Note: This document was written by Prof. Richard Haydon, a previous lecturer for this course.

0. Preliminaries

This preliminary chapter is an attempt to bring together important results from earlier courses which we shall need this term. It is unlikely that I shall manage to include everything at the first attempt. So the chapter may grow during the course of the term as I think of more things that ought to have been included.

0.1. **Topology.**

I shall assume students attending this course know the basics of topology, as to be found in the second-year Section A course. What follows are a few reminders, intended mainly to establish some notation.

Definition 0.1. A topology on a set X is a subset $\mathscr{T}$ of the power set $\mathscr{P}X$ [that is to say, a collection $\mathscr{T}$ of subsets of X] satisfying:

TOP1 $\emptyset \in \mathscr{T}$, $X \in \mathscr{T}$;
TOP2 $\mathscr{S} \subseteq \mathscr{T} \implies \bigcup \mathscr{S} \in \mathscr{T}$;
TOP3 $U, V \in \mathscr{T} \implies U \cap V \in \mathscr{T}$.

Of course, it follows from TOP3 that if $\mathscr{A}$ is a *finite* subset of $\mathscr{T}$ then $\bigcap A \in \mathscr{T}$. We may thus summarize TOP2 and TOP3 as stating that $\mathscr{T}$ is closed under arbitrary unions and finite intersections.

When $\mathscr{T}$ is a topology on X we shall often refer to the elements of $\mathscr{T}$ as "*open sets*". If we have more than one topology of interest on a given set X we may even prefer the circumlocution "let U be a $\mathscr{T}$-open subset of X" to the more succinct "let $U \in \mathscr{T}$". If $\mathscr{T}_1$ and $\mathscr{T}_2$ are topologies on X, and $\mathscr{T}_1 \subseteq \mathscr{T}_2$ (i.e. every $\mathscr{T}_1$-open set is $\mathscr{T}_2$-open) we say that $\mathscr{T}_2$ is *finer* than $\mathscr{T}_1$, or that $\mathscr{T}_1$ is *coarser* than $\mathscr{T}_2$).

We say that a subset F of X is *closed*, for a topology $\mathscr{T}$ on X, if $X \setminus F \in \mathscr{T}$. The collection of closed sets is closed (!) under arbitrary intersections and finite unions.

When $(X, \mathscr{T})$ is a topological space and A is a subset of X we define the *subspace topology* on A to be the set

$$\mathscr{T}_A = \{A \cap U : U \in \mathscr{T}\}.$$

It is often convenient to say that a subset V of A is open *in* A (or $\mathscr{T}$-open in A when there are several topologies of interest) if $V \in \mathscr{T}_A$.

I am in too much of a hurry to write out the (I hope familiar) definitions of continuity, homeomorphism, compactness, closure, the Hausdorff property and so on. As a matter of notation, I may use either of the notations $\overline{A}$ or $\operatorname{cl} A$ or, more precisely $\operatorname{cl}_X A$, for the closure of a subset A in a topological space X. Recall that a subset A of X is said to be *dense* if its closure is the whole of X.

Convergent sequences are more popular with analysts than with general topologists. It will be useful for us to be aware of the following general definition and the (easy but surprisingly useful) lemma that follows.

Let $X, \mathscr{T}$ be a topological space and let $x, x_n \in X$ $(n \in \mathbb{N})$. We say that x_n converges to x in $\mathscr{T}$, writing

$$x_n \to x \quad \text{or} \quad x = \mathscr{T}\text{-lim} x_n,$$

if every open neighbourhood U of x contains x_n for all but finitely many values of n
(i.e. $\forall$ open $U \ni x\ \exists N \in \mathbb{N}\ \forall n \geq N\ x_n \in U$).

Lemma 0.2. *Let $X, \mathscr{T}$ be a topological space, and let x, x_n $(n \in \mathbb{N})$ be elements of X. If every subsequence of (x_n) has a further subsequence that converges to x then $x_n \to x$.*

You should keep in mind the important theorem that can be paraphrased by saying that "*if $(X, \mathscr{T})$ is compact and $\mathscr{T}'$ is a coarser Hausdorff topology on X then $\mathscr{T}' = \mathscr{T}$.*

Theorem 0.3. *Let $(X, \mathscr{T})$ be compact, let $(X', \mathscr{T}')$ be Hausdorff and let $f : X \to X'$ be a continuous bijection. Then f^{-1} is continuous (i.e. f is a homeomorphism).*

Definition 0.4. A *metric* on a set X is a function $d : X \times X \to [0, \infty)$ satisfying:

D1 $d(x, y) = 0 \implies x = y$;
D2 $d(x, y) = d(y, x)$ for all $x, y \in X$;
D3 $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$.

We define the open ball $B(x; r)$ (writing $B^d_X(x; r)$ if there may be ambiguity about what the metric is and what the "ambient space" X is) and the closed ball $B[x; r]$ with centre x and radius r by

$$B(x; r) = \{y \in X : d(x, y) < r\}$$
$$B[x; r] = \{y \in X : d(x, y) \leq r\}$$

The open balls form a base for the topology $\mathscr{T}_d$. A topology $\mathscr{T}$ is said to be *metrizable* if there is some metric d with $\mathscr{T} = \mathscr{T}_d$.

Metrizable topologies have some nice properties, notably those allowing a description of topological properties like continuity, closure and compactness in terms of convergent sequences; I hope you are familiar with these. Be warned, however, that not all of the topologies that occur in functional analysis are metrizable. (See, for instance, the weak and weak* topologies of Chapter 5)

Here is one important feature of metrizable topologies that may have slipped your attention. Recall that a topological space X is said to be *separable* if there is a countable subset X which is dense in X.

Proposition 0.5. *If X is a separable metrizable space and Y is a subset of X then Y is separable (for the subspace topology).*

Look up a proof if you need to. Just note that the "obvious approach", which starts *let D be countable and dense in X, and consider $D \cap Y$*, does not work ($D \cap Y$ is frequently empty). For general topological spaces a subset of a separable space need not be separable. [1]

0.2. **Integration.** It is not my intention to give a course on Measure and Integration, but we shall have to be ready to use results from that subject in a rather more formal way than you will have experienced in the Section A course.

