Review of Essential Mathematics

- Vector spaces
- Linear independence
- Bases and dimensions
- Norms
- Inner products
- Hilbert space
- Decomposition theorem

Review of Essential Mathematics I.1.1

- Algebra Terms
- Complex numbers
- Rules and Exponents
- Radicals
- Linear Equations
- [illegible]
- Pythagorean Theorem

Brief review of essential mathematics

The proper mathematical framework for geophysical inverse theory requires the tools of linear algebra. We therefore need to refresh ourselves about the basic properties of linear vector spaces, norms, inner products..

We begin with the concept of a vector space. (Both model space and data space will be vector spaces)

Real Linear Vector Space.

A real linear vector space is a set V containing elements that can be related by two operations; addition and scalar multiplication. The operations are written as

$$f + g \qquad \text{and} \qquad \alpha f$$

where $f, g \in V$ and $\alpha \in \mathbb{R}$. For any (real) scalars α and β, the following sets of nine relations must hold

$$f + g \in V \tag{V1}$$
$$\alpha f \in V \tag{V2}$$
$$f + g = g + f \tag{V3}$$
$$f + (g+h) = (f+g) + h \tag{V4}$$
$$f + g = f + h \quad \text{if and only if} \quad g = h \tag{V5}$$
$$\alpha(f+g) = \alpha f + \alpha g \tag{V6}$$
$$(\alpha+\beta) f = \alpha f + \beta f \tag{V7}$$
$$\alpha(\beta f) = (\alpha\beta) f \tag{V8}$$
$$1 f = f \tag{V9}$$

A notation $-f$ means -1 times f. From these axioms it follows that every vector space contains a unique zero element

$$f + 0 = f \qquad f \in V$$

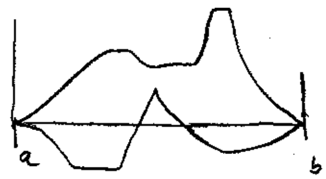
and whenever $\alpha f = 0$ we must have $\alpha = 0$ or $f = 0$.

Any collection of elements which satisfy the above properties then form a real linear vector space. Usually we will refer to this as a vector space.

Examples of real linear vector spaces.

(1) $\mathbb{R}^N$: ordered sequence of N numbers $(x_1, x_2 \cdots x_N)$
These are the familiar 'vectors' which use every day
Often we work in 3 dimensions so $\mathbb{R}^3$
contains vectors $\bar{x} = (x_1, x_2, x_3)$

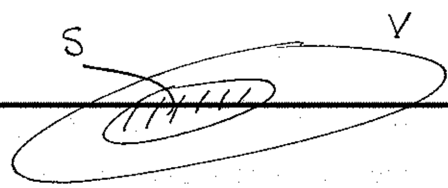
(2) $C^0[a,b]$: space of continuous functions defined on the interval $[a,b]$. This is a very large space.



(3) $C^n[a,b]$: space of functions which possess continuous derivatives up to and including order n.

Subspaces, Linear Combinations and Linear Independence.

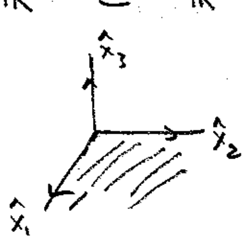
A linear subspace for V is a subset of V that forms a linear vector space under the rules of scalar multiplication and addition for V.



$S \in V$

eg. $C^m[a,b] \in C^n[a,b]$ for $m \geq n$

or $\mathbb{R}^M \in \mathbb{R}^N$ for $M \leq N$



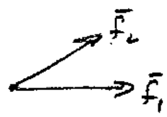
$\mathbb{R}^2$ is a subset of $\mathbb{R}^3$

Linear combination: of elements $f_1, f_2, \cdots f_n$ is a vector of the form

$$f = \alpha_1 f_1 + \alpha_2 f_2 + \cdots + \alpha_n f_n$$

The set of all linear combinations formed from a fixed collection of elements is a subspace of the original space; the fixed elements are said to span the subspace.

eg.



All vectors in a 2-D plane can be written as linear combinations of the vectors shown at the left. These two vectors therefore span a 2-D subspace of $\mathbb{R}^3$

Linear dependence

The elements $f_1, f_2, \cdots f_n$ are linearly dependent if there exists scalars $\alpha_1, \alpha_2 \ldots \alpha_n$ such that

$$\alpha_1 f_1 + \alpha_2 f_2 + \cdots \alpha_n f_n = 0 \qquad (*)$$

when not all α's are equal to zero. If this is so then effectively one (or more) of the elements can be written in terms of the others.

