HILBERT SPACE METHODS USED IN A FIRST COURSE IN QUANTUM MECHANICS

AN OUTLINE, WITH QUITE A FEW DETAILS, OF PERTINENT FACTS ABOUT LINEAR SPACES

PART ONE

LARRY SUSANKA

Contents

Date: May 19, 2013.

1. A Dip into the Pure Clear Waters of Mathematics . . .

Mathematicians (and physicists—they used to be the same people with different hats on) saw for centuries how useful the companion concepts of displacement and direct variation were in understanding the world. The prime source of examples, of course, is "space," the universe of displacements within which our personal corpi reside. Calculus itself is another example: displacements combined with a notion of limit from a definition of distance. Perpendicularity kept popping up as a useful concept.

These folks found themselves proving the same useful theorems repeatedly, but in contexts that were different enough so it was not totally obvious how to adapt the theorem in each situation.

Mathematicians looked at the most useful theorems about various aspects of "linearity" and decided, basically, to call a vector space any object for which the conclusions of these theorems were valid.

More specifically, they went over the proofs of these useful results and extracted the features that made the proofs "go." They then defined a vector space to be any object which possessed the necessary linearity features. And they defined a normed space to be a vector space with sufficient extra structure so limit-taking is convenient. And an inner product space is a normed space with a way of calculating angles: right angles will do, other angles then follow.

Though you do keep in mind the arrows in the air around you, or ordered triples of numbers on a piece of paper, you use those simply to guide intuition and as examples to show we are actually talking about *something*. We don't want to make unwarranted presumptions about what vectors *must* be.

Hermann Grassmann did this "extraction of key properties" before 1850, and later in 1888 Giuseppe Peano popularized the ideas and got most of the credit. Grassmann's work went largely unread, being regarded as obfuscatory and opaque. In this matter, as in the ideas themselves, Grassmann was ahead of his time.

In the following we give an outline, a listing of main theorems and definitions, of those purely mathematical results necessary to phrase and understand elementary quantum mechanics in the modern style. They are standard results, whose proofs are readily found in any book on functional analysis. Frequently, there will be some feature of the proof itself which is useful to emphasize here and in that case we do give a proof, but otherwise reference to a proof will suffice.

We want to approach these matters with sufficient generality so that the vector spaces used in the applications we have in mind can be seen in their natural *mathematical* context. We will strive to be very clear about what the theorems actually state: what the assumptions are, what the conclusions are . . . *exactly.*

This is not possible to do during a basic QM class; with all detail included, the amazing "physical" facts of quantum mechanics would not be revealed until a year of "non-physical" mathematical wrangling had played itself out.

But there is a danger in this informal approach. Theorems are not (usually) carefully stated. But this is not really the issue, since it is only rarely verified that the conditions required in the (unstated) statements of these theorems actually pertain. From the standpoint of the mathematician, it looks like the mathematics has been reduced to a mnemonic gimmick invoked by magic phrases, whose purpose is to help students remember which calculation they are supposed to do.

While this may be appropriate for a first course, to accustom folks to usage of vocabulary, it can't be pushed beyond carefully staged examples organized by those who have actually checked these things and do know what they mean.

We want to be able to do this for ourselves, and that is the goal.

The early sections of these notes are intended to serve as a reminder of some of the ideas and theorems from a linear algebra course. It is not intended to be a substitute for such a course, which (along with calculus) is assumed to be familiar to the reader. However, we do introduce, rather early, notation and examples that are not usually found in beginning linear algebra. For instance, we will use the following abbreviations without further comment, wherever convenient:

$$\exists \ = \text{“there exists”}, \quad \forall \ = \text{“for all”}, \quad \in = \text{“is an element of”},$$

$$\subset = \text{“is a subset of”}, \quad \bigcap \ = \text{“intersection”}, \quad \bigcup \ = \text{“union”}, \quad \varnothing \ = \text{“empty set.”}$$

2. Vector Spaces: The Primal Definition

A **field** $\mathbb{F}$ is a nonempty set together with two commutative and associative binary operations $+$ and $\cdot$ both of which have identities, denoted 0 and 1, respectively. Further, each nonzero element of $\mathbb{F}$ has a multiplicative inverse, and every element has an additive inverse. Finally, the usual distributive property of multiplication over addition holds.

In other words, a field is an object that acts very much like the real numbers $\mathbb{R}$ with ordinary arithmetic, and we will use the customs of order-of-operations, exponentiation and so on from arithmetic in $\mathbb{R}$.

The rational numbers $\mathbb{Q}$ form a field with the same operations. So does $\mathbb{Q}(\sqrt{2})$, which consists of all numbers of the form $a + b\sqrt{2}$ for rational a and b. So does $\mathbb{C}$, the complex numbers with complex arithmetic. If p is a prime, the set $\mathbb{Z}_p = \{\, 0, 1, \ldots, p-1 \,\}$ with "mod p arithmetic" (that is, remainder after division by p) is a finite field whose order (that is, cardinality or size) is p.

A **vector space over a field** $\mathbb{F}$ is a nonempty set V together with two operations called **vector addition** and **scalar multiplication** that satisfy a collection of **ten properties.** Vector addition "acts on" pairs of members of V. Scalar multiplication "acts on" a pair one of which is a number and the other of which is a member of V.

1 and 2: We require both vector addition and scalar multiplication to be **closed in V**, and by that we mean that the result of applying these operations on members of V or numbers *always* produces a member of V. You cannot leave V by doing these operations to members of V and numbers.

3: There must be a member of V, always denoted $\mathbf{0}$ and called **the zero vector**, for which $\mathbf{v} + \mathbf{0} = \mathbf{v}$ for all members $\mathbf{v}$ of V. You distinguish this member of V from the number 0 by context.

4: For each $\mathbf{v}$ in V there must be a member $\mathbf{u}$ of V for which $\mathbf{v} + \mathbf{u} = \mathbf{0}$. $\mathbf{u}$ can be denoted $-\mathbf{v}$, and is called the negative or opposite of $\mathbf{v}$.

5 and 6: Vector addition must be commutative and associative: that is,

$$\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v} \quad \text{and} \quad (\mathbf{v} + \mathbf{w}) + \mathbf{u} = \mathbf{v} + (\mathbf{w} + \mathbf{u})$$

for any members $\mathbf{v}$, $\mathbf{u}$ and $\mathbf{w}$ of V.

7 and 8: The two distributive laws must hold:

$$(r + s)\mathbf{v} = r\mathbf{v} + s\mathbf{v} \quad \text{and} \quad r(\mathbf{v} + \mathbf{w}) = r\mathbf{v} + r\mathbf{w}$$

for all numbers r and s and any members $\mathbf{v}$ and $\mathbf{w}$ of V.

9 and 10: Last, $(rs)\mathbf{v} = r(s\mathbf{v})$ and $1\mathbf{v} = \mathbf{v}$ for all $r, s \in \mathbb{F}$ and all $\mathbf{v}$ in V.

There is a procedure one goes through in order to qualify l'objet du jour as susceptible to vector methods.

First, pick your field $\mathbb{F}$; here, this will be $\mathbb{R}$ or $\mathbb{C}$ unless otherwise specified. Next, define V, your proposed vectors. (Functions, Arrows, Sequences, etc.)

At this point you must specify the two operations: scalar multiplication and vector addition. This could be quasi-physical, describing how to stretch arrows and add them, tail-to-tip. But usually it will involve some kind of arithmetic.

Now, one is obliged to show or deduce that these operations satisfy the ten qualities which must be possessed by vector spaces.

Mathematicians are pretty picky about this process: you must decide on one single representation of what you are talking about. Shifting from one representation or aspect to another in the course of an argument without setting up—explicitly—the appropriate framework is verboten. It leads to an endless list of statements which are false but kind-of-true, almost true, might be true, under the right circumstances true, well dammit they must be true—quit bothering me about it!

We're going to try to avoid that aggravation, substituting for it aggravation of a different type.

3. Vector Spaces: The Secondary Definitions

VS1 If V is a vector space over field $\mathbb{F}$ a **linear combination** of vectors is a *finite* sum (of any length) $a^1 v_1 + \cdots + a^n v_n$ where all the a^i are from $\mathbb{F}$ and all the v_i are from V.

If there is more than one field around, it might be necessary to specify which one is intended to supply the coefficients: viz. "$\mathbb{R}$-linear combination."

VS2 A set S of vectors from the vector space V is said **to span** (this is a verb) V if every member of V can be written as a linear combination of members of S.

VS3 If S is a set of vectors from V, the **span of S**, denoted **Span(S)** (this is a noun), is the set of all *finite* linear combination of members of S.

VS4 A nonempty set S of vector space V is called **linearly independent** if whenever $s_1, \dots, s_n$ is a *finite* list of distinct members of S, a sum

$$a^1 s_1 + \cdots + a^n s_n$$

where all the a^i are from $\mathbb{F}$ can never equal the zero vector unless *all* the a^i are 0.

This is equivalent to saying that no member of S can be written as a linear combination of *other* members of S. It is also equivalent to saying that every member of $Span(S)$ can be written in only *one* way as a linear combination of members of S.

If nonempty S is *not* linearly independent it is called **linearly dependent**.

VS5 A nonempty subset S of V is called a **subspace** of V if it is itself a vector space over the same field with the operations inherited from V.

To check that a subset is a subspace we need only demonstrate closure of the two vector operations in S. The other eight properties will follow from the (already checked, presumably) fact that V itself is a vector space.

VS6 A linearly independent set of vectors that spans the vector space V is called a **basis** for V.

You will usually want to identify a **convenient basis** for your space. Sometimes there is more than one useful basis! Often a vector space is prescribed from the outset as the span of a linearly independent set, which is then automatically a basis.

Sometimes a basis as defined here is called a **Hamel basis** to distinguish it from a different notion of basis we will use later in the context of Banach or Hilbert Spaces.

Very often it is useful to "line up" the members of a basis in a specific order, and one then speaks of an **ordered basis**: a basis with some explicit order prescribed. It follows from one of the foundational axioms of mathematics (the axiom of choice) that any set can be ordered, lined up as the members of the natural numbers $\mathbb{N}$ are, so that every nonempty subset has a least member. In practice there is no need to invoke the axiom: bases are usually presented with members in easily-ordered form.

VS7 You can use an explicit basis to define **coordinates** for vectors. Every vector v can be written in exactly one way as a finite linear combination of distinct basis vectors, and the coordinates of v are the numerical factors in such a sum for the vector v. This collection of numbers (one number specified for each basis vector) determines v. Even if the basis is infinite, there are only finitely many of these coordinates which are nonzero for each particular vector.

VS8 **Every spanning set can be pruned to a basis. Every linearly independent set can be fattened to a basis.** And the number of basis vectors (finite or not) for a vector space V is fixed. This cardinal number, denoted $\dim(\mathbf{V})$, is called the **dimension** of the space.

The proofs of these facts (both existence of basis and properties of cardinal numbers for infinite sets) require the Axiom of Choice. In practice one presents or is presented with a literal basis, one does not merely deduce its existence.

VS9 In a vector space of finite dimension n with *ordered basis* $b_1, \ \ldots, \ b_n$ the coordinates of a vector $v = v^1\, b_1 + \cdots + v^n\, b_n$ can be listed in an ordered n-tuple as

$$(v^1,\, v^2,\ \ldots,\ v^n).$$

The number v^i is called the ith coordinate of vector v in basis $\mathbf{b}$.

Usually (beyond the earliest introductions) these coordinates are represented *not in a row but as a column*

$$\sum_{i=1}^{n} v^i e_i$$

where e_i is the column vector with 1 in the ith row and zeroes elsewhere. The purpose of this is to coordinate the representation with ordinary matrix operations.

If the ordered basis $\mathbf{b}$ is infinite but **countable** (that is, we can list the members of $\mathbf{b}$ in an infinite or finite sequence) we can still define e_i as the sequence defined on the positive integers by $e_i(j) = \delta_i^j$ where δ_i^j is the **Kronecker delta function**, 1 when $i = j$ and 0 otherwise.

Actually, this agrees with the original definition: what did we intend when we said above

"Put a one in the ith row and zeroes in the rest?"

if not, implicitly, a visualization of the Kronecker delta with i some fixed position-number and the other argument restricted to lie between 1 and n?

In any case we associate vector

$$v = \sum_{i=1}^{\infty} v^i b_i \quad \text{with} \quad \sum_{i=1}^{\infty} v^i e_i$$

and we note explicitly that, though the sum indicated is infinite, only a finite number of terms are nonzero for each v. Again, we will have more to say about this later in the context of Banach and Hilbert spaces.

In either the finite or the infinite case, for each vector v, the column vector (or "infinite column sequence") indicated above will be called the **coordinate vector** for v in ordered basis $\mathbf{b}$ and denoted $[\mathbf{v}]_{\mathbf{b}}$.

VS10 If you change basis to new basis $\mathbf{c}$ in vector space V the coordinates of vector v change. If matrix $M_{\mathbf{c}\leftarrow\mathbf{b}}$ is the matrix whose columns contain the coordinates of the $\mathbf{b}$ vectors in terms of the $\mathbf{c}$ vectors

$$M_{\mathbf{c}\leftarrow\mathbf{b}} = \left(\ [b_1]_{\mathbf{c}}\ \ [b_2]_{\mathbf{c}} \cdots\ \ [b_n]_{\mathbf{c}}\ \right)$$

then these new coordinates can be calculated by matrix multiplication

$$[v]_{\mathbf{c}} = M_{\mathbf{c}\leftarrow\mathbf{b}}\, [v]_{\mathbf{b}}\,.$$

Note that **matrix of transition** $M_{\mathbf{c}\leftarrow\mathbf{b}}$ is invertible and

$$M^{-1}_{\mathbf{c}\leftarrow\mathbf{b}} = M_{\mathbf{b}\leftarrow\mathbf{c}}.$$

It is *important to emphasize that matrix operations are virtually never done by hand* beyond the first course in linear algebra, or in simple examples of small size used to illustrate some point. Humans simply are not good at doing hundreds of arithmetic operations, and it is just a distraction to try.

You *use* the matrices not by personally multiplying them together but by understanding what they are intended to do, and adopting a notation that reminds you of that. Think about what the matrix operation *means*. Then turn the calculation over to your calculator, or to MATLAB or Mathematica or Maple if the matrix is too big or symbolic calculation is required.

If you find yourself multiplying matrices by hand you are, at best, wasting your time.

4. Vector Spaces: Examples

VEx1 The set containing just the zero vector is a subspace of every vector space. Any one-element vector space is called a **trivial** vector space.

Also (obvious but worth mentioning) every vector space is a subspace of itself. A **proper subspace** is a subspace smaller than this.

$\mathbb{F}$ itself is a very simple vector space over $\mathbb{F}$. Also $\mathbb{C}$ is a vector space over $\mathbb{R}$, and both $\mathbb{C}$ and $\mathbb{R}$ are vector spaces over $\mathbb{Q}$.

Many vector spaces are given as function spaces, and we require a compact notation to specify name of function, domain and range.

We indicate that F is a function whose domain is the (nonempty) set V and whose range is *contained in* set W by

$$\mathbf{F}\colon \mathbf{V} \to \mathbf{W}.$$

VEx2 Suppose M is any nonempty index set and let V_m be any vector space over $\mathbb{F}$ for each $m \in M$.

