let $\phi_m(x) = \phi(x/m)$. Then $\phi_m f \to f$ in the strong topology on distributions as $m \to \infty$. It follows that $(\phi_m f)*(g*h) \to f*(g*h)$, and $((\phi_m f)*g)*h \to f*(g*h)$, the convergence being uniform on compact sets in both instances. The result for general f therefore follows from the result for compactly supported f. $\square$

We need to finish the proof of Lemma 10.6. To do this, note that the sum on the right is an approximation (coming from the trapezoidal rule) to the integral $\int h(x-y)k(y)dy$ which gives the convolution $h*k$. From the standard error estimate for the trapezoidal rule,

$$\sup|\Phi_p^{h,k} - h*k| \leqslant Cp^{-2}$$

where C is a constant depending on the size of the supports of h and k and on the bounds for the derivatives of order $\leqslant 2$ of these functions. In particular, $\Phi_p^{h,k} \to h*k$ uniformly. Since $D^n\Phi_p^{h,k} = \Phi_p^{D^nh,k}$ and $D^n(h*k) = (D^nh)*k$, this tells us that $D^n\Phi_p^{h,k} \to D^n(h*k)$ uniformly, so in fact $\Phi_p^{h,k} \to h*k$ in $\mathscr{D}(\mathbb{R})$ as asserted.

Exercise 10.7. Let $g \in \mathscr{D}(\mathbb{R}^n)$ be any smooth, positive function of compact support having $\int g = 1$, and let $g_k(x) = k^n g(kx)$. (Thus the $\{g_k\}$ form an approximate identity as in Definition 8.12.) Show that for any distribution f the smooth functions $f*g_k$ converge to f in the strong topology of $\mathscr{D}'(\mathbb{R}^n)$. (This proves our earlier assertion that $\mathscr{D}(\mathbb{R}^n)$ is *strongly* (not just weakly) dense in $\mathscr{D}'(\mathbb{R}^n)$.)

Example 10.8. The distribution δ at the origin is the identity for convolution ($\delta * f = f$). More generally, convolution by the derivative $D^\alpha\delta$ is equal to the differential operator D^α: that is,

$$(D^\alpha\delta)*f = \delta*(D^\alpha f) = D^\alpha f.$$

It follows that constant-coefficient differential operators are simply convolutions by distributions with support $\{0\}$.

If f is a fixed distribution of compact support, the operation "convolution by f" defines a continuous linear map $\mathscr{D} \to \mathscr{D}$. We can dualize this linear map, obtaining a continuous linear map $\mathscr{D}' \to \mathscr{D}'$. We regard this as an operation of *convolution by* f on *distributions*: if f is a compactly supported distribution and g is any distribution, then $f*g$ is defined as a distribution.

Remark 10.9. Notice that we now have *two* definitions of what it means to convolve a (compactly supported) distribution f by a (compactly supported) function g. The first is that provided by Proposition 10.5. The second is to regard g as a distribution and consider the operation "convolution by g" on *distributions*, defined by duality as in the previous paragraph, and apply it to the distribution f. *A priori*, this yields another distribution: however, it is not hard to check that this distribution is in fact defined by a smooth function, which is the same smooth function as is produced by the first method.

Proposition 10.10. *The operation of convolution defines separately continuous bilinear maps*

$$\mathscr{E} \times \mathscr{D} \to \mathscr{D}, \quad \mathscr{D}' \times \mathscr{D} \to \mathscr{E}, \quad \mathscr{E}' \times \mathscr{D}' \to \mathscr{D}'.$$

Proof. The first result (which doesn't involve distributions at all) follows from standard estimates on integrals.

Consider the second result. For fixed $g \in \mathscr{D}$ and $N \in \mathbb{N}$, and any compact $K \subseteq \mathbb{R}^n$, the functions $y \mapsto D^\alpha g(x-y)$, $|\alpha| \leqslant N$, $x \in K$, form a bounded subset of $\mathscr{D}$. It follows that for any $\epsilon > 0$ there is a 0-neighborhood U in $\mathscr{D}'$ (for the strong topology) such that, for all $f \in U$,

$$\sup\{|D^\alpha f * g(x)| : x \in K\} = \sup\{|f * (D^\alpha g)(x)| : x \in K\} < \epsilon.$$

These suprema are the seminorms defining the topology of $\mathscr{E}$. This shows that for fixed g, $f * g$ depends continuously on the distribution f (in the strong dual topology).

Similarly, the maps $\tau_x \colon \mathscr{D} \to \mathscr{D}$ defined by $\tau_x g(y) = g(x-y)$ form an equicontinuous set when x varies in a compact subset of $\mathbb{R}^n$. Therefore, for fixed f, and a fixed compact K, we can find a 0-neighborhood V in $\mathscr{D}$ such that if $g \in V$ then

$$\sup\{|D^\alpha f * g(x)| : x \in K\} = \sup\{|f(\tau_x(D^\alpha g))| : x \in K\} < \epsilon.$$

This shows that for fixed f, $f * g$ depends continuously on g.

The third result comes from the second by duality. □

Exercise 10.11. Verify that the convolution of distributions is associative if all the distributions involved have compact supports. Comment on the following example: let I be the constant function 1 (considered as a distribution), let H be the Heaviside function (also considered as a distribution), and let δ' be the derivative of the delta function at the origin. Verify that $1 * (\delta' * H) = \delta$, whereas $(1 * \delta') * H = 0$.

Remark 10.12. We can also define the convolution $f * g$ when $f \in \mathscr{S}'(\mathbb{R}^n)$, $g \in \mathscr{S}(\mathbb{R}^n)$; in that case the convolution is a smooth function and has polynomial growth at worst (so it can be considered as a temperate distribution). It is easy to see that the convolution need *not* belong to $\mathscr{S}(\mathbb{R}^n)$ in general, but it will do so if the distribution f has compact support.

Lemma 10.13. *Let $f \in \mathscr{S}'(\mathbb{R}^n)$, $g \in \mathscr{S}(\mathbb{R}^n)$, as above. Then the Fourier transform of $f * g$, considered as a temperate distribution, satisfies*

$$\mathscr{F}(f * g) = (2\pi)^{n/2}\mathscr{F}(g)\mathscr{F}(f).$$

The power of 2π is an artefact of our sign conventions. The expression on the right is to be interpreted in terms of the operation of multiplication of a temperate distribution by a Schwarz-class function, which is the dual of the operation of pointwise multiplication of Schwarz-class functions by a fixed Schwarz-class function: compare Proposition 9.12.

Lecture 11

Fourier Series and Local Solvability

We are now going to relate the theory of *Fourier series* to the distributional theory of the Fourier transform that we have developed. We identify the circle $\mathbb{T}$ with $\mathbb{R}/2\pi\mathbb{Z}$. The theory of Fourier series concerns the expansion of smooth functions on $\mathbb{T}$ (or $\mathbb{T}^n$) in terms of the characters $x \mapsto e^{ik\cdot x}$ of the group $\mathbb{T}^n$ (here $k \in \mathbb{Z}^n$).

Via the projection map $\mathbb{R}^n \to \mathbb{T}^n$ a function on the torus $\mathbb{T}^n$ can be identified with a *periodic* function on $\mathbb{R}^n$. To keep our notation straight we will use $\tilde{f}$ to denote the periodic function corresponding to the function f on the torus. Even if f (and therefore $\tilde{f}$) is smooth, $\tilde{f}$ does not belong to $\mathscr{S}(\mathbb{R}^n)$; however, being uniformly bounded, it can be identified with a linear functional on $\mathscr{S}(\mathbb{R}^n)$, that is, with a temperate distribution.

A *distribution* on $\mathbb{T}^n$ is a linear functional on the space $\mathscr{D}(\mathbb{T}^n) = \mathscr{E}(\mathbb{T}^n)$ of smooth functions on the torus (since $\mathbb{T}^n$ is a compact manifold, issues of supports do not arise). As with functions, a distribution Λ on $\mathbb{T}^n$ can be identified with a (periodic) distribution $\tilde{\Lambda}$ on $\mathbb{R}^n$. This (temperate) distribution is characterized as follows using the fact that $\pi \colon \mathbb{R}^n \to \mathbb{T}^n$ is a covering map: any point in $\mathbb{R}^n$ has an open neighborhood U such that the restriction of π to a map $U \to \pi(U)$ is a diffeomorphism: for $f \in \mathscr{D}(U)$ we define

$$\tilde{\Lambda}(f) = \Lambda(f \circ (\pi_{|U})^{-1}).$$

This defines $\tilde{\Lambda}$ locally. For a global definition, use a partition of unity to decompose a compactly supported function f into a finite sum of functions to each of which the local definition applies.

The key to understanding Fourier series is the following definition.

Definition 11.1. The *Dirichlet kernel* (in one dimension) is the sequence of smooth functions on $\mathbb{T}$ defined by

$$D_N(x) = \sum_{k=-N}^{N} e^{ikx}.$$

Proposition 11.2. *As $N \to \infty$ we have $D_N \to 2\pi\delta_0$ in the strong topology of distributions on $\mathbb{T}$.*

Thus, in the distributional sense, the Dirichlet kernel forms an approximate identity. Notice however that this is a generalization of our earlier use of the term "approximate identity"—the Dirichlet kernels are *not* everywhere positive. This manifests itself in the need to appeal to the *differentiability* of f, not just its continuity, in the following argument.

Proof. We must consider the limit of the integral

$$\int D_N(t)\phi(t)\,dt$$

for smooth test functions ϕ on $\mathbb{T}$. It follows by the formula for summation of a geometric series that

$$D_N(t) = \sum_{k=-N}^{N} e^{ikt} = \frac{\sin((N+\frac{1}{2})t)}{\sin(\frac{1}{2}t)}.$$

Noting that $\int D_N(t)dt = 2\pi$ for all N, we see that

$$\left(\int D_N(t)\phi(t)\,dt\right) - 2\pi\phi(0) = \int \sin((N+\tfrac{1}{2})x)\frac{\phi(t)-\phi(0)}{\sin\frac{1}{2}t}\,dt.$$

The function appearing on the right of the display, $\psi\colon t \mapsto \dfrac{\phi(t)-\phi(0)}{\sin\frac{1}{2}t}$, is *smooth* on the circle. Integration by parts shows that the display is equal to

$$\frac{1}{N+\frac{1}{2}}\int \cos((N+\tfrac{1}{2})x)\psi'(t)\,dt$$

so it tends to zero as $N \to \infty$ (uniformly on bounded subsets in $\mathscr{E}(\mathbb{T})$). □

Let $f \in \mathscr{E}(\mathbb{T})$. The *Fourier coefficients* of f are the numbers

$$c_k = (2\pi)^{-\frac{1}{2}} \int_{\mathbb{T}} f(x)e^{-ikx}\,dx.$$

The *Fourier series* of f is the formal series

$$(2\pi)^{-\frac{1}{2}} \sum_{k=-\infty}^{\infty} c_k e^{ikx}.$$

Proposition 11.3. *Let $f \in \mathscr{E}(\mathbb{T})$. Then the Fourier coefficients of f are rapidly decreasing (that is to say, for each $m = 0, 1, \ldots$ we have $|c_k| \leqslant C_m(1+|k|)^{-m}$), and the Fourier series for f converges to f in the topology of $\mathscr{E}(\mathbb{T})$.*

Proof. Integration by parts m times shows

$$c_k = (2\pi)^{-\frac{1}{2}}(-ik)^{-m}\int f^{(m)}(x)e^{-ikx}dx$$

so the Fourier coefficients are rapidly decreasing. It follows from Weierstrass' M-test that the Fourier series converges uniformly, together with all of its derivatives. To see that the limit is indeed the function f, notice that

$$(2\pi)^{-\frac{1}{2}}\sum_{k=-N}^{N} c_k e^{ikx} = (2\pi)^{-1}\int \sum_{k=-N}^{N} e^{ik(x-y)}f(y)dy = (2\pi)^{-1}D_N * f(x).$$

Since $D_N \to 2\pi\delta$ as $N \to \infty$, Proposition 10.10 shows that $(2\pi)^{-1}D_N * f \to f$ as required. □

Now we relate this to the Fourier transforms of periodic functions on $\mathbb{R}$ considered as temperate distributions. Let $f(x)$ be a smooth function on $\mathbb{T}$ with Fourier series $(2\pi)^{-\frac{1}{2}}\sum_{k=-\infty}^{\infty} c_k e^{ikx}$. This series converges to f in $\mathscr{E}(\mathbb{T})$. It follows easily that, when lifted to $\mathbb{R}$, it converges to $\tilde{f}$ in the topology of temperate distributions. However, we know that the Fourier transform of $x \mapsto e^{ikx}$ is just the temperate distribution $(2\pi)^{1/2}\delta_k$. Consequently, we obtain

Proposition 11.4. *Let f be a function on $\mathbb{T}$ and let $\tilde{f}$ be the corresponding periodic function on $\mathbb{R}$ (considered as a temperate distribution). Then the Fourier series of f and the (distributional) Fourier transform of $\tilde{f}$ are related by*

$$\mathscr{F}(\tilde{f}) = \sum_{k=-\infty}^{\infty} c_k\delta_k.$$

That is, the Fourier transform of $\tilde{f}$ is a "row of deltas" whose coefficients are the Fourier coefficients of f. □

It is instructive to apply this to the Dirichlet kernel itself. The Fourier transform of the Dirichlet kernel is simply

$$\mathscr{F}D_N = (2\pi)^{1/2}(\delta_{-N} + \cdots + \delta_N).$$

This gives half of

Proposition 11.5. *(a) As $N \to \infty$, the Fourier transforms $\mathscr{F}(D_N)$ converge (in the topology of temperate distributions) to the "comb" distribution*

$$(2\pi)^{1/2} \sum_{\ell=-\infty}^{\infty} \delta_\ell.$$

(b) As $N \to \infty$, the Dirichlet kernels $\tilde{D}_N(x)$ themselves converge to the "comb" distribution

$$2\pi \sum_{\ell=-\infty}^{\infty} \delta_{2\ell\pi}.$$

Thus (up to a scaling) the "comb" distribution is its own Fourier transform.

Proof. Part (a) was proved in the discussion before the statement of the proposition, and part (b) is just a restatement of Proposition 11.2 in the language appropriate to periodic distributions on $\mathbb{R}$ rather than distributions on $\mathbb{T}$. □

This gives us the Poisson summation formula.

Theorem 11.6. *Let $f \in \mathscr{S}(\mathbb{R})$. For any $r > 0$ we have*

$$\sum_{k=-\infty}^{\infty} f(2\pi(t+kr))e^{2\pi iky} = \frac{1}{r\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} \hat{f}((k-y)/r)e^{2\pi i(k-y)t/r}.$$

Proof. Let $\Lambda = \lim D_N$ be the comb distribution appearing in the previous proposition. The definition of the Fourier transform of a tempered distribution gives us $(\mathscr{F}\Lambda)(f) = \Lambda(\mathscr{F}f)$. Using the previous proposition, $\mathscr{F}(\Lambda)$ is also a comb, and we get

$$\sum_{k=-\infty}^{\infty} \hat{f}(k) = (2\pi)^{1/2} \sum_{k=-\infty}^{\infty} f(2k\pi).$$

This is the case $t = y = 0$, $r = 1$ of the Poisson summation formula. The general case follows by applying this special case to the function $x \mapsto f(rx + 2\pi t)e^{ixy}$. □

Suppose we know that the smooth function f is supported in the interval $(-r\pi, r\pi)$. Then the left side of the Poisson summation formula, considered as a function of t, restricts on $(-\frac{1}{2}, \frac{1}{2})$ to $t \mapsto f(2\pi t)$, and thus it determines f completely. We deduce that, if we know a priori that the support of f is restricted to an interval of length less than $2\pi r$, we can reconstruct f completely by sampling its Fourier transform at the points k/r, $k \in \mathbb{Z}$ (or at the points of any lattice in $\mathbb{R}$ with spacing $1/r$). The more restricted we know the support of f is, the sparser the lattice on which we need to sample $\hat{f}$.

Remark 11.7. There are natural n-variable generalizations of all this discussion, but to save notation we have only written out the one-variable case. The only tricky point is this: in n dimensions we may define the Dirichlet kernel to be the sequence of functions

$$D_N(x_1, \ldots, x_n) = \sum_{p_1, \ldots, p_n = -N}^{N} e^{i(p_1 x_1 + \cdots + p_n x_n)}.$$

(In other words we sum over "cubes" not "balls".) With this definition, all the above arguments go through more or less word for word.

Lecture 12

Applications of Fourier Series

Now we make an application to linear differential operators. Let T be such a differential operator, acting on smooth functions (or on distributions) on $\mathbb{R}^n$. We will say that the equation $Tu = v$ is *locally solvable* if every point of $\mathbb{R}^n$ has an open neighborhood U with the following property: if $v \in \mathscr{D}(U)$ is compactly supported then there exists $u \in \mathscr{E}(\mathbb{R}^n)$ such that $tu = v$ on U. We are going to prove that every *constant coefficient* differential operator has this property. The method is taken from Dadok and Taylor, *Local solvability of constant coefficient PDE as a small divisor problem*, Proc AMS **82**(1981),58–60.

Remark 12.1. C^∞ local solvability does *not* hold in general for non-constant coefficients (a surprising fact!). See H. Lewy, *An example of a smooth linear PDE without solution*, Annals of Math **66**(1958),155-158.

Let P be a polynomial in n variables, and consider the constant coefficient differential operator $T = P(iD)$ on (smooth) functions on $\mathbb{R}^n$. Then $Tu = v$ implies

$$P(\xi)\hat{u}(\xi) = \hat{v}(\xi).$$

The natural approach to solve $Tu = v$ is therefore to form $\hat{v}(\xi)/P(\xi)$ and hope that this makes sense (as a tempered distribution for example).

Example 12.2. Suppose that T=$1 + \Delta$, where Δ is the Laplacian $-\sum_{j=1}^n \partial^2/\partial x_j^2$. Then $P(\xi) = 1 + |\xi|^2$. The function $1/P(\xi)$ is of slow growth and multiplication by $1/P(\xi)$ sends $\mathscr{S}(\mathbb{R}^n)$ to $\mathscr{S}(\mathbb{R}^n)$. Therefore, whenever $v \in \mathscr{S}(\mathbb{R}^n)$, we can find a global solution $u \in \mathscr{S}(\mathbb{R}^n)$ to the equation $Tu = v$.

We were able to construct a global solution here because $1/P(\xi)$ was globally well-behaved. The first idea needed to get the general proof is to use the Poisson summation formula to show that, if we only want to find a *local* solution, then we only need $1/P(\xi)$ to be well behaved on a *lattice*.

Lemma 12.3. *Suppose that the polynomial P has the property that there is some lattice L in $\mathbb{R}^n$, of the form $\{(N_1k_1 + r_1, N_2k_2 + r_2, \ldots) : (k_1, k_2, \ldots) \in \mathbb{Z}^n\}$ such that*

$$|P(\xi)|^{-1} \leqslant C(1 + |\xi|^2)^d$$

for some C, d and all $\xi \in L$. Then $P(iD)$ is locally solvable.

Proof. Using the Poisson summation formula, choose $r > 0$ so small that any smooth function on $\mathbb{R}^n$ supported within a ball of radius r is completely determined by the restriction of its Fourier transform to the lattice L. Suppose now that v has support within a ball B of radius r, and consider the function

$$\xi \mapsto \hat{v}(\xi)/P(\xi)$$

defined on L. Since $\hat{(v)} \in \mathscr{S}$, whereas $1/P$ is polynomially bounded on L, the function is of rapid decrease on L. Let ϕ be a smooth bump function on $\mathbb{R}^n$, with $\phi(0) = 1$ and having support much smaller than the period of the lattice L. Then

$$t \mapsto \sum_{\xi \in L} \phi(t - \xi)\hat{u}(\xi)/P(\xi)$$

belongs to the Schwarz class; let $f \in \mathscr{S}$ be its inverse Fourier transform. By construction, $P(\xi)\hat{f}(\xi) = \hat{v}(\xi)$ for $\xi \in L$. The Poisson summation formula now produces a function u (the periodization of f) which has the property that $P(iD)u$ equals the periodization of v; in particular, $P(iD)u = v$ on the original ball B. □

Example 12.4. Suppose that the polynomial P has integer (or rational) coefficients. Then $T = P(iD)$ is locally solvable. To see this, write $P(\xi) = \sum_{|\beta| \leqslant m} a_\beta \xi^\beta$, with $a_\beta \in \mathbb{Z}$. Certainly there exist vectors v such that the one-variable polynomial

$$t \mapsto P(tv)$$

is not identically zero; the set of such vectors is open, and therefore it contains a vector with rational coefficients. Hence there exist rational numbers $r_i = p_i/q_i$, $i = 1, \ldots, n$, such that $P(r_1, \ldots, r_n) \notin \mathbb{Z}$. Now let $N = (q_1 \ldots q_n)^m$. Then for $(k_1, \ldots, k_n) \in \mathbb{Z}^n$,

$$P(Nk_1 + p_1/q_1, Nk_2 + p_2/q_2, \ldots) \equiv P(r_1, r_2, \ldots) \pmod 1.$$

In particular, $|P(Nk_1+p_1/q_1, Nk_2+p_2/q_2, \ldots)|$ is bounded away from zero. Thus $1/P(\xi)$ is bounded on the lattice $L = \{(Nk_1 + p_1/q_1, Nk_2 + p_2/q_2, \ldots)\}$.

The general case is handled by the following measure theoretic lemma.

Lemma 12.5. *Let P be a polynomial in n variables. There exists $d \in \mathbb{R}$ such that, for almost all $r \in [0, 1]^n$,*

$$|P(k + r)^{-1}| \leqslant C_r(1 + |k|^2)^{-d} \quad \forall k \in \mathbb{Z}^n. \tag{1}$$

In particular, $|P(\xi)^{-1}|$ is bounded by a polynomial in ξ on some lattice L.