If the only way that (*) can be satisfied is to have all $\alpha_i = 0$ then the elements $\{f_i\}$ are linearly independent.

If the elements $\{f_i\}$ are linearly independent and if

$$\sum_i^n \alpha_i f_i = \sum_i^n \beta_i f_i$$

then $\alpha_1 = \beta_1$; $\alpha_2 = \beta_2$; $\cdots$ $\alpha_n = \beta_n$. That is, the coefficients are unique in the expansion of an element in terms of a linear combination of linearly independent elements.

Bases and Dimension.

A basis for a vector space V is any set of linearly independent elements which span the space. (An analogous statement holds if V is thought of as a subspace of larger vector space)

The dimension of V is equal to the number of basis elements that are required to span V.

eg: A usual basis for $\mathbb{R}^N$ is

$$e_1 = (1, 0, \dots 0)$$
$$e_2 = (0, 1, 0 \dots 0)$$
$$\vdots$$
$$e_N = (0, 0, \dots, 0, N)$$

The linear independence is immediate. They also span the space since any vector $\bar{x} = (x_1, x_2, \dots x_N)$ can be written as

$$\bar{x} = \sum_{i=1}^{N} \alpha_i e_i$$

where $\alpha_i = x_i$. $(i = 1, N)$.

A basis for $C^0[0, a]$ is

$$\psi_m(x) = \sin \frac{m\pi x}{a}$$

Any continuous function on $[0, a]$ can be written as

$$f(x) = \sum_{m=1}^{\infty} c_m \sin \frac{m\pi x}{a}$$

Note that an infinite number of basis f^{ns} are (possibly) required to represent an arbitrary $f(x)$. The dimension of $C^0[0, a]$ is infinity. We shall see that this fact has important implications for the nonuniqueness inherent in the solution of inverse problems where a function is to be found.

Norms

Our vector space is a collection of elements. We need a way of distinguishing between them. One property to quantify is the size or length of the vector. Our ruler, which quantifies this length is called a norm.

The concept of a norm is vital to the solution to the inverse problem. Given the nonuniqueness that is inherent in the inverse problem, we will design our algorithms to pick out the "smallest length" model. The quantification of smallest is in terms of the norm.

If we equip our vector space with a norm then we have a normed vector space.

A normed vector space is a linear vector space in which every element f has a norm written as $\|f\| \in \mathbb{R}$. A norm is a real valued functional satisfying the following conditions: for every $f, g, h \in V$ and $\alpha \in \mathbb{R}$

$$\|f\| \geq 0 \qquad (N1)$$

$$\|\alpha f\| = |\alpha| \, \|f\| \qquad (N2)$$

$$\|f+g\| \leq \|f\| + \|g\| \qquad (N3)$$

$$\|h\| = 0 \quad \text{iff} \quad h = 0 \qquad (N4)$$

If the last condition is omitted, the functional is called a seminorm.

Common Norms for $\mathbb{R}^N$

$$\|x\|_1 = \sum_{i=1}^{N} |x_i| \qquad \text{1-norm.}$$

$$\|x\|_2 = \Big(\sum_{i=1}^{N} x_i^2\Big)^{1/2} \qquad \text{Euclidean norm}$$

$$\|x\|_\infty = \max_i |x_i| \qquad \infty\text{-norm};$$

$$\|x\|_p = \Big(\sum_{i=1}^{N} |x_i|^p\Big)^{1/p} \qquad 1 \leq p < \infty \quad \text{(general p-norm)}$$

Remark: A vector space on which the above norms are defined is l_p (or l_p^N ?)

Norms for Infinite dimensional subspaces.

Consider the vector space $C^0[a,b]$. (space of continuous f^{cns} on $[a,b]$).

$$\|f\|_1 = \int_a^b |f(x)|\,dx \qquad \text{1-norm}$$

$$\|f\|_2 = \left(\int_a^b |f(x)|^2\,dx\right)^{1/2}$$

$$\|f\|_\infty = \max_{a\le x\le b} |f(x)| \qquad \text{uniform norm.}$$

$$\|f\|_p = \left(\int_a^b |f(x)|^p\,dx\right)^{1/p}$$

Remark: In many of our inverse problems we will be concerned about wiggliness or roughness of the model. We want to find a function which has a small amount of structure (if possible) so that interpretation of features seen on the constructed model are more straightforward.