We will define a very general type of vector space, a function space denoted

$$\prod_{\mathbf{m}\in\mathbf{M}} \mathbf{V_m}$$

and called a **product vector space.**

The set of vectors for this space is the collection of functions

$$f\colon M \to \bigcup_{m\in M} V_m \quad \text{for which} \quad f_m \in V_m \quad \text{for every } m \in M.$$

We define **pointwise addition** and **pointwise scalar multiplication** on this set in the obvious way:

$$\text{If } f, g \in \prod_{m\in M} V_m \text{ and } c \in \mathbb{F}$$

$$\text{define } f+g \text{ to be the function } (f+g)_m = f_m + g_m$$
$$\text{and } (cf)_m = cf_m \text{ for all } m \in M.$$

$$\prod_{m\in M} V_m \text{ is a vector space with these pointwise operations.}$$

This one is worth thinking about. *All* of the "abstract" vector spaces you will actually use for something, both here and almost anywhere else, are given explicitly as subspaces of product spaces.

Once you identify a nonempty subset of such a space, all you need to do is verify closure of the two operations and you can conclude your subset is a vector space in its own right.

VEx3 If V_1 and V_2 are two vector spaces over field $\mathbb{F}$ the product space

$$\prod_{m\in\{1,2\}} V_m \quad \text{is normally denoted} \quad V_1 \times V_2.$$

There is, however, just a little more going on in the second notation. There is a specific order on this particular index set, and we honor that fact by listing the vector spaces involved, called the **factor spaces**, in subscript order. Many product spaces have no obvious or natural (or intended) order on their indices, while in others an order or some other structure on the indices is crucial.

VEx4 Suppose M is any nonempty index set and V is any vector space over $\mathbb{F}$. Define $\mathbf{V^M}$ to be the set of functions with domain M and range in V. This is a special case of the example above, where $V_m = V$ for all m.

$$V^M \text{ is a vector space with pointwise operations.}$$

The vector space $\mathbb{R}^{\{1,2,\ldots,n\}} = \mathbb{R} \times \cdots \times \mathbb{R}$ is an example of this type of product, normally denoted $\mathbb{R}^n$.

$\mathbb{R}^{\mathbb{R}}$ and $\mathbb{R}^{[0,1]}$ are two more examples, the sets of all real valued functions defined on the real line and the unit interval, respectively.

The set of continuous functions, the set of differentiable functions, the set of integrable functions, the set of polynomials and the set of polynomials of third degree or less (with appropriate domain) are subspaces of these vector spaces.

VEx5 The set of matrices of a specific shape with entries in $\mathbb{F}$ with usual operations forms a vector space. We can denote these $m \times n$ matrices by $\mathbb{M}_{m\times n}(\mathbb{F})$. These matrices can be conceived of (actually, *are*) the set of $\mathbb{F}$-valued functions with domain consisting of the block of integer pairs

$$\{\,(a,b) \mid a,b \text{ are integers with } 1 \le a \le m \text{ and } 1 \le b \le m\,\}.$$

The big rectangular symbol used to denote this function is convenient but is nothing more than a visualization of this by placing the matrix entries on grid points labeled by these pairs of integers.

The diagonal, upper triangular and lower triangular matrices form subspaces of $\mathbb{M}_{m\times n}(\mathbb{F})$. If $m = n$, the symmetric, skew symmetric, traceless and Hermitian matrices form subspaces.

VEx6 A real sequence is a function $f\colon \mathbb{N} \to \mathbb{R}$ and the set $\mathbb{R}^{\mathbb{N}}$ of all such sequences is a real vector space with pointwise operations. The set of convergent real sequences is a subspace.

More generally, if V is a vector space of functions then $V^{\mathbb{N}}$, the set of *sequences* of functions, is too. If there is a concept of convergence you have in mind for these sequences of functions, the set of convergent sequences can form a subspace of $V^{\mathbb{N}}$.

VEx7 If $f \in \left(\mathbb{F}^S\right)^{\mathbb{N}}$ is a *particular (fixed) sequence* of $\mathbb{F}$-valued functions with shared domain S and a is a sequence of numbers then af is also a sequence of functions. We define

$$Seq(f) = \left\{\, a\, f \mid a \in \mathbb{F}^{\mathbb{N}} \,\right\}.$$

$Seq(f)$ is a vector space, a subspace of $\left(\mathbb{F}^S\right)^{\mathbb{N}}$.

If g is any member of $\left(\mathbb{F}^S\right)^{\mathbb{N}}$ define $S(g)$ to be the new sequence of functions given by

$$S(g)_n = \sum_{i=0}^{n} g_i \quad \text{for any } n \in \mathbb{N}.$$

We now define

$$Ser(f) = \left\{\, S(af) \mid a \in \mathbb{F}^{\mathbb{N}} \,\right\}.$$

$Ser(f)$ is a vector space too, an important one.

For instance, if $f_n(t) = t^n$ for each n, the sequence of real valued monomials defined on the real line, a typical member of $Seq(f)$ might be

$$(3,\ -3t,\ 6\,t^2,\ 8t^3,\ \ldots\ ).$$

Using the same numerical sequence a would give

$$S(af) = \big(3,\ 3-3t,\ 3-3t+6\,t^2,\ 3-3t+6\,t^2+8\,t^3,\ \ldots\ \big).$$

In other words, this is the sequence of partial sums of a power series.

As another example, if $g_n(t) = \cos(nt)$ for $n = 0, 1, \ldots$ then $Ser(g)$ would be the sequences of partial sums of all Fourier Cosine series.

Subspaces of these series spaces (such as those whose members converge in various ways) are of interest to both mathematicians and physicists for *many* reasons.

5. A Comment on Skew Fields and Modules

It is possible to do most of ordinary linear algebra using a **skew field** rather than a field as scalars. A **skew field** satisfies all the axioms for a field *except* commutativity of the multiplication. The skew field most people have heard of is the **quaternions**, $\mathbb{H}$. This example is important to physicists, particularly particle physicists, but we will not explore them here.

If you retain commutativity but dispense with the existence of multiplicative inverses, a field becomes a **(unitary) commutative ring**, and linear algebra becomes the study of **modules** over that ring. These are extremely important in classical mechanics and relativistic mechanics. We hint at an application in the following paragraphs.

Suppose M is a surface (such as a sphere or torus) or more generally a **manifold**, an object that locally looks like $\mathbb{R}^n$ for some fixed $n \in \mathbb{N}$. If V is a real vector space V^M is called a **vector bundle** on M.

The vector space V can be $\mathbb{R}$, in which case we have nothing more than the real valued functions on M. (These real functions on M, $\mathbb{R}^M$, form a unitary commutative ring.) Or we could have $V = \mathbb{R}^n$. Members of $(\mathbb{R}^n)^M$ provide a selection of a vector for each $m \in M$.

Manifolds typically have enough structure to define differentiability for elements of these vector bundles; this is used to tie the vector space to the geometry of the manifold near its "point of attachment." And the resulting structures constitute the language, for instance, of classical Hamiltonian and general relativistic mechanics. Symplectic forms, the semi-Riemannian metric tensors or other kinds of tensors and tensor-related objects such as the Ricci tensor and Christoffel symbols are from this universe.

If $T \in (\mathbb{R}^n)^M$ and $g \in \mathbb{R}^M$ then gT is also a member of $(\mathbb{R}^n)^M$. So $(\mathbb{R}^n)^M$ is not only a vector space over $\mathbb{R}$ but also a module over the ring $\mathbb{R}^M$.

We will not explore any of this in these notes.

However ... it should be pointed out that a real function on a manifold is frequently called a number field, a vector assignment at each point is usually called a vector field, a tensor given at at each point is called a tensor field and so on. This is not incorrect usage of vocabulary. It is just a *different* usage, coming from a different context. There is occasional opportunity for confusion, so be aware.

6. Vector Spaces: Linear Functions

LF1 Suppose F is a function[1] whose domain is the vector space V and whose range is contained in vector space W, both over the same field. This function is called **linear** (or a **linear transformation**) if for v, u in V and field member k we have

$$F(v + ku) = F(v) + kF(u).$$

When a function F is linear and no confusion can result the notations

$$F(v) \quad \text{and} \quad Fv$$

[1]The words **map** and **function** are used synonymously.

will be used interchangeably. Though customary, it should be used with caution; parentheses to indicate function evaluation should be restored wherever there seems to be too much going on.

LF2 Quite a few different properties will be discussed in combinations, and we adopt the following notational contrivance in the interest of brevity. If we want to ascribe a property to an object X we might, in the first instance that X is encountered in a discussion, list that property along with X as a superscript. For example we could indicate that Y is a real vector subspace of a real vector space X by

$$Y^{\substack{\text{Real Vector}\\ \text{Subspace}}} \subset X^{\substack{\text{Real Vector}\\ \text{Space}}},$$

or that P is a real linear function by

$$P^{\substack{\text{Real}\\ \text{Linear}}}\colon V \to W.$$

LF3 A generic linear transformation is also called a (vector space) **homomorphism.** The set of linear transformations from $\mathbb{F}$-vector space V to $\mathbb{F}$-vector space W is denoted $\mathbf{Hom}_{\mathbb{F}}(\mathbf{V}, \mathbf{W})$. It is a vector space too, a (small) subspace of the $\mathbb{F}$-vector space W^V.

When $V \subset W$ a linear transformation is often called a **linear operator**, though it should be pointed out that this is by no means universal: in many sources the terms linear transformation and linear operator are synonymous.

When W is just the number space a linear transformation is also called a **linear functional.**

The collection of all these functionals is called the **algebraic dual** of V and denoted $\mathbf{V}^*$. If it is necessary to specify the field, the notation $\mathbf{V}^*_{\mathbb{F}}$ will be used.

LF4 For linear $F\colon V \to W$ we define $\mathcal{K}\mathbf{er}(\mathbf{F})$ to be the set of those $v \in V$ for which $Fv = 0$. It is called the **kernel** or **nullspace** of F and is a subspace of the domain space V.

$\dim(\mathcal{K}er(F))$ is called the **nullity** of F and this cardinal number is denoted $\mathbf{nullity}(\mathbf{F})$.

We define $\mathcal{R}\mathbf{an}(\mathbf{F})$ to be the set of all $F(v)$ for $v \in V$. This set is called the **image** of F and is also a subspace, this time of W.

$\dim(\mathcal{R}an(F))$ is called the **rank** of F, and denoted $\mathbf{rank}(\mathbf{F})$.

We will have occasion to consider functions whose domains are various subspaces of V and not all of V. In that context we will refer to the domain of definition of a linear function F by $\mathcal{D}\mathbf{om}(\mathbf{F})$.

It is a fact that

$$rank(F) + nullity(F) = \dim(\mathcal{D}om(F))$$

which can be quite useful, particularly when all these spaces have finite dimension.

LF5 Linear $F\colon V \to W$ is called an **isomorphism**, and the two spaces are called **isomorphic**, if F is invertible as a function. The inverse of an isomorphism is also linear, and so is itself an isomorphism.

Two vector spaces over the same field are isomorphic exactly when they have the same dimension. Explicit isomorphisms are sometimes used to encourage identification of the two spaces involved: for raw "vector space" purposes they are, essentially, identical. Usually there must be more than shared dimension to make an identification worth the effort it takes to establish it.

For instance, if V is an n dimensional $\mathbb{F}$-vector space with ordered basis $\mathbf{b}$ then the isomorphism

$$[\,\cdot\,]_{\mathbf{b}} : V \to \mathbb{F}^n$$

is often very useful. It may be much easier to think about a standard example, a column of numbers for instance, than whatever form the members of V have in their original setting.

LF6 The algebraic dual V^* of V is, itself, a vector space. So *it too* has an algebraic dual, $(V^*)^*$.

Define the **evaluation map** $E\colon V \to (V^*)^*$ by

$$E(x)(f) = f(x) \quad \text{for each } x \in V \text{ and } f \in V^*.$$

E is linear, and also one-to-one: that is, $E(x) = E(y)$ exactly when $x = y$. So if E is onto then it is invertible and an isomorphism. In that case, $(V^*)^*$ can be identified with (i.e. it "is") V.

If V is finite dimensional, V and V^* have the same dimension. And it follows that $(V^*)^*$ has the same dimension as does V. So E must be onto. The infinite dimensional case is much more delicate and we will consider the extent to which we can recover this important identification later.

LF7 A linear operator $G\colon V \to V$ can have **eigenvalues** and associated **eigenvectors.** A nonzero vector v is called an eigenvector of G for eigenvalue λ if $G(v) = \lambda\, v$. Eigenvectors for **different** eigenvalues cannot be dependent: a set consisting of a finite number of eigenvectors for **different** eigenvalues is linearly independent.

LF8 Every linear function $F\colon V \to W$ is determined by what it does to a basis of its domain. Specifically, if $v = v^1\, b_1 + \cdots + v^n\, b_n$ for numbers v^i and basis vectors b_i of V then

$$F(v) = F\left(\sum_{i=1}^{n} v^i\, b_i\right) = \sum_{i=1}^{n} v^i\, F(b_i).$$

If you know all the $F(b_i)$ then you know any $F(v)$.

Conversely, you can define a function in any way you like on a basis of V and extend by linearity to all of V.

We will see, however, that in infinite dimensional spaces with more structure the resulting linear function may lack the important property of *continuity.* Some choices in such situations are forbidden if we require this property.

LF9 If F is a linear function as above with ordered basis **b** of its domain and ordered basis $\mathbf{c} = \{ c_1, \ \ldots \ c_m \}$ of its range W then each $F(b_i)$ can be written as

$$F(b_i) = \sum_{j=1}^{m} A_i^j c_j$$

for certain unique numbers A_i^j. These numbers are aligned vertically in the coordinate column $[F(b_i)]_{\mathbf{c}}$ and if you line up these columns consecutively in row you get the $m \times n$ matrix

$$A = \left(A_i^j \right) = (\ [F(b_1)]_{\mathbf{c}} \ \ [F(b_2)]_{\mathbf{c}} \cdots \ [F(b_n)]_{\mathbf{c}} \ ) \, .$$

Expansion and comparison of left and right sides below verifies that for any vector v in V the effect of F can be calculated on coordinates as

$$[F(v)]_{\mathbf{c}} = A \, [v]_{\mathbf{b}} \, .$$

So, essentially, *any* linear transformation can be handled through coordinates, and *calculated by left multiplication by this matrix.*

The matrix is, of course, dependent on the basis used in both domain and range. The matrix A, called the **matrix of F** in bases **b** in domain and **c** in range, will be denoted $[F]_{\mathbf{c}\leftarrow\mathbf{b}}$. The entries of this matrix are called the **coordinates of F in these two bases.**

Using this notation we have

$$[F(v)]_{\mathbf{c}} = [F]_{\mathbf{c}\leftarrow\mathbf{b}} \, [v]_{\mathbf{b}}$$

Finally, we note that if V has dimension n and W has dimension m then

$$[\,\cdot\,]_{\mathbf{c}\leftarrow\mathbf{b}} : Hom_{\mathbb{F}}(V, W) \to \mathbb{M}_{m\times n}(\mathbb{F}) \quad \text{is an isomorphism.}$$

Most of the time, working with something concrete like a matrix is easier to think about, and certainly easier to calculate using machines, than some abstract homomorphism. The results here say that, *as long as you are careful to keep track of basis in domain and range, you can do just that.*

LF10 With F as above and if $G\colon W \to X$ is also linear and **d** is a basis of X then

$$\begin{aligned} [G \circ F]_{\mathbf{d}\leftarrow\mathbf{b}} \, [v]_{\mathbf{b}} &= [G \circ F(v)]_{\mathbf{d}} = [G(F(v))]_{\mathbf{d}} \\ &= [G]_{\mathbf{d}\leftarrow\mathbf{c}} \, [F(v)]_{\mathbf{c}} = [G]_{\mathbf{d}\leftarrow\mathbf{c}} \, [F]_{\mathbf{c}\leftarrow\mathbf{b}} \, [v]_{\mathbf{b}} \end{aligned}$$

Our conclusion is that the matrix of a composition of linear functions is the product of matrices of the functions involved.