Proof. It suffices to show that there exists d such that the set

$$S_C = \{\xi \in \mathbb{R}^n : |P(\xi)^{-1}| \geqslant C(1 + \|\xi\|^2)^d\} \tag{2}$$

has finite Lebesgue measure for sufficiently large C. If this is so, then the continuity property of Lebesgue measure shows that, given any $\epsilon > 0$, $\lambda(S_C) < \epsilon$ for sufficiently large C. The set of values of $r \in [0, 1]^n$ of which some integer translate meets S_C then has measure less than ϵ, and the desired estimate 1 holds on the complement of this set.

To establish that the set S_C defined by 2 has finite Lebesgue measure for some d, it is enough to show that for some $\delta > 0$,

$$\int_{|\xi|\geqslant 1} |P(\xi)|^{-\delta}|\xi|^{-2n} d\xi < \infty.$$

Let m be the degree of P. Change variable from ξ to $\zeta = \xi/|\xi|^2$ to get

$$\int_{|\xi|\geqslant 1} |P(\xi)|^{-\delta}|\xi|^{-2n} d\xi = \int_{|\zeta|\leqslant 1} |\zeta|^{2m\delta}|Q(\zeta)|^{-\delta} d\zeta$$

where $Q(\zeta) = |\zeta|^{2m}P(\zeta/|\zeta|^2)$ is a polynomial of degree $2m$. We see finally that it is enough to prove that $|Q(\zeta)|^{-\delta}$ is a locally integrable function for some $\delta > 0$.

Rotating if necessary, we may assume that Q has the form

$$Q(\zeta_1, \ldots, \zeta_n) = c\zeta_1^r + q_{r-1}(\zeta_2, \ldots, \zeta_n)\zeta_1^{r-1} + \cdots ,$$

where c is a nonzero constant, $r \leqslant 2m$, and $q_{r-1}, \ldots, q_0$ are polynomials in $\zeta_2, \ldots, \zeta_n$. We may write

$$Q(\zeta_1, \ldots, \zeta_n) = c(\zeta_1 - a_1) \cdots (\zeta_1 - a_r)$$

where the $a_1, \ldots, a_r$ depend on $\zeta_2, \ldots, \zeta_n$ and remain bounded for $|\zeta_j| \leqslant 1$, $j = 2, \ldots, n$. It follows that there is a constant C such that

$$\lambda\{\zeta_1 : |Q(\zeta_1, \ldots, \zeta_n)| < \eta\} \leqslant C\eta^{-1/r}$$

for all $|\zeta_j| \leqslant 1$, $j = 2, \ldots, n$. Consequently,

$$\lambda\{\zeta_1 : |Q(\zeta_1, \ldots, \zeta_n)|^{-\delta} > \eta\} \leqslant C\eta^{-1/\delta r}.$$

This shows that as soon as $\delta < 1/r$, $|Q|^{-\delta}$ is uniformly integrable as a function of ζ_1. Fubini's theorem now shows that $|Q|^{-\delta}$ belongs to $L^1([-1, 1]^n$. This completes the proof. □

Theorem 12.6. *Every constant coefficient linear partial differential operator is locally solvable.* $\square$

Our second application of Fourier series techniques will be to the proof of the *Schwarz kernels theorem*. Let's begin by describing this result.

Early in a functional analysis or PDE course one learns about *integral operators*. Let $X = \mathbb{R}^n$ (or, X could be any smooth manifold equipped with a smooth measure). Let k be a function (say a smooth function) on $X \times X$. Then one defines the *integral operator* with *kernel* k to be the linear map $T\colon \mathscr{D}(X) \to \mathscr{E}(X)$ defined by

$$Tu(y) = \int_X k(x,y)u(x)dx.$$

This is a "continuous version of the law of matrix multiplication". We will have $T\colon \mathscr{D}(X) \to \mathscr{D}(X)$ if T is *properly supported*, meaning that the intersection of $\mathrm{Support}(T)$ with $K \times X$ or $X \times K$, K compact in X, is itself compact.

What happens when k is a distribution on $X \times X$? In this case the integral above does not define a function, but it *does* make sense as a distribution: Tu is by definition the distribution such that

$$\langle Tu, v\rangle = \iint k(x,y)u(x)v(y)dxdy = \langle k, u \otimes v\rangle, \quad v \in \mathscr{D}(X)$$

where $u \otimes v$ denotes the function $(x,y) \mapsto u(x)v(y)$ on $X \times X$. In this way the operator T can be defined for the *distributional kernel* k.

Proposition 12.7. *The operator T defined by a distributional kernel k, as above, is* continuous *as a linear map $\mathscr{D}(X) \to \mathscr{D}'(X)$ (here $\mathscr{D}'(X)$ is equipped with the strong dual topology).*

Proof. Let T be defined by a distributional kernel k and let V be an 0-neighborhood in $\mathscr{D}'(X)$, which we may take to be the polar of a bounded subset B of $\mathscr{D}(X)$. Notice that the family of maps $\mathscr{D}(X) \to \mathscr{D}(X \times X)$ defined by $\{u \mapsto u \otimes v : v \in B\}$ is equicontinuous. Thus there is an 0-neighborhood U in $\mathscr{D}(X)$ such that

$$\{k(u \otimes v) : v \in B, u \in U\} \subseteq D(0;1).$$

But this means that $T(U) \subseteq V$, so T is continuous as required. $\square$

The Schwarz kernels theorem is the converse of the above proposition.

Theorem 12.8. *Let T be a continuous linear map from $\mathscr{D}(X)$ to $\mathscr{D}'(X)$. Then T is given by a distributional kernel.*

Lecture 13

Proof of the Kernels Theorem

Recall that the Schwarz Kernels Theorem 12.8 says that every continuous linear map $\mathscr{D}(X) \to \mathscr{D}'(X)$, where X is a smooth manifold, is given by a distributional kernel. We are going to prove the theorem by making use of some ideas about Hilbert-Schmidt operators and Sobolev spaces.

Let H and K be separable Hilbert spaces, and choose complete orthonormal systems $\{e_i\}$ and $\{f_j\}$ in H and K respectively. Let $T\colon H \to K$ be a bounded linear map. Consider the sum of the squares of the matrix coefficients of T, that is

$$S = \sum_{i,j} |\langle Te_i, f_j\rangle|^2.$$

From the Parseval equation, $S = \sum_i \|Te_i\|^2$ does not depend on the choice of the orthonormal system $\{f_j\}$; on the other hand, we can take adjoints to see that

$$S = \sum_{i,j} |\langle e_i, T^* f_j\rangle|^2 = \sum_j \|T^* f_j\|^2$$

and therefore does not depend on the norm of the the system $\{e_i\}$. We conclude that S is an intrinsic invariant of the operator T (possibly $+\infty$) and that it is the same for T and T^*.

Definition 13.1. The *Hilbert-Schmidt norm* of T is the quantity $\|T\|_{HS}$ defined by

$$\|T\|_{HS}^2 = \sum_{i,j} |\langle Te_i, f_j\rangle|^2.$$

The operator T is a *Hilbert-Schmidt operator* if its Hilbert-Schmidt norm is finite.

Lemma 13.2. *For any Hilbert-Schmidt operator T we have $\|T\| \leqslant \|T\|_{HS}$ (the norm on the left is the operator norm).*

Proof. Let x be a unit vector in E. The fact that $\{x\}$ can be extended to a complete orthonormal set shows that $\|Tx\|^2 \leqslant \|T\|_{HS}^2$. But $\|T\|$ is defined to be $\sup\{\|Tx\| : \|x\| = 1\}$, so the result follows. $\square$

Proposition 13.3. *The space of Hilbert-Schmidt operators is a Banach space (in fact, a Hilbert space) in the Hilbert-Schmidt norm. Moreover, it is an ideal in the* *bounded operators: if $T\colon H \to K$ is a Hilbert-Schmidt operator and $S\colon H' \to H$, $R\colon K \to K'$ are bounded operators then RTS is a Hilbert-Schmidt operator and*

$$\|RTS\|_{HS} \leqslant \|R\|\|T\|_{HS}\|S\|.$$

Proof. The map which sends a Hilbert-Schmidt operator T to its list of matrix coefficients $\langle Te_i, f_j\rangle$ exhibits an isomorphism of the space of Hilbert-Schmidt operators with the sequence space ℓ^2. To prove the statement about ideals it is enough to show that $\|RT\|_{HS} \leqslant \|R\|\|T\|_{HS}$ (the statement about S follows on taking adjoints). And we prove this by writing

$$\|RT\|_{HS}^2 = \sum_i \|RTe_i\|^2 \leqslant \|R\|^2 \sum_i \|Te_i\|^2 = \|R\|^2\|T\|_{HS}^2.$$

□

Exercise 13.4. Show that every Hilbert-Schmidt operator is compact.

Suppose now that $H = L^2(X,\mu)$ and $K = L^2(Y,\nu)$ for suitable measure spaces. Let $k \in L^2(X \times Y, \mu \otimes \nu)$ be a square-integrable kernel on the product space. Attempt to define an integral operator $T\colon H \to K$ by

$$Tu(y) = \int_Y k(x,y)u(x)d\mu(x).$$

Proposition 13.5. *Given $u \in L^2(X,\mu)$, the above integral converges for almost all y and defines a function $Tu \in L^2(Y,\nu)$. The operator $T\colon L^2(X,\mu) \to L^2(Y,\nu)$ so defined is a Hilbert-Schmidt operator (in particular, it is bounded) with Hilbert-Schmidt norm $\|T\|_{HS} = \|k\|_{L^2(X\times Y)}$.*

Proof. By Fubini's theorem the function $x \mapsto k(x,y)$ belongs to L^2 for almost all y, which shows that the integral exists almost everywhere. Now write

$$\|Tu\|^2 = \int \left| \int k(x,y)u(x)d\mu(x) \right|^2 d\nu(y).$$

Consider an orthonormal basis $\{e_i\}$ for $L^2(X,\mu)$. By the Parseval equation,

$$\sum_i \left| \int k(x,y)e_i(x)d\mu(x) \right|^2 = \int |k(x,y)|^2 d\mu(x).$$

Thus, summing the previous equation as u runs over the orthonormal basis $\{e_i\}$, and using the monotone convergence theorem, we get

$$\|T\|_{HS}^2 = \sum_i \|Te_i\|^2 = \iint |k(x,y)|^2 d\mu(x) d\nu(y)$$

as asserted. $\square$

The following easy converse to the above proposition is a measure-theoretic prototype of the Schwarz kernels theorem.

Proposition 13.6. *Every Hilbert-Schmidt operator from $L^2(X,\mu)$ to $L^2(Y,\nu)$ is represented by an L^2 kernel as in the previous proposition.*

Proof. Let $\{e_i\}$ and $\{f_j\}$ be orthonormal bases in $H = L^2(X,\mu)$ and $K = L^2(Y,\nu)$ respectively. Notice that the functions $\bar{e}_i \otimes f_j$ form an orthonormal set in $L^2(X \times Y, \mu \otimes \nu)$.

Let T be a Hilbert-Schmidt operator from H to K. Consider the sum

$$\sum_{i,j} \langle Te_i, e_j \rangle \bar{e}_i \otimes f_j$$

which converges in $L^2(X \times Y)$ by Bessel's inequality, say to a function k. The function k has the property that

$$\begin{aligned} \int k(x,y) e_m(x) d\mu(x) &= \sum_{i,j} \langle Te_i, f_j \rangle \langle e_i, e_m \rangle f_j(y) \\ &= \sum_j \langle Te_m, f_j \rangle f_j(y) = Te_m(y), \end{aligned}$$

so the kernel k defines the Hilbert-Schmidt operator T. $\square$

We are going to use some special internal structure of spaces of smooth functions (what will later be called their *nuclear* structure) to reduce the general Schwarz kernel theorem to this elementary version.

For the moment we are going to assume that the smooth manifold X on which we are trying to prove the kernels theorem is the n-torus $X = \mathbb{T}^n$. Later, we will use a partition of unity to reduce the general case to this special one. Recall that the normalized characters

$$e_k(x) = (2\pi)^{-n/2} e^{ik \cdot x}, \quad k \in \mathbb{Z}^n$$

form an orthonormal basis for $L^2(X)$. In particular, for each $f \in L^2(X)$ the series $\sum_k c_k e_k$ converges in $L^2(X)$ to f, where the Fourier coefficients c_k are defined by

$$c_k = \langle f, e_k \rangle = (2\pi)^{-n/2} \int_X f(x) e^{-ik\cdot x}\, dx.$$

We have $\|f\|^2_{L^2} = \sum_k |c_k|^2$ (Parseval's theorem). The Fourier coefficients can also be defined for any distribution on X, and the Fourier coefficients of a distribution are at most of polynomial growth.

Definition 13.7. For each integer s (positive or negative) the *Sobolev space* $H^s(X)$ is the space of all distributions whose Fourier coefficients satisfy

$$\sum_k (1 + |k|^2)^s |c_k|^2 < \infty.$$

We make $H^s(X)$ into a Hilbert space by using the expression in the display to define the square of the norm.

Thus $H^0 = L^2$, and if $s > 0$ then H^s is the space of those L^2 functions whose distributional derivatives up to order s also coincide with L^2 functions. A distribution is in H^{-s} precisely when it defines a continuous linear functional on H^s. Notice that if $s < s'$ then $H^{s'} \subseteq H^s$.

Clearly $C^s(X) \subseteq H^s(X)$; a function that is s times continuously differentiable belongs to H^s. There is a sort of approximate converse to this.

Proposition 13.8. *(Sobolev embedding theorem) If $r \geqslant 0$ and $s > r + n/2$, then $H^s(X) \subseteq C^r(X)$ (and the inclusion map is continuous). Every distribution of order $\leqslant r$ on X belongs to $H^{-s}(X)$.*

Proof. It suffices to consider the case $r = 0$. Suppose that $s > n/2$, and let $\{c_k\}$ be the sequence of Fourier coefficients of a function in $H^s(X)$. Notice that

$$\sum_{k \in \mathbb{Z}^n} (1 + |k|^2)^{-s} < \infty.$$

Consequently

$$c_k = (1 + |k|^2)^{-s/2} \cdot \left[(1 + |k|^2)^{s/2} c_k\right]$$

is the product of two elements of $\ell^2(\mathbb{Z}^n)$, so it belongs to $\ell^1(\mathbb{Z}^n)$ (by Cauchy-Schwarz). But then the Fourier series

$$f(x) = (2\pi)^{-n/2} \sum c_k e^{ik\cdot x}$$

converges absolutely and uniformly to a continuous function. Continuity of the inclusion map follows from the closed graph theorem (or a direct calculation), and the last statement of the proposition comes from duality. ☐

Notice that as a result of this proposition, $\bigcap_s H^s(X) = \bigcap_r C^r(X) = \mathscr{D}(X)$ is the space of smooth functions. Dually, the inductive limit $\bigcup_s H^s(X)$ is the space $\mathscr{D}'(X)$ of all distributions.

A similar computation proves

Proposition 13.9. *For $t > n/2$, the inclusion $H^{s+t} \to H^s$ is a Hilbert-Schmidt operator.*

Proof. An orthonormal basis for the space H^s is given by the functions $\{(1 + |k|^2)^{-s/2} e_k\}$. The matrix entries with respect to these standard bases of the inclusion map from H^{s+t} to H^s are therefore $(1 + |k|^2)^{-t/2}$ along the diagonal, and 0 elsewhere. As we just observed, these form a square-summable sequence provided $t > n/2$. ☐

Now suppose that we have a continuous linear map $T\colon \mathscr{D}(X) \to \mathscr{D}'(X)$, as in the statement of the kernels theorem.

Proposition 13.10. *If $T\colon \mathscr{D}(X) \to \mathscr{D}'(X)$ is a continuous linear map, then there is some $r > 0$ such that $\operatorname{Im}(T)$ consists entirely of distributions of order $\leqslant r$.*

Proof. Consider the bilinear form

$$B(u, v) = \langle Tu, v \rangle$$

defined on $\mathscr{D}(X) \times \mathscr{D}(X)$. It is continuous in u for fixed v (because of the continuity of T) and in v for fixed u (because Tu is a distribution); that is, it is separately continuous. By Proposition 5.13 it is jointly continuous; in particular, there are 0-neighborhoods U and V in $\mathscr{D}(X)$ such that $|\langle Tu, v\rangle| < 1$ for $u \in U$ and $v \in V$.

In other words, T maps U into the polar of V. Supposing now that V contains a basic 0-neighborhood of the form $\{f : |D^\alpha f(x)| < \epsilon \; \forall x \in X, |\alpha| \leqslant N\}$, each distribution $\Lambda \in V^\circ$ satisfies

$$|\Lambda(f)| \leqslant \epsilon^{-1} \sup\{|D^\alpha f(x)| : x \in X, |\alpha| \leqslant N\},$$

and is therefore of order $\leqslant N$. ☐

Using Proposition 13.8 we conclude that T defines a map from $\mathscr{D}(X)$ to the Banach space $H^{-s}(X)$, for some X. Since T is continuous when considered as a map into $\mathscr{D}'(X)$, it is also continuous as a map into $H^{-s}(X)$, by the closed graph theorem (Remark 5.19). By definition of the topology on $\mathscr{D}(X)$, we see that T extends to a continuous map $C^r(X) \to H^{-s}(X)$ for some r. Therefore, by proposition 13.8 again, it extends to a continuous map $H^t(X) \to H^{-s}(X)$ for some s, t.

Proposition 13.9 shows that, by increasing s and t if necessary, we may assume that the map $T\colon H^t(X) \to H^{-s}(X)$ is in fact a Hilbert-Schmidt operator. So, according to the matrix representation of Hilbert-Schmidt operators, there exist $\{c_{k,k'}\} \in \ell^2(\mathbb{Z}^n \times \mathbb{Z}^n)$ such that

$$T((1+|k|^2)^{-t/2}e_k) = \sum_{k'}(1+|k'|^2)^{s/2}c_{k,k'}e_{k'}.$$

The sum

$$k(x,y) = (2\pi)^{-n}\sum_{k,k'}(1+|k|^2)^{t/2}(1+|k'|^2)^{s/2}c_{k,k'}e^{i(k'y-kx)}$$

converges in $\mathscr{D}'(X \times X)$ to a distribution k. The operator T' defined by this kernel agrees with T on each basis vector e_k (by construction). Since the span of the $\{e_k\}$ is dense in $\mathscr{D}(X)$ (by Fourier theory), and both T and T' are continuous, we have $T = T'$ and the proof of the Kernels Theorem is complete.

Lecture 14

Topological Tensor Products

It should be clear from our discussion that the Schwarz kernels theorem has something to do with tensor products. In fact, it should be a "topological" manifestation of the isomorphism

$$(E \otimes F)^* \cong \mathrm{Hom}(E, F^*)$$

(valid for finite-dimensional vector spaces E and F). If we take $E = F = \mathscr{D}(X)$, interpret $*$ as the strong dual and Hom as continuous linear maps, and hope that $\mathscr{D}(X) \otimes \mathscr{D}(X) \cong \mathscr{D}(X \times X)$, this displayed identity exactly becomes the Schwarz kernels theorem.

To make this precise involves giving a suitable interpretation of the tensor product of two LCTVS. There are several different ways of doing this, each of which shares a different subset of the desirable properties of the tensor product in the purely algebraic context. The *nuclear spaces* of Grothendieck (which include the spaces of smooth functions and of distributions) are those for which all possible definitions of the topological tensor product coincide (and, consequently, the topological tensor product has the maximum possible number of desirable properties.)

In the next lecture or two I'll sketch some aspects of this theory, which is something that all functional analysts should know these days, but which is poorly covered in existing texts. For the sake of completeness let us begin by reviewing the ordinary algebraic theory of the tensor product (this is part of standard linear algebra).

Definition 14.1. Let E and F be vector spaces (more generally, they could be modules over any commutative ring). A *tensor product* of E and F is a vector space $E \otimes F$ equipped with a bilinear map $E \times F \to E \otimes F$, $(x, y) \mapsto x \otimes y$, having the property that any other bilinear map $B\colon E \times F \to V$ factorizes uniquely through this one: there is a unique linear map $E \otimes F \to V$ making the diagram

$$\begin{tikzcd}
E \times F \arrow[d, "\otimes"] \arrow[dr, "B"] & \\
E \otimes F \arrow[r, dotted] & V
\end{tikzcd}$$

commute.

Standard universal nonsense shows that a tensor product is unique up to isomorphism, so one usually refers to "the" tensor product $E \otimes F$. The existence of the tensor product is proved in a course on Algebra: a simple approach is to consider the free vector space $\mathfrak{V}$ on $E \times F$, that is the set of all finite linear combinations of formal symbols $x \diamond y$, $x \in E$, $y \in F$, and divide by the subspace spanned by all the elements

$$\{(\lambda x + \lambda' x') \diamond y - \lambda(x \diamond y) - \lambda'(x' \diamond y), \\ x \diamond (\mu y + \mu' y') - \mu(x \diamond y) - \mu'(x \diamond y')\}.$$

The quotient space is a tensor product, with $x \otimes y$ being the coset of $x \diamond y$.

The tensor product $E \times F$ is spanned by the *elementary tensors* $x \otimes y$ (exercise: why?). A general element of $E \times F$ can be represented in many different ways as a linear combination of elementary tensors; it cannot always be represented as a single elementary tensor.

Lemma 14.2. *Let $u \in E \otimes F$. The following are equivalent:*

(a) $u \neq 0$ *in* $E \times F$.