Norms which are affected by roughness might be

$$\|f\|_s = \left(\int_a^b \left(w_0(x)f(x)^2 + w_1(x)f'(x)^2\right)dx\right)^{1/2} \qquad f\in C^1[a,b]$$

$$\|f\|'' = \left[f(a)^2 + f'(a)^2 + \int_a^b f''(x)\,dx\right]^{1/2} \qquad f\in C^2[a,b]$$

$$\|f\|_I = \left(\int_a^b \left\{\int_a^x f(t)\,dt\right\}^2 dx\right)^{1/2}$$

The first norm is an example of a Sobelev norm. The second norm essentially measures curvature. The last norm is very insensitive to high frequency behavior (ie you can add a great deal in without penalty).

Remark: The introduction of a norm on a vector space immediately allows one to compute the "distance" between two objects.

The distance between two elements f and g is given by

$$d(f,g) = \| f - g \|$$

This concept is important since in data space we want our predicted data to be "close" to the observations. In model space we may also want our constructed model to be close to a particular predetermined model.

Inner Product

The norm of a vector is a mathematical abstraction of length, the inner product is a generalization of a dot product. This allows us to compute an "angle" between two elements of the vector space and leads to the concept of orthogonality.

A real inner product space, I, (sometimes referred to as a pre-Hilbert space) is a real linear vector space in which there is defined for every pair of elements f, g a functional, the inner product (f,g) with the following properties. If $f, g, h \in I$ and $\alpha \in \mathbb{R}$

$$(f,g) = (g,f) \qquad (I1)$$
$$(f+g, h) = (f,h) + (g,h) \qquad (I2)$$
$$(\alpha f, g) = \alpha (f,g) \qquad (I3)$$
$$(f,f) > 0 \quad \text{if } f \neq 0 \qquad (I4)$$

Remark: If the elements of the space are complex, then the definition for the inner product needs to be altered as $(f,g) = (g,f)^*$. Complex inner product spaces are called unitary spaces.

★ Remark: An inner product space comes equipped with a norm. Define

$$\| f \| = (f,f)^{1/2} \qquad (I5)$$

Ass One can show that (I5) satisfies all of the properties of a norm.

Examples of inner product spaces. $E^N \equiv (\mathbb{R}^N + \text{inner product})$

Elements have N-tuples of numbers

$$(x, y) = x \circ y = \sum_{i=1}^{N} x_i y_i$$

then $\|x\| = (x, x)^{1/2} = \left(\sum_{i=1}^{N} x_i^2 \right)^{1/2}$ is the usual Euclidean length.

Remark: We will often want to have a generalization of this. Let A be a positive definite and symmetric matrix.

A symmetric $\Rightarrow$ $A = A^T$ (so is square)

A positive definite. For any vector $x \in E^N$ then $x \cdot Ax > 0$

We write the generalized inner product as

$$\boxed{(x, y)_A = x \circ (Ay)}$$

or in full matrix notation we write

$$\boxed{(x, y)_A = x^T A y} \qquad (2)$$

Ass Remark: Might convince yourselves that (2) satisfies all of the requirements for an inner product.

Inner product on function space.

Consider the space $C^0[a,b]$ on which we define an inner product

$$\boxed{(f,g) = \int_a^b f(x)\, g(x)\, dx}$$

The norm induced by this inner product is

$$\boxed{\|f\| = (f,f)^{1/2} = \left(\int_a^b f(x)^2\, dx\right)^{1/2}}$$

There is one generalization that we will wish to make for some problems. Consider a positive function $w(x)$

$$w(x) > 0 \quad [a,b]$$

Define

$$\boxed{(f,g)_W = \int_a^b w(x)\, f(x)\, g(x)\, dx}$$

and hence the norm $\|f\| = (f,f)_W^{1/2} = \left(\int_a^b w(x)\, f(x)^2\, dx\right)^{1/2}$

Remark: Because the introduction of a norm and associated metric means that we have more tools to work with in our vector spaces, it is usual to rename the vector space to remind us that those tools are defined. The vector space defined with the above inner product is called $L_2[a,b]$. (Actually it is $CL_2[a,b]$ and then when completed it becomes the Hilbert space $L_2[a,b]$. We won't concern ourselves with the concepts and computations required to ensure that our space is complete)

Hilbert Space

A Hilbert space $\mathcal{H}$ is an inner product space that is complete under the norm $(f,f)^{1/2}$. The two Hilbert spaces we shall work most with are:

E^N : elements $\vec{x} = (x_1, x_2 \cdot x_N)$ $\qquad (x,y) = x^T y$

$L_2[a,b]$: $\qquad (f,g) = \int_a^b f(x)g(x)\,dx$

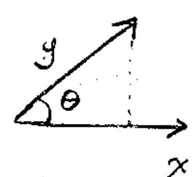
On a Hilbert space we can introduce the concept of orthogonality. Two elements of a Hilbert space are said to be orthogonal if their inner product is zero