LF11 Even if one or more of the ordered bases in the earlier material have infinite cardinality, all of the relevant individual linear combinations involved are finite. So all the calculations involving coordinates are unchanged so long as the indices of summation in each case include all nonzero terms. For linear transformations this means we are, effectively, working with matrices with an infinite number of rows and/or columns. However each specific column contains only a finite number of nonzero entries.

LF12 Suppose $F\colon V \to W$ is linear and $\mathbf{a}$ and $\mathbf{b}$ are two bases of vector space V. Suppose $\mathbf{c}$ and $\mathbf{d}$ are two bases of vector space W. Then

$$\begin{aligned}[F]_{\mathbf{d}\leftarrow\mathbf{a}}\,[v]_{\mathbf{a}} &= M_{\mathbf{d}\leftarrow\mathbf{c}}\,[F]_{\mathbf{c}\leftarrow\mathbf{a}}\,[v]_{\mathbf{a}} \\ &= M_{\mathbf{d}\leftarrow\mathbf{c}}\,[F]_{\mathbf{c}\leftarrow\mathbf{b}}\,[v]_{\mathbf{b}} \\ &= \left(M_{\mathbf{d}\leftarrow\mathbf{c}}\,[F]_{\mathbf{c}\leftarrow\mathbf{b}}\,M_{\mathbf{b}\leftarrow\mathbf{a}}\right)[v]_{\mathbf{a}}\,.\end{aligned}$$

In other words, the matrix of a linear transformation changes under change of bases by

$$[F]_{\mathbf{d}\leftarrow\mathbf{a}} = M_{\mathbf{d}\leftarrow\mathbf{c}}\,[F]_{\mathbf{c}\leftarrow\mathbf{b}}\,M_{\mathbf{b}\leftarrow\mathbf{a}}.$$

It is important to understand the meaning of this equation, and a description in terms of an imagined internal monologue helps me do this. I say to myself:

> The left hand side is a matrix that takes the **a**-coordinates of a vector, does the F-thing to them, and produces **d**-coordinates of the answer.

Looking at the right hand side we have instructions:

> Take the **a**-coordinates of a vector and translate to language-**b**, then do the F-thing to these **b**-coordinates, to produce **c**-coordinates of the answer. Finally, translate these **c**-coordinates to language-**d**.

I would like to emphasize one more time that doing matrix calculations of this kind with pencil and paper will not help you understand anything. *Keep firmly in mind what the matrices are doing for you. Do specific calculations with hardware.*

LF13 Suppose $\mathbf{b}$ is a basis in domain and $\mathbf{c}$ in range of isomorphism F. If I is the identity isomorphism on V, then $[I]_{\mathbf{b}\leftarrow\mathbf{b}}$ is the identity matrix. The computation $F^{-1}\circ F = I$ then requires $\left[F^{-1}\right]_{\mathbf{b}\leftarrow\mathbf{c}}$ to be the matrix inverse to $[F]_{\mathbf{c}\leftarrow\mathbf{b}}$.

7. Quotient and Sum

QS1 If N is a subspace of vector space V and $v \in V$ we define $v + N$ to be the "translates" of N by v. Specifically,

$$v + N = \{\, v + n \mid n \in N \,\}.$$

We now define the **quotient space** of V by N to be the set of all these translates of N. We use the notation **V/N** to denote this set of translates.

V/N has the structure of a vector space with operations defined for $a \in \mathbb{F}$ and $v, w \in V$ by

$$a(v+N) = (av) + N \quad\text{and}\quad (v+N) + (w+N) = (v+w) + N.$$

The cardinal numbers of the dimensions combine as

$$\dim(V/N) + \dim(N) = \dim(V).$$

QS2 If $f: V \to W$ is linear and if subspace N of V is contained in $\mathcal{K}er(f)$ then f induces a linear function

$$\tilde{f}: V/N \to W$$

given by $\tilde{f}(v + N) = f(v)$.

QS3 Considering a different matter, if Y and Z are two subspaces of vector space X, we write $\mathbf{Y} + \mathbf{Z}$ to denote the vector subspace of X spanned by the vectors in Y and Z. We call this vector space **the sum of Y and Z.**

If, in addition, $Y \cap Z = \{ 0 \}$ we write $\mathbf{Y} \bigoplus \mathbf{Z}$ and call this **the direct sum of Y and Z.** The importance of direct sum is that any vector in $Y \bigoplus Z$ can be written in a unique way as $y + z$ where $y \in Y$ and $z \in Z$.

A basis of $Y \bigoplus Z$ can be given that is a disjoint union of a basis for Y and a basis for Z.

QS4 We let $\mathbf{X} - \mathbf{Y}$ denote the members of X which are not in Y. If $w \in X - Y$ we define $\mathbb{F}\mathbf{w}$ to be the one dimensional subspace of X generated by w. Then $Y \cap \mathbb{F}w = \{ 0 \}$ and so the sum $Y + \mathbb{F}w$ is a direct sum $Y \bigoplus \mathbb{F}w$.

QS5 If $A \subset B$ for nonempty set A and $F: B \to C$ we will use the notation $\mathbf{F}|_{\mathbf{A}}$ to denote the function whose domain is A and which agrees with F for all members of A.

If $G = F|_A$ for some nonempty set A then F is called **an extension of G** and G is called **the restriction of F to A.**

QS6 Any linear $F: Y \bigoplus Z \to W$ defines, by restriction, unique linear maps $F|_Y: Y \to W$ and $F|_Z: Z \to W$.

Conversely, any two linear functions $H: Y \to W$ and $K: Z \to W$ can be used to create a linear function $F: Y \bigoplus Z \to W$ by $F(x) = H(y) + K(z)$ where $x = y + z$ for $y \in Y$ and $z \in Z$.

8. Snarky Editorial Comment

A functional on finite dimensional V can be represented in a basis by a $1 \times n$ row matrix. This looks a lot like the coordinates of a vector: an $n \times 1$ column matrix. But it behaves quite differently when there is change of basis.

Letting $\mathbf{e}$ be the basis $\{1\}$ of $\mathbb{F}$ and changing basis in domain from basis $\mathbf{a}$ to basis $\mathbf{b}$ in the domain gives

$$\begin{aligned} [F]_{\mathbf{e}\leftarrow\mathbf{a}} [v]_{\mathbf{a}} &= [F]_{\mathbf{e}\leftarrow\mathbf{a}} M_{\mathbf{a}\leftarrow\mathbf{b}} M_{\mathbf{b}\leftarrow\mathbf{a}} [v]_{\mathbf{a}} \\ &= ([F]_{\mathbf{e}\leftarrow\mathbf{a}} M_{\mathbf{a}\leftarrow\mathbf{b}}) (M_{\mathbf{b}\leftarrow\mathbf{a}} [v]_{\mathbf{a}}) = [F]_{\mathbf{e}\leftarrow\mathbf{b}} [v]_{\mathbf{b}} . \end{aligned}$$

Remember, the left factor is a row and the right factor is a column. The output is a number. You change the functional row to new basis by right-multiplying by $M_{\mathbf{a}\leftarrow\mathbf{b}}$ and you change the vector upon which it is being applied by left-multiplying by the inverse of that matrix, $M_{\mathbf{b}\leftarrow\mathbf{a}}$.

If you are in the (arguably convenient) habit of representing functionals by columns and evaluation of functionals by dot product, it is very mysterious *why* some vectors change in one way, while others change another way. It is mysterious, that is, unless you never use any type of coordinate changes[2] except those for which

$$M_{\mathbf{b}\leftarrow\mathbf{a}}^{-1} = M_{\mathbf{a}\leftarrow\mathbf{b}} = M_{\mathbf{b}\leftarrow\mathbf{a}}^{t},$$

where the "t" refers to the transpose of the matrix. In this case these functionals, which you are representing in coordinates as columns, transform in the same way as vector coordinates, and the differences are concealed from the user.

Concealed, that is, until you need to do something as simple as switching from centimeters to meters in your coordinates.

Now it is undeniable that this subterfuge can be convenient in $\mathbb{R}^n$, and even more so in the infinite dimensional Hilbert space settings to which we are headed, but the true nature of these functionals must be noted to avoid confusion.

9. Tensor Spaces

Suppose V is a finite dimensional vector space over field $\mathbb{F}$ and $r, s \in \mathbb{N}$ and not both 0. We define the product space

$$V_s^r = \underbrace{V^* \times \cdots \times V^*}_{r \text{ copies}} \times \underbrace{V \times \cdots \times V}_{s \text{ copies}}$$

where r is the number of times that V^* occurs (they are all listed first) and s is the number of times that V occurs. We will define $V_s^r(k)$ by

$$V_s^r(k) = \begin{cases} V^*, & \text{if } 1 \le k \le r; \\ V, & \text{if } r+1 \le k \le r+s. \end{cases}$$

A function $T \colon V_s^r \to \mathbb{F}$ is called **multilinear** if it is linear in each domain factor separately.

Making this precise is notationally awkward. Here is an example.

T is multilinear if whenever $v(i) \in V_s^r(i)$ for each i then the function

$$T(v(1), \dots, v(j-1), x, v(j+1), \dots, v(r+s))$$

in the variable $x \in V_s^r(j)$ is linear on $V_s^r(j)$ for $j = 1, \dots, r+s$.

Sometimes we say, colloquially, that T is linear in each of its **slots** separately. The independent variable x in the formula above is said to be in the jth slot of T.

It is, obviously, necessary to **tweak** this condition at the beginning and end of a subscript range, where $j-1$ or $j+1$ might be "out of bounds" and it is left to the reader to do the right thing here.

A **tensor**[3] **on V** is any multilinear function T as described above. $\mathbf{r} + \mathbf{s}$ is called the **order** or **rank** of the tensor T.

---

[2]These matrices of transition are called **orthogonal**.

[3]The word **tensor** is adapted from the French "tension" meaning "strain" in English. Understanding the physics of deformation of solids was an early application.

T is said to have **contravariant order r** and **covariant order s**. If $s = 0$ the tensor is called **contravariant** while if $r = 0$ it is called **covariant.** If neither r nor s is zero, T is called a **mixed** tensor.

Sometimes a covariant tensor of order 1 is called, depending on who is speaking and the context, a **1-form**, a **covariant vector** or a **covector**. They are functionals on V, simply members of V^*.

A contravariant tensor of order 1 is often called a **contravector** or, simply, a **vector**. They are functionals on V^* and are identified with (that is, *they are*) members of (finite dimensional) V; the identification is given by the evaluation isomorphism.

For fixed r and s in $\mathbb{N}$, the collection of all tensors defined on V_s^r constitute an important example of a vector space over the field of V, denoted $\mathcal{T}_{\mathbf{s}}^{\mathbf{r}}(\mathbf{V})$.

10. Complexification

Any complex vector space is also automatically a real vector space, and *we will consider it to be such whenever that is convenient.*

If B is a basis for complex vector space V and $i\,B = \{\, ib \mid b \in B \,\}$ then $B \bigcup iB$ is a basis for V as a real vector space. So if B is finite the real dimension of V is double its complex dimension.

The real algebraic dual of a complex vector space is not, however, the same as the complex algebraic dual of that same space, though they are closely related.

Any member Θ of the complex algebraic dual $V_{\mathbb{C}}^*$ of the complex vector space V can be written as

$$\Theta = \Theta_{\text{real}} + i\Theta_{\text{complex}}$$

where Θ_{real} and Θ_{complex} are real linear and real valued on V, called the **real and complex parts of $\boldsymbol{\Theta}$**, respectively.

The real and complex parts of Θ are related: for all $x \in V$

$$\Theta_{\text{complex}}\, x = -\Theta_{\text{real}}(ix) \quad \text{and} \quad \Theta_{\text{real}}\, x = \Theta_{\text{complex}}(ix).$$

So any member of $V_{\mathbb{C}}^*$ can be associated with a unique member of $V_{\mathbb{R}}^*$, namely its real part, from which the original functional can be recovered.

On the other hand, if Ψ is any member of the real algebraic dual $V_{\mathbb{R}}^*$ of the complex vector space V then, except in the trivial case where Ψ is the zero functional, it cannot be complex linear.

Still, if we define Θ by

$$\Theta x = \Psi x - i\Psi(ix)$$

then Θ is a member of $V_{\mathbb{C}}^*$ and our real functional Ψ is the real part of Θ.

Now let's think about the vector spaces themselves.

Any real vector space W can be "complexified" as follows.

Let $V = W \times W$ with the obvious addition and real scalar multiplication. We make V into a complex vector space by decreeing $i(x, y) = (-y, x)$ and, generally,

$$(a + bi)(x, y) = (ax - by, bx + ay) \text{ for real } a \text{ and } b \text{ and any } (x, y) \in V.$$

One must verify that this operation does, in fact, give V the structure of a complex vector space.

Also, it is easy to check that if $B \subset W$ is a basis for the real vector space W then $\overline{B} = \{\, (b, 0) \mid b \in B \,\}$ is a basis for the complex vector space V. So V, as a complex vector space, has the same dimension as does W, as a real vector space.

The complex functionals on V are related to the real functionals on W.

Suppose Θ is a member of the complex algebraic dual of V.

If B is a real basis for W and $\overline{B}$ is the associated complex basis for V then for each $(b, 0) \in \overline{B}$ we have

$$\begin{aligned} i\,\Theta(b,0) &= i\ (\ \Theta_{\text{real}}(b,0) + i\,\Theta_{\text{complex}}(b,0)\ ) \\ &= -\ \Theta_{\text{complex}}(b,0) + i\,\Theta_{\text{real}}(b,0). \end{aligned}$$

But also we have

$$i\,\Theta(b,0) = \ \Theta_{\text{real}}(0,b) + i\,\Theta_{\text{complex}}(0,b).$$

So if we know Θ on $\overline{B}$ then we know not just

$$\Theta_{\text{real}}(b,0) \quad \text{but also} \quad \Theta_{\text{real}}(0,b) = -\Theta_{\text{complex}}(b,0)$$

for all b in the real basis B for W. Using these numerical values, and the real linearity of Θ_{real}, we can calculate $\Theta_{\text{real}}(x,0)$ and $\Theta_{\text{real}}(0,y)$ for any x and y in W.

Recall from above that Θ_{real} determines Θ_{complex}.