Definition 0.6. Let Ω be a non-empty set. We say that a collection of subsets $\mathscr{F}$ of Ω is a *σ-field* (or σ-algebra) if the following hold:

σF1 $\emptyset \in \mathscr{F}$;
σF2 $F \in \mathscr{F} \implies \Omega \setminus F \in \mathscr{F}$;
σF3 $F_n \in \mathscr{F}$ $(n \in \mathbb{N}) \implies \bigcup_{n\in\mathbb{N}} F_n \in \mathscr{F}$.

Thus a σ-field $\mathscr{F}$ is closed under complementation and *countable* unions. It follows from de Morgan's laws that $\mathscr{F}$ is also closed under countable intersections. For any collection $\mathscr{A}$ of subsets of a set Ω there is a smallest σ-field $\sigma\mathscr{A}$ containing $\mathscr{A}$. In particular, when $(L, \mathscr{T})$ is a topological space, there is a smallest σ-field containing $\mathscr{T}$; this is called the *Borel* σ-field of L.

Definition 0.7. Let $\mathscr{F}$ be a σ-field of subsets of a set Ω. A function $\mu : \mathscr{F} \to [0, \infty]$ is said to be a *measure* on $\mathscr{F}$ if:

M1 $\mu(\emptyset) = 0$;
M2 $\mu\left(\bigcup_{n\in\mathbb{N}} F_n\right) = \sum_{n\in\mathbb{N}} \mu(F_n)$ whenever $F_1, F_2, \ldots$ are *pairwise disjoint* members of $\mathscr{F}$.

We then say that $(\Omega, \mathscr{F}, \mu)$ is a *measure space.* If $\mu(\Omega) = 1$ we say that μ is a *probability measure* and that $(\Omega, \mathscr{F}, \mu)$ is a probability space. In this case the notation $\mathbb{P}$ is usually used instead of μ.

Examples 0.8.

(1) On any set Ω we may take $\mathscr{F}$ to be the collection $\mathscr{P}(\Omega)$ of *all* subsets of Ω define the "counting measure" # by setting $\#(F) = n$ if F is a finite set with n elements and $\#(F) = \infty$ if F is infinite.
(2) On $\Omega = \mathbb{R}$ we may take $\mathscr{F}$ to be the collection $\mathscr{M}_{\mathrm{Leb}}$ of all Lebesgue measurable sets, and μ to be Lebesgue measure.
(3) Taking Ω to be the unit interval $(0, 1)$, and $\mathbb{P}$ to be the restriction of Lebesgue to the subsets of $(0, 1)$ we have an important example of a probability space.
(4) Let Ω be any uncountable set, take $\mathscr{F} = \mathscr{P}(\Omega)$ and define

$$\mu(F) = \begin{cases} 0 & \text{if } F \text{ is countable} \\ \infty & \text{if } F \text{ is uncountable.} \end{cases}$$

This is a nasty example and we usually make assumptions on our measure spaces to exclude such pathologies. The commonest such assumption is given below as Definition 0.9.

Definition 0.9. We say that a measure space $(\Omega, \mathscr{F}, \mu)$ is *σ-finite* if there exist $A_n \in \mathscr{F}$ with $\Omega = \bigcup_{n\in\mathbb{N}} A_n$ and $\mu(A_n) < \infty$ for all n.

From now on, we shall assume that all our measure spaces are σ-finite.

Definition 0.10. A real-valued function x on a measure space $(\Omega, \mathscr{F}, \mu)$ is said to be *measurable* ($\mathscr{F}$-measurable if we are being careful) if $\{t \in \Omega : x(t) \le \alpha\} = x^{-1}(-\infty, \alpha] \in \mathscr{F}$ for every real α. It follows that $x^{-1}[B] \in \mathscr{F}$ for every Borel subset of $\mathbb{R}$. Borrowing notation from probability theory, we shall write

$$[x \le \alpha] = x^{-1}(-\infty, \alpha], \quad [x \in B] = x^{-1}[B],$$

and so on. For any $F \in \mathscr{F}$ we define the *indicator function* $\mathbf{1}_F$ by setting

$$\mathbf{1}_F(t) = \begin{cases} 1 & \text{if } t \in F \\ 0 & \text{if } t \notin F. \end{cases}$$

[1]Indeed the study of *hereditarity separable* spaces is "a can of worms".

A function $g : \Omega \to \mathbb{R}$ is called an $\mathscr{F}$-*simple* function if we can write

$$g = \sum_{j=1}^{n} \alpha_j \mathbf{1}_{F_j},$$

where $\alpha_1, \dots, \alpha_n$ are real numbers and $F_1, \dots, F_n$ are members of $\mathscr{F}$ with $\mu(F_j) < \infty$.

Definition 0.11. For an $\mathscr{F}$-simple function $g = \sum_{j=1}^{n} \alpha_j \mathbf{1}_{F_j}$ we (well!) define

$$\int_\Omega g \, \mathrm{d}\mu = \sum_{j=1}^{n} \alpha_j \mu(F_j).$$

For a *non-negative* real-valued $\mathscr{F}$-measurable function x on Ω, we define

$$\int_\Omega x \, \mathrm{d}\mu = \sup\{ \int_\Omega g \, \mathrm{d}\mu : \ g \le x \text{ and } g \text{ is } \mathscr{F}\text{-simple}\} \in [0, \infty]$$

We say that x is *integrable* if this supremum is finite. For a general real-valued $\mathscr{F}$-measurable function x, we say that x is integrable, if both x^+ and x^- are, and we set

$$\int_\Omega x \, \mathrm{d}\mu = \int_\Omega x^+ \, \mathrm{d}\mu - \int_\Omega x^- \, \mathrm{d}\mu.$$

When $(\Omega, \mathscr{F}, \mathbb{P})$ is a probability space, it is usual to refer to refer to measurable functions as *random variables* and to use the notation $\mathbb{E}[x]$ for $\int x \, \mathrm{d}\mathbb{P}$; we call $\mathbb{E}[x]$ the *expectation* of x.