L_2 : $\qquad (f,g) = 0 \Rightarrow f \perp g$

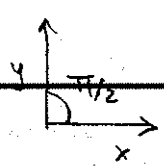
E^N : $\qquad (x,y) = 0 \Rightarrow x \perp y$

The concept of orthogonality is well ingrained in us for ordinary vectors. If x, y are two N-length Euclidean vectors

$$(x,y) = \|x\| \|y\| \cos\theta$$



so inner product multiplies $\|x\|$ by the projection of y onto the vector x. when $\theta = \pi/2$



There is no projection of y onto x (and vice versa) and the inner product is zero, and the functions are perpendicular.

The concept of orthogonality of functions is somewhat more difficult to visualize but is nevertheless something we are familiar with.

Consider an expansion of $f(x)$ on an interval $[a,b]$

eg. $$f(x) = \sum_{m=1}^{\infty} c_m \sin\left(\frac{m\pi x}{a}\right) \equiv \sum_{m=1}^{\infty} c_m \psi_m(x)$$

where we write the m^{th} basis function as

$$\psi_m(x) = \sin\left(\frac{m\pi x}{a}\right)$$

The basis functions are orthogonal to each other

$$(\psi_m, \psi_n) = \int_0^a \sin\left(\frac{n\pi x}{a}\right) \sin\left(\frac{m\pi x}{a}\right) dx = \begin{cases} 0 & m \neq n \\ b_n & m = n \end{cases}$$

Remark: If the basis functions are defined so $b_n = 1$ then the functions are <u>orthonormal</u>.

The expansion of one function in terms of another set of function is greatly facilitated by orthogonality of the basis f^{cn}. For example, suppose we need to expand $f(x)$ in a basis set where $(\psi_i, \psi_j) = b_n \delta_{ij}$ (δ_{ij} is Kronecker delta)

$$\boxed{f(x) = \sum_i^{\infty} c_i \psi_i(x)}$$

$(\;, \psi_j)$

$$(f, \psi_j) = \left(\sum_i^{\infty} c_i \psi_i(x), \psi_j(x) \right)$$

$$= \sum_{i=1}^{\infty} c_i (\psi_i, \psi_j) = c_j b_j \qquad j = 1, \cdots$$

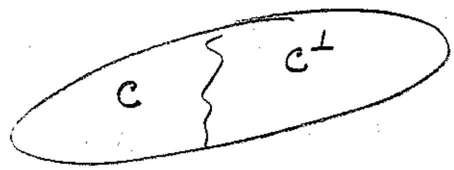
So the coefficients are $$\boxed{c_i = \frac{1}{b_i}(f, \psi_i)}$$

ie the coefficients are the projection of the function being expanded onto the i^{th} basis f^{cn}. (then scaled by b_i)

Decomposition Theorem.

With the foregoing work it is now possible to introduce the decomposition theorem for Hilbert spaces.

Let C be a (complete) subspace of H.



The orthogonal complement of C is called $C^\perp$. Every element in $C^\perp$ is orthogonal to every element in C.

ie

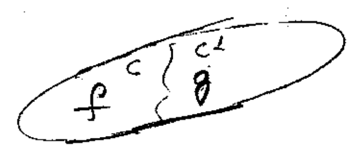
$$f \in C \quad \text{and} \quad g \in C^\perp \quad \text{then} \quad (f, g) = 0$$

Decomposition theorem:

Any element $h \in H$ can be written as a sum of a part in C and a part in $C^\perp$

$$h = f + g \qquad f \in C, \ g \in C^\perp$$

and the decomposition is unique.



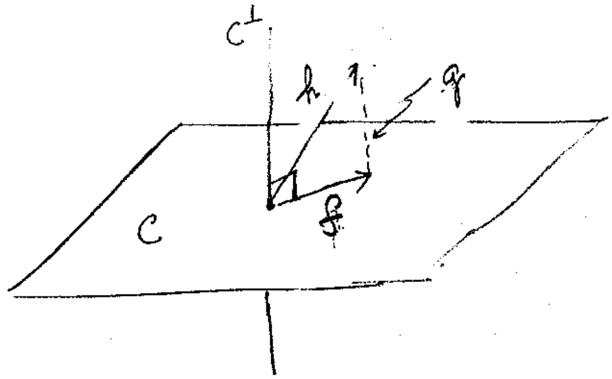
The element f is said to be the orthogonal projection of h onto the subspace C.