In our case,

$$\begin{aligned} \Theta(x,y) &= \Theta_{\text{real}}(x,y) + i\,\Theta_{\text{complex}}(x,y) = \Theta_{\text{real}}(x,y) - i\,\Theta_{\text{real}}\,(\,i\,(x,y)\,) \\ &= \Theta_{\text{real}}(x,y) + i\,\Theta_{\text{real}}\,(y,-x) \\ &= \Theta_{\text{real}}(x,0) + \Theta_{\text{real}}(0,y) + i\,\Theta_{\text{real}}(y,0) - i\,\Theta_{\text{real}}(0,x). \end{aligned}$$

We can now use real linearity of Θ_{real} and the fact that B is a basis for W to evaluate Θ on all of V.

Further, Θ_{real} is real linear when restricted to $W \times \{0\}$ and so is associated with exactly one member of the real algebraic dual $W^*_{\mathbb{R}}$.

Finally, the process can be reversed. Given a member Ψ of the real dual of W we can build a unique member of the complex algebraic dual of V as

$$\Theta(x,y) = \Psi x + \Psi y + i\ (\Psi y - \Psi x)\,.$$

11. The Hahn-Banach Theorem

This section is devoted to the possibility of extending a function with certain properties to a larger domain while preserving those properties.

One reason we want this theorem is that it allows us to conclude that there is a rich stock of continuous functionals whenever the theorem applies.

A nonempty subset S of a real vector space X is called **convex** if

$$tu + (1-t)v \in S \quad \forall t \in [0,1] \text{ and } u, v \in S.$$

In other words, all points on the line segment connecting u and v are in S whenever u and v are in S.

If X is any real vector space we say that a function $P\colon X \to \mathbb{R}$ is **convex** provided

$$P(tu + (1-t)v) \leq tP(u) + (1-t)P(v) \quad \forall t \in [0,1].$$

Geometrically, and in case $X = \mathbb{R}$, this means that the graph of a convex function always lies on or beneath the straight line connecting any two points on the graph. For this reason convex functions are also called **sublinear.**

We note in particular that any member P of $X^*_{\mathbb{R}}$ is convex, since in that case equality holds above for all u, v and t. So is $|P|$, defined in the obvious way by $|P|(x) = |P(x)|$ for each $x \in X$.

Theorem* 11.1. *The Hahn-Banach Theorem

If $Y^{\substack{\text{Real Vector}\\ \text{Subspace}}} \subset X^{\substack{\text{Real Vector}\\ \text{Space}}}$ *and* $P\colon X \to \mathbb{R}$ *is convex*

and $\Lambda \in Y^*_{\mathbb{R}}$ *satisfies* $\Lambda \leq P|_Y$

then $\exists \Psi \in X^*_{\mathbb{R}}$ *with* $\Lambda = \Psi|_Y$ *and* $\Psi \leq P$.

Proof. If $w \in X - Y$ and α, β are positive and $u, v \in Y$

$$\begin{aligned} \beta\,\Lambda u + \alpha\,\Lambda v \; &= \; (\alpha+\beta)\,\Lambda\left(\frac{\beta}{\alpha+\beta}u + \frac{\alpha}{\alpha+\beta}v\right) \\ &\leq (\alpha+\beta)\,P\left(\frac{\beta}{\alpha+\beta}(u-\alpha w) + \frac{\alpha}{\alpha+\beta}(v+\beta w)\right) \\ &\leq \beta\,P(u-\alpha w) + \alpha\,P(v+\beta w). \end{aligned}$$

$$\text{So} \quad \frac{1}{\alpha}\,[\ \Lambda u - P(u - \alpha\,w)\ ] \;\leq\; \frac{1}{\beta}\,[\ P(v+\beta\,w) - \Lambda v\ ].$$

The left side does not depend on v or β, while the right is independent of α and u. So there is a real number a with

$$\sup_{\substack{u\in Y\\ \alpha>0}} \frac{1}{\alpha}\,[\ \Lambda u - P(u-\alpha\,w)\ ] \;\leq\; a \;\leq\; \inf_{\substack{v\in Y\\ \beta>0}} \frac{1}{\beta}\,[\ P(v+\beta\,w) - \Lambda v\ ].$$

Define $\Theta\colon Y \bigoplus \mathbb{R}w \to \mathbb{R}$ by $\Theta(v + rw) = \Lambda v + ra$ for each $r \in \mathbb{R}$ and $v \in Y$. Considering the cases of r positive, negative or zero separately, the definition of a yields

$$\Theta(v+rw) = \Lambda v + ra \leq \Lambda v + P(v+rw) - \Lambda v = P(v+rw).$$

So Λ can be extended one dimension at a time while preserving its relationship with P.

Let S be the set of all linear extensions of Λ to subspaces of X which are dominated by P on their domain. Partially order this set of extensions by $\Theta \leq \Psi$ if Ψ is an extension of Θ. Chains in S have upper bounds in S and we invoke Zorn's lemma and assert that there is a maximal member Ψ of S. The domain of Ψ is X, else it could be extended by one dimension, contradicting maximality. □

When I first encountered this result, my instinct was to say its proof should be obvious: fatten a basis of Y to a basis of X and simply define Ψ to be 0 on those members of this basis not in Y, extending by linearity. Unfortunately this naive attempt fails: an inspection of the proof shows how a value of 0 on certain vectors may not be possible for Ψ. The point here is that there always *is* a value that is consistent with the other values of this function subject to domination by P.

Corollary* 11.2. *The Hahn-Banach Theorem

$$\textit{If } Y \overset{\textit{Complex Vector}}{\underset{\textit{Subspace}}{}} \subset X \overset{\textit{Complex Vector}}{\underset{\textit{Space}}{}} \textit{ and } P\colon X \to \mathbb{R} \textit{ satisfies}$$

$$P(\alpha v + \beta u) \leq |\alpha| P(v) + |\beta| P(u) \textit{ if } u, v \in X \textit{ and } |\alpha| + |\beta| = 1$$

$$\textit{and if } \Lambda \in Y^*_{\mathbb{C}} \textit{ satisfies } |\Lambda| \leq P|_Y$$

$$\textit{then } \exists \Psi \in X^*_{\mathbb{C}} \textit{ with } \Lambda = \Psi|_Y \textit{ and } |\Psi| \leq P.$$

Proof. Let L be the real part of Λ, thought of as a real linear functional. $\forall y \in Y,\ Ly \leq |\Lambda y| \leq P(y)$. Also, for real positive constants α and β the condition on P in the statement of this corollary reduces to convexity. So Theorem 11.1 applies: $\exists$ real linear $M\colon X \to \mathbb{R}$ extending L and with $Mx \leq P(x)\ \forall x \in X$.

Let $\Psi x = Mx - iM(ix)\ \forall x \in X$. Check that $\Psi(ix) = i\Psi x$, so Ψ is a complex linear functional and extends Λ to all of X. It remains only to show that $\Psi \leq P$.

Pick $x \in X$. Find angle θ so that $\Psi x = |\Psi x| e^{i\theta}$.

$$\begin{aligned} \text{Then } |\Psi x| \ = \ (\Psi x)e^{-i\theta} &= M\left(e^{-i\theta}x\right) - iM\left(ie^{-i\theta}x\right) \\ &= M\left(e^{-i\theta}x\right) \quad \text{(the complex part must be zero)} \\ &\leq P\left(e^{-i\theta}x\right) \ \leq \left|e^{-i\theta}\right| P(x) = P(x). \end{aligned}$$

□

12. Metric Spaces

A **pseudometric** on a nonempty set V is a function $d\colon V \times V \to [0, \infty)$ with the following properties:

$d(x,x) = 0\ \forall x \in V.$

$d(x,y) = d(y,x)\ \forall x, y \in V.$ (**symmetry**)

$d(x,z) \leq d(x,y) + d(y,z)\ \forall x, y, z \in V.$ (**triangle inequality**)

If, further, $d(x,y) = 0$ implies $x = y$ the pseudometric is called a **metric.**

A **pseudometric space** is a set V endowed with a pseudometric. And a **metric space** is a set V endowed with a particular metric.

A nonempty subset S of a pseudometric space is called **bounded** if there is a real number c for which $d(x,y) < c$ for any $x, y \in S$.

In a pseudometric space you can talk about limits. In particular, if $v_0,\ v_1,\ v_2, \ldots$ is a sequence in V we say that the sequence converges to $w \in V$ if the numbers $d(v_i, w)$ converge to 0. The point w is called a **limit** of the sequence. We indicate this situation by

$$\mathbf{v_i \to w}.$$

A sequence in a pseudometric space can converge to more than one limit.

A sequence as above is called **Cauchy** if for any $\varepsilon > 0$ there is an integer k for which $d(v_i, v_j) < \varepsilon$ whenever both i and j exceed k. Due to the triangle inequality, every Cauchy sequence is bounded: that is, the set consisting of sequence members is bounded.

Also every convergent sequence is Cauchy.

An **open ball** of radius ρ centered at v is the collection of all members of V whose distance from v is less than ρ. This ball is denoted $\mathbf{B}_\rho(\mathbf{v})$. The **closed ball** of radius ρ centered at v is the collection of all members of V whose distance from v is less than or equal to ρ. The closed ball is denoted $\mathbf{C}_\rho(\mathbf{v})$. A **sphere** of radius ρ centered at v is the collection of all members of V at distance ρ away from v. This sphere is denoted $\mathbf{S}_\rho(\mathbf{v})$.

$$C_\rho(v) = B_\rho(v) \bigcup S_\rho(v).$$

A subset S of pseudometric space V is called **closed** if S contains all limits of every convergent sequence comprised solely of members of S. It is not guaranteed that Cauchy sequences converge, and this definition says nothing about these potential nonconvergent Cauchy sequences.

The **closure** of a set S of members of a pseudometric space V is the set of all limits of Cauchy sequences of members of S. The closure of S is denoted $\overline{\mathbf{S}}$, and is a closed set. It is the smallest closed set containing S.

It is *not* true in general that $\overline{B_\rho(v)} = C_\rho(v)$, though this *will* be true in the metric vector spaces we work with later.

A subset S of pseudometric space V is called **open** if the set $V - S$ is closed. So for each $v \in S$ you *cannot* get arbitrarily close to v using only points that are not in S. This can be rephrased as follows: S is open if for every $v \in S$ $\exists \rho > 0$ for which $B_\rho(v) \subset S$.

Open sets can, therefore, all be represented as a union of open balls, and any such set is open.

The pseudometric is a metric exactly when every one-point set $\{ v \}$ is closed.

A metric space V is called **complete** if every Cauchy sequence in V converges to some member of V.

A subset S of a metric space V is called **sequentially compact** if every sequence in S has a subsequence which converges to a member of S. This means that sequentially compact subsets of a metric space are closed. They also must be bounded. Also, any closed subset of a sequentially compact set is sequentially compact.

The **Bolzano-Weierstrass theorem** states that in a finite dimensional Euclidean space (i.e. $\mathbb{R}^n$) the converse is true. Every closed and bounded subset of $\mathbb{R}^n$ is sequentially compact. This is an extremely important property.

If a subset S of metric V is contained in the union of a subsets M_α of V, where α is drawn from some generic index set A, this collection of sets is called a **cover** of S. A **subcover** of this cover is a (possibly) smaller selection of the M_α that nevertheless *still* covers S.

If every cover of S by open sets has a finite subcover we say S is **compact.** Since every open set is a union of open balls in a metric space we can, equivalently, restrict attention to covers by open balls in this definition.

In a metric space the notions of compactness and sequential compactness are equivalent.

A subset S of a metric space v is called **totally bounded** if for each $\alpha > 0$ there is a finite collection of balls of radius α which covers S.

The **Heine-Borel theorem** states that in a *complete* metric space, a set S is compact exactly when it is closed and totally bounded.

A set is bounded in $\mathbb{R}^n$ exactly when it is totally bounded, so in the finite dimensional setting (given the equivalence of sequential compactness and compactness in metric spaces) the Heine-Borel theorem adds nothing to Bolzano-Weierstrass. But in an infinite dimensional space boundedness and total boundedness are **not** the same and we require the more general total boundedness condition.

A function $f\colon M \to N$ between two metric spaces is called an **isometry** if

$$d(f(u), f(v)) = d(u, v) \quad \forall u, v \in M$$

and where the metric in use in the line above is deduced from context.

If f is an invertible isometry, the two spaces M and N are called **isometric.** Isometric spaces can be identified for any purpose involving just the metric.

A subset S of V is called **dense** if every member of V is the limit of members of S. In other words, $V = \overline{S}$. This is equivalent to saying that every open ball contains a point in S.

V is called **separable** if there is a countable dense subset of V. (Countable means it can be "counted" using the members of $\mathbb{N}$.) Separability, when present, is a very convenient property.

A function $F\colon V \to W$ between two pseudometric spaces is called **continuous at point v** $\in V$ if, for any $\varepsilon > 0$ there is a $\delta > 0$ so that $d(F(w), F(v)) < \varepsilon$ whenever $d(w, v) < \delta$. The function is called **continuous** if it is continuous at every v in the domain. The function is called **uniformly continuous** if δ can be chosen for each ε independently of the point v: that is, the same δ "works" for every v.

If V and W are pseudometric spaces, the function $F\colon V \to W$ is continuous at v exactly when the sequence $F(v_i)$ converges to $F(p)$ for every sequence in $v_0, v_1 \ldots$ converging to p in V.

This is an important feature of pseudometric spaces: continuity can be determined or refuted by examining function values on sequences.

It is not too hard to show that a function F as above is continuous exactly when $\mathbf{F^{-1}(B)}$, defined to be $\{\, x \in V \mid F(x) \in B \,\}$, is open whenever B is open in the range space W.

If V and W are two pseudometric spaces with pseudometrics d_1 and d_2 the product space $V \times W$ will be given pseudometric d defined by

$$d\left(\ (x, y),\ (z, w)\ \right) = d_1\left(\, x, z\,\right) + d_2\left(\, y, w\,\right) \quad \forall x, z \in V \text{ and } \forall y, w \in W.$$

Standard examples from multivariable calculus illustrate the following slightly unpleasant phenomenon.

Suppose $F\colon V \times W \to Z$ where Z is a third metric space. It is quite possible for the functions

$$F(v,\cdot)\colon W \to Z \quad \text{and} \quad F(\cdot,w)\colon V \to Z$$

to be continuous for every $v \in V$ and every $w \in W$ and yet F itself fails to possess this virtue. We will consider this matter later in the context of normed spaces.

Finally, we have the deceptively useful:

Theorem* 12.1. *Baire Category Theorem *If Y is a complete metric space every countable collection of open dense subsets has dense intersection.*

Proof. See the literature. □

In this work our metric spaces will usually be vector spaces, and we insist that our metrics honor this vector space structure to the extent that scalar multiplication and vector addition must be continuous on the appropriate product spaces. This will be obvious in the contexts that concern us.

13. Normed Spaces

NLS1 If you want to do calculus in your space you must have limits, and the easiest way to talk about limits in a vector space is through the explicit notion of distance provided through a norm.

If V is a vector space over $\mathbb{R}$ or $\mathbb{C}$, a seminorm on V is a function

$$\|\cdot\|\colon V \to [0,\infty)$$

with the property that for any number k and vectors v and w

$$\|kv\| = |k|\,\|v\|$$
$$\text{and} \quad \|\,v+w\,\| \le \|v\| + \|w\|$$

The seminorm is called **homogeneous** by virtue of the first line. The second of these is called the **triangle inequality**. The triangle inequality can be tweaked slightly to produce a lower limit for the norm of a sum too.

$$|\;\|v\| - \|w\|\;| \le \|\,v+w\,\| \le \|v\| + \|w\|.$$

A **seminormed linear space**, abbreviated **SNLS**, is a real or complex vector space endowed with a seminorm.