In all cases, the set $\mathscr{L}_1(\Omega, \mathscr{F}, \mu)$ of μ-integrable functions is a vector space and the integral is a linear mapping. There are two important *convergence theorems.*

Theorem 0.12 (The Monotone Convergence Theorem)**.** *Let $(x_n)_{n\in\mathbb{N}}$ be an increasing sequence of μ-integrable functions with* $\sup_n \int x_n \, \mathrm{d}\mu < \infty$. *Then the sequence $(x_n(t))_{n\in\mathbb{N}}$ converges (in $\mathbb{R}$) for almost all values of t and the function x defined (almost everywhere) by $x(t) = \lim_n x_n(t)$ is integrable with $\int x \, \mathrm{d}\mu = \lim \int x_n \, \mathrm{d}\mu$.*

In the above statement we have used the useful expression “almost everywhere”, often shortened to a.e. In general, we say that a property $P(t)$ holds μ-almost everywhere if

$$\mu\{t \in \Omega : P(t) \text{ is false } \} = 0.$$

In particular, we say that two measurable functions x and y “agree almost everywhere” if $\mu[x \ne y] = 0$. Usually we do not need to distinguish two such functions (e.g. in the context of the space $L_1(\mu)$) and so it is not a problem for us if a function is defined only a.e. In the context of a probability space it is usual to say “almost surely”, rather than “$\mathbb{P}$-almost everywhere”.

Theorem 0.13 (The Dominated Convergence Theorem)**.** *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence of μ-integrable functions such that $x_n(t)$ converges to a real limit $x(t)$ almost everywhere. Assume further that there exists a “dominating function” $d \in \mathscr{L}_1(\mu)$ such that $|x_n(t) \le d(t)$ for all n and almost all t. Then x is integrable and $\int x_n \, \mathrm{d}\mu \to \int x \, \mathrm{d}\mu$ as $n \to \infty$.*

Proposition 0.14. *If x is a nonnegative μ-integrable function and $\int x(t) \, d\mu(t) = 0$ then $x(t) = 0$ μ-almost everywhere.*

Proof. Here are two proofs of this important result; each one is instructive in its own way.

(1) For each $n \in \mathbb{N}$, let $E_n = [x \ge n^{-1}]$ and let $g_n = n^{-1}\mathbf{1}_{E_n}$. Then g_n is a simple and $g_n \le x$; hence by definition of the integral $n^{-1}\mu(E_n) = \int g_n \, \mathrm{d}\mu \le 0$. So $\mu\left(\bigcup_n E_n\right) = 0$ and $x(t) = 0$ a.e.
(2) Let $x_n = nx$ and apply the MCT.

□

Often it will suit us to regard two integrable functions that agree almost everywhere as being the same. This corresponds to taking equivalence classes modulo the relation of a.e. equality. Another way of saying this is to say we are looking at the quotient space $\mathscr{L}_1/\mathscr{N}$, where $\mathscr{N}$ is the subspace of all functions x of the type considered in Proposition 0.14. We write L_1 for this quotient space. The “full notation” is of course $L_1(\Omega, \mathscr{F}, \mu)$, and we shall sometimes write this. At other times we may leave out some (or all) of the arguments if this does not lead to ambiguity. For instance, we may sometimes write $L_1(\Omega)$ if it is clear from the context what $\mathscr{F}$ and μ are. In particular, if we write $L_1(\mathbb{R})$ and $L_1(0,1)$ we are talking about Lebesgue measure and the σ-field of Lebesgue-measurable sets.

0.3. **Normed Spaces.** In this section, we briefly recall material from the B4 course, establishing notation that will be used later, and providing easy references for key results. It is hoped that the missing proofs are familiar to you.

Definition 0.15. Let X be a vector space over the field $\mathbb{K}$ (which is either $\mathbb{R}$ or $\mathbb{C}$. A *norm* on X is a mapping $\nu : X \to \mathbb{R}$ satisfying:

N1 $\nu(x) \geq 0$ for all $x \in X$, and $\nu(x) = 0 \implies x = 0$;
N2 $\nu(\lambda x) = |\lambda|\nu(x)$ for all $x \in X$ and $\lambda \in \mathbb{K}$;
N3 $\nu(x + y) \leq \nu(x) + \nu(y)$ for all $x, y \in X$.

We usually use a notation like $\|x\|$ (with subscripts when necessary) for $\nu(x)$ when ν is a norm. In this course, most of the results do not depend upon which of the two fields $\mathbb{K}$ is. However, some of the inequalities need to be modified, or at least are harder to prove; for simplicity we shall assume $\mathbb{K} = \mathbb{R}$.

If X is an inner-product space, we may define a norm by

$$\|x\| = \sqrt{\langle x, x\rangle},$$

but it is important to remember that not all norms are associated with inner products.

Examples 0.16.

(1) On $\mathbb{R}^n$ we may define norms $\|\cdot\|_p$ $(1 \leq p \leq \infty)$ by

$$\|(a_1, a_2, \dots, a_n)\|_p = \left(\sum_{j=1}^{n} |a_j|^p\right)^{1/p} \quad (1 \leq p < \infty)$$

$$\|(a_1, a_2, \dots, a_n)\|_\infty = \max_{1 \leq j \leq n} |a_j|.$$

The space $\mathbb{R}^n$, when equipped with $\|\cdot\|_p$ will be denoted ℓ_p^n.

(2) We define ℓ_∞ to be the space of all bounded scalar sequences $\mathbf{a} = (a_1, a_2, a_3, \dots)$, equipped with the supremum norm

$$\|\mathbf{a}\|_\infty = \sup_n |a_n|.$$

The subspace c_0 (resp. c) of ℓ_∞ consists of those sequences that converge to 0 (resp. to some limit).

(3) For $1 \leq p < \infty$ the space ℓ_p is defined to consist of all p-summable sequences (i.e. sequences $(a_n)_{n \in \mathbb{N}}$ such that the series $\sum_{n=1}^{\infty} |a_n|^p$ converges. For such a sequence we define

$$\|(a_1, a_2, a_3, \dots)\|_p = \left(\sum_{j=1}^{\infty} |a_j|^p\right)^{1/p}.$$

(4) If L is a topological space we define $\mathscr{C}^{\mathrm{b}}(L)$ to be the space of all *bounded* continuous scalar-valued functions on L, and equip it with the supremum norm

$$\|x\|_\infty = \sup_{t \in L} |x(t)|.$$

When L is compact, all functions on L are bounded so we may leave out the superscript b and write simply $\mathscr{C}(L)$.

The norm axioms are easy to verify for all of the cases in 0.16 except for the ℓ_p-norms where $p \notin \{1, 2, \infty\}$. For these, we need Minkowski's Inequality

$$\left(\sum_j |a_j + b_j|^p\right)^{1/p} \leq \left(\sum_j |a_j|^p\right)^{1/p} + \left(\sum_j |b_j|^p\right)^{1/p}.$$

Another important class of Banach spaces are the *Lebesgue spaces* L_p. These may already be familiar to you, but we shall give a detailed treatment including a proof of Minkowski's Inequality in subsection 0.5.