(b) There is a representation $u = \sum_{i=1}^{n} x_i \otimes y_i$*, where the* $\{x_i\}$ *are linearly independent and one of the* $\{y_i\}$ *is nonzero.*

(c) There is a representation $u = \sum_{i=1}^{n} x_i \otimes y_i$*, where the* $\{x_i\}$ *and the* $\{y_i\}$ *are linearly independent.*

Proof. (a) implies (c). Among the representations of u as a sum of elementary tensors there is one of minimal length; choose that one to be $u = \sum i = 1^n x_i \otimes y_i$. If there were a dependence relation among the $x's$, say $x_1 = \sum_{i=1}^{n} \lambda_i x_i$, then we get the shorter expression

$$u = \sum_{i=1}^{n} x_i \otimes (y_i + \lambda_i y_1).$$

By minimality this can't happen, so the x's are independent, and so are the y's by the same argument.

(c) implies (b). This is obvious.

(b) implies (a). Suppose that $y_k \neq 0$. Let $\phi \in E^*$ be a linear functional sending x_k to 1 and all the other x_i, $i \neq k$, to 0 (this uses the independence of the $\{x_i\}$). Let $\psi \in F^*$ be a linear functional sending y_k to 1. Then $(x, y) \mapsto \phi(x)\psi(y)$ is a bilinear functional on $E \times F$, so it passes to a linear functional on $E \otimes F$ by the universal property. This functional sends u to $\sum \phi(x_i)\psi(y_i) = 1$, so $u \neq 0$. ☐

Corollary 14.3. *If $\{x_i\}$ is a basis for E and $\{y_j\}$ is a basis for F then $\{x_i \otimes y_j\}$ is a basis for $E \otimes F$.*

Proof. They clearly span $E \otimes F$, and the lemma shows that they are linearly independent. □

Let E, F be vector spaces. There are canonical linear maps

$$E^* \otimes F \to \operatorname{Hom}(E, F), \quad E^* \otimes F^* \to B(E, F)^*$$

where $B(E, F)$ denotes the space of bilinear forms on $E \times F$. To construct the first of these, consider the bilinear map $E^* \times F \to \operatorname{Hom}(E, F)$ defined by sending (ϕ, y) to the "rank one" homomorphism $x \mapsto \phi(x)y$; by the universal property of the tensor product, this bilinear map factors through $E^* \otimes F$. The second is constructed in a similar way. By working with bases for E, F and their dual bases for E^*, F^* one sees

Proposition 14.4. *If E, F are finite-dimensional vector spaces the above maps are isomorphisms.* □

Now pass to the case where E and F are locally convex *topological* vector spaces; we would like to define their tensor product as an LCTVS also. There are two issues here.

(a) What topology shall we put on the tensor product? There is more than one natural choice.

(b) Once we have chosen a topology, we will find that the "algebraic" tensor product (defined above) is not usually complete. We must therefore form its completion.

To help discuss (b), we make a change of notation: from now on use $\odot$ to denote the "algebraic" tensor product of two vector spaces (as defined above); we will use $\otimes$ to denote a suitable completion.

Let p and q be seminorms on E and F respectively. The *projective tensor product seminorm* is defined on $E \odot F$ by

$$(p \odot q)(u) = \inf\{\sum_{i=1}^{n} p(x_i)q(y_i) : u = \sum_{i=1}^{n} x_i \odot y_i\}$$

(the infimum is taken over all possible ways of representing u as a sum of elementary tensors). It is clear that this is indeed a seminorm.

Exercise 14.5. (For C^*-algebraists only.) If E, F are C^*-algebras and p, q are their norms, show that $p \odot q$ (as defined above) need not be a C^*-norm on $E \odot F$. Thus the tensor product that we are going to construct (in the category of Banach spaces, or LCTVS) does not restrict to the tensor product in the category of C^*-algebras.

Definition 14.6. Let E and F be (complete, Hausdorff) LCTVS. The *projective topology* on $E \odot F$ is the topology defined by the seminorms $p \odot q$, as p, q range over the continuous seminorms on E, F. The *projective tensor product* $E \otimes_\pi F$ is the completion of $E \odot F$ in the projective topology.

Proposition 14.7. *Let E, F be LCTVS. Then*

(a) If E, F are Hausdorff, so is $E \otimes_\pi F$.

(b) For all $x \in E, y \in F$ and seminorms p, q we have $(p \odot q)(x \odot y) = p(x)q(y)$.

(c) The tensor product map $E \times F \to E \odot F$ is continuous, and in fact the projective topology is the strongest locally convex topology on $E \odot F$ for which this map is continuous.

(d) The universal property holds: if $B\colon E \times F \to G$ is a (jointly) continuous bilinear map to another LCTVS, then there is a unique continuous linear map $E \otimes_\pi F \to G$ making the diagram

$$\begin{tikzcd}
E \times F \arrow[d, "\otimes"] \arrow[dr, "B"] & \\
E \otimes_\pi F \arrow[r, dotted] & G
\end{tikzcd}$$

commute.

Proof. We will start with (d). Suppose that $B\colon E \times F \to G$ is a continuous bilinear map: this means that for each seminorm r for G there are seminorms p for E and q for F such that

$$r(B(x, y)) \leqslant p(x)q(y).$$

There is a linear map $\beta\colon E \odot F \to G$ defined by $b(u) = \sum_{i=1}^n B(x_i, y_i)$ for any representation $u = \sum_{i=1}^n x_i \odot y_i$; we must show that b is continuous. But the triangle inequality shows that for any such representation, $r(b(u)) \leqslant \sum p(x_i)q(y_i)$;

taking the infimum over all such representations gives $r(b(u)) \leqslant (p \odot q)(u)$, giving the required continuity.

Now we consider $(p \odot q)(x \odot y)$. Clearly this is $\leqslant p(x)q(y)$; we must prove the opposite inequality. By the Hahn-Banach theorem there exist continuous linear functionals $\phi \in E^*$, $\psi \in F^*$ with $|\phi| \leqslant p$, $|\psi| \leqslant q$, $\phi(x) = p(x)$, $\psi(y) = q(y)$. The bilinear functional

$$B(x, y) = \phi(x)\psi(y)$$

is continuous on $E \times F$, so (by the first part) it gives rise to a continuous $b\colon E \odot F \to \mathbb{C}$ having $|b(u)| \leqslant (p \odot q)(u)$. Apply this to $u = x \odot y$ to get

$$p(x)q(y) = b(u) \leqslant (p \odot q)(x \odot y)$$

as required.

This calculation shows that the tensor product map $E \times F \to E \odot F$ is continuous. The universal property of the projective tensor product (part (d)) shows that the projective tensor product topology is the strongest topology with this property.

Finally we must prove that the projective tensor product is Hausdorff. It is enough to prove that $E \odot F$ is Hausdorff. Suppose that $u \neq 0$ in $E \odot F$. By Lemma 14.2, we may write $u \sum x_i \odot y_i$ with the x_i linearly independent and $y_1 \neq 0$. By the Hahn-Banach theorem (and the properties of finite-dimensional Hausdorff TVS) we can find continuous linear functionals $\phi \in E^*$, $\psi \in F^*$ with $\phi(x_1) \neq 0$, $\psi(y_1) \neq 0$, $\phi(x_2) = \ldots = \phi(x_n) = 0$. Then the continuous bilinear map $(x, y) \mapsto \phi(x)\psi(y)$ passes to a continuous linear map on $E \odot F$ mapping u to a nonzero element. □

If E, F are Banach spaces it suffices to consider the standard norms for p, q. Thus $E \otimes_\pi F$ becomes a Banach space also.

Exercise 14.8. Let E, F be Banach spaces. Show that every $u \in E \otimes_\pi F$ can be represented as the sum of an absolutely convergent series $\sum_{k=1}^{\infty} x_k \otimes y_k$, with $x_k \in E$, $y_k \in F$, and that

$$\|u\|_\pi = \inf \left\{ \sum_{k=1}^{\infty} \|x_k\| \|y_k\| : u = \sum_{k=1}^{\infty} x_k \otimes y_k \right\}.$$

Exercise 14.9. Suppose that (X, μ) and (Y, ν) are measure spaces. Show that $L^1(X, \mu) \otimes_\pi L^1(Y, \nu) = L^1(X \times Y, \mu \otimes \nu)$. (This result is not true for L^p spaces, $p \neq 1$.)

The projective tensor product is functorial. Suppose that E, E', F are LCTVS and $T\colon E \to E'$ is a continuous linear map. The map $E \times F \to E' \otimes_\pi F$ defined by $(x, y) \mapsto Tx \otimes y$ is bilinear and continuous, so by the universal property it factors through a continuous linear

$$E \otimes_\pi F \to E' \otimes_\pi F$$

which it is natural to denote $T \otimes_\pi 1$. Similarly for functoriality in F, or in both variables together.

Proposition 14.10. *With notation as above, suppose that E, E', F are Fréchet spaces, and that T is surjective. Then $T \otimes_\pi 1$ is surjective also.*

Proof. Note that all spaces involved are Fréchet spaces. The dual of the tensor product $E \otimes F$ is the space $B(E, F)$ of bilinear forms on $E \times F$. Consider the dual map

$$S = (T \otimes_\pi 1)^*\colon B(E', F) \to B(E, F).$$

According to Proposition 7.8 and Theorem 7.11, to show that $T \otimes_\pi 1$ is surjective, it suffices to show that S is injective and has weak-$*$ closed range. We have

$$(SB)(x, y) = B(Tx, y).$$

Since T is surjective, it is clear that $SB = 0$ implies $B = 0$, that is, S is injective. I claim that the range of S is exactly the set W those bilinear forms C for which $C(x, y) = 0$ for all $x \in \ker T$ and $y \in F$; it is clear that W is a weak-$*$ closed set.

Plainly $\operatorname{Im} S \subseteq W$. Suppose that $C \in W$. By a corollary of the open mapping theorem, T is a *homomorphism*: this implies that for each $y \in Y$ there is a continuous linear map $B(\cdot, y)$ on E' such that $C(x, y) = B(Tx, y)$. Now B is a separately continuous bilinear form on $E' \times F$; since these are Fréchet spaces, it is jointly continuous, and $SB = C$. □

By contrast it need *not* be the case that if T is injective, then $T \otimes_\pi 1$ is injective (not even if T is an isomorphism onto its image).

Lecture 15

Notes on Nuclearity

In this second lecture we will sketch the abstract formulation of the kernels theorem in terms of topological tensor products.

Let X be a manifold (compact, to keep matters simple; for instance, the torus $\mathbb{T}^n$) and let E be the Fréchet space $\mathscr{D}(X)$. A continuous linear map $T\colon E \to E^*$ gives rise to a bilinear form B on E, $B(u,v) = \langle Tu, v\rangle$; this form is separately continuous, hence jointly continuous (compare the proof of Proposition 13.10). By the universal property of the (projective) tensor product, the space of such continuous bilinear forms on $E \times E$ is isomorphic to the dual space $(E \otimes_\pi E)^*$. It is thus clear that the Schwarz kernels theorem is equivalent to the statement that the canonical map (coming from the universal property)

$$\mathscr{D}(X) \otimes_\pi \mathscr{D}(X) \to \mathscr{D}(X \times X)$$

is an *isomorphism*.

How might we go about proving this? First notice that the canonical map on the *algebraic* tensor product

$$\mu\colon \mathscr{D}(X) \odot \mathscr{D}(X) \to \mathscr{D}(X \times X)$$

is an *injection*. Indeed, suppose that $u \neq 0$ in $\mathscr{D}(X) \odot \mathscr{D}(X)$. According to Lemma 14.2 we can write $u = \sum i = 1^n f_i \odot g_i$ with the $\{f_i\}$ linearly independent and at least one g nonzero. But if $\mu(u) = 0$ then $\sum_{i=1}^n f_i(x)g_i(y) = 0$ for all $x, y \in X$; by linear independence of the $\{f_i\}$ this means that for each fixed y all of the $g_i(y) = 0$; this is a contradiction.

We can therefore regard $\mathscr{D}(X) \odot \mathscr{D}(X)$ as a *subspace* of $\mathscr{D}(X \times X)$. According to the universal property of the projective tensor product, the Fréchet topology of $\mathscr{D}(X \times X)$ induces on this subspace a topology which is *weaker* than the projective tensor product topology. (It is instructive also to prove this by directly computing with the seminorms defining this topology.) If we can also show that the induced subspace topology is *stronger* than the projective tensor product topology, we will be finished, since then the completed projective tensor product $\mathscr{D}(X) \otimes_\pi \mathscr{D}(X)$ will be identified with the *closure* in $\mathscr{D}(X \times X)$ of the dense subspace $\mathscr{D}(X) \odot \mathscr{D}(X)$.

To get some feeling for the issues here consider the analogous case of $E = C(X)$, the continuous functions on X. As remarked in an earlier exercise, there is a continuous (norm-1) map of Banach spaces

$$C(X) \otimes_\pi C(X) \to C(X \times X)$$

but this map is *not onto* in general. The issue is that even if a function k of two variables is bounded (say by 1), there may not be an efficient way to express it as a sum $\sum_{i=1}^\infty f_i(x)g_i(y)$, where the functions $\{f_i\}$ and $\{g_i\}$ themselves have their sup norms under control.

However, if X is the torus for instance, and if k is sufficiently differentiable, there *is* an efficient way to write it as such a sum: use Fourier series! Specifically, if $k \in C^r$, where $r > 2 \dim X$, then the Fourier series

$$k(x, y) = \sum_{p,q \in \mathbb{Z}^n} c_{pq} e^{i(px+qy)}$$

converges absolutely. Now the function $e^{i(px+qy)}$ is of the form $f \odot g$, and it has projective tensor norm 1 by Proposition 14.7. Thus the Fourier series converges in projective tensor norm and we find

$$C^r(X \times X) \subseteq C(X) \otimes_\pi C(X) \subseteq C(X \times X).$$

Notice the "loss of derivatives" that appears here, analogous to our earlier discussion of the Kernels Theorem. It is quite easy now to generalize the discussion and to observe that if $k \in \mathscr{D}(X \times X)$, then the Fourier series for k converges with respect to each of the seminorms defining the projective tensor product $\mathscr{D}(X) \otimes_\pi \mathscr{D}(X)$. This shows that $\mathscr{D}(X) \otimes_\pi \mathscr{D}(X) = \mathscr{D}(X \times X)$ as required.

To interpret this argument in a framework of abstract functional analysis, one does the following:

(a) Define a second natural topology in which to complete the algebraic tensor product of two LCTVS. This topology (the "injective topology") is weaker than the projective topology, and it will be constructed in such a way that the topology of $\mathscr{D}(X \times X)$ is obviously *stronger* than the injective topology on $\mathscr{D}(X) \odot \mathscr{D}(X)$.

(b) Give an abstract condition ("nuclearity") on a TVS which encodes the "loss of derivatives" phenomenon above, and prove that for a nuclear space the projective and injective tensor topologies agree. The definition is arranged so that $\mathscr{D}(X)$ is a nuclear space.

To carry out item (a) of this program, it helps to know a bit about the theory of reflexivity for LCTVS. The point is that we will define the injective topology on $E \odot F$ in terms of bilinear functionals on $E^* \times F^*$ — a kind of "double duality". For motivation, it helps to think about how E is related to linear functionals on E^*; this is reflexivity.

Proposition 15.1. *Let E be a LCTVS. Then the original topology of E is the same as the topology of uniform convergence on equicontinuous subsets of E^*. Consequently, if E is barreled, then E can be identified with a closed subspace of E^{**}, the dual of E^* equipped with its strong topology,*

Proof. This is just a fancy way of saying that the prepolar of an equicontinuous subset of E^* is a 0-neighborhood, a fact that we have already used several times. The last assertion follows from the Banach-Steinhaus theorem. ☐

If E is any LCTVS, there is a continuous injection $E \to E^{**}$ sending each $x \in E$ to the functional "evaluation at x".

Definition 15.2. Let E be an LCTVS. It is *semireflexive* if the natural map $E \to E^{**}$ is bijective. It is *reflexive* if the natural map is an isomorphism of TVS.

Clearly, if E is barreled and semireflexive, it is reflexive.

The following result is stated as an exercise because we won't be needing it, but it is a key result in the theory and you should be aware of it.

Exercise 15.3. Show that a LCTVS E is semireflexive iff every weakly closed bounded subset of E is weakly compact.

Exercise 15.4. Use the previous exercise to show that $\mathscr{D}(X)$ is reflexive. Deduce that the subspace spanned by the delta-functions is strongly dense in $\mathscr{D}'(X)$.

Now we define the injective tensor product topology. Let E and F be LCTVS. If $x \in E$, $y \in F$ the map

$$(\phi, \psi) \mapsto \phi(x)\phi(y)$$

is a (weak-$*$) continuous bilinear form on $E^* \times F^*$, and this process gives rise to an isomorphism

$$E \odot F \to B(E^*_w, F^*_w)$$

where B denotes the continuous bilinear forms, and the subscript w refers to the weak-$*$ topologies. Let us retopologize $B(E^*_w, F^*_w)$ with the topology of *uniform* *convergence on biequicontinuous sets*: that is, a basis for the 0-neighborhoods in this topology is provided by the sets

$$U(H, K, \epsilon)\text{‘} := \{B : |B(\psi, \psi)| < \epsilon \, \forall \, \phi \in H, \, \psi \in K\}$$

as H, K run over the equicontinuous subsets of E^*, F^* respectively. The corresponding topology on $E \odot F$ is called the *injective topology*, and the completion of $E \odot F$ in this topology is denoted $E \odot_\epsilon F$. Note that our remarks about reflexivity above amount to saying that $E \odot_\epsilon \mathbb{C} = E$ as a TVS.

Remark 15.5. It is necessary to check that the family $\mathfrak{S}$ of biequicontinuous sets has the property that $B(S)$ is bounded for all $S \in \mathfrak{S}$ and each continuous bilinear form B; but this is easy.

By construction, the natural bilinear map $E \times F \to E \odot F$ is continuous with respect to the injective topology. It follows (from the universal property of the projective topology) that the injective topology on $E \odot F$ is weaker than the projective.

In the special case of $\mathscr{D}(X)$, notice the bilinear map

$$\mathscr{D}'(X) \times \mathscr{D}'(X) \to \mathscr{D}'(X \times X)$$

which sends a pair of distributions (Λ, Θ) on X to the distribution

$$k \mapsto \Lambda(x \mapsto \Theta(y \mapsto k(x, y))) = \Theta(y \mapsto \Lambda(x \mapsto k(x, y)))$$

on $X \times X$. This map sends a product of equicontinuous subsets of $\mathscr{D}'(X)$ to an equicontinuous subset of $\mathscr{D}'(X \times X)$. Thus we see that the topology induced on $\mathscr{D}(X) \odot \mathscr{D}(X)$ as a subspace of $\mathscr{D}(X \times X)$ is *stronger* than the injective topology. Since we already know that it is weaker than the projective topology, we are interested to know when these two topologies are the same. This leads to the notion of *nuclearity*.

Definition 15.6. Let E, F be Banach spaces and recall that $L(E, F)$ denotes the Banach space of bounded linear maps from E to F. There is a natural bounded linear map

$$E^* \otimes_\pi F \to L(E, F)$$

coming from the bilinear map $E^* \times F \to L(E, F)$ that sends the pair (ϕ, y) to the linear map $x \mapsto \phi(x)y$. The image of $E^* \otimes_\pi F$ in $L(E, F)$ is called the space of *nuclear* maps from E to F, and is denoted $L^1(E, F)$. The *trace norm* or *nuclear norm* on $L^1(E, F)$ is the norm induced from $E \otimes_\pi F$.

In view of Exercise 14.8, an operator $T\colon E \to F$ is nuclear if and only if there exist sequences $\{\phi_n\}$ in E^* and $\{y_n\}$ in F with $\sum \|\phi_n\|\|y_n\| < \infty$ and

$$Tx = \sum_{n=1}^{\infty} \phi_n(x) y_n.$$

Example 15.7. On a Hilbert space, the nuclear operators coincide with the *trace-class* operators (that is, the linear span of the products of two or more Hilbert-Schmidt operators). Let us show that the product $T = RS$ of two Hilbert-Schmidt operators is nuclear.

Hilbert-Schmidt operators are compact, so S^*S is a compact self-adjoint operator. By the spectral theorem for such operators, there is an orthonormal basis $\{e_i\}$ for H consisting of eigenvectors for S^*S. This implies that the vectors $\{Se_i\}$ are mutually orthogonal, and the Hilbert-Schmidt property tells us that $\sum \|Se_i\|^2 < \infty$.

Since the $\{Se_i\}$ are orthogonal, the Hilbert-Schmidt property of R tells us that we can write $\|RSe_i\| = \lambda_i \|Se_i\|$, with $\sum \lambda_i^2 < \infty$. We conclude that $\sum \|RSe_i\| < \infty$ by the Cauchy-Schwarz inequality. Now the equation

$$Tx = RSx = \sum \langle x, e_i \rangle RSe_i$$

expresses T as a nuclear operator.

Exercise 15.8. Complete the above example by showing that every nuclear operator on H can be expressed as a product of Hilbert-Schmidt operators.

Exercise 15.9. Show by contrast that if E is a Hilbert space the *injective* tensor product $E^* \otimes_\epsilon E$ is the space of all compact operators on E.

Recall (cf. Proposition 6.6) that if E is a LCTVS and p a continuous seminorm, then E_p denotes the Banach space obtained as follows: take $E/\ker(p)$, which is a normed space with p as its norm, and complete to a Banach space. For example, the Sobolev spaces $H^s(X)$ are obtained from $\mathscr{D}(X)$ in exactly this way. Notice that if p, q are two such seminorms and $p \leqslant q$ then there is a natural continuous linear injection $E_q \to E_p$.

Definition 15.10. A LCTVS E is a *nuclear space* if for every continuous seminorm p one can find another continuous seminorm $q \geqslant p$ such that the natural injection $E_q \to E_p$ is a nuclear map.

Our calculations with Sobolev spaces exactly establish that $\mathscr{D}(X)$ has this property (at least when X is a torus).

Theorem 15.11. *If E is a nuclear space and F any LCTVS, then the natural map*

$$E \otimes_\pi F \to E \otimes_\epsilon F$$

is an isomorphism.