A seminorm satisfies

$$\|\,\alpha v + \beta u\,\| \le |\alpha|\,\|\,v\,\| + |\beta|\,\|\,u\,\| \quad \text{if } \; u,v \in X \text{ and } \alpha,\beta \in \mathbb{F}.$$

So a seminorm is an example (the most important example) of a sublinear function as found in the statement of Theorem 11.1 and its Corollary.

If you add the condition

$$\|v\| = 0 \;\text{ when and only when }\; v = 0$$

the seminorm is called a norm.

If $G\colon V \to F$ is any linear functional, the map $|G|\colon V \to [0,\infty)$ given as $|G|(v) = |G(v)|$ is a seminorm, and a common source of them too. This seminorm can never be a norm unless V has dimension 1.

A **normed linear space**, abbreviated **NLS**, is a real or complex vector space endowed with a norm.

NLS2 The **distance** between vectors v and w in a SNLS V is defined by

$$d(x, y) = \|v - w\|.$$

The distance notion is a pseudometric on V, and is a particularly nice one, having the properties

$d(x + z, y + z) = d(x, y)\ \forall x, y, z \in V$ (**translation invariance**)

$d(a\,x, a\,y) = |a|\, d(x, y)\ \forall x, y \in V$ and $a \in \mathbb{F}$ (**homogeneity**)

not required of a general metric or pseudometric.

Whenever a notion of distance is used in an SNLS it is this pseudometric which will be intended.

This pseudometric is a metric exactly when the seminorm is a norm.

NLS3 If a sequence v_0, v_1, $v_2, \ldots$ converges in a SNLS using this pseudometric we say that the sequence **converges in seminorm (or norm).** Sometimes this is also called **strong convergence**, particularly when we have a norm from an inner product. (There is a weaker concept of convergence which is also useful there.) In case there might be confusion about the type of convergence involved, we might indicate intent by

$$v_i \underset{\text{strong}}{\longrightarrow} w.$$

If V is an SNLS the set $\mathbf{N} = \{\, x \in V \mid \|x\| = 0 \,\}$ is a vector subspace of V, sometimes called the set of **null vectors** for the seminorm (not to be confused with vectors from the nullspace of a linear transformation.)

In an SNLS a sequence can converge to more than one point. In fact, if $v_i \to w$ then $v_i \to x$ exactly when $x = w + y$ for some $y \in \mathbf{N}$.

This implies that $\mathbf{N}$ is a closed set.

The seminorm is a norm exactly when $\{\,0\,\}$ is closed.

Completeness is a very important property for us. Normed linear spaces which are complete are called **Banach spaces.**

$\mathbb{R}$ and $\mathbb{C}$ are themselves Banach spaces, a critical fact that is used often and assumed without discussion in most first-year calculus classes.

In fact, $\mathbb{R}$ is usually built using the linear order and distance notion in the rational numbers $\mathbb{Q}$ by adjoining elements to $\mathbb{Q}$ corresponding to all the "holes" in $\mathbb{Q}$. So $\mathbb{R}$ is ordered too, and complete by its very definition.

If S is a vector subspace of V then $\overline{S}$ is also a SNLS space, a subspace of V. If V is Banach, so is $\overline{S}$.

NLS4 For V and W both SNLSs, a function $F\colon V \to W$ is continuous at $p \in V$ exactly when

$$\forall \varepsilon > 0\ \exists \delta > 0 \text{ so that if } \|\, p - v \,\| < \delta \text{ then } \|\, F(p) - F(v) \,\| < \varepsilon.$$

In general, δ will depend on both ε and p, but for linear functions it does not. Using linearity, one shows that continuity at *any point*, that is for just one v, implies continuity at *every point*, and the δ chosen for a particular ε does not depend on *which* point.

Linear functions between two SNLSs
which are continuous at a point are uniformly continuous.

In fact if you can find $\delta_1 > 0$ for $\varepsilon = 1$ and $p = 0$ then F is continuous. That is because for any $\varepsilon > 0$ we have

$$\|v\| < \delta_1 \varepsilon \;\Rightarrow\; \frac{\|v\|}{\varepsilon} < \delta_1 \Rightarrow \left\| F\left(\frac{v}{\varepsilon}\right)\right\| < 1 \Rightarrow \frac{1}{\varepsilon}\; \|F(v)\| < 1$$

and so $\|F(v)\| < \varepsilon$. Linearity then allows us to translate the argument away from the origin.

A useful equivalent condition, again using linearity of F, is that if $F^{-1}(B)$ is open for even *one open ball* B then $F^{-1}(B)$ will be open for *every* open subset B contained in W and therefore F will be continuous.

For continuous linear F, the set $\mathcal{K}er(F)$ is closed.

NLS5 If $\mathbf{N}$ is the set of null vectors for SNLS V with seminorm $\|\cdot\|$ we can form the quotient space $V/\mathbf{N}$. The function $\|\cdot\|'$ defined on $V/\mathbf{N}$ by

$$\| v + \mathbf{N}, \|' = \| v \|$$

is a norm on $V/\mathbf{N}$.

Further, if $F^{\text{Linear}}\colon V^{\text{SNLS}} \to W^{\text{SNLS}}$ and if $\mathbf{N} \subset \mathcal{K}er(F)$ the function

$$\widetilde{F}\colon (V/\mathbf{N})^{\text{NLS}} \to W^{\text{SNLS}} \quad \text{given by} \quad \widetilde{F}(v+\mathbf{N}) = F(v)$$

is well-defined and linear. $\widetilde{F}$ is continuous when F is continuous.

NLS6 If V and W are two SNLSs with seminorms $\|\cdot\|_1$ and $\|\cdot\|_2$, respectively, the product vector space $V \times W$ can be given the **product seminorm** defined by

$$\|(u,v)\| = \|u\|_1 + \|v\|_2.$$

This function is, in fact, a seminorm on the product space, and a norm exactly when both $\|\cdot\|_1$ and $\|\cdot\|_2$ are norms.

NLS7 A function $F^{\text{Linear}}\colon V^{\text{NLS}} \to W^{\text{NLS}}$ is an **isometry** provided that $\| F(x) \| = \| x \|\ \forall x \in V$. Isometries are continuous.

Also, an isometry is automatically one-to-one, so isometry F implements an isomorphism between V and $\mathcal{R}an(F)$. But this is not just any isomorphism: it preserves the notion of distance provided by the norm on the domain.

Using the restriction of the norm on W to $\mathcal{R}an(F)$, the NLS V and the NLS $\mathcal{R}an(F)$ are identified. If F is onto, we have an identification of V and W with norm structure preserved.

NLS8 For $F^{\text{Linear}}\colon V^{\text{SNLS}} \to W^{\text{SNLS}}$ we define $\|\mathbf{F}\|$ by

$$\|F\| = \sup\{\, \|F(x)\| \mid x \in S_1(0)\,\}.$$

If $\|F\| < \infty$ we say F is **bounded**, and it turns out that $\|F\|$ is bounded exactly when F is continuous. This is important enough that we enshrine the result in:

A linear functions between two SNLSs
is continuous exactly when it is bounded.

B(V, W) is defined to be the set of bounded linear functions from V to W. The number $\|F\|$ defined for each $F \in B(V, W)$ is a norm on $B(V, W)$, generally called the **operator norm**.

If the range space W is a Banach space, so is $B(V, W)$.

If there are multiple norms floating around, we might distinguish this one by the notation $\| \cdot \|_{op}$.

NLS9 $A^{\text{Linear}}: X^{\substack{\text{Banach}\\ \text{Space}}} \to Y^{\substack{\text{Banach}\\ \text{Space}}}$ is called **compact** exactly when $\overline{A(S)}$ is a compact subset of Y whenever S is a bounded subset of X. If compact, a function A must be bounded and so compact linear functions from one Banach space to another are continuous.

There is an important equivalent condition, which reduces the question of whether a mapping is compact or not to an issue involving sequences.

Suppose (x_i) is a bounded sequence in X. Then compactness of A requires $\{ A(x_i) \mid i \in \mathbb{N} \}$ to have compact closure in Y. This implies that the sequence of image points $(A(x_i))$ has a Cauchy subsequence.

Conversely, if this last condition must necessarily hold for any bounded sequence in X then A is a compact mapping.

If a continuous map $A^{\text{Linear}}: X^{\substack{\text{Banach}\\ \text{Space}}} \to Y^{\substack{\text{Banach}\\ \text{Space}}}$ has finite rank, it is clearly compact. It is not too hard to show that the limit (in operator norm) of continuous finite rank maps is compact. It is also true that if Y is a Hilbert space or any Banach space with a Schauder basis, any compact linear function from Y to Y is the limit of continuous finite rank maps, a desirable feature referred to as the **approximation property** for these compact maps.

NLS10 The collection of bounded linear functionals on an NLS V is a particularly important Banach space. It is called the **continuous dual** of V, denoted $V' = B(V, \mathbb{F})$. The continuous dual is a subspace of the algebraic dual, $V^* = Hom_{\mathbb{F}}(V, \mathbb{F})$.

Suppose x is nonzero in NLS V. The linear transformation Λ defined on $\mathbb{F}x$ by $\Lambda(a\,x) = a\,\|x\|$ satisfies the condition of Theorem 11.1 (or its Corollary) where P is the norm on V. In this one dimensional case we have equality:

$$|\,\Lambda(a\,x)\,| = |a|\,\|x\| = \|a\,x\|.$$

So Λ can be extended to linear Ψ defined on all of V and for which

$$|\Psi(v)| \leq \|v\| \quad \text{for all} \quad v \in V.$$

This means that the operator norm $\|\Psi\|$ cannot exceed 1.

But $\Psi(x/\|x\|) = \Lambda(x/\|x\|) = 1$, So $\|\Psi\| = 1$.

To recap, for each $x \in V$ there is a functional Ψ in V' with

$$\|\Psi\| = 1 \quad \text{and} \quad \Psi x = \|x\|.$$

NLS11 Since V' is a NLS, it too has a an algebraic dual $(V')^*$ and continuous dual $V'' \subset (V')^*$.

Recall the evaluation map

$$E\colon V \to (V^*)^* \text{ given by } E(x)(f) = f(x).$$

Every member of $(V^*)^*$ produces a member of $(V')^*$ by restriction and $\widetilde{E}(x)$ defined to be $E(x)|_{V'}$ is such a member.

And if $f \in V'$ and $\|f\| = 1$ then $\left|\widetilde{E}(x)(f)\right| = |f(x)| \leq \|f\|\,\|x\| = \|x\|$.

So each $\widetilde{E}(x)$ is bounded as a function from Banach space V' with operator norm to $\mathbb{F}$. In other words,

$$\widetilde{E}\colon V \to V''.$$

But by the last result of NLS10 there is a functional $\Psi \in V'$ with operator norm 1 for which $|\Psi(x)| = \|x\|$. This means that $\left|\widetilde{E}(x)\right|$ actually attains its maximum value, $\|x\|$, on the members of V' with operator norm 1.

Even more, this means that

$$\widetilde{E}\colon V \to V'' \quad \text{is an isometry.}$$

So the image of $\widetilde{E}$ with operator norm and V itself with its norm are not only isomorphic as vector spaces but are interchangeable in any calculation involving norms as well.

Spaces for which $\widetilde{E}$ is onto V'' are very important, and are called **reflexive**. We will have occasion to refer to this property later. Hilbert spaces are reflexive.

14. The Open Mapping, Banach-Steinhaus and Closed Graph Theorems

A function $\Psi^{\text{Linear}}\colon V^{\text{SNLS}} \to W^{\text{SNLS}}$ is called **open** if $\Psi(A)$, defined to be $\{\,\Psi(x) \mid x \in A\,\}$, is an open set in W whenever A is open in V. Actually, applying linearity of Ψ, if $\Psi(A)$ is open for even *one* open ball A then $\Psi(A)$ will be open for *every* open subset A of the domain.

Proposition* 14.1. *The Open Mapping Theorem
If $\Psi \in B\left(V^{Banach}, W^{Banach}\right)$ *and* $\Psi(V) = W$ *then* Ψ *is an open map.*

Proof. See the literature. □

***Corollary* 14.2.** *If* $\Psi \in B\left(V^{Banach}, W^{Banach}\right)$ *is one-to-one and* $\Psi(V) = W$ *then* $\Psi^{-1} \in B(W, V)$.

Proof. The proof is an immediate consequence of the last proposition. □

Proposition* 14.3.** ***The Banach-Steinhaus Theorem*, also called* ***The Principle of Uniform Boundedness***

Suppose $\mathcal{A} \subset B\left(V^{Banach}, W^{NLS}\right)$.

Then either $\exists M < \infty$ *for which* $\|\Psi\| \leq M \;\forall \Psi \in \mathcal{A}$

or *the function* $\lambda \colon V \to [0, \infty]$ *given by* $\lambda(x) = \sup_{\Psi \in \mathcal{A}} \|\Psi x\|$ *is infinite on a dense subset of* V.

Proof. See the literature. □

Suppose given $F \in (V \times W)^*$ where V and W are vector spaces over the same field.

For each $v \in V$ and $w \in W$ define functions A_v and B_w by

$$A_v = F(v, \cdot\,) \colon W \to \mathbb{F} \quad \text{and} \quad B_w = F(\,\cdot\,, w) \colon V \to \mathbb{F}.$$

So A_v is in W^* and B_w is in V^* for each v and w.

We use this notation in the statement of the following result, giving the product of two Banach spaces the product norm.

Corollary* 14.4.** ***A linear functional on a product of two Banach spaces which is continuous in each factor separately is jointly continuous.
If $F \in \left(V^{Banach} \times W^{Banach}\right)^*$
and $A_v \in W'$ *and* $B_w \in V'$ *for every* $v \in V,\; w \in W$
then $F \in (V \times W)'$.

Proof. Follows from the Banach-Steinhaus Theorem. □

If $f \colon X \to Y$ where X and Y, the **graph of f** is the set

$$\gamma(\mathbf{f}) = \{\,(x, f(x)) \mid x \in X\,\} \subset X \times Y.$$

When X and Y are SNLSs, we give $X \times Y$ the product norm.

Proposition* 14.5.** ***The Closed Graph Theorem

Suppose $\Psi^{Linear} \colon V^{Banach} \to W^{Banach}$.

$\Psi \in B(V, W)$ *if and only if*
$\gamma(\Psi)$ *is a closed subset of* $V \times W$.

Proof. This follows from the Open Mapping Theorem and the last Corollary. □

15. Schauder Bases in a Banach Space

The Baire category theorem implies that no complete infinite dimensional space can have a countable basis. However, if we can't have a countable *basis*, we can do almost as well with a Schauder basis, defined below, which uses concepts of limit and continuity provided by a norm to get most of what a true basis provides in the finite dimensional setting.

The ideas to follow make sense in more general settings but we will confine consideration here to Banach spaces.