Given a norm on a space X, we can define a metric by $d(x, y) = \|x - y\|$ and introduce the associated topology, thus giving meaning to such notions as continuity of mappings, as well as completeness, separability, compactness, precompactness for X and its subsets. We retain the ball notation, putting $B_X(x; r) = \{y \in X : \|x - y\| < r\}$ and $B_X[x; r] = \{y \in X : \|x - y\| \leq r\}$. It has become standard to write B_X for $B_X[0; 1]$ and to refer to this as "the" unit ball of X. The unit sphere $\{x \in X : \|x\| = 1\}$ is denoted S_X.

Given two norms $\|\cdot\|$ and $\|\cdots\|'$ on the same space X, we say that $\|\cdot\|$ *dominates* $\|\cdot\|'$ if there exists a positive constant C such that $\|x\|' \leq C\|x\|$ for all x. If in addition there exists C' such that $\|x\| \leq C'\|x\|'$, we say that the two norms are *equivalent*. On the space $\mathbb{R}^n$ we may (fairly) easily see that

$$\|(a_1, a_2, \dots, a_n)\|_q \leq \|(a_1, a_2, \dots, a_n)\|_p \leq C\|(a_1, a_2, \dots, a_n)\|_q,$$

where $C = n^{1-p/q}$ when $p \leq q$. In particular, all the ℓ_p norms are equivalent on $\mathbb{R}^n$ (though with constants that get worse as n increases.

In fact, more is true.

Proposition 0.17. *On a finite dimensional space X all norms are equivalent.*

Corollary 0.18. *If X is finite-dimensional then closed bounded subsets of X are compact.*

It is very important to remember that closed bounded sets are not compact in general. Indeed we have the following converse to the above corollary.

Proposition 0.19. *If X is a normed space and the unit ball B_X is precompact then X is finite-dimensional.*

Exercise 0.20. Let c_{00} denote the vector space of sequences $\mathbf{a} = (a_1, a_2, a_3, \dots)$ such that all but finitely many of the terms a_j are zero. Then c_{00} may be equipped with any of the ℓ_p-norms of 0.16(2) and (3). We see (fairly) easily that $\|\cdot\|_p$ dominates $\|\cdot\|_q$ if and only if $p \leq q$. These norms are thus non-equivalent when $p \neq q$.

Definition 0.21. A normed space which is *complete* is called a Banach space. All of the spaces in 0.16 are Banach spaces. By Proposition 0.17 every finite-dimensional normed space is complete. A complete inner-product space is called a *Hilbert space*; ℓ_2 ad L_2 are Hilbert spaces.

Definition 0.22. Let x_n ($n \in \mathbb{N}$) be elements of a normed space X. We say that the series $\sum_{n=1}^{\infty} x_n$ *converges* in X if there exists $s \in X$ such that $\|s - \sum_{j=1}^{n} x_j\| \to 0$ as $n \to \infty$. We say that the series $\sum_{n=1}^{\infty} x_n$ *converges absolutely* if the scalar series $\sum_{n=1}^{\infty} \|x_n\|$ is convergent.

It is a familiar fact from Mods Analysis that absolute convergence implies convergence for scalar series. Here is the normed space version of that result.

Proposition 0.23. *Let X be a normed space. Then X is a Banach space if and only if every absolutely convergent series in X is convergent in X.*

0.4. **Bounded linear operators and functionals.**

Proposition 0.24. *Let X and Y be normed spaces and let $T : X \to Y$ be a linear mapping. Then the following are equivalent:*

(1) *T is continuous;*
(2) *T is continuous at 0;*
(3) *there exists a constant M such that $\|T(x)\| \leq M\|x\|$ for all $x \in X$.*

Definition 0.25. A linear mapping T satisfying the equivalent conditions of Proposition 0.24 will be called a *bounded linear operator* (often abbreviated to BLO). The smallest constant M that works in 0.24(3) will be called the *operator norm* of T and will be denoted $\|T\|$. The space of all bounded linear operators from X to Y will be denoted $\mathcal{L}(X;Y)$; $\mathcal{L}(X;X)$ is usually abbreviated to $\mathcal{L}(X)$; $\mathcal{L}(X;\mathbb{K})$ is denoted X^* and is called the *dual* space of X. The elements of X^* are usually referred to as bounded (or continuous) linear "functionals" on X.

The reader should be warned that there are other notations in popular use: some authors prefer $\mathcal{B}(X)$ rather than $\mathcal{L}(X)$ (emphasizing "boundedness", rather than linearity); others would use X' for our X^*.

Proposition 0.26. *When X and Y are normed spaces, $\mathcal{L}(X;Y)$ is a normed space when equipped with the operator norm. If Y is a Banach space so is $\mathcal{L}(X;Y)$. In particular X^* is always a Banach space.*

Examples 0.27.

(1) The dual of c_0 is, in a natural way, isometrically isomorphic to ℓ_1.
(2) For $1 \leq p < \infty$ the dual of ℓ_p is, in a natural way, isometrically isomorphic to ℓ_q with $1/p + 1/q = 1$.
(3) A similar result holds for L_p and we shall prove it in Chapter 2.
(4) For any real Hilbert space there is a natural isometric isomorphism between H and H^* (something similar happens for complex Hilbert spaces but complex conjugation is involved);
(5) There is a representation theorem for the dual space of $\mathscr{C}(K)$ when K is compact in terms of "signed measures" the Borel σ-field of K. We do not have room for this useful result in the present course.

Definition 0.28. Let X and Y be normed spaces and let $T : X \to Y$ be linear. We say that T is an *isomorphism* if both T and T^{-1} are continuous. [Thus an isomorphism is a linear homeomorphism.] An equivalent condition is that T be surjective and that there exist positive constants A, B such that

$$A^{-1}\|x\| \leq \|T(x)\| \leq B\|x\|$$

for all $x \in X$. If T satisfies such an inequality, but is not necessarily surjective, we say that T is an isomorphic *embedding*. We say that T is an *isometric* isomorphism (or embedding) if $\|T(x)\| = \|x\|$ for all x.

A powerful technique in functional analysis is the use of dense linear subspaces. When A is a subset of a normed space X, we write $\mathrm{sp}\langle A\rangle$ for the linear subspace generated by A and $\overline{\mathrm{sp}}\langle A\rangle$ for its closure (also a linear subspace). A normed space is separable if there exists a countable subset A with $\overline{\mathrm{sp}}\langle A\rangle = X$.