As we've already remarked, this result applied to $\mathscr{D}(X)$ will give us the Schwarz kernels theorem.

Proof. (Just a sketch) A preliminary reduction allows one to consider only the case when F is a Banach space. The key claim is then the following: *if $T\colon E_1 \to E_2$ is a nuclear map of Banach spaces, then $T \odot 1 :: E_1 \odot F \to E_2 \odot F$ extends to a continuous linear map*

$$E_1 \otimes_\epsilon F \to E_2 \otimes_\pi F.$$

Notice the different tensor product topologies on the left and the right! Once we have something like this we can use the definition of nuclearity to argue that for any seminorm p on E there is a seminorm q such that there are continuous maps

$$\begin{tikzcd}
E_p \otimes_\pi F \arrow[r] & E_p \otimes_\epsilon F \\
E_q \otimes_\pi F \arrow[u] \arrow[r] & E_q \otimes_\epsilon F \arrow[u] \arrow[ul]
\end{tikzcd}$$

so that the sequence of projective tensor seminorms and the sequence of injective tensor seminorms define the same topology in the limit.

Why is the italicized statement true? Well, recall the formula

$$Tx = \sum_{n=1}^{\infty} \phi_n(x) y_n$$

for a nuclear map between Banach spaces. What this says is that a nuclear map is the limit (in the nuclear norm) of maps that factor through *finite-dimensional* spaces

$$E_1 \longrightarrow \mathbb{C}^N \longrightarrow E_2\ .$$

It suffices then to check the assertion for maps T of this sort. But for such maps there is nothing to check, because the projective and injective tensor products coincide for finite-dimensional spaces

$$E \otimes_\pi \mathbb{C}^N = E \otimes_\epsilon \mathbb{C}^N.$$

(We proved this for $N = 1$ but the case for general finite N follows immediately.)

$\square$

Lecture 16

Differentiation in Topological Vector Spaces

The basic idea of differentiation is that of 'best linear approximation'. Topological vector spaces provide a systematic way to express this. We begin with normed spaces.

Notation 16.1. Let f be a function defined on some ball $B(0;\epsilon)$ in a normed space E and having values in a locally convex topological vector space F. We shall write "$f(h) = o(\|h\|)$" to mean that the limit $\lim_{h\to 0}\big(\|h\|^{-1}f(h)\big)$ exists (in F) and equals zero. Similarly for expressions like "$f(h) = g(h) + o(h)$".

Exercise 16.2. Suppose that $T\colon E \to F$ is a linear map. Show that $T(h) = o(\|h\|)$ if and only if $T = 0$.

Now let $f\colon E \to F$ be a continuous (need *not* be linear) map from a normed space to a LCTVS. In fact, we need only suppose f is defined on some open subset Ω of E.

Definition 16.3. With above notation, let $x \in E$. We say that f is *differentiable* at x if there is a continuous linear map $T\colon E \to F$ such that

$$f(x+h) = f(x) + T\cdot h + o(\|h\|).$$

By the exercise, T is unique if it exists. It is called the *derivative* of f at x and written $Df(x)$.

This definition of the derivative is sometimes called the "Fréchet derivative" to distinguish it from the more general "Gâteaux derivative" which we shall define later. It is then an annoying terminological fact that the Fréchet derivative is usually used in Banach spaces, whereas the Gâteaux derivative is usually used in Fréchet spaces.

Example 16.4. If $E = F = \mathbb{R}$ then every linear map $E \to F$ is multiplication by a scalar, i.e. we identify $\mathcal{L}(E, F) = \mathbb{R}$. Under this identification our definition of the derivative corresponds to the usual one from Calculus I.

Example 16.5. If $E = \mathbb{R}^m$ and $F = \mathbb{R}^n$ then $\mathcal{L}(E; F)$ is the space of $n \times m$ matrices. The matrix entries of the derivative of f are the *partial derivatives* of the components of f as defined in Calculus III. However, the existence of the partial derivatives does not by itself imply differentiability in the sense of our definition above.

Exercise 16.6. Show that the derivative of a *linear* map is always equal to the map itself.

Proposition 16.7. *(Mean value theorem) Suppose that the function f is differentiable throughout a ball $B(x_0; r) \subseteq E$, and that $\|Df(x)\| \leqslant k$ for all $x \in B(x_0; r)$. Then*

$$\|f(x) - f(x_0)\| \leqslant k\|x - x_0\|$$

for all $x \in B(x_0; r)$.

Proof. Let $\epsilon > 0$. By definition of the derivative each $y \in B(x_0; r)$ has a neighborhood $U_y \subseteq B(x_0; r)$ such that

$$\|f(y') - f(y)\| \leqslant (k + \epsilon)\|y' - y\| \quad \forall y' \in U_y.$$

Consider the line segment $[x_0, x] \subseteq B(x_0; r)$. The $\{U_y\}$ form an open cover of this compact metric space, so there is a Lebesgue number $\delta > 0$ for this covering. Now subdivide the line segment into finitely many subintervals $[x_0, x_1]$, $[x_1, x_2]$ and so on with $x_N = x$ for some N and $\|x_i - x_{i+1}\| < \delta$. Using the triangle inequality

$$\|f(x) - f(x_0)\| \leqslant \sum_{i=0}^{N-1} \|f(x_i) - f(x_{i+1})\|$$
$$\leqslant (k + \epsilon) \sum_{i=0}^{N-1} \|x_i - x_{i+1}\| = (k + \epsilon)\|x - x_0\|.$$

Since ϵ was arbitrary, the proof is complete. ☐

Remark 16.8. From the proof one sees that the theorem is true if we replace the ball by any region that is *star-shaped* about x_0.

It is easy to see that the derivative of $\lambda f + \mu g$ is $\lambda Df + \mu Dg$. Another familiar calculus result that generalizes easily is the chain rule.

Proposition 16.9. *(Chain rule) Let E, F, G be normed vector spaces, and let $x \in E$. Let $f\colon E \to F$ be differentiable at x and let $g\colon F \to G$ be differentiable at $y = f(x)$. Then $g \circ f$ is differentiable at x and*

$$D(g \circ f)(x) = Dg(y) \circ Df(x).$$

Note that f need only be defined near x, and g need only be defined near y.

Proof. Let $k(h) = f(x+h) - f(x) = Df(x) \cdot h + o(\|h\|)$. Note that there is a constant A such that $\|k(h)\| \leqslant A\|h\|$ for small $\|h\|$. Now write

$$\begin{aligned} (g \circ f)(x+h) - (g \circ f)(x) &= g(y + k(h)) - g(y) \\ &= Dg(y) \cdot k(h) + o(\|k(h)\|) \\ &= Dg(y) \cdot Df(x) \cdot h + o(\|h\|) \end{aligned}$$

giving the result. □

Remark 16.10. Note that if E, F are both *normed* spaces then the derivative of f, Df, is itself a function having values in a normed space (namely the space $\mathcal{L}(E, F)$). Thus we can differentiate *it* and define second and higher derivatives if we require.

Exercise 16.11. (Open-ended) Formulate precisely the advanced calculus result that “the mixed derivatives are symmetric”, that is $\partial^2 f/\partial x \partial y = \partial^2 f/\partial y \partial x$, and prove it (under suitable hypotheses) for maps between normed vector spaces.

Definition 16.12. Let X be a metric space. A mapping $f\colon X \to X$ is a *(strict) contraction* if there is a constant $a < 1$ such that $d(f(x), f(x')) \leqslant ad(x, x')$ for all $x, x' \in X$.

Note that a contraction must be continuous.

Theorem 16.13. *(Banach) A strict contraction on a complete metric space has a unique* fixed point *(a point x such that $f(x) = x$).*

Proof. A fixed point is unique because if x, x' are two such then $d(x, x') = d(f(x), f(x')) \leqslant ad(x, x')$, which implies $d(x, x') = 0$.

To prove existence, start with any $x_0 \in X$ and define $x_1 = f(x_0)$, $x_2 = f(x_1)$ and so on. If $d(x_0, x_1) = r$ then $d(x_n, x_{n+1}) \leqslant a^n r$ and so

$$d(x_n, x_{n+k}) \leqslant (a^n + \cdots + a^{n+k-1})r \leqslant \frac{a^n r}{1-a}$$

which tends to 0 as $n \to \infty$. Thus (x_n) is a Cauchy sequence, which converges to a point x. We have

$$f(x) = \lim f(x_n) = \lim x_{n+1} = x$$

so x is a fixed point. □

Exercise 16.14. A (non-strict) contraction on X is defined to be a map $f\colon X \to X$ such that $d(f(x), f(x')) < d(x, x')$ for all $x \neq x'$. Show that there exist non-strict contractions on *complete* metric spaces that have no fixed points, but that any non-strict contraction on a *compact* metric space has a unique fixed point.

Let E, E' be normed vector spaces. The space $\mathcal{L}(E, E')$ is then a normed vector space also. Say $T \in \mathcal{L}(E, E')$ is *invertible* if it has an inverse which is a *bounded* linear map from E' to E. (If E, E' are Banach spaces the Closed Graph Theorem shows that this is just the same as saying that T is bijective.)

Proposition 16.15. *Let E, E' be* Banach *spaces. Then the invertibles form an open subset of $\mathcal{L}(E, E')$, and the inverse operation $T \mapsto T^{-1}$ is continuous (on this subset) from $\mathcal{L}(E, E')$ to $\mathcal{L}(E', E)$.*

Proof. Without loss of generality $E' = E$. Look first at a neighborhood of the invertible operator I. Suppose $\|S\| < 1$. For $y \in E$ consider the map

$$\phi_y\colon E \to E, \qquad x \mapsto y - Sx.$$

Since $\|S\| < 1$ this map is a strict contraction, so it has a unique fixed point x (Theorem 16.13). This fixed point is an x such that $(I + S)x = y$. By the triangle inequality,

$$\|y\| \geqslant (1 - \|S\|)\|x\|, \quad \text{so } \|x\| \leqslant (1 - \|S\|)^{-1}\|y\|$$

and the map $y \mapsto x$ is bounded. We have shown that if $\|S\| < 1$, $I+S$ is invertible. Moreover

$$\|y - (I + S)^{-1}y\| = \|S(1 + S)^{-1}y\| \leqslant \|S\|(1 - \|S\|)^{-1}\|y\|$$

which shows that the map $S \mapsto (I + S)^{-1}$ is continuous at $S = 0$.

We have shown that the identity is an interior point of the set of invertibles and that the inverse is continuous there. However, left multiplication by a fixed invertible is a homeomorphism from the set of invertibles to itself; so the same results apply to any point of the set of invertibles. ☐

Exercise 16.16. Let E be a normed space and let F be the normed space $\mathcal{L}(E, E)$. Show that the mapping $i\colon T \mapsto T^{-1}$ is differentiable (where defined) on F, and that its derivative is

$$Di(T) \cdot H = -T^{-1}HT^{-1}.$$

Definition 16.17. A map f from an open subset of a normed space E to a normed space F is *continuously differentiable* or C^1 if it is differentiable (everywhere) and the map $x \mapsto Df(x)$ is continuous.

The inverse function theorem says that under suitable conditions, if $Df(x)$ is invertible then f is 'locally invertible' near x.

Theorem 16.18. *Let f be as above, defined near $x \in E$, and suppose that f is continuously differentiable and that $Df(x) \in \mathcal{L}(E, F)$ is invertible. Suppose moreover that E, F are* Banach *spaces. Then there is an open set U containing x such that f is a bijection of U onto $f(U)$, which is an open set containing $f(x)$, and such that its inverse $g\colon f(U) \to U$ is also continuously differentiable.*

Proof. It is an application of Banach's fixed point theorem 16.13. We will construct the inverse function by looking at the fixed points of a suitable map.

Let $A = Df(x)^{-1}$. For $y \in F$ define a map $\phi_y\colon E \to E$ by

$$\phi_y(z) = z + A \cdot (y - f(z)).$$

A fixed point of ϕ_y is a solution to $f(z) = y$. We have

$$D\phi_y(z) = I - A \circ Df(z) = A \circ \big(Df(x) - Df(z)\big).$$

Since Df is continuous there is $r > 0$ such that if $z \in U = B(x; r)$ then $\|D\phi_y(z)\| < \frac{1}{2}$.

Fix $z_0 \in U$ and let $y_0 = f(z_0)$. Choose $\delta > 0$ such that the closed ball $\overline{B}(z_0; \delta)$ is contained in U. Choose $\epsilon = \frac{1}{2}\|A\|^{-1}\delta$. If $z \in \overline{B}(z_0; \delta)$ and $y \in B(y_0; \epsilon)$ then we may compute

$$\begin{aligned} \|\phi_y(z) - z_0\| &= \|\phi_y(z) - \phi_{y_0}(z_0)\| \\ &\leqslant \|\phi_y(z) - \phi_y(z_0)\| + \|\phi_y(z_0) - \phi_{y_0}(z_0)\| \\ &\leqslant \tfrac{1}{2}\|z - z_0\| + \|A(y - y_0)\| \leqslant \tfrac{1}{2}\delta + \tfrac{1}{2}\delta = \delta \end{aligned}$$

using the mean value theorem 16.7. We conclude that, for these y, the map ϕ_y is a contraction of the *complete* metric space $\overline{B}(z_0; \delta)$ and thus has a (unique) fixed point there. We have shown that each $y_0 \in f(U)$ has an ϵ-neighborhood contained in $f(U)$; thus, $f(U)$ is open. The contraction property of the ϕ_y ensures that each $y \in f(U)$ has only one inverse image in U; thus f is a bijection of U onto $f(U)$.

Now to show that the inverse function $g = f^{-1} \colon f(U) \to U$ is differentiable at $y = f(x)$ write $g(y + h) - g(y) = u(h)$, say. The calculations of the previous paragraph show that $\|u(h)\| \leqslant 2\|A\|\|h\|$ for $\|h\|$ small. Then

$$h = f(g(y + h)) - y = f(g(y) + u(h)) - y = Df(x) \cdot u(h) + o(\|h\|).$$

Apply $A = Df(x)^{-1}$ to get

$$u(h) = A \cdot h + o(\|h\|).$$

This gives differentiability of g with $Dg(f(x)) = Df(x)^{-1}$. *Continuous* differentiability now follows from the continuity of the inverse operation on the space of bounded linear maps (Proposition 16.15). ☐

Remark 16.19. The chain rule shows that the derivative of the (local) inverse map to f is the inverse of the derivative of f: this is the multi-variable counterpart of the one-variable fact that dx/dy is the reciprocal of dy/dx. It follows from this that if f is *smooth* (infinitely differentiable), then its local inverse is smooth also.

Lecture 17

Applications and generalizations

(The first discussion is about the jacobian formula for change of variable in multiple integrals. I don't really know whether I'll lecture on this material, but it is a nice application of the inverse function theorem.)

Proposition 17.1. *Let $T\colon \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation. Then, for every Lebesgue measurable set E, the set $T(E)$ is also Lebesgue measurable, and*

$$\lambda(T(E)) = |\det T|\lambda(E).$$

Proof. If the result holds for linear transformations T_1 and T_2, it also holds for their composite T_1T_2. Now, by the theory of elementary row operations, every linear transformation can be written as a composite of elementary operations: multiplying a coordinate by a scalar, interchanging two coordinates, or adding one coordinate to another one. For each of these the result follows from Fubini's theorem together with the translation-invariance and scaling properties of one-dimensional Lebesgue measure.

For example, check this for elementary operations of the third type, such as $(x_1, \dots, x_n) \mapsto (x_1 + x_2, x_2, \dots, x_n)$. If $f(x_1, \dots, x_n)$ is the characteristic function of E, then $f(x_1 - x_2, x_2, \dots, x_n)$ is the characteristic function of $T(E)$. According to Fubini's theorem, then, we need to check that

$$\int \cdots \int f(x_1, \dots, x_n) dx_1 \cdots dx_n = \int \cdots \int f(x_1 - x_2, \dots, x_n) dx_1 \cdots dx_n.$$

This is true because for any positive measurable function g

$$\int g(x_1) dx_1 = \int g(x_1 - x_2) dx_1$$

by the translation invariance of Lebesgue measure. □

Now consider more general coordinate transformations. Let U, U' be open subsets of $\mathbb{R}^n$. A *diffeomorphism* from U to U' is a continuously differentiable bijection $f\colon U \to U'$ whose inverse is also continuously differentiable.

Definition 17.2. Let $f\colon U \to U'$ be a diffeomorphism. The *Jacobian* of f is the real-valued function J_f on U defined by $J_f(x) = \det Df(x)$, where $Df(x)$ is the derivative of f at x (which is a linear transformation $\mathbb{R}^n \to \mathbb{R}^n$).

We are going to prove

Theorem 17.3. *Let $f\colon U \to U'$ be a diffeomorphism. Then*

$$\lambda(f(U)) = \int_U |J_f(x)| d\lambda(x).$$

More generally, for any $g \in L^1(U')$,

$$\int_{U'} g(y) d\lambda(y) = \int_U g(f(x))|J_f(x)| d\lambda(x).$$

Example 17.4. For $n = 1$ (functions of a single variable), and taking f to be increasing (for simplicity of notation) the theorem says

$$\int_{f(a)}^{f(b)} g(y)dy = \int_a^b g(f(x))f'(x)dx$$

which is the usual formula for ‘integration by substitution’. Recall: this can be proved for continuous g by showing that both sides have the same derivative with respect to b, then extended to all $g \in L^1$ by the usual approximation.

Example 17.5 (Polar coordinates)**.** For an integrable function g on $\mathbb{R}^2$,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)dxdy = \int_0^{2\pi}\int_0^{\infty} g(r\cos\theta, r\sin\theta)rdrd\theta.$$

Exercise 17.6. Let $I = \int_{-\infty}^{\infty} e^{-x^2}dx$. By writing

$$I^2 = \iint e^{-x^2-y^2}dxdy$$

in polar coordinates, show that $I = \sqrt{\pi}$.

Proof of Theorem 17.3. If the theorem is true for two diffeomorphisms f_1 and f_2, it is also true for their composite. As with the special case of linear transformations (Proposition 17.1) this allows us to prove the result in special cases and build up the general case by composition. A key role is played by the Inverse Function Theorem 16.18.

Notice also that it is sufficient to prove the theorem *locally*: that is, to show that every point $x \in U$ has a neighborhood U_x such that the theorem holds for all functions g supported in U_x. This is because any nonnegative function g can be broken up (using a partition of unity) into a countable sum of nonnegative functions g_j each of which is supported in one of the sets U_{x_j}. By the monotone convergence theorem, the truth of the theorem for each g_j implies its truth for g.

Suppose that f is a diffeomorphism of the special form

$$(x_1, \ldots, x_n) \mapsto \big(h(x_1, \ldots, x_n), x_2, \ldots, x_n\big)$$

where $h\colon \mathbb{R}^n \to \mathbb{R}$ is a smooth function. (We will call such a diffeomorphism 'triangular'.) The Jacobian of f is then just the partial derivative $\partial h/\partial x_1$. According to Fubini's theorem, then, we need to check that

$$\int \cdots \int_{U'} g(x_1, \ldots, x_n) dx_1 \cdots dx_n = \\ \int \cdots \int_{U} g(h(x_1, \ldots, x_n), \ldots, x_n) |\partial h/\partial x_1| dx_1 \cdots dx_n.$$

This follows from the formula for integration by substitution (Example 17.4) applied in the innermost iterated integral. Thus Theorem 17.3 holds for triangular diffeomorphisms.

To complete the proof it suffices to show that any diffeomorphism is locally a composite of triangular diffeomorphisms and linear transformations. (*Locally* means that every $x \in U$ has an open neighborhood on which this is true.) This follows from the inverse function theorem. We may assume that we are looking locally near the point 0, and that $f(0) = 0$. By composing with the inverse of the linear transformation $Df(0)$, we may further assume that $Df(0) = I$, the identity transformation.

Write $f(x_1, \ldots, x_n) = (y_1, \ldots, y_n)$, where each $y_i\colon \mathbb{R}^n \to \mathbb{R}$ is a smooth function of n variables, and $\partial y_i/\partial x_j(0) = \delta_{ij}$. By the inverse function theorem, the smooth maps

$$\phi_i\colon (x_1, \ldots, x_n) \mapsto (y_1, \ldots, y_i, x_{i+1}, \ldots, x_n)$$

are all local diffeomorphisms near x. The diffeomorphisms

$$\phi_1, \quad \phi_2 \circ (\phi_1)^{-1}, \quad \phi_3 \circ (\phi_2)^{-1}, \ \cdots$$

are all triangular, and their composite is f. Thus, Theorem 17.3 holds for f, as required. □

Weaker notions of differentiation

Our notion of differentiability for a map $E \to F$ from a normed space to a LCTVS is the strongest reasonable one. Weaker ones can be devised in the following way. Let Ω be an open subset of a normed space E, let $x_0 \in \Omega$, let $f\colon \Omega \to F$ be a continuous map to a LCTVS, and consider the various composite maps that can be formed in the diagram

$$\begin{tikzcd} \Bbbk \arrow[r, "\phi_v"] & \Omega \arrow[r, "f"] & F \arrow[r, "\psi"] & \Bbbk \end{tikzcd}$$

Here $\phi_v\colon \Bbbk \to E$ is a map of the form $t \mapsto x_0 + tv$, $v \in E$, which maps a 0-neighborhood in $\Bbbk$ into Ω, and ψ is a linear functional in F^*. Differentiability in our (Fréchet) sense implies differentiability of all these composite maps.

Definition 17.7. Let E, F be LCTVS and let f map an open subset $\Omega \subseteq E$ to F. Let $T\colon E \to F$ be a continuous linear map. One says that T is the *Gâteaux derivative* of f at $x_0 \in E$ if, for each $v \in E$, the map $\Bbbk \to F$ defined by

$$t \mapsto f(x_0 + tv)$$

is differentiable with derivative $T \cdot v$.