A **Schauder basis** for Banach X is a countable *ordered* set of vectors $v_0, v_1, \ldots$ for which every member x of X can be written in a *unique* way as

$$x = \sum_{n=0}^{\infty} a^n(x)\, v_n$$

for $a^n(x) \in \mathbb{F}$. The uniqueness refers to the values of the **coordinate functionals** a^n, which are therefore linear, and the convergence of the sequence of partial sums is in norm:

$$\sum_{n=0}^{k} a^n(x)\, v_n \underset{\text{strong}}{\longrightarrow} x \quad \forall x \in X.$$

When referring to a sequence of vectors in vector space X the notation $\mathbf{v} = (\mathbf{v_n}) \subset X$ will be used. Thus, for Schauder basis as above we have the paired sequences

$$\mathbf{v} \subset X \quad \text{with unique associated coordinate functionals} \quad \mathbf{a} \subset X^*.$$

Any Banach space with a Schauder basis is separable, so there are Banach spaces without Schauder bases. In fact, there are Banach spaces for which no infinite dimensional subspace has a Schauder basis. Still, Banach spaces that have these bases are common in practice.

Uniqueness of coefficients implies that 0 is not among the vectors in a Schauder basis, and in fact the vectors in a Schauder basis must constitute a linearly independent list of vectors.

If $(\mathbf{v_n})$ is a Schauder basis, so is $(\mathbf{v_n}/\|\mathbf{v_n}\|)$ and the latter Schauder basis is called **normalized.**

Normalized or not, the members of the sequence of linear functionals $(\mathbf{a^n})$ are all continuous. In fact, (though it takes a bit of work to prove) their norms satisfy

$$1 \le \|a^n\|\, \|v_n\| \le K \quad \forall n \in \mathbb{N}$$

for a positive constant K that will vary with the basis.

So for Schauder basis $\mathbf{v} \subset X$ we actually have $\mathbf{a} \subset X'$, not just $\mathbf{a} \subset X^*$.

The sum $P^k = \sum_{n=0}^{k} a^n\, v_n$ is bounded and linear from X onto the finite dimensional subspace of x spanned by $v_0, \ldots, v_k$. It is a **projection** onto that subspace: $P^k \circ P^k = P^k$ and, more generally, if $i \le j$ then $P^i \circ P^j = P^j \circ P^i = P^i$. The sequence of projections converges to the identity operator in operator norm.

This sequence is used to show that a compact operator K is the limit of finite rank operators in a Banach space with a Schauder basis: examine the sequence $P^i \circ K$.

A Schauder basis is called **unconditional** if, for any $x \in X$ the series obtained by any permutation of the terms in the series representation for x in this Schauder basis also converges to x.

A Schauder basis is called **bounded** if there are positive constants A and B for which

$$A \leq \|v_n\| \leq B \quad \forall n \in \mathbb{N}.$$

Since each vector in a Schauder basis $\mathbf{v}$ can be conceived of as (that is, *it is*) a member of X'' the possibility arises that $\mathbf{a}$ could be a Schauder basis for the Banach space $\overline{Span(\mathbf{a})}$ with operator norm, with coordinate functionals $\mathbf{v}$ (the domain of each vector in $\mathbf{v}$ restricted, of course, to $\overline{Span(\mathbf{a})}$.)

This is, in fact, the case. Even more, we have the following.

Theorem* 15.1. *The Dual Basis Theorem

Suppose $\mathbf{v} \overset{Schauder}{\underset{Basis}{\;}} \subset X^{Banach}$ *with associated coordinate functionals* $\mathbf{a} \subset X'$.

Then $\mathbf{a} \overset{Schauder}{\underset{Basis}{\;}} \subset \overline{Span(\mathbf{a})}$ *with associated coordinate functionals* $\mathbf{v}$.

Moreover, if $\mathbf{v}$ *is unconditional or bounded, so is* $\mathbf{a}$.

Proof. See the literature. □

The most interesting case, of course, is when X is reflexive. Then $X' = \overline{Span(\mathbf{a})}$ and so we have a correspondence between Schauder bases with their coefficient sequences for X and those for X'.

16. The Banach Adjoint

Suppose $F \in B\left(V^{\text{Banach}}, W^{\text{Banach}}\right)$.

Define for each $w' \in W'$ the member $F^*(w') \in V^*$ given by

$$F^*(w')(v) = w'(F(v)).$$

Note that $\|F^*(w')(v)\| \leq \|w'\|\,\|F\|\,\|v\|$.

So, in fact

$$\|F^*(w')\| \leq \|w'\|\,\|F\|$$

and this means $F^*(w')$ is actually in V', not just V^*.

Looking again we see that $F^* \in B(W', V')$ and $\|F^*\| \leq \|F\|$.

Application of the Hahn-Banach theorem proves more: that $\|F^*\| = \|F\|$.

It is also a fact that if F is an isometry, so is F^*.

The map $F^*\colon W' \to V'$ is called the **Banach adjoint** of F and the **Banach adjoint operator**

$$^* \;:\; B(V, W) \to B(W', V') \quad \text{is an isometry.}$$

Finally, if V and W are reflexive, we note that $F^{**} = F$, and the adjoint operator is an isomorphism.

17. Integral Tools and Some Key Examples

We presume in the propositions below that Γ is a Lebesgue integral for a positive σ-finite measure μ with integrable real valued functions $\mathcal{L}$.

An example of this would be the usual Lebesgue integral defined on the real line, where $\mathcal{L}$ would denote the real valued measurable functions with finite integral. So in this case for $f \in \mathcal{L}$ we have, using the usual notation,

$$\Gamma(f) = \int_{-\infty}^{\infty} f \, dx.$$

The phrase σ-finite measure means that the underlying domain set (in this case $\mathbb{R}$) can be broken into a countable number of pieces of finite measure. So in our example $\mathbb{R}$ could be broken up into the pieces $[n, n+1]$ for whole numbers n, each of measure 1.

If the underlying domain has finite measure to begin with, such as a bounded interval in the real line with usual Lebesgue measure there, a σ-finite measure is usually just called a measure. Measures are typically a little better-behaved than non-finite σ-finite measures. The prototypical example is, of course, the usual Lebesgue integral on $[0, 1]$ given by

$$\Gamma(f) = \int_{0}^{1} f \, dx.$$

If you don't know what a Lebesgue integral is, feel free to think of the Riemann integral. The Lebesgue integral agrees with the Riemann integral wherever the latter is defined and is much better behaved under limit-taking, which is of central importance to us. However if you are willing to believe certain convergence facts without proof, your intuition from the Riemann integral will not (often) lead you astray.

The members of $\mathcal{L}$ form a real vector space of functions with finite integral.

There are three very important features of this space which we bring up now.

First, $\mathcal{L}$ is a **lattice** in addition to being a vector space.

By that we mean that if $f, g \in \mathcal{L}$ then so are $f \vee g$ and $f \wedge g$, where we define these functions for x in the domain of integration by

$$(f \vee g)(x) = \max\{\, f(x), g(x)\,\} \quad \text{and} \quad (f \wedge g)(x) = \min\{\, f(x), g(x)\,\}.$$

Second, any nonincreasing sequence $f_0, f_1, \ldots$ of non-negative members of $\mathcal{L}$ converges (pointwise) to a member g of $\mathcal{L}$, and $\Gamma(f_n) \to \Gamma(g)$.

This is called **the monotone convergence theorem.**

Finally, if $f_0, f_1, \ldots$ is *any* sequence of members of $\mathcal{L}$ and if there is a member h of $\mathcal{L}$ with $|f_n| \leq h$ for all n, and if $f_n(x) \to g(x)$ for each x in the domain of integration, then $g \in \mathcal{L}$ and $\Gamma(|g - f_n|) \to 0$ from which it follows that $\Gamma(f_n) \to \Gamma(g)$.

This is called **the dominated convergence theorem.**

We define $\overline{\mathcal{L}} \vee 0$ to be the non-negative functions which are increasing limits (in the extended sense: ∞ is a potential limit value) of members of $\mathcal{L}$.

Not only might these functions have infinite values, they could also have infinite integrals, defined as the limit of the integrals of the increasing sequence of functions used to define them, and we allow this in the inequalities found below.

Proposition* 17.1. *Hölder's Inequality

Suppose $f, g \in \overline{\mathcal{L}} \vee 0,\ p > 1$ *and* $\dfrac{1}{p} + \dfrac{1}{q} = 1.$

Then $\quad \Gamma(fg) \leq (\Gamma(f^p))^{\frac{1}{p}} (\Gamma(g^q))^{\frac{1}{q}}.$

Proof. See the literature. □

Proposition* 17.2. *Minkowski's Inequality

Suppose $p \geq 1$ *and* $f, g \in \overline{\mathcal{L}} \vee 0$. *Then*

$$(\ \Gamma\left((f+g)^p\right)\ )^{1/p} \leq (\ \Gamma(f^p)\ )^{1/p} + (\ \Gamma(g^p)\ )^{1/p}.$$

Proof. See the literature. □

For $p \geq 1$ and any function g for which $|g| \in \overline{\mathcal{L}} \vee 0$ define the number

$$\|\, g\,\|_p = (\Gamma\,(\,|g|^p\,))^{1/p}\,.$$

$\boldsymbol{\mathcal{L}_p}$ consists then, by definition, of those real valued g for which

$$g \vee 0 \in \overline{\mathcal{L}} \vee 0 \quad \text{and} \quad (-g) \vee 0 \in \overline{\mathcal{L}} \vee 0 \quad \text{and} \quad \|\, g\,\|_p < \infty.$$

For those g for which $|g| \in \overline{\mathcal{L}} \vee 0$ define the number

$$\|\, g\,\|_\infty = \inf\{\, C \mid \mu(\,\{x \mid |g(x)| > C\}\,) = 0\,\}.$$

In other words, this is the least number for which the measure of the set where $|g|$ exceeds that number is zero. As far as the measure can tell, $|g|$ never exceeds this number. The infinum of the empty set is ∞, so functions that become arbitrarily large on non-negligible sets have $\|\, g\,\|_\infty = \infty$.

We then define $\boldsymbol{\mathcal{L}_\infty}$ to be those real valued functions g for which

$$g \vee 0 \in \overline{\mathcal{L}} \vee 0 \quad \text{and} \quad (-g) \vee 0 \in \overline{\mathcal{L}} \vee 0 \quad \text{and} \quad \|\, g\,\|_\infty < \infty.$$

Each $\boldsymbol{\mathcal{L}_p}$ is a vector space and contains $\mathbf{N}$, defined to be the set of those real valued functions g for which $\|\, g\,\|_p = 0$ for any, and hence every, $p \geq 1$. Members of $\mathbf{N}$ are referred to as **null functions.**

Finally, we define $\boldsymbol{L^p}$ to be the quotient space $\boldsymbol{\mathcal{L}_p}/\mathbf{N}$. So each member of this quotient space consists of a set of functions which differ in a way that the integral cannot distinguish: they differ on "a set of measure zero" and we think of them all as the same function, essentially.

If $f, g \in \overline{\mathcal{L}} \vee 0$ then (using norm notation) the Hölder's and Minkowski inequalities become:

(i) For $p > 1$ and $\frac{1}{p} + \frac{1}{q} = 1 \quad \|\, fg\,\|_1 \leq \|\, f\,\|_p\, \|\, g\,\|_q$

(ii) $\|\, fg\,\|_1 \leq \|\, f\,\|_1\, \|\, g\,\|_\infty$

(iii) For $p \geq 1 \quad \|\, f+g\,\|_p \leq \|\, f\,\|_p + \|\, g\,\|_p$

For $\infty \geq p \geq 1$ and $f + \mathcal{N} \in \boldsymbol{L^p}$ we define the number $\| f + \mathcal{N} \|_p$ to be $\| f \|_p$. An easy calculation shows this to be a good definition: that is, if $f + \mathcal{N} = g + \mathcal{N}$ then $\| g + \mathcal{N} \|_p = \| f \|_p$.

Using the same "norm" symbol for two different functions can produce confusion, but the custom is to simply pay attention and deal with it. In fact, many sources make little or no distinction between a function and the equivalence class of that function beyond a few perfunctory phrases when these classes are introduced, leaving it to context to distinguish intent. That usually works just fine.

(i) $\| \cdot \|_p$ is a seminorm on $\boldsymbol{\mathcal{L}_p}$ and a norm on $\boldsymbol{L^p}$. The norm is referred to as the **p-norm on $\boldsymbol{L^p}$** if $1 \leq p < \infty$, while $\| \cdot \|_\infty$ is called the **essential sup norm**.
The underlying measure is understood and made explicit in the setup of the situation.

(ii) If $1 \leq p \leq \infty$ define for $\tilde{f} = f + \mathcal{N}$ and $\tilde{g} = g + \mathcal{N}$ in $\boldsymbol{L^p}$ the number

$$d_p(\tilde{f}, \tilde{g}) = \| f - g \|_p.$$

d_p is a metric on $\boldsymbol{L^p}$.

Theorem* 17.3. *The Riesz-Fisher Theorem
$\boldsymbol{L^p}$ is complete with metric d_p for $1 \leq p \leq \infty$.

Proof. See the literature. □

For $p \geq 1$ we will generally use the notation $\tilde{f}$ when $f + \mathcal{N} \in \boldsymbol{L^p}$, as we did above. The notation tends to make discussions easier to look at (and a lot shorter.) Be aware though that we could have $\tilde{f} = \tilde{g}$ without $f = g$. The two functions f and g may differ by a null function.

Suppose $\tilde{f}_0, \tilde{f}_1, \ldots$ is a Cauchy sequence with respect to the metric defined above for some $p \geq 1$. In the proof of the Riesz-Fisher Theorem, a subsequence $\tilde{h}_0, \tilde{h}_1, \ldots$ is produced and a function $H \in \boldsymbol{\mathcal{L}_p}$ to which $h_0, h_1, \ldots$ converges at each point ("pointwise" convergence) and for which $\tilde{h}_0, \tilde{h}_1, \ldots$ converges to $\widetilde{H}$ in norm. Since the original sequence is Cauchy, $\tilde{f}_0, \tilde{f}_1, \ldots$ converges to $\widetilde{H}$ in norm also.

As a cautionary note, we remark that if we let $p = 1$ and if Γ is the usual Lebesgue integral on the unit interval $[0, 1]$, we can produce an example of a sequence of real functions $f_0, f_1, \ldots$ which does not, itself, converge at any point yet $\tilde{f}_0, \tilde{f}_1, \ldots$ is Cauchy in norm.

The Riesz-Fisher Theorem as proved originally in 1907 was a specific result about Fourier series and the completeness of $\boldsymbol{L^2}$, but the essential point of their proof holds more generally and the name has come to be applied to the fact of $\boldsymbol{L^p}$ completeness with these metrics for more general p.

Now that we have learned that $\boldsymbol{L^p}$ is a Banach space with p-norm for $p \geq 1$ we are interested in finding the continuous functionals on $\boldsymbol{L^p}$.

A combination of the **Riesz Representation Theorem** and the **Radon-Nikodým Theorem** gives us the following result in this direction.

***Theorem* 17.4. *For* $1 > p > \infty$ *The Continuous Dual of* $\boldsymbol{L^p}$ *is* $\boldsymbol{L^q}$.**

Suppose $1 < p < \infty$*. Consider the Banach space* $\boldsymbol{L^p}$ *for integral* Γ*.*
For each continuous linear functional Ψ *on* $\boldsymbol{L^p}$
there is a unique member g *of* $\boldsymbol{L^q}$ *where* $\frac{1}{p} + \frac{1}{q} = 1$

$$\text{for which} \quad \Psi(f) = \Gamma\left(\overline{g}\, f\right) \quad \forall f \in \boldsymbol{L^p}.$$

Proof. See the literature. □

The function $\overline{g}$ is called a **Radon-Nikodým derivative** of the measure induced by the continuous functional on the underlying domain of the function space.