Examples 0.29.

(1) If $X = \ell_p$ ($1 \leq p < \infty$ or c_0 the unit vectors
$$\mathbf{e}_n = (0, 0, 0, \dots, 0, 1, 0, \dots),$$
(with 1 in the n^{th} coordinate) form a subset with dense linear span, i.e. $\overline{\mathrm{sp}}\langle \mathbf{e}_n : n \in \mathbb{N}\rangle = X$. Note that this is not the case if $X = \ell_\infty$ or $X = c$.
(2) If $X = L_p(\mathbb{R})$ ($1 \leq p < \infty$) then we may take A to consist of all indicator functions $\mathbf{1}_I$ with I a bounded interval. Then $\mathrm{sp}\langle A\rangle$ is the space of all *step functions* (more precisely their equivalence classes modulo null functions) and $\overline{\mathrm{sp}}\langle A\rangle = L_p(\mathbb{R})$. If we restrict attention to intervals I with rational end-points we obtain a countable subset A with $\overline{\mathrm{sp}}\langle A\rangle = L_p(\mathbb{R})$, thus establishing separability of $L_p(\mathbb{R})$.
(3) In the space $\mathscr{C}[0,1]$ the polynomials form a dense linear subspace (though this has not yet been proved in a course you have attended). Another interesting dense linear subspace is formed by the *piecewise-linear* functions. In this case, density is easy to prove (using uniform continuity).

Theorem 0.30 (Uniqueness theorem)**.** *Let X and Y be normed spaces, let T_1 and T_2 be bounded linear operators from X to Y and let A be a subset of X with $\overline{\mathrm{sp}}\langle A\rangle = X$. If $T_1 \restriction_A = T_2 \restriction_A$ then $T_1 = T_2$.*

Proof. Given our hypotheses, $\ker(T_2 - T_1)$ is a closed linear subspace containing A and hence containing $\overline{\mathrm{sp}}\langle A\rangle = X$. □

Theorem 0.31 (Continuous extension)**.** *Let X be a normed space, let Y be a Banach space, let Z be a dense linear subspace of X and let $T : Z \to Y$ be a bounded linear operator. Then there is a unique bounded linear operator $U : X \to Y$ extending T. We have $\|U\| = \|T\|$. If T is an isomorphic (resp. isometric) embedding then so is U.*

Theorem 0.32 (Convergence theorem)**.** *Let X be a normed space, let A be a subset of X with dense linear span, let Y be a Banach space and let $(T_n)_{n\in\mathbb{N}}$ be a sequence of bounded linear operators from X to Y. Assume*

(1) $\sup_{n\in\mathbb{N}} \|T_n\| < \infty$*;*
(2) *$T_n(x)$ converges to a limit in Y for every $x \in A$.*

Then there is a bounded linear operator $T : X \to Y$ such that $\|T_n(x) = T(x)\| \to 0$ for all $x \in X$.

Theorem 0.33 (Hahn–Banach Extension Theorem)**.** *Let Z be a linear subspace of a normed space X and let $g \in Z^*$. Then there exists $f \in X^*$ with $\|f\| = \|g\|$ and $f \restriction_Z = g$.*

Corollary 0.34. *If Z is a normed space and $0 \neq z \in Z$ then there exists $h \in Z^*$ with $\|h\| = 1$ and $h(z) = \|z\|$.*

Definition 0.35. If A is a subset of a normed space X, we define

$$A^\perp = \{f \in X^* : f(x) = 0 \text{ for all } x \in A\},$$

and when $B \subseteq X^*$ we define

$$B_\perp = \{x \in X : f(x) = 0 \text{ for all } f \in B.$$

These are called *annihilators.*

Proposition 0.36 (Another corollary to the Hahn–Banach theorem)**.** *Let X be a normed space and let A be a subset of X. Then*

$$\left(A^\perp\right)_\perp = \overline{\mathrm{sp}}\langle A\rangle.$$

Definition 0.37. Let X and Y be normed spaces and let $T : X \to Y$ be a bounded linear operator. Then

$$T^*(g) = g \circ T \quad (g \in Y^*)$$

defines a linear mapping $T^* : Y^* \to X^*$, called the *dual operator.* The words "transpose" and "adjoint" are also used, but the latter is best reserved for the Hilbert-space adjoint.

Proposition 0.38. *If $T \in \mathcal{L}(X;Y)$ then $T^* \in \mathcal{L}(Y^*, X^*)$ and $\|T^*\| = \|T\|$.*

Proof. Easy, but notice that Corollary 0.34 is needed for the equality of operator norms. □

0.5. **L_p-spaces.** I hope you have come across all this material already, but just in case I have included full proofs of everything.

Definition 0.39. Let $(\Omega, \mathscr{F}, \mu)$ be a measure space and let p be a positive real number. We define $\mathscr{L}_p(\Omega, \mathscr{F}, \mu)$ to be the space of all μ-measurable functions x for which the p^{th} power $|x|^p$ is μ-integrable. Sometimes we abbreviate notation to $\mathscr{L}_p(\mu)$, or even just $\mathscr{L}_p$. Let $\mathscr{N}$ be the subspace consisting of those functions that are almost everywhere zero [i.e. $\mu\{t \in \Omega : x(t) \neq 0\} = 0$]. The quotient space $\mathscr{L}_p/\mathscr{N}$ is denoted $L_p(\mu)$ and for a coset $x^\bullet = x + \mathscr{N} \in L_p$ we (well!)-define

$$\|x^\bullet\|_p = \left(\int_\Omega |x(t)|^p \,\mathrm{d}\mu\right)^{1/p}.$$

Note that if μ is the counting measure $\#$ on $\mathbb{N}$ (resp. $\{1, 2, \ldots, n\}$ then $L_p(\mu)$ is naturally identifiable with ℓ_p (resp. ℓ_p^n). It is often useful to use "function" notation even in these familiar sequence spaces, regarding an element of ℓ_p as a function $x : \mathbb{N} \to \mathbb{R}$, rather than as a sequence $(x_1, x_2, \ldots)$. This means that we write $x(n)$ instead of x_n for the n^{th} co-ordinate, and simplifies things when (as often) we want to consider a sequence (x_k) each of whose terms is an element of ℓ_p. Generalizing slightly, we write $\ell_p(\Gamma)$ for the L_p-space associated with the counting measure on an arbitrary set Γ.