The advantage of this definition is that it makes sense when E is not a normed space (and, in particular, when E is a Fréchet space). However the definition does not enjoy all the good properties of the Fréchet derivative. In particular, a function that is Gâteaux differentiable at x_0 need not even be *continuous* at x_0.

Example 17.8. Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$f(x,y) = \frac{x^3 y}{x^6 + y^2}, \qquad f(0,0) = 0.$$

This map is Gâteaux differentiable at $(0,0)$, with derivative zero. In fact, for (x,y) fixed, $f(tx,ty) = 0$ if $y = 0$, and $f(tx,ty) = O(t^2)$ if $y \neq 0$. But $f(t,t^3) = \frac{1}{2}$ for all $t \neq 0$, so f is not continuous at the origin.

Example 17.9. Let E be an infinite-dimensional normed space and let $\phi\colon E \to \Bbbk$ be a discontinuous linear functional. Then $f(x) = \|x\|\phi(x)$ is Gâteaux differentiable (with derivative 0) but not Fréchet differentiable.

There is some disagreement about what it should mean for a function to be *continuously Gâteaux differentiable.* We shall say that f is continuously Gâteaux differentiable in Ω if it is differentiable at every point of Ω and the Gâteaux derivative Df is continuous as a function from $\Omega \times E$ (product topology) to F. Note that if E, F are normed this is a weaker continuity condition than requiring that Df be continuous as a map from Ω to $\mathcal{L}(E, F)$ (see the exercise below).

Exercise 17.10. Let E be the Banach space of continuous functions on the circle, and let $L_t\colon E \to E$ be the operator of translation by $t \in \mathbb{T}$. Show that the map $(t, f) \mapsto L_t f$ is continuous from $\mathbb{T} \times E$ to E, but the map $t \mapsto L_t$ is *not* continuous from $\mathbb{T}$ to $\mathcal{L}(E, E)$.

We say that f is r *times Gâteaux differentiable* at x_0 if there is a symmetric r-linear map $T\colon E \times \cdots \times E \to F$ with the property that,for each $v \in E$, the map $\Bbbk \to F$ defined by

$$t \mapsto f(x_0 + tv)$$

is r times differentiable at x_0 and has derivative $T(v, \ldots, v)$. (By the polarization identity, there can be at most one T with this property.) Similarly for the definition of a *Gâteaux* C^r map defined in a domain—that's a map that is s times Gâteaux differentiable for all $s \leqslant r$ and such that all the Gâteaux derivatives are continuous as maps $\Omega \times R \times \cdots \times E \to F$.

Proposition 17.11. *A Gâteaux C^r map between normed vector spaces is Fréchet C^{r-1}.*

Proof. We'll just consider the case $r = 2$. Let the map be $f\colon \Omega \to F$, using the same notation as in the discussion above.

The second (Gâteaux) derivative $D^2 f$ is a continuous map $\Omega \times E \times E \to F$, symmetric bilinear in the last two variables. By definition of continuity and the product topology, there is a convex neighborhood U of x_0 and an $\epsilon > 0$ such that $D^2 f$ maps $U \times B(0; \epsilon) \times B(0; \epsilon)$ into the unit ball in F. It follows that for all $x \in U$, the symmetric bilinear map $D^2 f(x)$ is bounded ($\|D^2 f(x)(v, w)\| \leqslant \epsilon^{-2}\|v\|\|w\|$).

Now let $x \in U$ and apply the mean value theorem to the differentiable map $t \mapsto Df(x_0 + th)$, where $h = x - x_0$, $t \in [0, 1]$, to see that

$$\|Df(x) - Df(x_0)\| \leqslant \epsilon^{-2}\|x - x_0\|.$$

In particular, this gives us the continuity of Df (at x_0) as a map $U \to \mathcal{L}(E, F)$. Finally, apply the mean value theorem once again, this time to the map

$$t \mapsto f(x_0 + th) - f(x_0) - tDf(x_0) \cdot h, \quad t \in [0, 1]$$

The derivative of this map is $(Df(x_0+th)-Df(x_0))\cdot h$, so the mean value theorem gives

$$\|f(x_0+h)-f(x_0)-Df(x_0)\cdot h\| \leqslant \epsilon^{-2}\|h\|^2$$

which proves Fréchet differentiability. $\square$

Instead of precomposing with standard maps into E, we can postcompose with standard maps out of F. This leads to the definition of *weak differentiability*.

Definition 17.12. Let E be a normed space, F a LCTVS, and let f map an open subset $\Omega \subseteq E$ to F. Let $T\colon E \to F$ be a continuous linear map. One says that T is the *weak (Fréchet) derivative* of f at $x_0 \in E$ if, for each $\psi \in F^*$, the map $\psi \circ f\colon \Omega \to \Bbbk$ is differentiable at x_0, with derivative $\psi \circ T$.

The hypothesis of local convexity ensures that F^* separates points of F, so the weak derivative is unique if it exists. You can also combine all the previous ideas and define *weak Gâteaux differentiability*, *weak* C^r and so on if you want.

Lemma 17.13. *Let E be a finite-dimensional normed space, let F be a complete LCTVS, let $\Omega \subseteq E$ be open, and let $x_0 \in \Omega$. Let $f\colon \Omega\setminus\{x_0\} \to F$ be a continuous map having the property that, for each $\psi \in F^*$, the map $\psi \circ f$ extends to a continuously differentiable map on all of Ω. Then one can extend f to a continuous map defined on all of Ω.*

Proof. Let $K \subseteq \Omega$ be a compact neighborhood of x_0. Let $\psi \in F^*$. Since $D(\psi\circ f)$ is continuous, hence bounded, on K, the Mean Value Theorem shows that the function

$$(x,y) \mapsto \frac{\psi(f(x))-\psi(f(y))}{\|x-y\|}$$

is bounded on $K \times K$. Thus, by definition, the function

$$(x,y) \mapsto \frac{f(x)-f(y)}{\|x-y\|}$$

is weakly bounded on $(K\setminus\{x_0\})\times(K\setminus\{x_0\})$.

It follows by Proposition 6.6 that this function is bounded for the original topology of F. That is to say, for any 0-neighborhood U in F, there exists $r>0$ such that

$$f(x)-f(y) \in r\|x-y\|U$$

for all $x, y \in (K \setminus \{x_0\})$. Now let $\{x_n\}$ be any sequence in K converging to x_0. The above calculation shows that the $\{f(x_n)\}$ form a Cauchy sequence in F, converging to a limit a that is independent of the sequence $x_n \to x_0$ chosen. Defining $f(x_0) = a$ now gives the desired continuous extension. □

Proposition 17.14. *Suppose that a map $f\colon \Omega \to F$, with F a complete LCTVS, is weakly Gâteaux C^k; then it is Gâteaux C^{k-1}.*

Proof. Take $k = 2$, as before. Since the entire discussion is about Gâteaux differentiation we may as well assume that E is the ground field $\Bbbk$. Consider the difference quotient

$$g\colon x \mapsto \frac{f(x) - f(x_0)}{x - x_0}$$

defined on $\Omega \setminus \{x_0\}$. By the hypothesis that f is weakly C^2, $\psi \circ g$ extends to a C^1 function on Ω for each $\psi \in F^*$. Thus, by the lemma, g extends to a continuous function on Ω. That is to say, f is differentiable at x_0. □

Lecture 18

Degree Theory

Banach's contraction mapping principle is one important way to prove existence theorems of various kinds. But there are also fixed point theorems which can be applied to mappings that are *not* contractive. Generally, these depend on more sophisticated topological ideas such as those of *degree theory*.

In its simplest form, degree theory applies to continuous maps $f\colon S^n \to S^n$ from the n-sphere to itself. To each such map is associated an integer $\deg(f)$, the *degree* of f, with the following properties:

(a) The degree of the identity map is 1; the degree of a constant map is 0.

(b) Homotopic maps have the same degree (two maps $f_0, f_1 : S^n \to S^n$ are *homotopic* if there is a map $F\colon S^n \times [0,1] \to S^n$ with $F(\cdot, t) = f_t(\cdot)$ for $t = 0, 1$).

(c) The degree is additive: if $f, g\colon S^n \to S^n$ are maps and $f + g$ is defined to be the composite

$$\begin{tikzcd} S^n \arrow[r] & S^n \vee S^n \arrow[r, "f\vee g"] & S^n \end{tikzcd}$$

(where the first map collapses the equator of S^n to a point) then $\deg(f+g) = \deg(f) + \deg(g)$.

These statements can be summarized by saying that the degree gives a surjective homomorphism of groups from $\pi_n(S^n)$ to $\mathbb{Z}$. A more sophisticated result of degree theory is that this homomorphism is actually an *isomorphism*. We shall not need this more sophisticated result.

The existence of the degree can be proved by the methods of algebraic topology. A map $f\colon S^n \to S^n$ induces a homomorphism $f_*\colon H_n(S^n;\mathbb{Z}) \to H_n(S^n;\mathbb{Z})$ on homology groups. Since $H_n(S^n;\mathbb{Z}) \cong \mathbb{Z}$, this homomorphism is given by multiplication by a well-defined integer. That integer is the degree of f. The properties (a)–(c) above follow from standard facts about homology theory.

An alternative approach to degree theory can be based on the general position properties of differentiable maps. We will discuss this in the next lecture. For now, we concentrate on making use of degree theory in functional analysis.

Theorem 18.1. *(Brouwer fixed point theorem) Let $K \subseteq \mathbb{R}^n$ be a compact convex set and $f\colon K \to K$ a continuous map. Then there exists $x \in K$ such that $f(x) = x$.*

Proof. First consider the case where K is a closed ball B centered at the origin and bounded by an $(n-1)$-sphere S. Supposing that f has no fixed point, for each $x \in B$ there is a unique ray originating at $f(x)$ and passing through x, and this ray meets S at a unique point which we denote $g(x)$. The map $g\colon B \to S$ is continuous and (by construction) $g(x) = x$ if $x \in S$. But then the family of maps $[0,1] \times S \to S$, $(t,x) \mapsto g(tx)$, give a homotopy between the identity map on S and a constant map, contradicting properties (a) (b) of the degree.

Now for the general case, since $\mathbb{R}^n$ is a Hilbert space, for any $x \in \mathbb{R}^n$ there is a unique nearest point $h(x) \in K$. The map h is continuous (in fact $d(h(x), h(y)) \leqslant d(x,y)$), and it is the identity on K. Enclose K in a closed ball B of sufficiently large radius, and consider the map $f \circ h\colon B \to B$. It has a fixed point x (by the special case already proved). We must have $x \in K$, so $h(x) = x$ and x is a fixed point of f as well. ☐

Notice that we can replace $\mathbb{R}^n$ by any finite-dimensional Hausdorff LCTVS here, since each such space is linearly homeomorphic to $\mathbb{R}^n$.

Theorem 18.2. *(Schauder fixed point theorem) If K is a compact convex subset of a Hausdorff LCTVS E, and $f\colon K \to K$ is continuous, then f has a fixed point.*

Proof. It suffices to show that for every convex, balanced, open 0-neighborhood V there exists $x_V \in K$ such that $f(x_V) \in x_V + V$. (The $\{x_V\}$ then form a net, which has a convergent subnet by compactness, and a limit of such a subnet is a fixed point of f.)

Let p be the Minkowski functional of V and let $q(x) = \max\{0, 1 - p(x)\}$ so that $q(x) > 0$ iff $x \in V$. Choose points $x_1, \dots, x_n \in K$ so that the open sets $x_1 + V, \dots, x_n + V$ form a finite cover of K and define $q_i(x) = q(x - x_i)$ and

$$g_i(x) = \frac{q_i(x)}{q_1(x) + \cdots + q_n(x)}.$$

The map

$$g(x) = \sum_{i=1}^{n} g_i(x) x_i$$

maps K continuously into the convex hull H of the points $x_1, \dots, x_n$, which is a convex subspace of a finite-dimensional Hausdorff LCTVS. Moreover,

$$x - g(x) = \sum_{i=1}^{n} g_i(x)(x - x_i)$$

is a convex combination of points of V, hence must itself belong to V.

Thus, $g \circ f$ has a fixed point $x \in H$, by the Brouwer fixed point theorem. As observed above, $g(f(x)) - f(x) \in V$; that is, $f(x) \in x + V$, as required. $\square$

Example 18.3. Here is a very sketchy discussion of how the Schauder fixed point theorem can be applied to prove existence theorems for nonlinear PDE (for more details consult Gilbarg and Trudinger, *Elliptic Partial Differential Equations of Second Order*, Chapter XI). Let Ω be a smoothly bounded domain in $\mathbb{R}^n$ and consider the Dirichlet problem for a *quasilinear elliptic equation*

$$Qu := \sum_{i,j} a^{ij}(x,u,Du)D_{ij}u + b(x,u,Du) = 0, \quad u_{|\partial\Omega} = \phi$$

where the coefficient matrix $a^{ij}(x,z,p)$ is positive for all $(x,z,p) \in \overline{\Omega}\times\mathbb{R}\times\mathbb{R}^n$. A solution to this problem may be regarded as a fixed point of the nonlinear operator T which sends a function v to the unique solution of the *linear* Dirichlet problem

$$\sum_{i,j} a^{ij}(x,v,Dv)D_{ij}u + b(x,v,Dv) = 0, \quad u_{|\partial\Omega} = \phi.$$

The theory of linear elliptic equations can be used to show that T is a compact map (that is, it maps bounded sets to relatively compact sets) on a suitable Banach space E (for example the Hölder space $C^{1,\beta}(\overline{\Omega})$. Suppose additionally that one can establish an *a priori estimate* that every solution of $Tu = \lambda u$, $\lambda \in [1,\infty)$, satisfies $\|Tu\|_E \leqslant M$. Define then $S \colon E \to E$ by

$$Su = Tu/\max\{1, M^{-1}\|Tu\|\}.$$

S maps the closed ball of radius M in E into itself, and the closure of the range of S is compact. Hence by Schauder's fixed point theorem, there exists a fixed point u of S. If we assume that $\|Tu\| > M$ then the identity $Tu = M^{-1}\|Tu\|Su = M^{-1}\|Tu\|u$ gives a contradiction to the a priori estimate. Hence $\|Tu\| \leqslant M$ and $Tu = u$, as required.

We will now prove the fixed point theorem of Darbo which is a common generalization of the Banach and Schauder fixed point theorems.

Definition 18.4. Let C be a bounded subset of a complete metric space X. The *Kuratowski measure (of noncompactness)* of C, denoted $\mathfrak{k}(C)$, is the infimum of all those r for which C admits a finite cover by sets of diameter r.

Note that $\mathfrak{k}(C) = 0$ iff C is precompact (whence compact if it is closed, since X is complete).

Exercise 18.5. Show that the Kuratowski measure of C is the same as the Kuratowski measure of its closure $\overline{C}$.

Exercise 18.6. In a normed vector space prove that $\mathfrak{k}(A + B) \leqslant \mathfrak{k}(A) + \mathfrak{k}(B)$.

Proposition 18.7. *The Kuratowski measure of a bounded subset C of a Banach space is the same as the Kuratowski measure of its convex hull.*

Proof. Let $\mathfrak{k}(C) = r$. Let $\epsilon > 0$. There is a finite cover of C by sets $U_1, \ldots, U_n$ of diameter less than $r + \epsilon$. We may as well assume that the sets U_i are convex, because the diameter of a subset of a normed vector space is the same as the diameter of its convex hull (proof?).

Let Δ be the simplex $\{(\lambda_1, \ldots, \lambda_m) : \lambda_i \geqslant 0, \ \sum \lambda_i = 1\} \subseteq \mathbb{R}^n$. For each $\lambda = (\lambda_1, \ldots, \lambda_n) \in \Delta$ let

$$U_\lambda = \lambda_1 U_1 + \cdots + \lambda_n U_n.$$

Each U_λ is convex and of diameter less than $r + \epsilon$. Moreover $\bigcup_{\lambda \in \Delta} U_\lambda$ is convex and contains C, so it contains the convex hull of C.

Let V_λ be the ϵ-neighborhood of U_λ (i.e., $V_\lambda = \bigcup_{x \in U_\lambda} B(x; \epsilon)$. Each V_λ has diameter $\leqslant r + 3\epsilon$. Moreover, there is $\delta > 0$ such that if $\|\lambda - \lambda'\| < \delta$ then $U_{\lambda'} \subseteq V_\lambda$. By compactness of the simplex Δ there are finitely many sets V_λ that cover all the U'_λ, and therefore cover the convex hull of C. Thus the convex hull of C is covered by finitely many sets of diameter at most $r + 3\epsilon$. Since ϵ is arbitrary this completes the proof. □

Proposition 18.8. *Suppose that $C_0 \supseteq C_1 \supseteq C_2 \supseteq \ldots$ is a decreasing sequence of nonempty closed bounded subsets of a complete metric space X with $\mathfrak{k}(C_n) \to 0$ as $n \to \infty$. Then $C_\infty := \bigcap C_n$ is compact and nonempty.*

This generalizes the *finite intersection property* for compact sets.

Proof. Choose a sequence of points $x_n \in C_n$, and let $\{y_i\}_{i \in I}$ be a universal subnet. Notice that there is a sequence $r_n \to 0$ such that each C_n has a finite cover $\mathscr{U}_n$ by sets of diameter $\leqslant r_n$. For each such n, the properties of universal nets show that $\{y_i\}$ is eventually in one of the members of the cover $\mathscr{U}_n$. It follows that $\{y_i\}$ is a Cauchy net, hence it converges to a member of C_∞ which is therefore nonempty. Since C_∞ is the intersection of closed sets, it is closed; and since $\mathfrak{k}(C_\infty) = 0$, it is compact. □

Exercise 18.9. We're in a countable situation in this proof, so we don't really need to use the existence of universal nets. Rewrite the proof to use an inductive construction of subsequences and a diagonal argument.

Definition 18.10. Let X be a metric space and let $F\colon X \to X$ be a (continuous) map. The map F is called *condensing* if $\mathfrak{k}(F(B)) \leqslant \mathfrak{k}(B)$ for all bounded subsets B, with strict inequality if $\mathfrak{k}(B) > 0$.

Notice that any strict contraction, as well as any compact map, is condensing. This shows that the following result simultaneously generalizes the fixed point theorems of Banach and Schauder.

Theorem 18.11. *(Darbo's fixed point theorem) Let E be a Banach space and C a nonempty closed bounded convex subset of E. Let $F\colon C \to C$ be a condensing map. Then F has a fixed point.*

Proof. Suppose first that F is setwise contractive: that is, there is a constant $a < 1$ with $\mathfrak{k}(F(B)) \leqslant a\mathfrak{k}(B)$. Define a decreasing sequence of closed sets as follows: $C_0 = C$, and inductively C_{n+1} is the closed convex hull of $F(C_n)$. By Propositions 18.7 and 18.8, the intersection $C_\infty = \bigcap C_n$ is a nonempty closed convex compact set. The map F sends C_∞ to itself, so by Schauder's theorem (18.2) it has a fixed point there.

In the general case, assume without loss of generality that $0 \in C$ and let $F_n(x) = (1 - 1/n)F(x)$. The first part of the proof applies to F_n and shows that it has a fixed point $x_n \in C$. We have

$$F(x_n) - x_n = F(x_n) - F_n(x_n) = (1/n)F(x_n) \to 0.$$

Let A be the set $\{-F(x_n) + x_n\}$; clearly $\mathfrak{k}(A) = 0$. Let $S = \{x_n\}$; then

$$S \subseteq F(S) + A$$

and so $\mathfrak{k}(S) \leqslant \mathfrak{k}(F(S)) + \mathfrak{k}(A) = \mathfrak{k}(F(S))$. Since F is condensing, this implies that S is precompact, so the sequence $\{x_n\}$ has a Cauchy subsequence. Let x be the limit of such a subsequence: then x is a fixed point of F. $\square$

Lecture 19
Degree Theory and General Position

In this lecture we are going to prove the results of degree theory by the methods of differential topology. Begin with a lemma which is a simple version of the multivariable Taylor's Theorem (the general version can be proved in the same way).

Lemma 19.1. *Suppose that E, F are Banach spaces, with E finite-dimensional, and that $f\colon \Omega \to F$ is of class C^{k+1}, where Ω is an open subset of E. Suppose that, at a point $x \in \Omega$, $D^r f(x) = 0$ for $0 \leqslant r \leqslant k$. Then for each closed ball $B = \overline{B}(x; r) \subseteq \Omega$ there is a constant $C > 0$ such that*

$$\|f(y)\| \leqslant C\|x - y\|^{k+1}$$

for all $y \in B$.

Proof. By compactness the $(k+1)$st derivative of f is bounded on B (as a multilinear map $E^{k+1} \to F$), say by M. By the Hahn-Banach theorem there is $\phi \in F^*$ of norm 1 such that $\phi(f(y)) = \|f(y)\|$. Apply the usual (one-variable) version of Taylor's theorem to the function $g\colon t \mapsto \phi(f((1-t)x + ty))$, $t \in [0, 1]$. We get

$$\|f(y)\| = |g(1)| \leqslant \frac{1}{(k+1)!} \sup_{t\in[0,1]} g^{(k+1)}(t) \leqslant \frac{M}{(k+1)!}$$

which gives the result. $\square$

The key result here is *Sard's Theorem.*

Theorem 19.2. *Let $f\colon \mathbb{R}^m \to \mathbb{R}^n$ be a smooth (C^∞) map and let C_f (the* critical set*) be the set of those points $x \in \mathbb{R}^m$ for which the linear map $Df(x)$ is not surjective. Then $f(D_f)$ has measure 0 in $\mathbb{R}^n$.*

For example, if f is a constant map, then C_f is the whole of $\mathbb{R}^m$, but $f(C_f)$ is a single point of $\mathbb{R}^n$.