The operator norm of Ψ is the q-norm of g. The association is an isometry from the continuous dual of $\boldsymbol{L^p}$ with operator norm onto $\boldsymbol{L^q}$ with q-norm. So the function g "is" (for all practical purposes) the functional Ψ.

With this understanding, we have also that the continuous dual of $\boldsymbol{L^q}$ is $\boldsymbol{L^p}$, so if $\infty > p > 1$ we find that $\boldsymbol{L^p}$ is reflexive.

It is also true that the continuous dual of $\boldsymbol{L^1}$ is $\boldsymbol{L^\infty}$ and the association is with the Radon-Nikodým derivative as above. However the continuous dual of $\boldsymbol{L^\infty}$ is larger than $\boldsymbol{L^1}$, so we have found a Banach space $\boldsymbol{L^1}$ which is *not*, generally, reflexive.

Suppose $1 < a < b < c < \infty$. Then $\boldsymbol{L^a} \cap \boldsymbol{L^c} \subset \boldsymbol{L^b}$, and $\boldsymbol{L^a} \cap \boldsymbol{L^c}$ is Banach with norm $\|\cdot\| = \|\cdot\|_a + \|\cdot\|_c$.

We also note that if Γ is the integral for a *finite* measure (and not otherwise) the following containments obtain for $1 < a < b < \infty$:

$$\boldsymbol{L^\infty} \subset \boldsymbol{L^b} \subset \boldsymbol{L^a} \subset \boldsymbol{L^1}.$$

As a source of examples, we define $C([0,1])$ and $C(\mathbb{R})$ to be the sets of real valued continuous functions on $[0,1]$ and $\mathbb{R}$, respectively. $B(\mathbb{R})$ consists of the bounded members of $C(\mathbb{R})$. And we define $D([0,1])$ to be the set of continuously differentiable functions on the unit interval. The first four are Banach spaces with sup norm. The last is not Banach with sup norm, but is Banach with norm given by

$$\| f \| = \| f \|_\infty + \| f' \|_\infty .$$

Finally, we consider important special cases involving sequences.

For $f \in \mathbb{R}^{\mathbb{N}}$ and $p \geq 1$ define

$$\| f \|_p = \left(\sum_{k=0}^{\infty} | f_k |^p \right)^{\frac{1}{p}} .$$

The space $\boldsymbol{\ell^p}$ is defined to be those members of $f \in \mathbb{R}^{\mathbb{N}}$ for which $\| f \|_p < \infty$.

This is an example of the situation above, where the σ-finite measure is counting measure on $\mathbb{N}$.

$\boldsymbol{\ell}^1$ is the set of sequences whose terms converge absolutely.

We know from above that $\boldsymbol{\ell}^p$ with p-norm is a Banach space, and reflexive if $1 < p < \infty$.

The case of $p = 2$ will be particularly important to us, both in this sequence space example and in the case of more general measures. $\boldsymbol{\ell}^2$ is the set of "square summable" sequences, and $\boldsymbol{L}^2$ encompasses the "square integrable functions." They form a Hilbert spaces. They are "self-dual" since $1 - \frac{1}{2} = \frac{1}{2}$.

We define $\| f \|_\infty = \sup\{ |f_0|, |f_1|, |f_2|, \dots \}$ where sup indicates the least upper bound (possibly infinite) of the numbers in a set. This is called the **sup norm**, pronounced as "soup norm."

$\boldsymbol{\ell}^\infty$ is defined to be those members of $f \in \mathbb{R}^{\mathbb{N}}$ for which $\| f \|_\infty < \infty$.

Sup norm is a norm on $\boldsymbol{\ell}^\infty$, and $\boldsymbol{\ell}^\infty$ is complete with this norm. We note that this particular Banach space is not separable, and so has no Schauder basis.

It is a fact that the continuous dual of $\boldsymbol{\ell}^1$ is $\boldsymbol{\ell}^\infty$. However the continuous dual of $\boldsymbol{\ell}^\infty$ is *not* $\boldsymbol{\ell}^1$. It is the space of *finitely* additive, not *countably* additive measures on $\mathbb{N}$, with total variation norm. The point is that the Banach space $\boldsymbol{\ell}^1$ *is not* reflexive.

Investigating this interesting issue would take us pretty far afield so we will not explore it here.

18. Inner Product Spaces

IPS1 An **inner product** on a vector space V is a function

$$\langle \cdot , \cdot \rangle \colon V \times V \to \mathbb{F}$$

with the following properties, which vary slightly in appearance depending on whether our numbers are real or complex. First[4], it must be **"linear in the second slot."** This means that for all numbers k and vectors u, v and w

$$\langle u , v + kw \rangle = \langle u , v \rangle + k \langle u , w \rangle.$$

Second, it must have a **symmetry property**: for vectors u and v

$$\langle u , v \rangle = \overline{\langle v , u \rangle}$$

where the "overbar" indicates complex conjugation. If the numbers we use are real, this conjugation does nothing and we have, for instance, the usual symmetry of the dot product. But if the numbers are complex, this means that a complex inner product is **"conjugate linear in the first slot,"** i.e.

$$\langle u + kw , v \rangle = \langle u , v \rangle + \overline{k} \langle w , v \rangle.$$

Finally, we must have a **positivity** condition: $\langle u , u \rangle$ must always be real, non-negative, and can only equal 0 when u is the zero vector.

If the field is real, an inner product is a covariant tensor of order 2. But if the field is complex it is conjugate linear in the first slot, and two-slot functions of pairs of vectors with this property are referred to as **sesquilinear forms**. (In Latin sesqui means "one and a half.")

[4]A minority (recently) of authors reverse this, requiring linearity in the first slot rather than the second. This does change the appearance of many results slightly.

An **inner product space**, abbreviated **IPS**, is a vector space with an associated inner product.

The **norm associated with an inner product** is given by

$$\| v \| = \sqrt{\langle v , v \rangle}.$$

The fact that this actually *is* a norm, that it has the properties required of a norm, are immediate.

So an inner product space is also an NLS and, through the norm, a metric space.

IPS2 The norm from an inner product has two additional properties: for any vectors v and w

$$| \langle u , v \rangle | \leq \|v\| \, \|w\|$$

and also

$$2 \, \|u\|^2 + 2 \, \|v\|^2 = \| u + v \|^2 + \| u - v \|^2.$$

The first of these is the **Schwarz inequality**. Or the **Cauchy-Schwarz inequality**. Or the **Bunyakovsky-Cauchy-Schwarz inequality**. Depends on who you talk to. The second is called the **parallelogram law.**

We will prove the **BCS** inequality.

First, if either u or v is the zero vector the result is obvious, so presume neither are 0. We note that if

$$w = u - \frac{\langle u , v \rangle}{\langle v , v \rangle} v \quad \text{then } \langle w , v \rangle = 0.$$

We write $u = w + \frac{\langle u , v \rangle}{\langle v , v \rangle} v$ and expand $\| u \| = \langle u , u \rangle$ to produce the desired inequality.

Norms can come from various sources but only norms that come from (or *could* have come from) an inner product satisfy the parallelogram law. And an inner product can be built from any norm satisfying the parallelogram law.

If $\mathbb{F} = \mathbb{R}$ then a norm satisfying the parallelogram law comes from the inner product given by

$$\langle u , v \rangle = \frac{\| u + v \| - \| u - v \|}{4}.$$

If $\mathbb{F} = \mathbb{C}$ then a norm satisfying the parallelogram law comes from the inner product given by

$$\langle u , v \rangle = \frac{\| u + v \| - \| u - v \|}{4} + i \, \frac{\| i \, u + v \| - \| i \, u - v \|}{4}.$$

IPS3 The ordinary **dot product** on $\mathbb{R}^3$ is the inner product everyone knows about. Here $\langle u , v \rangle = u \cdot v$.

It is important to note that the properties of dot product not only let us define **length of a vector**, as

$$\|v\| = \sqrt{v \cdot v}$$

but also let us define **angle between two vectors** by

$$u \cdot v = \|u\|\|v\| \cos(\theta).$$

Angle is defined in an analogous way for *any* inner product: it is given by the equation

$$\langle\, u\,,\, v\,\rangle = \|u\|\|v\| \cos(\theta)$$

and the Schwartz inequality guarantees that θ can be defined by this equation. That it behaves as an angle should, that it corresponds to our intuition provided by our experience with Euclidean geometry, is a different matter. You can satisfy yourself about this by looking at the two dimensional space $Span(\{\, u, v\,\})$ and the lengths of the edges of the triangle in this space with corners at 0, u and v.

The notion of angle is the additional element the inner product supplies beyond the homogeneous and translation invariant "length" idea given by the norm: in particular, a notion of perpendicularity.

Two vectors u and v are said to be **orthogonal** if $\langle\, u\,,\, v\,\rangle = 0$. If both vectors are nonzero and they are orthogonal we sometimes call them **perpendicular.** A set of vectors is called **orthonormal** if each vector in the set has norm 1 and any two vectors from the set are orthogonal.

IPS4 In an inner product space, the **Gram-Schmidt orthonormalization procedure** can be applied to any nonempty set S of vectors to create a maximal orthonormal set M of vectors in $Span(S)$. However, $Span(M)$ will not in general equal $Span(S)$.

19. Hilbert Spaces

HS1 A **Hilbert Space** is a complete inner product space, and so is a special type of Banach space. The square summable sequence space $\boldsymbol{\ell}^2$ forms an example.

HS2 Suppose $\boldsymbol{\Gamma}$ is a Lebesgue integral for a positive σ-finite measure with square integrable functions $\boldsymbol{\mathcal{L}_2}$ and quotient space $\boldsymbol{L^2} = \boldsymbol{\mathcal{L}_2}/\boldsymbol{\mathfrak{N}}$.

Define $\langle f, g\rangle = \Gamma(fg)$ for f, g in $\boldsymbol{\mathcal{L}_2}$.

Hölder's inequality reduces to the BCS inequality

$$|\langle f, g\rangle| \leq \|\, f\,\|_2\, \|\, g\,\|_2.$$

The function $\langle\,\cdot\,,\cdot\,\rangle$ satisfies all the conditions for an inner product on $\boldsymbol{\mathcal{L}_2}$ except the positivity condition: it is possible for $\langle\, u\,, u\,\rangle$ to be 0 for $u \neq 0$.

Define $\langle \tilde{f}, \tilde{g}\rangle = \Gamma(fg)$ for $\tilde{f} = f + \boldsymbol{\mathfrak{N}}$ and $\tilde{g} = g + \boldsymbol{\mathfrak{N}}$ in $\boldsymbol{L^2}$.

The function $\langle\,\cdot\,,\cdot\,\rangle$, defined for pairs of functions in $\boldsymbol{L^2}$, *is* an inner product on complete $\boldsymbol{L^2}$, which is therefore a Hilbert space. The norm associated with the inner product is just $\|\,\cdot\,\|_2$.

The functions in $\boldsymbol{\mathcal{L}_2}$ are real valued, and we often want to work with complex valued functions. By separating a complex measure into its real and imaginary parts[5] and a complex function into its real and imaginary parts we define the complex square integrable functions and the quotient space $\boldsymbol{L}^{\mathbf{2}}_{\mathbb{C}}$ in the obvious way.

Defining $\langle f + \mathfrak{N}, g + \mathfrak{N} \rangle$ on $\boldsymbol{L}^{\mathbf{2}}_{\mathbb{C}}$ by

$$\langle f + \mathfrak{N}, g + \mathfrak{N} \rangle = \Gamma\left(\overline{f}\, g\right)$$

we have a complex inner product on complete $\boldsymbol{L}^{\mathbf{2}}_{\mathbb{C}}$, which is therefore a *complex* Hilbert space.

Going from $\boldsymbol{L}^{\mathbf{2}}$ with its inner product to $\boldsymbol{L}^{\mathbf{2}}_{\mathbb{C}}$ with its inner product is not a large step: we will go from $\boldsymbol{L}^{\mathbf{2}}$ with its functionals to the complexified space $\boldsymbol{L}^{\mathbf{2}}_{\mathbb{C}}$ with the associated complex functionals without much comment. It is easier to think in $\boldsymbol{L}^{\mathbf{2}}$ (fewer terms to deal with) but more useful for our applications to work in $\boldsymbol{L}^{\mathbf{2}}_{\mathbb{C}}$.

HS3 An orthonormal subset B of Hilbert space $\mathcal{H}$ is called a **Hilbert basis** if the span of B is dense in $\mathcal{H}$. In other words, you can get arbitrarily close to any member of $\mathcal{H}$ with a *finite* linear combination of members of B. This implies that for each individual $w \in \mathcal{H}$ there is a *countable* subset $b_0, b_1, \ldots$ of B and numbers $a^0, a^1, \ldots$ so that

$$w = \sum_{n=0}^{\infty} a^n b_n.$$

Every orthonormal subset of a Hilbert space can be extended to a Hilbert basis. And every Hilbert space has a Hilbert basis. This follows, as in the existence of Hamel basis, from the Axiom of Choice.

A Hilbert basis S can be characterized as a "maximal" orthonormal subset. By this we mean that it is orthonormal and is contained in no larger orthonormal subset. We reason as follows: Since S is a linearly independent set it can be fattened to a Hamel basis T. But every member of T not in S must be the limit of vectors in the span of S, and so *cannot be* orthogonal to all the members of S.

HS4 Every *continuous* linear transformation between two Hilbert spaces is determined by what it does to any Hilbert basis. So in describing the effect of continuous functions, a Hilbert basis can serve the same purpose in an infinite dimensional setting that a basis does in finite dimensions.

HS5 Suppose S is a subspace of a Hilbert space $\mathcal{H}$. We define $\mathbf{S}^{\perp}$ (pronounced "S perp") to be the set of all members of $\mathcal{H}$ which are orthogonal to S. This set is a vector subspace of $\mathcal{H}$ also, and is complete.

---

[5]Actually, we break a complex measure μ into four parts,

$$\mu = \mu_r + i\,\mu_i = \mu_r^+ - \mu_r^- + i\left(\mu_i^- - \mu_i^-\right)$$

where μ_r and μ_i are real σ-finite measures and $\mu_r^+, \mu_r^-, \mu_i^+$ and μ_i^- are four positive measures with at least one of the real pair finite, and at least one of the complex pair finite. Then we find the integrable functions for these four measures and define

$$\boldsymbol{\mathcal{L}}_{\mathbf{2}} = \boldsymbol{\mathcal{L}}_{\mathbf{2}}(\mu_r^+) \cap \boldsymbol{\mathcal{L}}_{\mathbf{2}}(\mu_r^-) \cap \boldsymbol{\mathcal{L}}_{\mathbf{2}}(\mu_i^+) \cap \boldsymbol{\mathcal{L}}_{\mathbf{2}}(\mu_i^-).$$

When subspace T is contained in $S^{\perp}$ we say S and T are **orthogonal.**

If S is itself complete then $(S^{\perp})^{\perp} = S$ and every member v of $\mathcal{H}$ can be written in a unique way as $v = a + b$ where a is in S and b is in $S^{\perp}$.