Proposition 0.40 (The inequalities of Hölder and Minkowski)**.** *Let p be a real number with $p > 1$ and let q denote the "conjugate index" given by $p^{-1} + q^{-1} = 1$. Then*

- HS *if $\mathbf{a} \in \ell_p$ and $\mathbf{b} \in \ell_q$ then $(a_i b_i)_{i\in\mathbb{N}} \in \ell_1$ and $|\sum_i a_i b_i| \leq \|\mathbf{a}\|_p \|\mathbf{b}\|_q$ whenever $\mathbf{a} \in \ell_p$ and $\mathbf{b} \in \ell_q$;*
- HI *if $x \in \mathscr{L}_p(\mu)$ and $y \in \mathscr{L}_q(\mu)$ then $xy \in \mathscr{L}_1(\mu)$ and $|\int_\Omega x(t)y(t)\,\mathrm{d}\mu(t)| \leq \|x\|_p \|y\|_q$ whenever $x \in L_p(\mu)$ and $y \in L_q(\mu)$.*
- MS $\|\mathbf{a} + \mathbf{b}\|_p \leq \|\mathbf{a}\|_p + \|\mathbf{b}\|_p$ *whenever $\mathbf{a}, \mathbf{b} \in \ell_p$;*
- MI $\|x + y\|_p \leq \|x\|_p + \|y\|_p$ *whenever $x, y \in L_p(\mu)$.*

The proof of these inequalities is based on an elementary lemma.

Lemma 0.41 (A special case of Young's Inequality)**.** *Let $t, u > 0$ and $0 < \lambda < 1$ be real numbers. Then*

$$t^\lambda u^{1-\lambda} \leq \lambda t + (1-\lambda)u,$$

with equality only when $t = u$. If we write $x = t^\lambda$, $y = u^{1-\lambda}$, $p = 1/\lambda$, $q = 1/(1-\lambda)$ we have

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q}.$$

Proof. The function $\phi(t) = t^\lambda$ is strictly concave on $[0, \infty)$ (because $\phi''(t) < 0$ for all positive t). Thus the graph of $y = \phi(t)$ is everywhere below the tangent at $(u, \phi(u))$. That is to say

$$t^\lambda = \phi(t) \leq \phi(u) + (t-u)\phi'(u) = u^\lambda + \lambda(t-u)u^{\lambda-1}.$$

Multiplying by $u^{1-\lambda}$ yields the result. □

Proof of Proposition 0.40. We shall prove [HI] and [MI], since the inequalities for sums may be regarded as special cases (by considering the counting measure). So let $x \in L_p$ and $y \in L_q$ and assume initially that $\|x\|_p = \|y\|_q = 1$. For each $s \in \Omega$ we apply the lemma to the positive real numbers $|x(s)|, |y(s)|$ obtaining

$$|x(s)y(s)| \leq \frac{|x(s)|^p}{p} + \frac{|y(s)|^q}{q}.$$

Integrating, we obtain

$$|\int_\Omega x(s)y(s)\,\mathrm{d}\mu(s)| \leq \int_\Omega |x(s)y(s)|\,\mathrm{d}\mu(s) \leq \int_\Omega \left(\frac{|x(s)|^p}{p} + \frac{|y(s)|^q}{q}\right)\mathrm{d}\mu(s) = \frac{\|x\|_p^p}{p} + \frac{\|y\|_q^q}{q} = 1$$

For general x and y we introduce the normalized versions $\hat{x} = \|x\|_p^{-1}x$ and $\hat{y} = \|y\|_q^{-1}y$ for which we have

$$|\int \hat{x}\hat{y}\,\mathrm{d}\mu| \leq 1.$$

This immediately yields [HI].

We now pass to [MI]. Let x and y be in $L_p(\mu)$. We notice that

$$\|x+y\|_p^p = \int |x(s)+y(s)|^p \mathrm{d}\mu(s) = \int (x(s)+y(s))z(s)\,\mathrm{d}\mu(s) = \int x(s)z(s)\,\mathrm{d}\mu(s) + \int y(s)z(s)\,\mathrm{d}\mu(s),$$

where $z(s) = \mathrm{sign}\,(x(s)+y(s))|x(s)+y(s)|^{p-1}$. Now this function z is in L_q with

$$\|z\|_q^q = \int |z|^q \mathrm{d}\mu = \int |x+y|^{q(p-1)} = \int |x+y|^p = \|x+y\|_p^p.$$

[This is part of the magic of conjugate indices.] Applying [HI] twice we obtain

$$\|x+y\|_p^p = \int |x(s)+y(s)|^p \mathrm{d}\mu(s) = \int x(s)z(s)\,\mathrm{d}\mu(s) + \int y(s)z(s)\,\mathrm{d}\mu(s) \leq \|x\|_p\|z\|_q + \|y\|_p\|z\|_q,$$

which, after dividing by $\|z\|_q = \|x+y\|_p^{p/q}$, becomes

$$\|x+y\|_p = \|x+y\|_p^{p-p/q} = \|x+y\|_p^p\|z\|_q^{-1} \leq \|x\|_p + \|y\|_p,$$

as required. □

Exercise 0.42. Investigate the conditions under which equality occurs in the inequalities of Hölder and Minkowski.

Proposition 0.43. *When $1 \leq p < \infty$, the space $L_p(\mu)$ is a Banach space under the norm $\|\cdot\|_p$.*

Proof. We have already remarked in Definition 0.39 that $\|x^\bullet\|_p$ is well-defined. Let's just check on this:

$$x^\bullet = y^\bullet \implies x(t) = y(t)\ \mu - \text{a.e.} \implies \int |x|^p \mathrm{d}\mu = \int |y|^p \mathrm{d}\mu.$$

The norm axiom N2 is obvious and N3 is Minkowski's Inequality. We are left with N1. Suppose then that $\|x^\bullet\|_p = 0$. The non-negative integrable function $|x|^p$ thus satisfies

$$\int |x|^p \mathrm{d}\mu = 0.$$

So $x(t) = 0$ μ-a.e. by Proposition 0.14, $x \in \mathscr{N}$ and $x^\bullet = 0$.