Proof. The proof is by induction on m, the base case $m = 0$ being trivial. Assume the result true for maps $\mathbb{R}^k \to \mathbb{R}^n$ with $k \leqslant m - 1$.

Define nested subsets $C_f = C_0 \supseteq C_1 \supseteq C_2 \supseteq \ldots$ as follows: for $r \geqslant 1$, C_r is the set of those x such that $D^k f(x) = 0$ for all $k \leqslant r$ (in other words, all the partial derivatives of f of order up to r vanish at x). We'll prove that $f(C_r \setminus C_{r+1})$ has measure 0 for all r and that $f(C_r)$ has measure 0 for r sufficiently large. The first of these assertions uses the inductive hypothesis, and the second does not.

Proof that $f(C_0 \backslash C_1)$ has measure 0 It suffices to show that each $x \in C_0 \backslash C_1)$ has a neighborhood U_x such that $f(C_f \cap U_x)$ has measure 0. Since $x \notin C_1$, $f = (f_1, \ldots, f_m)$ has some partial derivative that does not vanish at x; reorder coordinates so that it is $\partial f_1/\partial x_1$. By the inverse function theorem, the map

$$h \colon (x_1, \ldots, x_m) \mapsto (f_1(x_1, \ldots, x_m), x_2, \ldots, x_m)$$

is a diffeomorphism of some neighborhood U of x onto an open subset $U' \subseteq \mathbb{R}^m$. It suffices to prove the result for $g = f \circ h^{-1} \colon U' \to \mathbb{R}^n$, since g (restricted to U') has exactly the same critical values as f (restricted to U). But, by construction, g is of the form

$$(x_1, \ldots, x_m) \mapsto (x_1, g_2(x_1, \ldots, x_m), \ldots, g_n(x_1, \ldots, x_m))$$

and a point $(x_1, \ldots, x_m)$ is critical for g iff the point $(x_2, \ldots, x_m)$ is critical for the map $h_{x_1} \colon \mathbb{R}^{m-1} \to \mathbb{R}^{n-1}$ defined by the last $n-1$ components of g with x_1 fixed. It follows by induction that the intersection of $g(C_g)$ with any hyperplane $x_1 = \text{constant}$ is a measure 0 subset of that hyperplane. But $g(C_g)$ is certainly measurable (it's the continuous image of a closed set), so Fubini's theorem tells us that it has measure 0.

Proof that $f(C_r \setminus C_{r+1})$ has measure 0 This is a similar argument to the previous one, but does not require Fubini's theorem. For $x \in C_r \setminus C_{r+1}$ we can find a function u that is some r'th partial derivative of a component of f, such that u vanishes at x but some partial derivative of u, say $\partial u/\partial x_1$, does not. As before the map

$$h(x_1, \ldots, x_m) = (u(x_1, \ldots, x_m), x_2, \ldots, x_m)$$

is a diffeomorphism of some neighborhood U of x onto an open subset $U' \subseteq \mathbb{R}^m$. In this case h maps $C_r \cap U$ into a hyperplane in U' with first coordinate 0. Again, the critical values of f are the same as those of $g = f \circ h^{-1} \colon U' \to \mathbb{R}^n$. But since all these critical values come from the restriction of g to a hyperplane, the inductive hypothesis shows once again that they have measure 0.

Proof that $f(C_r)$ has measure 0 for r large For $r > m/n - 1$ we will show that $f(C_r \cap B)$ has measure 0, where B is any cube with unit sides. The $(r+1)$st derivatives of f are bounded on B (by compactness), and so by Taylor's Theorem (Lemma 19.1) there is a constant a such that

$$\|f(x+h) - f(x)\| \leqslant a\|h\|^{r+1}$$

whenever $x \in C_r \cap B$. Now subdivide B into k^m cubes of side $1/k$. By the estimate above, any one of these cubes that contains some point of C_r is mapped into a ball of radius $ak^{-(r+1)}$, and hence of volume $bk^{-n(r+1)}$ where b is another constant. There are at most k^m such cubes so the total volume of their images is at most $bk^{m-n(r+1)}$, and if r is sufficiently large this tends to zero as $k \to \infty$. This completes the proof. □

Sard's theorem has been proved only for maps between Euclidean spaces but an obvious argument shows that it is also true for maps between compact manifolds.

Let us now use this technology to define the degree of a map $f\colon S^n \to S^n$. We will only define the degree of a *smooth* map and prove its invariance under *smooth* homotopy. This is actually no restriction, since any continuous map is homotopic to a smooth map and, if two smooth maps are continuously homotopic, they are smoothly homotopic. (Proving these facts is an easy exercise, especially if you restrict attention only to maps between spheres.)

So, let $f\colon S^n \to S^n$ be a smooth map. By Sard's theorem, there exist (many) $a \in S^n$ such that $Df(x)$ is an isomorphism for all $x \in f^{-1}\{a\}$. (Such an a is called a *regular value* of f.) By the inverse function theorem, f is a local diffeomorphism in a neighborhood of each such x. In particular the set $f^{-1}\{a\}$ is discrete, hence finite. To each $x \in f^{-1}\{a\}$ is associated a sign ± 1 which is the sign of the determinant of $Df(x)$ at that point. (This assumes that we have chosen an orientation for S^n, but the particular orientation is not relevant.) We *define* $\deg(f)$ to be the sum

$$\deg f = \sum_{x \in f^{-1}\{a\}} \operatorname{sgn}\det Df(x).$$

It is easy to see that this definition possesses the properties (a) and (c) listed in the previous lecture (the degree of the identity map is 1; the degree of a constant map is 0; the degree is additive.)

Of course it is necessary to show that this is a good definition, i.e. independent of the regular value a that is chosen. This independence is in fact a special case of the homotopy invariance property (b): it is easy to show by composing with a suitable rotation that given regular values a and b, f is homotopic to a map g such that $f^{-1}\{b\} = g^{-1}\{a\}$ (including signs). Everything therefore boils down to proving the homotopy invariance property.

Proposition 19.3. *Let $f, g\colon S^n \to S^n$ be smoothly homotopic smooth maps such that a is a regular value for both f and g. Then*

$$\sum_{x\in f^{-1}\{a\}} \operatorname{sgn}\det Df(x) = \sum_{x\in g^{-1}\{a\}} \operatorname{sgn}\det Dg(x).$$

It is helpful (although not strictly necessary) to make use of the language of *cobordism*. Let M be a closed oriented manifold of dimension n. The notation $-M$ will refer to the same manifold considered with the opposite orientation. Two closed oriented n-manifolds M and M' are (oriented) *cobordant* if the disjoint union $M \sqcup (-M')$ is the (oriented) boundary of a compact oriented $(n+1)$-manifold with boundary.

We will only need the trivial case $n = 0$. A closed oriented 0-manifold is then a finite union of points, each of which is provided with a sign ± 1. The *degree* of a closed oriented 0-manifold is the sum of the signs of its points.

Lemma 19.4. *Two closed oriented 0-manifolds are cobordant iff they have the same degree.*

Proof. It is equivalent to prove that a closed oriented 0-manifold is the boundary of a compact oriented 1-manifold iff it has degree zero. The only compact connected oriented 1-manifolds are the circle (which has no boundary) and the interval (which has boundary two points with opposite signs). For each of these the index of the boundary is zero. A general compact oriented 1-manifold is the union of finitely many connected components each of which are as described above, so its boundary has degree zero.

Conversely, a closed oriented 0-manifold of degree zero has exactly as many points with sign $+1$ as with sign -1. Connect these points in pairs by arcs to obtain a nullcobordism. □

Our definition of the degree of a map is compatible with our use of "degree" for 0-manifolds above. In fact, the degree of $f\colon S^n \to S^n$ is exactly the degree of the oriented 0-manifold $f^{-1}\{a\}$ for a regular value a. Thus to complete the proof of Proposition 19.3 it will suffice to show that if $f, g\colon S^n \to S^n$ are smoothly homotopic smooth maps and a is a regular value for both f and g, then $f^{-1}\{a\}$ and $g^{-1}\{a\}$ are cobordant.

Lemma 19.5. *Let M and N be smooth manifolds of dimensions m and n respectively, $m \geqslant n$. Let $f\colon M \to N$ be a smooth map and $a \in N$ a regular value of f. Then $f^{-1}\{a\}$ is a smooth submanifold of M, of dimension $m - n$. If M and N are oriented, then so is $f^{-1}\{a\}$.*

Proof. The question is local, so assume $M = \mathbb{R}^m$ and $N = \mathbb{R}^n$. Let $x \in f^{-1}\{a\}$; then $Df(x)$ is surjective. By standard linear algebra we can make linear changes of coordinates in $\mathbb{R}^m$ and $\mathbb{R}^n$ so that $Df(x)$ is represented by the $m \times n$ matrix with 1 in position (i,i), $1 \leqslant i \leqslant n$, and 0 elsewhere. Now define $g\colon \mathbb{R}^m \to \mathbb{R}^m$ by $g(x_1, \ldots, x_n) = (f(x_1, \ldots, x_m), x_{n+1}, \ldots, x_m)$. Then $Dg(x) = I$, so the inverse function theorem shows that g is a local diffeomorphism near x, and it carries $f^{-1}\{a\}$ onto $\{a\} \times \mathbb{R}^{m-n} \subseteq \mathbb{R}^m$. It follows that $f^{-1}\{a\}$ is an $(m-n)$-dimensional submanifold. Its tangent space at x is identified with the kernel of the surjective map $Df(x)\colon \mathbb{R}^m \to \mathbb{R}^n$ and therefore acquires an orientation from the orientations of $\mathbb{R}^m$ and $\mathbb{R}^n$. $\square$

Now let $h\colon S^n \times I \to S^n$ be a smooth homotopy with $h(\cdot, 0) = f$ and $h(\cdot, 1) = g$. Assume for a moment that h has the following good properties:

(a) $h(\cdot, t)$ is independent of t for $t < \epsilon$ and $1 - t < \epsilon$.

(b) a is a regular value of the map h.

Later we will show using Sard's theorem that if there is any smooth homotopy at all from f to g, then there is a smooth homotopy that has these good properties.

Assuming (a) and (b), consider the set $h^{-1}\{a\}$. A point $(x,t) \in h^{-1}\{a\}$ is either a "boundary point" ($t \in \{0,1\}$) or an "interior point" ($0 < t < 1$). In the first case, property (a) shows that (x,t) has a neighborhood $\{x\} \times I$, where I is a half-open interval ending at (x,t). In the second case, Lemma 19.5 shows that $h^{-1}\{a\}$ is a 1-dimensional submanifold near (x,t). We conclude that $h^{-1}(a)$ is a compact 1-dimensional manifold with boundary $f^{-1}\{a\} \sqcup (-g^{-1}\{a\})$. In other words, $f^{-1}\{a\}$ and $g^{-1}\{a\}$ are cobordant, and the result now follows from Lemma 19.4.

It remains to show that given any smooth homotopy h from f to g, there exists one that has the good properties (a) and (b) above. First, replace h by a homotopy h' that is constant in t near 0 and 1, by defining

$$h'(x,t) = h(x, \phi(t))$$

where ϕ is a smooth, nondecreasing surjective function $[0,1] \to [0,1]$ which is constant for $t < 3\epsilon$ and $1 - t < 3\epsilon$.

By Sard's theorem, the set of regular values for h' is dense. Since regular values form an open set, there exists a regular value b for h' having the property that each point c on the great circle arc $[a,b]$ is a regular value for both f and g. Let R_s be a family of rotations of S^n having the property that R_0 is the identity,

$R_1(b) = a$, and $R_s(b)$, $s \in [0,1]$, is the great circle arc joining b and a. Then define

$$h''(x,t) = h'(R_{\psi(t)}(x), t)$$

where $\psi\colon [0,1] \to [0,1]$ is smooth, $\psi(t) = 0$ for $t < \epsilon$ and $1-t < \epsilon$, and $\psi(t) = 1$ for $t \in (2\epsilon, 1-2\epsilon)$. Then a is a regular value of h'', h'' is constant in t near 0 and 1, and h'' is a smooth homotopy between f and g. This finishes the proof of Proposition 19.3. □

Notice that the same argument allows us to define the degree of a map from any closed oriented n-manifold M to S^n. Indeed, we can go further. Suppose that Ω is a bounded open subset of $\mathbb{R}^n$ and that $f\colon \overline{\Omega} \to \mathbb{R}^n$ is continuous on $\overline{\Omega}$ and smooth on Ω. Let $a \in \mathbb{R}^n$ be such that $a \notin f(\partial\Omega)$. We can now define $\deg(f,a)$ to equal the degree of $f^{-1}\{a\}$ if a is a regular value (and in general to equal the degree of $f^{-1}\{b\}$ where b is a regular value arbitrarily close to a). This "degree" now depends on a, of course, but it is constant along any homotopy h such that $a \notin h(\partial\Omega \times I)$. The proof is the same as that given above.

To extend the degree to infinite dimensions, we can make use of a "suspension" property. Let Ω be as above and suppose that f has the form $f(x) = x - g(x)$, where g maps $\overline{\Omega}$ into an m-dimensional subspace $\mathbb{R}^m$, $m < n$. Then for $a \in \mathbb{R}^m$ the degree of f at a (considered as a map $\Omega \to \mathbb{R}^n$) as the same as the degree of the restriction of f to a map $\Omega \cap \mathbb{R}^m \to \mathbb{R}^m$. The proof is easy: both degrees are counting the points of exactly the same oriented 0-manifold.

Exercise 19.6. Those who prefer the homological definition of degree are invited to prove the suspension result by making use of the Mayer-Vietoris sequence. By induction, it suffices to consider the case $m = n-1$.

Suppose now that E is a real Banach space, $\Omega \subseteq E$ is open and bounded, and $f\colon \overline{\Omega} \to E$ is continuous. Suppose moreover that f is of the form $f(x) = x - g(x)$ where g maps bounded sets to precompact sets. Let $a \in E$ with $a \notin f(\partial\Omega)$. An easy argument shows that f is homotopic (through maps that do not send the boundary to a) to $f' = 1 - g'$ where $g(\overline{\Omega})$ is contained in a finite-dimensional subspace $\mathbb{R}^m$. Define

$$\deg(f,a) := \deg(f'_{|\mathbb{R}^m} \to \mathbb{R}^m.$$

The suspension property above shows that this definition is independent of the choice of finite dimensional susbspace. The resulting definition is called the *Leray-Schauder* degree of f at a, and this infinite-dimensional notion of degree can be applied directly to the study of existence theorems.

Lecture 20

Banach Algebras

A *Banach algebra* is a Banach space A which is also equipped with a bilinear, continuous multiplication operation which makes it into a ring. We will take it that the multiplication has norm 1, i.e. $\|xy\| \leqslant \|x\|\|y\|$. We will also restrict attention to Banach algebras with a unit element 1 such that $1x = x = x1$. We assume that $\|1\| = 1$.

A *Banach $*$-algebra* is a Banach algebra equipped with an antilinear involution $x \mapsto x^*$ which satisfies $(xy)^* = y^*x^*$. A complex Banach algebra is a C^*-*algebra* if $\|x^*x\| = \|x\|^2$ for all $x \in A$. (There is also a notion of real C^*-algebra, but this is more complicated and we shall not go into it.)

Example 20.1. Let X be a compact Hausdorff space. Then $C(X)$, with the sup norm and pointwise operations, is a commutative C^*-algebra. We will shortly prove the *Gelfand-Naimark theorem* which says that every commutative C^*-algebra is of this kind for some X.

Example 20.2. Let H be a Hilbert space. The collection of all bounded linear operators on H, $\mathcal{L}(H)$, is a C^*-algebra, as is any closed subalgebra.

Example 20.3. Let D be the closed unit disk in $\mathbb{C}$. The *disk algebra* is the subalgebra of $C(D)$ comprised of those functions that are holomorphic in the interior of D. Standard results about holomorphic functions show that this algebra is closed in the sup norm, so it is a (commutative) Banach algebra. It is not a C^*-algebra.

Definition 20.4. Let A be a Banach algebra. The *spectrum* $\mathrm{sp}(a)$ of $a \in A$ is the set of complex numbers λ such that $a - \lambda 1$ does *not* have an inverse in A. The complement of the spectrum is called the *resolvent set*.

We'll use the notation $\mathrm{sp}_A(a)$ if it is necessary to specify the algebra A.

Lemma 20.5. *The resolvent set of $a \in A$ is open (and so the spectrum is closed).*

Proof. It suffices to show that every element sufficiently close to $1 \in A$ is invertible, which follows from the absolutely convergent power series expansion

$$(1 - x)^{-1} = 1 + x + x^2 + \cdots ,$$

which is valid for $\|x\| < 1$ in any Banach algebra. □

Corollary 20.6. *Maximal ideals in a Banach algebra A are closed. Every algebra homomorphism $\alpha\colon A \to \mathbb{C}$ is continuous, of norm $\leqslant 1$.*

Proof. Let $\mathfrak{m}$ be a maximal ideal in A. Then $\mathfrak{m}$ does not meet the open set of invertible elements in A. The closure $\overline{\mathfrak{m}}$ then does not meet the set of invertibles either, so it is a proper ideal, and by maximality $\overline{\mathfrak{m}} = \mathfrak{m}$.

Since the kernel of α is a maximal ideal, it is closed. Thus α is continuous, by the usual continuity criterion for linear functionals. The norm estimate is obtained by refining the argument: if $\|\alpha\| > 1$ then there is $a \in A$ with $\|a\| < 1$ and $\alpha(a) = 1$. But then $1 - a$ is invertible, so $1 - \alpha(a)$ is invertible too, and this is a contradiction. ☐

Proposition 20.7. *Let A be a unital Banach algebra, and let $a \in A$. Then the spectrum* $\mathrm{sp}(a)$ *is a nonempty compact subset of* $\mathbb{C}$*. Moreover, the* spectral radius

$$\mathrm{spr}(a) = \inf\{r \in \mathbb{R}^+ : \mathrm{sp}(a) \subseteq D(0; r)\}$$

is given by the formula

$$\mathrm{spr}(a) = \lim_{n\to\infty} \|a^n\|^{1/n}.$$

Proof. The power series expansion

$$(\lambda - a)^{-1} = \lambda^{-1} + \lambda^{-2}a + \lambda^{-3}a^2 + \cdots$$

converges for $|\lambda| > \|a\|$ and shows that the spectrum of a is contained within $D(0; \|a\|)$ (note in particular that $\mathrm{spr}(a) \leqslant \|a\|$). We already showed that the spectrum is closed, so it is compact. It is nonempty since otherwise the function

$$\lambda \mapsto (\lambda - a)^{-1}$$

would be a bounded, entire function (with values in A), contradicting Liouville's Theorem from complex analysis. Finally let us prove the spectral radius formula. Since by hypothesis the resolvent function $\lambda \mapsto (\lambda - a)^{-1}$ is holomorphic for $|\lambda| > \mathrm{spr}(a)$, elementary complex analysis (Cauchy's n'th root test) yields the estimate

$$\limsup \|a^n\|^{1/n} \leqslant \mathrm{spr}(a)$$

for the coefficients of the Laurent expansion above. On the other hand, rewriting the expansion as

$$(\lambda - a)^{-1} = (\lambda^{-1} + \lambda^{-2}a + \cdots + \lambda^{-n}a^{n-1})(1 + \lambda^{-n}a^n + \lambda^{-2n}a^{2n} + \ldots)$$

shows that if $\|\lambda^{-n}a^n\| < 1$, then λ belongs to the resolvent set; so

$$\operatorname{spr}(a) \leqslant \|a^n\|^{1/n}$$

for each n, and thus

$$\operatorname{spr}(a) \leqslant \inf \|a^n\|^{1/n}.$$

Putting together our two estimates for the spectral radius of a, we complete the proof. □

Lemma 20.8. *(Gelfand-Mazur theorem) The only Banach algebra which is also a field is* $\mathbb{C}$.

Proof. Let A be such an algebra, $a \in A$. Then a has nonempty spectrum, so there is some $\lambda \in \mathbb{C}$ such that $a - \lambda 1$ is not invertible. But in a field the only non-invertible element is zero; so $a = \lambda 1$ is scalar. □

Proposition 20.9. *Let A be a unital C^*-algebra. If $a \in A$ is normal (that is, commutes with a^*) then $\|a\| = \operatorname{spr}(a)$. For every $a \in A$, $\|a\| = \operatorname{spr}(a^*a)^{1/2}$. Thus the norm on A is completely determined by the algebraic structure!*

Proof. If a is normal then

$$\|a\|^2 = \|a^*a\| = \|a^*aa^*a\|^{1/2} = \|a^2a^{*2}\|^{1/2} = \|a^2\|$$

using the C^*-identity several times. Thus the spectral radius formula gives $\|a\| = \operatorname{spr}(a)$. Whatever a is, the element $b = a^*a$ is normal (indeed selfadjoint), so the second part follows from the first applied to b, together with the C^*-identity again. □

Gelfand used the results above to give a systematic analysis of the structure of *commutative* unital Banach algebras. If A is such an algebra then its *Gelfand dual* $\widehat{A}$ is the space of maximal ideals of A; equivalently, by 20.8 and 20.6 above, it is the space of algebra-homomorphisms $A \to \mathbb{C}$. Every such homomorphism is of norm $\leqslant 1$, as we saw, so $\widehat{A}$ may be regarded as a subset of the unit ball of the Banach space dual A^* of A (namely, it is the subset comprising all those linear functionals α which are also *multiplicative* in the sense that $\alpha(xy) = \alpha(x)\alpha(y)$). This is a weak-star closed subset of the unit ball of A^*, and so by the Banach-Alaoglu Theorem it is a compact Hausdorff space in the weak-star topology. If $a \in A$ then we may define a continuous function $\hat{a}$ on $\widehat{A}$ by the usual double dualization:

$$\hat{a}(\alpha) = \alpha(a).$$

In this way we obtain an algebra-homomorphism $\mathcal{G}\colon a \mapsto \hat{a}$, called the *Gelfand transform*, from A to $C(\widehat{A})$, the algebra of continuous functions on the Gelfand dual.