So when S is a closed subspace $\mathcal{H}$ is a direct sum $\mathcal{H} = S \oplus S^{\perp}$.

Also, $\mathcal{H}$ is isometric and isomorphic to $S \times S^{\perp}$, where $S \times S^{\perp}$ is given inner product $\langle \cdot, \cdot \rangle_{\text{prod}}$

$\langle (u, w), (a, b) \rangle_{\text{prod}} = \langle u, a \rangle + \langle w, b \rangle \quad$ for $u, a \in S$ and $w, b \in S^{\perp}$.

HS6 If B is a Hilbert basis and F is a continuous functional then for each $\varepsilon > 0$ the function values $F(s)$ can exceed ε for only finitely many s. This means $F(s)$ is nonzero for only countably many members $s \in B$.

Let $s_0, s_1 \dots$ be a listing of these members of orthonormal B and let

$$S = \overline{Span\,(\{s_0, s_1, \dots\})}.$$

We saw above that any member v of $\mathcal{H}$ can be written in a unique way as $v = a + b$ where a is in S and b is in $S^{\perp}$.

By continuity we have $F(b) = 0$ and $F(v) = F(a)$. So let's restrict our attention to S and the effect of F on this closed subspace.

Suppose $a = \sum_{n=0}^{\infty} a^n s_n$. So

$$F(a) = \sum_{n=0}^{k} a^n F(s_n) + F\left(\sum_{n=k+1}^{\infty} a^n F(s_n)\right).$$

The magnitude of the vector inside F on the right converges to 0 so continuity of F requires that the sequence of partial sums to its left converges and

$$F(a) = \sum_{n=0}^{\infty} a^n F(s_n).$$

Prompted by this, we examine the sequence of partial sums

$$\sum_{n=0}^{k} F(s_n) s_n.$$

A version of the **Riesz Representation Theorem** implies that this sum converges, and the vector

$$w = \sum_{n=0}^{\infty} \overline{F(s_n)} s_n$$

can be used to represent the continuous linear functional F in

$$F(v) = \langle w\,,\, v \rangle.$$

The vector w is unique.

HS7 This theorem allows (even encourages) one to conflate the vector w with corresponding continuous functional F. The identification $\mathfrak{R}$ (short for $\boldsymbol{\mathfrak{R}eisz}$) taking F to $\mathfrak{R}(F) = w$ is not *quite* linear if $\mathbb{F} = \mathbb{C}$:

$$\mathfrak{R}(aF + J) = \overline{a}\,\mathfrak{R}(F) + \mathfrak{R}(J)$$

and we refer to this property as **conjugate linearity** in the complex case.

If $w = 0$ then F is the zero functional and the operator norm $\|F\|$ is zero too. On the other hand, if F is not the zero functional, then $|F(v)|$ attains its maximum among $v \in S_1(0)$ for $v = w/\|w\|_2$ and we find $\|F\| = \|w\|_2$.

So the map

$$\mathfrak{R}\colon \mathfrak{H}' \to \mathfrak{H} \quad \text{is a conjugate linear isometry.}$$

Because of this, it is customary to "forget" the distinction between F and w when $\mathfrak{R}(F) = w$.

Do remember, however, that the functional aF is associated with $\overline{a}w$ and not aw for scalar a.

We will call the inverse map $\mathfrak{Z}$. (Short for . . . well, guess. There will be a quiz.)

$$\mathfrak{Z}\colon \mathfrak{H} \to \mathfrak{H}' \quad \text{is also a conjugate linear isometry.}$$

HS8 Physicists use both $\mathfrak{R}$ and $\mathfrak{Z}$ to maneuver around during many of their favorite quantum calculations.

The symbols $\langle\,\cdot\,|$ and $|\,\cdot\,\rangle$, pronounced "bra" and "ket" respectively. They are used in a variety of ways, but at root, usages are all related somehow to

$$\langle\, f\,| = \mathfrak{Z}(f) \quad \text{and} \quad |\, f\,\rangle = f.$$

So for $f \in \mathfrak{H}$, $\langle\, f\,|$ is the functional which can be calculated using inner product as $\langle\, f,\,\cdot\,\rangle$.

It follows then that

$$\langle\, f\,| = \mathfrak{Z}\,(\,|\, f\,\rangle\,) \ \text{ and } \ |\, f\,\rangle = \mathfrak{R}\,(\,\langle\, f\,|\,)\,.$$

Rather than use a comma to separate entries in an inner product these folks use a vertical line, as $\langle\, f,\, h\,\rangle = \langle\, f\,|\, h\,\rangle$.

And then $\langle\, f\,|\,|\, h\,\rangle = \langle\, f\,|(h) = \langle\, f\,|\, h\,\rangle = \langle\, f,\, h\,\rangle$.

Being practical folk with business to attend to this user group finds it convenient to create several variations on this definition for special cases, particularly eigenfunctions for operators on $\mathfrak{H}$, and we will attend to this in later sections.

HS9 A Hilbert space has a Hilbert basis that can be arranged in a sequence (that is, the basis is countable) exactly when it is **separable.**

A listing of a countable Hilbert basis is an unconditional and bounded Schauder basis and we will usually use the notation developed for Schauder bases whenever we have a countable Hilbert basis.

Any separable Hilbert space is isomorphic to $\boldsymbol{\ell}^2$, and the isomorphism is an isometry. In other words, if you understand $\boldsymbol{\ell}^2$ you understand any separable Hilbert space.

We remark that, though $\boldsymbol{\ell}^2$ has a countable Hilbert basis, any Hamel basis for $\boldsymbol{\ell}^2$ is uncountable.

Separability is a technical convenience and will usually be assumed. As we saw above, functionals are zero off the closure of the span of a countable Hilbert basis for a subspace. Vectors have zero coordinates off the closure of the span of a countable Hilbert basis for a subspace. Hilbert spaces can be broken into the direct sum of various orthogonal Hilbert subspaces. Often one of these will be separable and all the action will take place there. So most theorems that are true for separable spaces are also true more generally.

But proofs of the main theorems are easier to organize assuming a countable Hilbert basis, and the most common Hilbert spaces *are* separable anyway. In some older texts, separability was assumed as part of the definition of a Hilbert space.

HS10 If $\mathcal{H}$ is a separable Hilbert space with Hilbert basis $\mathbf{b}$ and $v \in \mathcal{H}$

$$v = \sum_{i=0}^{\infty} \langle b_i, v \rangle \, b_i.$$

The numbers $\langle b_i, v \rangle$ can be referred to as the **Fourier coefficients** of v with respect to this Hilbert basis.

The sequence $(\langle b_i, \cdot \, \rangle)$ is the sequence of coordinate functionals for the Schauder basis $\mathbf{b}$.

HS11 If $\mathcal{H}$ is any Hilbert space and $b_0, b_1 \dots$ is any orthonormal sequence (not necessarily a Hilbert basis) then for each $v \in \mathcal{H}$

$$\sum_{i=0}^{\infty} |\langle b_i, v \rangle|^2 \leq \|v\|^2.$$

This is known as **Bessel's inequality.**

But if $\mathcal{H}$ is a separable Hilbert space with Hilbert basis $\mathbf{b}$ and v and w are in $\mathcal{H}$ we have more:

$$\langle v, w \rangle = \sum_{i=0}^{\infty} \langle v, b_i \rangle \, \langle b_i, w \rangle$$

This is called **Parseval's identity**.

20. Weaker Notions of Convergence

In a Banach space B (such as any Hilbert space) we have strong convergence of sequences $x_0, x_2, \dots$ to vector y when

$$\|y - x_n\| \to 0 \text{ and we write } x_i \underset{\text{strong}}{\longrightarrow} y$$

in that case.

But if, for *every* functional $\phi \in B'$ we have

$$\phi(x_i) \to \phi(y)$$

we have another notion of convergence called **weak convergence**. The sequence is converging to y *as far as any continuous functional can detect.* We write

$$x_i \underset{\text{weak}}{\longrightarrow} y$$

for this kind of convergence.

Weak convergence is in fact "weak" in the colloquial sense. For instance if $\mathbf{b}$ is a Hilbert basis for Hilbert space B then $b_i \underset{\text{weak}}{\longrightarrow} 0$. Not a very satisfying outcome.

However the following important addition to the condition improves matters for any sequence $v_0, v_1, \ldots$ in Banach B.

$$\text{If} \quad v_i \underset{\text{weak}}{\longrightarrow} w \quad \text{and} \quad \|v_i\| \to \|w\| \quad \text{then} \quad v_i \underset{\text{strong}}{\longrightarrow} w.$$

If B is a Hilbert space then functionals can be given as inner product against members of B, so the weak convergence criteria can be rewritten as

$$\langle v, x_i \rangle \to \langle v, y \rangle \ \forall v \in B.$$

Since B' is Banach with operator norm, there is the usual concept of convergence of sequences of functionals using this norm, strong convergence.

$$\phi_n \underset{\text{strong}}{\longrightarrow} \mu \quad \text{exactly when}$$

$$\|\mu - \phi_n\|_{op} = \sup\{\, |\mu(x) - \phi_n(x)| \mid x \in S_1(0)\,\} \to 0.$$

Weak* convergence (read "weak star convergence") checks the sequence of functionals on members of B one at a time, rather than all together.

$$\phi_n \underset{\text{weak}^*}{\longrightarrow} \mu \quad \text{exactly when} \quad |\mu(x) - \phi_n(x)| \to 0 \quad \text{for every } x \in B.$$

We note that testing against only those $x \in S_1(x)$ produces an identical condition. Weak* convergence is just pointwise convergence of the sequence of functionals.

Now if the evaluation map $\widetilde{E} \colon B \to B''$ is onto (i.e. B is reflexive, as is every Hilbert space) then every continuous functional on B' is given by evaluation at some $x \in B$ so weak* convergence and weak convergence are the same. But if B is not reflexive weak* convergence tests against fewer functionals so it is even weaker than weak convergence.

We consider an interesting example involving weaker convergence.

Suppose Γ is an integral for a *finite* measure.

Recall that the continuous dual of $\boldsymbol{L^1}$ is $\boldsymbol{L^\infty}$, and suppose $g_n \underset{\text{weak}^*}{\longrightarrow} f$ in $\boldsymbol{L^\infty}$.

That means

$$\Gamma(g_n h) \to \Gamma(fh) \text{ for every } h \in \boldsymbol{L^1}.$$

But for any p and q with $1 < q \le p < \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$

$$\boldsymbol{L^\infty} \subset \boldsymbol{L^p} \subset \boldsymbol{L^q} \subset \boldsymbol{L^1}.$$

So $\Gamma(g_n h) \to \Gamma(fh)$ for every $h \in \boldsymbol{L^p}$. Also, $\boldsymbol{L^p}$ is the continuous dual of $\boldsymbol{L^q}$.

We conclude that $g_n \underset{\text{weak}^*}{\longrightarrow} f$ in $\boldsymbol{L^\infty}$ implies $g_n \underset{\text{weak}}{\longrightarrow} f$ in $\boldsymbol{L^q}$.

21. The Hermitian Adjoint

If $G\colon \mathcal{H} \to \mathcal{H}$ is any continuous operator on a Hilbert space, an operator $G^\dagger$ called the **Hermitian adjoint** of G can be defined.

For each z in $\mathcal{H}$ the function $F_z(\cdot) = \langle\, z\,,\, G(\cdot)\,\rangle$ is a member of the continuous dual. So there is a unique vector w with

$$\langle\, z\,,\, G(\cdot)\,\rangle = \langle\, w\,,\, \cdot\,\rangle.$$

Define $\mathbf{G}^\dagger$ on the vector z by $G^\dagger(z) = w$. That is, $G^\dagger(z)$ is defined to be the unique vector for which

$$\langle\, z\,,\, G(v)\,\rangle = \langle\, G^\dagger(z)\,,\, v\,\rangle \text{ for every } v \text{ in } \mathcal{H}.$$

There is an easy relationship between the generic Banach adjoint G^* and the Hermitian adjoint $G^\dagger$.

Recall that $G^*\colon \mathcal{H}' \to \mathcal{H}'$ is defined by $G^*(\phi) = \phi \circ G$ for each $\phi \in \mathcal{H}'$. But here each functional ϕ corresponds to $\mathfrak{R}(z)$ for a unique $z \in \mathcal{H}$ so we could define G^* just as well by

$$G^*\left(\mathfrak{R}(z)\right) = \left(\mathfrak{R}(z)\right) \circ G.$$

But we define $G^\dagger$ by

$$\mathfrak{R}\left(G^\dagger(z)\right) = \left(\mathfrak{R}(z)\right) \circ G.$$

So we find that G^* and $G^\dagger$ are *almost* the same:

$$G^\dagger = \mathfrak{Z} \circ G^* \circ \mathfrak{R}.$$

So $G^\dagger$ is the composition of isometries with an operator whose operator norm is $\|G\|$. The two isometries are conjugate linear, so the result is linear. So $G^\dagger$ is continuous operator with operator norm $\|G\|$.

A calculation shows that $G^{\dagger\dagger} = G$.

An operator G on $\mathcal{H}$ is called a **self-adjoint** or **Hermitian** whenever

$$G = G^\dagger.$$

This has important consequences. For instance if $v \in \mathcal{H}$ is an eigenvector of Hermitian G, so that $G(v) = k\,v$ for some number (the eigenvalue) k then

$$\begin{aligned} k\,\|v\|^2 =& \langle\, v\,,\, k\,v\,\rangle = \langle\, v\,,\, G(v)\,\rangle = \langle\, G^\dagger(v)\,,\, v\,\rangle \\ =& \langle\, G(v)\,,\, v\,\rangle = \langle\, k\,v\,,\, v\,\rangle = \overline{k}\,\|v\|^2. \end{aligned}$$

So k is real: **Hermitian operators cannot have complex eigenvalues.**

Suppose G is a Hermitian operator and u and v are eigenvectors for G for eigenvalues s and t, respectively. We saw above that s and t are real. So

$$s\,\langle\, u\,,\, v\,\rangle = \langle\, s\,u\,,\, v\,\rangle = \langle\, G(u)\,,\, v\,\rangle = \langle\, u\,,\, G(v)\,\rangle = \langle\, u\,,\, t\,v\,\rangle = t\,\langle\, u\,,\, v\,\rangle.$$

Conclusion: $t = s$ or $\langle\, u\,,\, v\,\rangle = 0$. So **eigenvectors for different eigenvalues for a Hermitian operator are orthogonal.**

Self-adjoint operators have many properties of interest.

For instance, any continuous operator G from Hilbert $\mathcal{H}$ to itself is called **positive** if $\langle G(v), v \rangle$ is a non-negative real number for every $v \in \mathcal{H}$. It is called **positive definite** if $\langle G(v), v \rangle$ is a positive real number for every nonzero $v \in \mathcal{H}$.

The notations $\mathbf{G} \geq \mathbf{0}$ and $\mathbf{G} > \mathbf{0}$ are used to describe these situations.

It is a fact that positive operators are self-adjoint.

This allows us to create a useful ordering on some pairs of operators. We write

$$F \geq G \text{ when } F - G \geq 0 \quad \text{and} \quad F > G \text{ when } F - G > 0.$$

Positive operators are featured in quantum mechanics.

Index