To prove completeness we shall use Proposition 0.23. Let $x_n^\bullet \in L_p(\mu)$ be such that $\sum_n \|x_n\|_p$ converges. Our first task is to show that the scalar series $\sum_n x_n(t)$ converges a.e. We set $s_n = \sum_{j=1}^n x_n(t)$, $S_n = \sum_{j=1}^n |x_n(t)|$ and note that $S_n \in \mathscr{L}_p(\mu)$ with

$$\left(\int S_n^p \mathrm{d}\mu\right)^{1/p} \leq \sum_{j=1}^n \|x_j^\bullet\|_p$$

by Minkowski's inequality. Setting $u_n = S_n^p$, we have an increasing sequence of μ-integrable functions satisfying

$$\int u_n \,\mathrm{d}\mu \le M = \left(\sum_{n=1}^{\infty} \|x_n^\bullet\|_p\right)^p.$$

By the Monotone Convergence Theorem, $u_n(t)$ tends to a finite limit $u(t)$ almost everywhere, and the function u is integrable with $\int u \,\mathrm{d}\mu \le M$. It follows that the scalar series $\sum_n |x_n(t)|$ converges almost everywhere to a finite limit $S(t) = u(t)^{1/p}$ and that $S \in \mathscr{L}_p(\mu)$ with $\int S^p \mathrm{d}\mu \le M$.

Thus the series $\sum_n x_n(t)$ converges a.e. (since absolute convergence implies convergence for scalar series) and the limit function s satisfies

$$s \le S \quad \text{a.e.}$$

So $s^\bullet \in L_p$ with

$$\|s^\bullet\|_p \le \|S^\bullet\|_p \le \sum_{n=1}^{\infty} \|x_n^\bullet\|_p.$$

To finish the completeness proof we need to check that $\|s - s_m\|_p \to 0$ as $m \to \infty$. This is easy since the above argument shows that

$$\|s^\bullet - s_n^\bullet\|_p \le \sum_{n=m+1}^{\infty} \|x_n^\bullet\|_p.$$

□

For $p = \infty$ things are a little different, since the space $\mathscr{L}_\infty$ of bounded measurable functions is already a Banach space for the supremum norm. (It's a closed subspace of the Banach space $\ell_\infty(\Omega)$ of *all* bounded functions on Ω. However, we still have to take a quotient if we want to identify functions that differ only on a null set. This time the quotient is just a special case of Definition 1.4.

Definition 0.44. When $(\Omega, \mathscr{F}, \mu)$ is a measure space we define $\mathscr{L}_\infty(\Omega, \mathscr{F}, \mu)$ to be the space of all bounded μ-measurable functions, equipped with the supremum norm. Let $\mathscr{N}_\infty$ be the closed subspace consisting of those bounded functions that are almost everywhere zero [so $\mathscr{N}_\infty = \mathscr{N} \cap \mathscr{L}_\infty$]. The quotient space $\mathscr{L}_\infty/\mathscr{N}_\infty$ is denoted $L_\infty(\mu)$ and is a Banach space equipped with the quotient norm.

Proposition 0.45. *For $x \in \mathscr{L}_\infty$ the quotient norm is given by*

$$\|x^\bullet\|_\infty = \inf\{\lambda \in \mathbb{R} : |x(t)| \le \lambda \text{ a.e.}\}.$$

By abuse of notation, we shall generally leave out the "blob" when referring to a coset $x^\bullet \in L_p$. This convention requires a little care (particularly when $p = \infty$) but to do otherwise would leave us open to accusations of pedantry.

When $0 < p < 1$ the function $\|\cdot\|_p$ is not a norm (the triangle inequality fails). We should not assume from this that such values of p are without interest to analysts and probabilists. Losing the triangle inequality is not the end of the world.

0.6. **Probability.** Probability theory is far from being "just" a special case of measure theory. As well as its important applications in the "real world", probability has a lot to offer analysts, including those of us who are interested in Banach spaces and operators.

Thinking about elements of $L_1[0,1]$ as random variables allows us to introduce notions like independence, distribution functions and characteristic functions and to do useful calculations. I shall repeat only a few of the standard definitions here.

With any random variable x [2] we can associate the *distribution function* $F = F_x$ defined by

$$F(t) = \mathbb{P}[x \le t] \quad (t \in \mathbb{R}).$$

Some random variables admit a *density function*, i.e. a Lebesgue-integrable function $f = f_x$ satisfying

$$\mathbb{P}[x \in B] = \int_B f(t) \,\mathrm{d}t$$

for all Borel sets B.

As mentioned earlier, the integral $\int_\Omega x \,\mathrm{d}\mathbb{P}$ of an integrable random variable is denoted $\mathbb{E}[x]$ and is called the expectation, or mean, of x. (Integrable random variables are often referred to as having "finite mean".) A useful formula for calculating the expectation of a non-negative random variable is

$$\mathbb{E}[x] = \int_0^\infty \mathbb{P}[x > t] \,\mathrm{d}t.$$

[2] Notice that we do not use upper case letters for random variables in this course.

When x has a density function f we have (for Borel-measurable functions h)

$$\mathbb{E}[h(x)] = \int_{-\infty}^{\infty} h(t)f(t)\,\mathrm{d}t.$$

A random variable in $L_2(\Omega)$ is said to have "finite variance" and we define

$$\operatorname{var} x = \mathbb{E}[x^2] - \mathbb{E}[x]^2 = \mathbb{E}\left[(x - \mathbb{E}[x])^2\right].$$

An important example is the *standard normal distribution* $\mathscr{N}(0,1)$, where the density function is

$$\phi(t) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}t^2}$$

and the corresponding distribution is denoted Φ. Such a random variable is p-integrable for all $p < \infty$ with

$$\|z\|_p^p = \mathbb{E}[|z|^p] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} |t|^p \mathrm{e}^{-\frac{1}{2}t^2}\mathrm{d}t = \sqrt{\frac{2^p}{\pi}}\Gamma\left(\frac{p+1}{2}\right).$$

Fortunately we shall not need to remember the exact formula for $\|z\|_p$, except for the special case $p = 2$ where $\|z\|_2^2 = \operatorname{var} z = 1$.

Theorem 0.46. *If $z_1, z_2, \ldots, z_n$ are independent $\mathscr{N}(0,1)$ random variables and $\alpha_1, \ldots, \alpha_n$ are real numbers with $\sum_{j=1}^n \alpha_j^2 = 1$ then $\sum_{j=1}^n \alpha_j z_j$ has the $\mathscr{N}(0,1)$ distribution.*

A random variable u has the uniform distribution $\mathscr{U}(a,b)$ on a bounded interval $(a,b) \subset \mathbb{R}$ if it has density function $f(t) = (b-a)^{-1}\mathbf{1}_{(a,b)}$. In particular, the identity function on our favourite probability space $(0,1)$ has the $\mathscr{U}(0,1)$ distribution.