Theorem 20.10. *Let A be a commutative unital Banach algebra. An element $a \in A$ is invertible if and only if its Gelfand transform $\hat{a} = \mathcal{G}a$ is invertible. Consequently, the Gelfand transform preserves spectrum: the spectrum of $\mathcal{G}a$ is the same as the spectrum of a.*

Proof. If a is invertible then $\mathcal{G}a$ is invertible, since $\mathcal{G}$ is a homomorphism. On the other hand, if a is *not* invertible then it is contained in some maximal ideal $\mathfrak{m}$, which corresponds to a point of the Gelfand dual on which $\mathcal{G}a$ vanishes. So $\mathcal{G}a$ isn't invertible either. ☐

The algebra $C(\widehat{A})$ has an involution, given by pointwise complex conjugation. One might thus ask: if A is a Banach $*$-algebra, must $\mathcal{G}$ be a $*$-homomorphism, that is, preserve the involutions? To see that the answer is no, look at the disk algebra again which we may equip with the involution $f^*(z) = \bar{f}(\bar{z})$. The Gelfand dual of the disk algebra is the closed disk (exercise!) and the Gelfand transform is the identity map, but the involution on the disk algebra is of course not pointwise complex conjugation.

It turns out that this is related to the existence in the disk algebra of "selfadjoint" elements whose spectrum is not real (one says that the disk algebra is not *symmetric*). For instance the function z is "selfadjoint" in the disk algebra, but its spectrum is the whole disk. One has a simple

Lemma 20.11. *Let A be a commutative unital Banach $*$-algebra. Then the Gelfand transform is a $*$-homomorphism if and only if the spectrum of every "selfadjoint" element of A is real.*

Proof. In view of the Theorem above, the hypothesis implies that a "selfadjoint" element of A has real-valued Gelfand transform. Now split a general $a \in A$ into real and imaginary parts:

$$a = \frac{a + a^*}{2} + \frac{a - a^*}{2}$$

and apply the Gelfand transform separately to each. ☐

Proposition 20.12. *A selfadjoint element of a C^*-algebra has real spectrum. Consequently, the Gelfand transform for a commutative C^*-algebra is a $*$-homomorphism.*

Proof. Let $a \in A$ with $a = a^*$ where A is a unital C^*-algebra. Let $\lambda \in \mathbb{R}$. Then from the C^*-identity

$$\|a \pm \lambda i\|^2 = \|a^2 + \lambda^2 1\| \leqslant \|a\|^2 + \lambda^2.$$

Consequently the spectrum of a lies within the lozenge-shaped region

$$\{z : |z \pm \lambda i|^2 \leqslant \|a\|^2 + \lambda^2\}.$$

As $\lambda \to \infty$ the intersection of all these lozenge-shaped regions is the interval $[-\|a\|, \|a\|]$ of the real axis. ☐

Corollary 20.13. *The Gelfand transform for a commutative C^*-algebra is an* isometric *$*$-homomorphism (in particular, it is injective).*

Proof. This follows from the facts that $\mathcal{G}$ preserves the spectrum and the involution, and that the norm on a C^*-algebra is determined by the spectral radius. ☐

What about the *range* of the Gelfand transform? To identify this we need the Stone-Weierstrass theorem.

Theorem 20.14. *(Stone-Weierstrass) Let $C(X)$ be the C^*-algebra of continuous, complex-valued functions on a compact Hausdorff space X. Any $*$-subalgebra $\mathscr{A}$ which contains the constant functions and separates points of X is dense in $C(X)$.*

"Separates points" means that for any $x', x'' \in X$ there exists $f \in \mathscr{A}$ with $f(x') \neq f(x'')$. I will not prove the Stone-Weierstrass theorem here.

Lemma 20.15. *The Gelfand transform for a commutative C^*-algebra is surjective.*

Proof. The range of $\mathcal{G}$ is a $*$-subalgebra of $C(\widehat{A})$ which separates points of $\widehat{A}$ (for tautological reasons). By the Stone-Weierstrass theorem, then, the range of $\mathcal{G}$ is dense. But the range is also *complete*, since it is isometrically isomorphic (via $\mathcal{G}$) to a Banach algebra; hence it is closed, and thus is the whole of $C(\widehat{A})$. ☐

Putting it all together.

Theorem 20.16. *If A is a commutative unital C^*-algebra, then A is isometrically isomorphic to $C(X)$, for some (uniquely determined up to homeomorphism) compact Hausdorff space X. Any unital $*$-homomorphism of commutative unital*

C^*-algebras is induced (contravariantly) by a continuous map of the corresponding compact Hausdorff spaces. Consequently, the Gelfand transform gives rise to an equivalence of categories between the category of commutative unital C^*-algebras and the opposite of the category of compact Hausdorff spaces.

This is the *Gelfand-Naimark Theorem*. There is a non-unital version as well, which is proved by unitalization: every commutative C^*-algebra is of the form $C_0(X)$ for some locally compact Hausdorff space X.

Let A be a unital C^*-algebra, not necessarily commutative, and let $a \in A$ be a normal element. Then the unital C^*-subalgebra $C^*(a) \subseteq A$ generated by a is commutative, so according to Gelfand-Naimark it is of the form $C(X)$ for some compact Hausdorff space X. We ask: What is X? Notice that every homomorphism $\alpha\colon C^*(a) \to \mathbb{C}$ is determined by the complex number $\alpha(a)$. Thus $\hat{a}$ defines a continuous injection $\widehat{C^*(a)} \to \mathbb{C}$.

Proposition 20.17. *Let a be a normal element of a unital C^*-algebra A. Then the Gelfand transform identifies $C^*(a)$ with $C(\mathrm{sp}(a))$, the continuous functions on the spectrum (in A) of a. Under this identification, the operator a corresponds to the identity function $z \mapsto z$.*

Notice that it is a consequence of this proposition that, if $a \in A \subseteq B$, where A and B are C^*-algebras, then $\mathrm{sp}_A(a) = \mathrm{sp}_B(a)$. It is for this reason that we do not usually need to specify algebra when discussing the spectrum of an element of a C^*-algebra.

Proof. What is the image of the injection $\hat{a}\colon \widehat{C^*(a)} \to \mathbb{C}$ defined above? It is simply the spectrum of the function $\hat{a}$ considered as an element of the commutative Banach algebra $C(\widehat{C^*(a)})$. By Gelfand's theorem above, this is the same thing as the spectrum of a, considered as an element of the commutative Banach algebra $C^*(a)$. Thus $\hat{\alpha}$ is a homeomorphism[2] from $\widehat{C^*(a)}$ to $\mathrm{sp}_{C^*(a)}(a)$.

It remains to prove that $\mathrm{sp}_{C^*(a)}(a) = \mathrm{sp}_A(a)$. Clearly, if $x \in C^*(a)$ is invertible in $C^*(a)$, then it is invertible in A. On the other hand suppose that $x \in C^*(a)$ *fails* to be invertible in $C^*(a)$. Then $\hat{x}$ is a continuous function on a compact Hausdorff space which takes the value zero somewhere, say at $\alpha \in \widehat{C^*(a)}$. Using bump functions supported in neighborhoods of α one can produce, for each $\epsilon > 0$, an element $y_\epsilon \in C^*(a)$ such that $\|y_\epsilon\| = 1$ and $\|xy_\epsilon\| < \epsilon$. But if x were invertible in A, this would be impossible for sufficiently small ϵ (less than $\|x^{-1}\|^{-1}$). ☐

---

[2] A continuous bijective map of compact Hausdorff spaces is a homeomorphism.

Lecture 21
The Spectral Theorem

There are several slightly different results that go under the name of the "spectral theorem". What they all have in common is that they allow one to express certain operators T (in our case, bounded normal operators on a Hilbert space) in a *canonical form* which allows one to compute with them easily. A consequence of this canonical form — indeed, a more or less equivalent statement — is that one can define *functions* $f(T)$ of the operator T, not just for polynomials f where the definition is obvious, but for functions of a much more general sort. This is called the *functional calculus*.

Let T be a bounded normal operator on a Hilbert space H. We will use the notation $C^*(T)$ for the norm closure in $\mathcal{L}(H)$ of the polynomials in T and T^*; it is a commutative unital C^*-algebra, and the Gelfand transform gives an isomorphism

$$\mathcal{G}\colon C^*(T) \to C(\mathrm{sp}(T)).$$

Definition 21.1. (Functional calculus) For $f \in C(\mathrm{sp}(T))$ the operator $\mathcal{G}^{-1}(f) \in \mathcal{L}(H)$ will be denoted $f(T)$.

Notice that it follows from Proposition 20.17 that if f is a polynomial, then the $f(T)$ defined using the functional calculus is the same as the $f(T)$ defined "naively". The same applies if f is a rational function whose poles do not lie in $\mathrm{sp}(T)$. The functional calculus procedure has all the properties which the notation would lead you to expect (so long as you confine yourself to thinking about only a single normal element). For instance, $(f+g)(T) = f(T) + g(T)$, $(f \cdot g)(T) = f(T)g(T)$, $(f \circ g)(T) = f(g(T))$ and so on. The proofs all follow from the fact that the Gelfand transform is an isomorphism, except the last which uses polynomial approximation.

Here is a very simple consequence of the functional calculus.

Proposition 21.2. *A normal operator is self-adjoint if and only if its spectrum is real. It is unitary if and only if its spectrum lies in the unit circle.*

Proof. Let $X \subseteq \mathbb{C}$ be the spectrum of T. Then $T = T^*$ iff the function z, restricted to X, satisfies $z = \bar{z}$; similarly $T^*T = I$ iff the function z, restricted to X, satisfies $\bar{z}z = 1$. □

Definition 21.3. Let T be a a bounded normal operator on a Hilbert space H. A closed subspace K on H is a *cyclic subspace* for T if $C^*(T)K \subseteq K$ and there is a cyclic vector $\xi \in K$ (called a *cyclic vector*) such that $C^*(T)\xi$ is dense in K.

Proposition 21.4. *Let T be a bounded normal operator on a Hilbert space H. Then H can be written as an orthogonal direct sum (possibly infinite) of cyclic subspaces.*

Proof. Begin with a simple but crucial observation: if K is a $C^*(T)$-invariant subspace of H, then $K^\perp$ is $C^*(T)$-invariant too. Indeed if $\xi \in K$, $\eta \in K^\perp$ we have

$$\langle T\eta, \xi\rangle = \langle \eta, T^*\xi\rangle = 0, \quad \langle T^*\eta, \xi\rangle = \langle \eta, T\xi\rangle = 0$$

so $T\eta$, $T^*\eta$ belong to $K^\perp$. It follows that $K^\perp$ is $C^*(T)$-invariant.

Now for each unit vector $u \in H$, define H_u to be the closure of $C^*(T)u$; clearly H_u is a cyclic subspace. Let $\mathcal{S}$ denote the collection of sets S of unit vectors in H with the property that for distinct $u, u' \in S$ the cyclic subspaces $H_u, H_{u'}$ are orthogonal. By Zorn's lemma there exists a maximal $S \in \mathcal{S}$. I claim that the corresponding cyclic subspaces sum to all of H. If not, then the intersection

$$K = \bigcap_{u \in S} H_u^\perp$$

is a nonzero $C^*(T)$-invariant subspace, which will contain a unit vector v that can be added to S, contradicting maximality. □

Theorem 21.5. *Every bounded normal operator on a separable Hilbert space is unitarily equivalent to a multiplication operator.*

This is the *multiplication operator form of the spectral theorem*. In more detail, what it says is that given any such operator $T \in \mathcal{L}(H)$, we can find a σ-finite measure space (X, μ), and a function $f \in L^\infty(X, \mu)$, and an isomorphism of Hilbert spaces $U\colon H \to L^2(X, \mu)$, such that

$$(UTU^*)g(x) = f(x)g(x)$$

for all $g \in L^2(X, \mu)$.

Proof. Under the separability hypothesis, H is an at most countable sum of cyclic subspaces. It suffices then to prove the theorem for one of them, i.e. we may assume wlog that H is cyclic for T. Let ξ be a cyclic vector.

Consider the linear functional

$$\sigma\colon f \mapsto \langle f(T)\xi, \xi\rangle$$

defined on $C(\operatorname{sp}(T))$. If f is a positive function, then $f = |g|^2 = g^*g$ for some continuous g, and then $\sigma(f) = \|g(T)\xi\|^2 \geqslant 0$. Thus σ is a positive linear functional. By the Riesz representation theorem, there is a regular Borel measure μ on $X = \operatorname{sp}(T) \subseteq \mathbb{C}$ such that

$$\sigma(f) = \int f\, d\mu.$$

By construction, if $g \in C(X)$, then

$$\|g(T)\xi\|^2 = \int |g|^2 d\mu = \|g\|^2_{L^2(X,\mu)}.$$

Therefore the linear map $g \mapsto g(T)\xi$ extends to an isomorphism of Hilbert spaces $U\colon L^2(X,\mu) \to H$ (here we use the fact that ξ is a cyclic vector). Moreover since

$$Tg(T)\xi = f(T)\xi \quad \text{where } f(x) = xg(x),$$

we see that UTU^* is simply the operator of multiplication by x on $L^2(X,\mu)$. $\square$

Example 21.6. An operator (necessarily self-adjoint) T on a Hilbert space H is called *positive* if $\langle T\xi, \xi\rangle \geqslant 0$ for all $\xi \in H$. A self-adjoint operator T is positive if and only if its spectrum is contained in the positive real line $[0,\infty)$; to see this, by the spectral theorem it suffices to consider the case where T is the operator of multiplication by $f \in L^\infty(X,\mu)$ acting on $H = L^2(X,\mu)$; but in this case both conditions are equivalent to saying that $f(x) \in [0,\infty)$ for almost all x.

Remark 21.7. Notice that for any positive operator T a positive *square root* $T^{1/2}$ can be defined using the functional calculus. This positive square root is *unique*. To see this, note that if S is positive and $S^2 = T$ then S and T commute; let $A \cong C(X)$ be the commutative C^*-algebra that they generate. In this C^*-algebra, S and T are represented by positive functions s and t on X satisfying $s^2 = t$. It follows that $s = t^{1/2}$ (positive square root) and therefore $S \in C^*(T)$ and is the square root defined by the functional calculus.

Here is an extended example, the *polar decomposition*.

Definition 21.8. Let H be a Hilbert space. A *partial isometry* on H is an operator $V \in \mathcal{L}(H)$ which satisfies one of the following four (equivalent) conditions:

(a) V^*V is a projection (the 'initial projection'),

(b) VV^* is a projection (the 'final projection'),

(c) $V = VV^*V$,

(d) $V^* = V^*VV^*$.

It is easy to see that these conditions are equivalent, since (c) and (d) are adjoint are one another, (c) implies (a) (multiply on the left by V^*), and if (a) is true then $I - V^*V$ is a projection with range $\mathrm{Ker}(V)$, so that $V(I - V^*V) = 0$ which is (c). One should think of V as giving an isomorphism from the range of the initial projection to the range of the final projection.

Definition 21.9. Let $T \in \mathcal{L}(H)$. A *polar decomposition* for T is a factorization $T = VP$, where P is a positive operator and V is a partial isometry with initial space $\overline{\mathrm{ran}}(P)$ and final space $\overline{\mathrm{ran}}(T)$.

Here we are using the notation $\overline{\mathrm{ran}}(T)$ for the closure of the range of the operator T. Note that V and P need not commute, so that there are really two notions of polar decomposition (left and right handed); we have made the conventional choice.

Proposition 21.10. *Every operator $T \in \mathcal{L}(H)$ has a unique polar decomposition $T = VP$.*

Proof. Let $T = VP$ be a polar decomposition. Then $T^*T = P^*V^*VP = P^2$, so P is the unique positive square root of T^*T, which we denote by $|T|$. Fixing this P let us now try to construct a polar decomposition. For $v \in H$,

$$\|Pv\|^2 = \langle P^*Pv, v\rangle = \langle T^*Tv, v\rangle = \|Tv\|^2$$

so an isometry $\overline{\mathrm{ran}}(P) \to \overline{\mathrm{ran}}(T)$ is defined by $Pv \mapsto Tv$. Extending by zero on the orthogonal complement we get a partial isometry V with the property that $T = VP$; and it is uniquely determined since the definition of polar decomposition requires that $V(Pv) = Tv$ and that $V = 0$ on the orthogonal complement of $\overline{\mathrm{ran}}(P)$. □

The multiplication operator form of the spectral theorem allows us to extend the functional calculus to *Borel* functions. Notice in the construction above that UTU^* is a multiplication operator by a function f whose range is contained in the spectrum of T. Thus, if q is any bounded Borel function on the spectrum of T, we may define $q(T)$ by saying that $Uq(T)U^*$ is the operator of multiplication by $q(f)$. This agrees with our earlier definition if q is continuous. Warning: It may not be the case that $q(T)$ belongs to $C^*(T)$ (it *does* belong to the von Neumann algebra $W^*(T)$ but I don't want to get into that here.)

Remark 21.11. Notice that $\|f(T)\| \leqslant \|f\|_\infty$; this follows from the construction.

The final version of the spectral theorem is the *spectral measure* version of the theorem. This is traditional, and looks cool, but usually anything you can do with it you can do with one of the other versions too. So I will be brief.

Definition 21.12. Let X be a measurable space (i.e., a set equipped with a σ-algebra $\mathfrak{M}$ of subsets called *measurable subsets*). A *spectral measure* on X is a function $E\colon \mathfrak{M} \to \mathcal{L}(H)$ (for some Hilbert space H) such that

(a) Each $E(A)$ is a self-adjoint projection.

(b) E is a morphism of lattices. That is to say, $E(\emptyset) = 0$, $E(X) = I$, $E(A \cap B) = E(A)E(B)$, and if $A \cap B = \emptyset$ then $E(A \cup B) = E(A) + E(B)$.

(c) For every $\xi \in H$ the function $\mathfrak{M} \to \mathbb{C}$ defined by $A \mapsto \langle E(A)\xi, \xi\rangle$ is a positive measure on $(X, \mathfrak{M})$ of total mass $\|\xi\|^2$. (We denote this measure by E_ξ.)

Example 21.13. Let T be a normal operator, let $X = \operatorname{sp}(T)$ (equipped with its Borel σ-algebra), and for a Borel subset $A \subseteq X$ let

$$E(A) = \chi_A(T)$$

where χ_A denotes the characteristic function of A and the operator $\chi_A(T)$ is defined using the Borel functional calculus. Then E is a spectral measure.

In fact, this is in some sense the *only* example. To make sense of this, let us discuss integration with respect to a spectral measure. Suppose that E is a spectral measure on $(X, \mathfrak{M})$ and $f\colon X \to \mathbb{C}$ is a bounded, measurable function. If f is a simple function, of the form

$$f(x) = \sum_{i=1}^{n} a_i \chi_{A_i}(x)$$

with $a_i \in \mathbb{C}$, $A_i \in \mathfrak{M}$, then we define

$$\int_X f\,dE = \sum a_i E(A_i).$$

We must check that this definition is independent of the representation of the simple function f as a linear combination of characteristic functions: the proof is exactly the same as in the case of the standard Lebesgue integral.

This integral for simple functions has the property that $f \mapsto \int_X f dE$ is a $*$-homomorphism from the algebra of simple functions to $\mathcal{L}(H)$. For instance we check that it is multiplicative. Let f and g be simple functions, and write them as

$$f(x) = \sum_{i=1}^{n} a_i \chi_{A_i}(x), \quad g(x) = \sum_{i=1}^{n} b_i \chi_{A_i}(x)$$

where the sets A_i are mutually disjoint (this can always be arranged). Then

$$E(A_i)E(A_j) = \delta_{ij} E(A_i)$$

and so

$$\left(\int fdE\right)\left(\int gdE\right) = \sum_{i,j} a_i b_j E(A_i)E(A_j) = \sum_i a_i b_i E(A_i) = \int fgdE.$$

Let $T = \int fdE$ for some simple function f. Then $T^*T = \int |f|^2 dE$ (by the $*$-homomorphism property) and thus

$$\|T\xi\|^2 = \langle T^*T\xi, \xi\rangle = \int |f|^2 dE_\xi \leqslant \|f\|_\infty^2 \|\xi\|^2$$

since the measure E_ξ has total mass $\|\xi\|^2$. Thus

$$\left\|\int fdE\right\| \leqslant \|f\|_\infty.$$

Now let f be any (essentially) bounded measurable function on X. Then there exists a sequence $\{f_n\}$ of measurable simple functions converging uniformly to f. By the inequality displayed above, the integrals $\int f_n dE$ form a Cauchy sequence in the norm of $\mathcal{L}(H)$, convergent to an operator that we denote $\int fdE$. Our conclusion is that $\int fdE$ can be defined for every $f \in L^\infty(X)$ and that the assignment $f \mapsto \int fdE$ is a $*$-homomorphism.

Theorem 21.14. *For every bounded normal operator T on a Hilbert space, there exists a spectral measure E on $X = \operatorname{sp}(T) \subseteq \mathbb{C}$ (equipped with its Borel σ-algebra) such that*

$$T = \int_X z dE.$$

We then have

$$f(T) = \int_X f(z) dE$$

for every bounded Borel function f on X.

Proof. Define E, as above, by $E(A) = \chi_A(T)$. The equality $f(T) = \int f dE$ is then a tautology when f is the characteristic function of a Borel subset of X. By linearity, the equality folds good whenever f is a simple function. Any bounded Borel function is a uniform limit of simple functions, and the result for such functions (and in particular for $f(z) = z$) follows by passing to the limit. ☐

Lecture 22
Unbounded Operators

In the final two lectures we shall begin the study of *unbounded operators* on a Hilbert space. the study of such operators is a tricky business. An unbounded operator will not be defined everywhere, but only on a dense subspace called its *domain*. This means that the usual arithmetical operations might not make sense for unbounded operators (because their domains might not have large enough intersection). Similarly, a subtle distinction between *symmetric* and *self-adjoint* operators will make its appearance, arising from the fact that the "natural" domain of the adjoint T^* may be larger than the originally given domain of T.