If u has the distribution $\mathscr{U}(0,1)$ and we define $z = \Phi^{-1}(u)$ then z is a standard normal random variable (because $\mathbb{P}[z \le \alpha] = \mathbb{P}[u \le \Phi(\alpha)] = \alpha$ when $0 < \alpha < 1$). Thus, on our favourite probability space $(0,1)$, $z(t) = \Phi^{-1}(t)$ defines a standard normal random variable. It is slightly less obvious that we can define an infinite sequence (z_n) of independent $\mathscr{N}(0,1)$ random variables on $(0,1)$.

One way to do this is first to introduce a sequence of "Bernouilli" random variables. For $t \in (0,1)$ and $n = 1, 2, 3, \ldots$ let $x_n(t)$ be the n^{th} digit in the (non-terminating) binary expansion of t. Thus

$$x_n(t) = \begin{cases} 0 & (0 < t \le \frac{1}{2}) \\ 1 & (\frac{1}{2} < t < 1). \end{cases}$$

It is easy to check that the x_n are independent with $\mathbb{P}[x_n = 0] = \mathbb{P}[x_n = 1] = \frac{1}{2}$ for all n. If $M = \{m_1 < m_2 < m_3 < \cdots\}$ is any infinite subset of $\mathbb{N}$ then $u_M = \sum_{j=1}^{\infty} 2^{-j} x_{m_j}$ is uniformly distributed on $(0,1)$. Finally if we take disjoint infinite subsets M_n of $\mathbb{N}$ and set $z_n = \Phi^{-1}(u_{M_n})$ we have our desired independent standard normal random variables.

0.7. **Baire's "Category" Theorem, and its applications.** You encountered Baire's theorem last year; there will be further applications in the present course.

Theorem 0.47 (Baire's "Category" theorem)**.** *Let X, d be a complete metric space and let G_n ($n \in \mathbb{N}$) be a sequence of dense open subsets of X. Then the intersection $H = \bigcap_{n\in\mathbb{N}} G_n$ is also dense in X.*

Corollary 0.48. *Let X be a complete metric space and, for each $n \in \mathbb{N}$, let F_n be a closed subset of X. If $X = \bigcup_{n\in\mathbb{N}} F_n$ then the union of the interiors $\bigcup_{n\in\mathbb{N}} \operatorname{int} F_n$ is dense in X. In particular (provided $X \ne \emptyset$) there is some n such that $\operatorname{int} F_n \ne \emptyset$.*

Corollary 0.49. *Let X be a Banach space and let D_n ($n \in \mathbb{N}$) be a sequence of closed convex symmetric sets such that $X = \bigcup_{n\in\mathbb{N}} D_n$. Then there exist n and $\epsilon > 0$ such that $D_n \supseteq B_X(0;\epsilon)$.*

Proof. By the preceding corollary, there exist $n \in \mathbb{N}$, $x \in X$ and $\epsilon > 0$ such that $B(x;\epsilon) \subseteq D_n$. By symmetry of D_n, $B(-x;\epsilon) \subseteq D_n$ and by convexity, $B(0;\epsilon) = \frac{1}{2}B(x;\epsilon)_B(-x;\epsilon) \subseteq D_n$. □

Theorem 0.50 (The Banach–Steinhaus "Uniform Boundedness" theorem)**.** *Let X be a Banach space and, for each $n \in \mathbb{N}$ let T_n be a bounded linear operator from X into a normed space Y_n. If for each $x \in X$ there is a constant M_x such that $\|T_n(x)\| \le M_x$ for all $n \in \mathbb{N}$, then there is a constant M such that $\|T_n\| \le M$ for all $n \in \mathbb{N}$.*

Proof. For each $N \in \mathbb{N}$ let

$$D_N = \{x \in X : \|T_n(x)\| \le N \text{ for all } n \in \mathbb{N}\}.$$

This is easily seen to be convex symmetric and closed. Moreover, our hypothesis tells us that $X = \bigcup_{N\in\mathbb{N}} D_N$. So there exist N and $\epsilon > 0$ such that $D_N \supseteq B(0;\epsilon)$; this is easily seen to imply that $\|T\| \le n/\epsilon$ for all n. □

Theorem 0.51 (The Open Mapping Theorem)**.** *Let X and Y be Banach spaces and let $T : X \to Y$ be a bounded linear surjection. Then T is an open mapping (i.e. $T[U]$ is open in Y whenever U is open in X).*

Corollary 0.52. *If T is a bounded linear bijection from Banach space X onto a Banach space Y then T^{-1} is bounded (i.e. T is an isomorphism).*

Corollary 0.53. *Let X be a real vector space which is a Banach space for each of two norms $\|\cdot\|$ and $\|\cdot\|'$. If $\|x\| \leq \|x\|'$ for all $x \in X$ then the two norms are equivalent.*

In the following theorem, we consider the graph $G_T = \{(x, y) \in X \times Y : y = T(x)\}$ of a linear mapping $T : X \to Y$. It is easy to check that this is a subspace of the (exterior) direct sum $X \oplus Y$.

Theorem 0.54 (Closed Graph Theorem)**.** *Let X and Y be Banach spaces and let $T : X \to Y$ be linear. Then T is bounded if and only if G_T is closed in $X \oplus Y$.*

Proof. It is easy to see that the graph of a continuous mapping is always closed.

So consider a linear mapping $T : X \to Y$ where both X and Y are Banach spaces and G_T is closed in $X \oplus Y$. As a closed subspace of a Banach space, G_T is itself a Banach space. The operator $P_1 : X \oplus Y \to X$ defined by $P_1(x, y)$ is bounded and so $R = P_1 \restriction_{G_T}: G_T \to X$ is bounded. THis operator R is thus a bounded linear bijection between Banach spaces, so that R^{-1} is bounded by Corollary 6.7. The operator $T : X \to Y$ can be written as $T = P_2 \circ R^{-1}$, where $P_2 : X \oplus Y \to Y; (x, y) \mapsto y$ is another bounded linear operator. We have shown that T is bounded. □

The Closed Graph Theorem can sometimes be used to simplify continuity proofs dramatically. Normally, to prove T is continuous you need to show that if $x_n \to x$ then $T(x_n) \to T(x)$. Using this theorem it is enough to show that y must equal $T(x)$ when we have $x_n \to x$ and $T(x_n) \to y$. (The simplification is that we are allowed to assume $T(x_n)$ converges to *something*.)