Why get involved in all these technicalities? There is one overriding reason, which is that *differential operators* are unbounded when considered as acting on the natural Hilbert spaces of functions (L^2 spaces). A satisfactory Hilbert space analysis of differential operators—including an appropriate version of the spectral theorem—must therefore face the problem of unboundedness. For the differential operators appearing in quantum mechanics (such as the Schrödinger operator) the "spectrum" that appears in the spectral theorem has a direct relationship to the physical notion. Indeed, this is the origin of the "spectral" terminology.

Definition 22.1. Let H be a Hilbert space. A (densely defined) *unbounded operator* on H is a pair (T, D), where D is a (dense) subspace of H and T is a linear map from D to H. Usually we abbreviate the operator simply as T and we refer to D (the *domain* of T) by the notation $\operatorname{dom}(T)$.

The domain is an essential part of the definition. To put this another way, writing $S = T$ for unbounded operators means that $\operatorname{dom}(S) = \operatorname{dom}(T)$ *and* $Sx = Tx$ for all $x \in \operatorname{dom}(S) = \operatorname{dom}(T)$. If $\operatorname{dom}(S) \subseteq \operatorname{dom}(T)$ and we still have $Sx = Tx$ for all $x \in \operatorname{dom}(S)$ we say that T is an *extension* of S and write $S \subseteq T$.

Note that there is *no continuity assumption* on T — it is just a linear map. However, the following condition often provides a weak substitute for continuity.

Definition 22.2. The *graph* of T is the subset $\{(x, Tx) : x \in \operatorname{dom}(T)\}$ of $H \times H$ and T is *closed* if its graph is closed.

Example 22.3. Let (X, μ) be a σ-finite measure space and let f be a measurable complex-valued function on X, not necessarily bounded. The *unbounded multiplication* operator M_f is defined by $M_f g = fg$ and is defined on the domain

$$D = \{g \in L^2(X, \mu) : fg \in L^2(X, \mu)\}.$$

This is a closed operator.

We know from the closed graph theorem that an operator whose domain is the whole of H is closed iff it is continuous. An operator is called *closeable* if it has a closed extension: equivalently, if the closure of its graph is the graph of an operator. This need not happen. One important case in which it is automatic is described below.

Definition 22.4. Let T be a (densely defined) unbounded operator. The *adjoint* T^* is the unbounded operator defined as follows: the domain of T^* is the collection of all those $y \in H$ for which we have an estimate

$$|\langle Tx, y\rangle| \leqslant C_y \|x\|$$

for all $x \in \operatorname{dom}(H)$. In these circumstances, the densely defined linear functional $x \mapsto \langle Tx, y\rangle$ extends from $\operatorname{dom}(T)$ to a continuous linear functional on H, which (by standard Hilbert space theory) is given by inner product with some element of H, denoted T^*y.

Thus, by definition, we have

$$\langle Tx, y\rangle = \langle x, T^*y\rangle$$

and the domain of T^* is the largest possible for which this makes sense.

Exercise 22.5. In general the domain of T^* need not be dense. Prove this by considering the following example: $H = L^2(\mathbb{R})$, and T is defined on the dense domain $C_c^\infty(\mathbb{R})$ by setting $Tf = (\int f)g$, where $g \in L^2$ is some fixed nonzero element. Show that the domain of T^* is the orthogonal complement of the span of g.

Lemma 22.6. *The operator T^* is closed.*

Proof. A pair (y, z) belongs to the graph of T^* iff it is orthogonal (in $H \oplus H$) to the set $G' = \{(-Tx, x) : x \in \operatorname{dom}(T)\}$ (the “rotation” of the graph of T). Thus the graph of T^* is the orthogonal complement of a subspace, hence closed. □

Definition 22.7. The operator T is *symmetric* if $T \subseteq T^*$ (which is to say that $\langle Tx, y\rangle = \langle x, Ty\rangle$ for all $x, y \in \operatorname{dom}(T)$. It is *self-adjoint* if $T = T^*$.

As we will see, it is only for self-adjoint operators that the beautiful results of spectral theory extend effectively to the unbounded case. For differential operators, the distinction between "symmetric" and "self-adjoint" is related to the choice of boundary conditions.

Example 22.8. An unbounded multiplication operator (22.3) by a *real*-valued function is self-adjoint.

Exercise 22.9. Let $H = L^2[0,1]$ and define three unbounded operators T_1, T_2, T_3 on H as follows. Let D_1 be the Sobolev space W^1, which in this case is comprised of those absolutely continuous $f \in L^2$ whose derivative also belongs to L^2. The operator T_1 has domain D_2, and on this domain, $T_1 f = idf/dx$. The operators T_2 and T_3 are the restrictions of T_1 to the domains $D_2 = \{f \in D_1 : f(0) = f(1)\}$ and $D_3 = \{f \in D_1 : f(0) = f(1) = 0\}$ respectively. Thus $T_3 \subseteq T_2 \subseteq T_1$. Show that $T_r^* = T_{4-r}$ for $r = 1, 2, 3$. Thus T_2 is self-adjoint; T_3 is symmetric, but not self-adjoint (though it has a self-adjoint extension); and T_1 is not symmetric.

Let u be a complex number of modulus 1. Show that the restriction of T_1 to the domain

$$\{f \in D_1 : f(0) = uf(1)\}$$

is also self-adjoint. Thus the operator T_1 has uncountably many distinct self-adjoint extensions.

Notice in the above example that $[0,1]$ is a *manifold* (with boundary). There is some special terminology for differential operators on manifolds. If M is a smooth manifold (possibly with boundary) and D is a differential operator thought of as acting on $L^2(M)$, then D is said to be *formally self-adjoint* if it is symmetric when considered as an unbounded operator with domain $\mathscr{D}(M \setminus \partial M)$ (the smooth, compactly supported functions on the interior of M). When this is the case, the *minimal domain* of D is the closure of $\mathscr{D}(M \setminus \partial M)$ in the norm $u \mapsto \|u\|_{L^2} + \|Du\|_{L^2}$, and the *maximal domain* of D is the collection of all $u \in L^2$ such that the distribution Du can be identified with an L^2 function. By considering these domains, we obtain two closed extensions $D_{\min}$ and $D_{\max}$ of D. In the example above, $D_{\min}$ is D_3 and $D_{\max}$ is D_1. As the example illustrates, $D_{\max}$ is the Hilbert space adjoint of $D_{\min}$; this follows from the way (by duality) in which differential operators act on distributions. If the minimal and maximal domains coincide, then D has only one possible self-adjoint extension; in this case we say that it is *essentially self-adjoint*.

The basic criterion for self-adjointness is the following.

Proposition 22.10. *Let T be a symmetric operator on a Hilbert space. The following are equivalent.*

(a) T is self-adjoint.

(b) T is a closed operator and $\operatorname{Ker}(T^* \pm iI) = 0$.

(c) The operators $T \pm iI$ map the domain of T onto H.

Proof. Suppose (a). The graph of T then equals the graph of T^*, so is closed. Then an element x of $\operatorname{Ker}(T^*+iI)$ must belong to $\operatorname{dom}(T)$ and satisfy $Tx = -ix$. But then the equality $\langle Tx, x\rangle = \langle x, Tx\rangle$ implies $i\|x\|^2 = -i\|x\|^2$, so $x = 0$. This is (b).

Suppose (b). Let V be the range of the operator $T - iI$. The exact same argument as in the bounded operator case shows that the orthogonal complement of V is the kernel of $T^* + iI$; since this is 0 by hypothesis, V is dense. It therefore suffices to show that V is closed. But the identity

$$\|(T - iI)x\|^2 = \|Tx\|^2 + \|x\|^2$$

is valid for all $x \in \operatorname{dom}(T)$ and shows that $T - iI$ is an *isometry* from the graph of T onto V. By hypothesis the graph of T is closed, hence complete. Thus V is complete, hence closed. This gives (c).

Finally, suppose (c) and let $x \in \operatorname{dom}(T^*)$. By (c) there exists $y \in \operatorname{dom}(T)$ such that $(T - iI)y = (T^* - iI)x$. We will show that $x = y$. Let $x - y = z$; then $z \in \operatorname{Ker}(T^* - iI)$. But this space is the orthogonal complement of the range of $T + iI$, so it is zero. We deduce that every $x \in \operatorname{dom}(T^*)$ in fact belongs to $\operatorname{dom}(T)$, and this proves (a). □

Suppose now that T is a *closed, symmetric* operator on a Hilbert space H. The calculations above make it clear that the ranges of the operators $T \pm iI$ are closed subspaces of H, and that

$$\|(T + iI)x\|^2 = \|Tx\|^2 + \|x\|^2 = \|(T - iI)x\|^2$$

for all $x \in \operatorname{dom}(T)$. This shows that the formula

$$(T + iI)x \mapsto (T - iI)x$$

gives a well-defined *isometry* from $\operatorname{Im}(T + iI)$ to $\operatorname{Im}(T - iI)$. By composing with the orthogonal projection onto $\operatorname{Im}(T + iI)$ and the inclusion of $\operatorname{Im}(T - iI)$

into H, we get a *partial isometry* $H \to H$ which is called the *Cayley transform* of the closed symmetric operator T. From the discussion above we see that T is self-adjoint if and only if its Cayley transform is unitary.

Proposition 22.11. *A unitary operator U is the Cayley transform of a closed, self-adjoint operator if and only if $\operatorname{Ker}(I-U)=0$ and $\operatorname{Im}(I-U)$ is dense. When this is so, the formula*

$$T = i(I+U)(I-U)^{-1}$$

defines a closed self-adjoint operator, with domain $\operatorname{Im}(I-U)$, whose Cayley transform is U.

Proof. Suppose that U is the Cayley transform of T, If $y = Tx + ix$ then $Uy = Tx - ix$, so $(I-U)y = 2ix$. This shows that $I-U$ is injective and has dense range equal to $\operatorname{dom}(T)$. Moreover $(I+U)y = 2Tx$, so $i(I+U)(I-U)^{-1}$ sends $2ix$ to $2iTx$, i.e. it is the operator T.

If conversely U is unitary and $(I-U)$ is injective with dense range D, then the formula $T = i(I+U)(I-U)^{-1}$ defines an unbounded operator with domain D. Let us show that it is self-adjoint. First, we prove symmetry. By construction, if $x = z - Uz$ then $Tx = i(z+Uz)$. Let x bas as stated and let $x' = z' - Uz'$. Then

$$\langle Tx, x'\rangle = \langle i(z+Uz), z' - Uz'\rangle = i\langle z, Uz'\rangle - i\langle Uz, z'\rangle.$$

The symmetry of this expression shows that it equals $\langle x, Tx'\rangle$ also. So T is symmetric. Its graph consists of pairs $(z - Uz, i(z+Uz))$ and this is closed: if $(z_n - Uz_n, i(z_n + Uz_n))$ is a sequence that converges in the graph, then z_n and Uz_n converge to z and Uz, so the limit is also a point of the graph. Finally we use part (c) of proposition 22.10 to show that T is self-adjoint: if $x = z - Uz$, then $(T+iI)x = 2iz$ and $(T-iI)z = 2iUz$, so the ranges of these operators are all of H. □

Remark 22.12. In the case of a closed symmetric operator that is not self-adjoint, the codimensions of the ranges of $T \pm iI$ are called the *deficiency indices* of T. A symmetric operator has an extension to a self-adjoint operator if and only if its deficiency indices are equal (this is because a partial isometry can be extended to a unitary if and only if its initial and final projections have the same codimension).

Here is the spectral theorem for unbounded self-adjoint operators.

Theorem 22.13. *Every (unbounded) self-adjoint operator is unitarily equivalent to a real-valued (unbounded) multiplication operator.*

Proof. Let T be such an operator and let $U = (T - iI)(T + iI)^{-1}$ be its Cayley transform. This is unitary and thus, by the spectral theorem for bounded normal operators, it is unitarily equivalent to a multiplication operator M_h on $L^2(X, \mu)$, where h is a measurable function of modulus 1 on the measure space (X, μ). Because 1 is not an eigenvalue of U, the set $\{x \in X : h(x) = 1\}$ has measure zero. Now let $f(x) = i(1+h(x))(1-h(x))^{-1}$, which is a real-valued (unbounded) measurable function defined almost everywhere on X. The corresponding unbounded operator on $L^2(X, \mu)$ is self-adjoint and its Cayley transform is M_h; which is to say that the unitary equivalence $H \to L^2(X, \mu)$ carries T to M_f. □

"Functional calculus" and "spectral measure" versions of the spectral theorem for unbounded self-adjoint operators can be deduced from this version just as in the bounded case.

Lecture 23

Theorems on Unbounded Self-Adjoint Operators

Let H be a separable Hilbert space.

Definition 23.1. A *strongly continuous one-parameter unitary group* on H is a group homomorphism $t \mapsto U(t)$ from $\mathbb{R}$ to the group of unitary operators on H, which has the property that for each fixed $\xi \in H$, the map

$$t \mapsto U(t)\xi$$

is continuous from $\mathbb{R}$ to H.

Let A be an unbounded self-adjoint operator on H. According to the spectral theorem, there is an isomorphism of H to an L^2 space $L^2(X, \mu)$ that carries A to an unbounded multiplication operator M_f, where f is real-valued. The exponentials $U(t) = e^{itA}$ can therefore be defined as multiplications by the functions e^{itf}, and they form a one-parameter unitary group which is strongly continuous since, as $t \to t_0$,

$$\|(U(t) - U(t_0))\xi\|^2 = \int_X \left|e^{itf(x)} - e^{it_0 f(x)}\right|^2 |\xi(x)|^2 \, d\mu(x) \to 0$$

by the dominated convergence theorem. Notice that $t \mapsto U(t)$ is continuous in norm iff $t \mapsto e^{itf}$ is uniformly continuous in t, which is true if and only if A is a *bounded* operator.

The operator A is called the *infinitesimal generator* of the unitary group $U(t)$.

Remark 23.2. In (nonrelativistic) quantum mechanics the time evolution of a quantum system is described by a one-parameter unitary group, whose infinitesimal generator is the *Hamiltonian* operator that quantizes the total energy of the system.

Theorem 23.3. *(Stone's theorem) Let $t \mapsto U(t)$ be a strongly continuous unitary group on a separable Hilbert space H. Then there is a self-adjoint operator A such that $U(t) = e^{itA}$. For $\xi \in \operatorname{dom}(A)$ we have*

$$A\xi = \lim_{t \to 0} \frac{U(t)\xi - \xi}{it},$$

and ξ belongs to $\operatorname{dom}(A)$ if and only if the limit on the right-hand side exists.

Proof. We will obtain A by differentiating $U(t)$ at $t = 0$. We need first to define a dense set of "smooth vectors" on which this can be done.

For $f \in \mathscr{S}(\mathbb{R})$ define an operator S_f by

$$S_f = \int_{-\infty}^{\infty} f(t)U(t)dt.$$

This integral is to be interpreted in the following sense: for $\xi, \eta \in H$ the integral

$$S(\xi, \eta) = \int_{-\infty}^{\infty} f(t)\langle U(t)\xi, \eta\rangle d\xi$$

is well-defined (as a complex number) and S is a sesquilinear form satisfying the inequality

$$|S(\xi, \eta)| \leqslant \|f\|_{L^1}\|\xi\|\|\eta\|.$$

It follows from the Riesz representation theorem that there is a bounded linear operator S_f such that $\langle S_f\xi, \eta\rangle = S(\xi, \eta)$.

Let D be the subspace of H spanned by vectors $S_f\xi$ for $f \in \mathscr{S}(\mathbb{R})$ and $\xi \in H$. I claim that D is dense. Indeed, let $\{f_\epsilon\}$ be an approximate identity in $\mathscr{S}(\mathbb{R})$ of the form $f_\epsilon(t) = \epsilon^{-1}f(\epsilon^{-1}t)$, f a fixed positive smooth function with support in $[-1, 1]$ and mass 1. Then

$$\|S_{f_\epsilon}\xi - \xi\| = \left\|\int_{-\infty}^{\infty} f_\epsilon(t)(U(t)\xi - \xi)\, dt\right\| \leqslant \sup_{t\in[-\epsilon,\epsilon]} \|U(t)\xi - \xi\|,$$

and this tends to 0 by the strong continuity. Moreover if $\xi \in D$ then the limit $\lim t \to 0(U(t)\xi - \xi)/it$ exists. In fact, for $\xi = S_f\eta$,

$$\frac{U(s)\xi - \xi}{is} = \int_{-\infty}^{\infty} f(t)\left(\frac{U(s+t) - U(t)}{is}\right)\eta\, ds =$$
$$\int_{-\infty}^{\infty} \frac{f(u-s) - f(u)}{is}U(u)\xi\, du \to i\int f'(u)U(u)\xi\, du = iS_{f'}\xi$$

using substitution and the Dominated Convergence Theorem. Let

$$B\xi = \lim_{t\to 0} \frac{U(t)\xi - \xi}{it}$$

on the dense domain D. A simple formal calculation shows that B is a symmetric operator.

Let A be the closure of B. We shall show that A is self-adjoint using the basic criterion 22.10. It suffices to show that there are no nonzero solutions to $A^*\xi = \pm i\xi$. Let ξ be such a solution and let $\eta \in D$. Then

$$\frac{d}{dt}\langle U(t)\eta, \xi\rangle = \langle iAU(t)\eta, \xi\rangle = \langle U(t)\eta, -iA^*\xi\rangle = \pm\langle U(t)\eta, \xi\rangle.$$

This differential equation shows that the complex-valued function $t \mapsto \langle U(t)\eta, \xi\rangle$ is exponential in t. Since it is obviously bounded, it must be zero. Taking $t = 0$ we conclude that ξ is orthogonal to D, so $\xi = 0$. Thus A is self-adjoint.

Finally we must prove that $e^{itA} = U(t)$. Indeed, $U(t)$ and $V(t)$ are now two strongly continuous one-parameter groups. If $\xi \in D$ then $U(t)\xi$ is a differentiable vector-valued function of t with derivative $iAU(t)\xi$, and similarly $V(t)\xi$ is a differentiable vector-valued function of t with derivative $iAV(t)\xi$. Thus if $\theta(t) = U(t)\xi - V(t)\xi$,

$$\frac{d}{dt}\|\theta(t)\|^2 = \langle iA\theta(t), \theta(t)\rangle + \langle\theta(t), iA\theta(t)\rangle = 0$$

so $\theta \equiv 0$ as required. □

Remark 23.4. Suppose that, instead of strong continuity, we assume merely that the group $U(t)$ is *weakly measurable* — that is, the functions $t \mapsto \langle U(t)\xi, \eta\rangle$ are measurable for all ξ, η. The above argument still works except for the proof that D is dense. To see this, suppose that η is orthogonal to D and let $\{e_n\}$ be an orthonormal basis for H. Fixing n, we have

$$\int f(t)\langle U(t)e_n, \eta\rangle dt = 0$$

for all $f \in \mathscr{S}(\mathbb{R})$, whence $\langle U(t)e_n, \eta\rangle = 0$ for almost all t, say for $t \notin N_n$ a set of measure zero. Let $t \notin \bigcup_n N_n$; then η is orthogonal to all the $U(t)e_n$; but these form a complete orthonormal set, so $\eta = 0$. Conclusion: a weakly measurable one-parameter unitary group is automatically strongly continuous. This extension of Stone's theorem is due to von Neumann.

For the final result of the course we shall prove the Friedrichs Extension Theorem. Let T be a symmetric operator on H. It is said to be *positive* if $\langle T\xi, \xi\rangle \geqslant 0$ for all $\xi \in \operatorname{dom}(T)$.

Theorem 23.5. *Let T be a positive symmetric operator. Then it has a positive self-adjoint extension.*

It is *not* asserted (and it is usually not true) that the self-adjoint extension constructed by this theorem (the *Friedrichs extension* of T) is the only self-adjoint extension that exists.

Proof. Equip $D = \operatorname{dom}(T)$ with the new inner product

$$Q(\xi, \eta) = \langle T\xi, \eta \rangle + \langle \xi, \eta \rangle.$$

This makes D into a pre-Hilbert space. Let H_1 denote its Hilbert space completion. The inclusion of D into H extends to a continuous linear injection $\iota \colon H_1 \to H$ with dense range.

Suppose that $\xi \in H$. Then the linear functional $y \mapsto \langle \xi, \iota(y) \rangle$ is continuous on H_1, so (by the Riesz representation theorem in H_1) there exists $C\xi \in H_1$ such that $Q(C\xi, y) = \langle \xi, \iota(y) \rangle$. Let $B = \iota C \colon H \to H$. The map $B \colon H \to H$ is linear and satisfies

$$Q(B\xi, y) = \langle \xi, y \rangle.$$

In particular

$$\langle \xi, B\eta \rangle = Q(B\xi, B\eta) = \overline{Q(B\eta, B\xi)} = \overline{\langle \eta, B\xi \rangle}$$

so B is symmetric. An everywhere-defined symmetric operator is automatically bounded, by the Closed Graph Theorem.

Suppose $x \in D$. Then

$$\langle Tx + x, y \rangle = Q(x, y)$$

which shows that $B(Tx + x) = x$. Thus the self-adjoint operator B has dense range, and therefore zero kernel. By the spectral theorem, B is unitarily equivalent to a multiplication by an almost everywhere nonvanishing function f on some measure space. Let $A = B^{-1}$ be the unbounded operator that is unitarily equivalent to multiplication by the (unbounded) function f^{-1}. A is self-adjoint and the calculation above shows that $Tx + x = Ax$ if $x \in D$. Thus $A - I$ is a self-adjoint operator that extends T. ☐