It is helpful to have a name for the function that maps an arbitrary member of the Hilbert space into its unique orthogonal projection in a fixed closed subspace.

The function is called the orthogonal projection operator.

Written as P_c

$$P_c : H \to C$$
$$f = P_c h$$



The diagram illustrates the concept and as well brings to light an important concept.

Question: Of all of the elements in the subspace C, which is the one that lies closest to h

Answer: There is a unique element that minimizes the distance and it is given by the orthogonal projection of h onto C. ie the unique element is $P_C h$

Remark: We now have all the tools necessary to formulate the solution to a linear inverse problem. But without going into explicit details lets take a quick overview.

(i) the data $\quad d_j = (g_j, m) \quad j = 1, N$

The j^{th} datum is equal to the projection of the model onto the j^{th} basis vector g_j

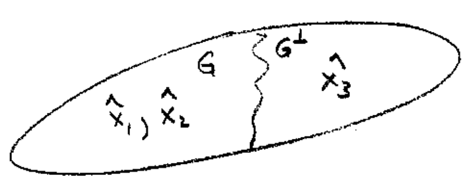
Suppose we were in E^3. $\quad \bar{m} = (2, 3, 4)$ model.
Have a basis

$$\hat{x}_1 = (1, 0, 0)$$
$$\hat{x}_2 = (0, 1, 0)$$
$$\hat{x}_3 = (0, 0, 1)$$

Experiment consists of shining a light in $-\hat{z}$ direction. This casts a shadow onto the $\hat{x}_1 - \hat{x}_2$ plane. Experimentalist measures

$$d_1 = (\hat{x}_1, m) \quad \text{that is} \quad d_1 = 2$$
$$d_2 = (\hat{x}_2, m) \quad \text{that is} \quad d_2 = 3$$

Activated bases $\quad \hat{x}_1, \hat{x}_2$



Question: What do we know about $\vec{m}$?

Clearly m_1 and m_2 are known precisely but nothing is known about m_3.

The experiment does not provide complete information about $\vec{m}$. Any model that fits the data can be written as:

$\vec{m} = (2, 3, \alpha)$ where α is any scalar.

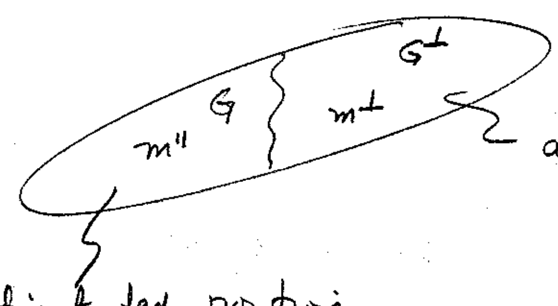
So what is the smallest model?

$$\|m\|^2 = \sum m_i^2 = 4 + 9 + \alpha^2$$
$$= 13 + \alpha^2$$

but α is arbitrary, thus $\|m\|$ is minimized by accepting that $\alpha = 0$. Equivalently, the constructed model is made up only of components of the activated space.

$$\boxed{\vec{m} = 2\hat{x}_1 + 3\hat{x}_2}$$

Model space is conveniently divided into two portions and any model can be written as $m = m^{\parallel} + m^{\perp}$



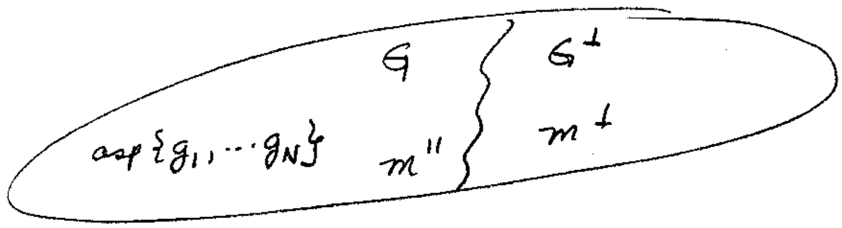
annihilator component
$(g_j, m^{\perp}) = 0$

activatated portion
$d_j = (g_j, m^{\parallel})$

The data supply total information about $m^{\parallel}$; ie $m^{\parallel}$ is known precisely. The data provide no information about $m^{\perp}$.

There is nothing in this reasoning which changes when the inner product pertains to a function space.

$$d_j = (g_j, m) \qquad j = 1, N$$



$$m = m^\parallel + m^\perp$$

The data supply direct information about $m^\parallel$ and no information about $m^\perp$.

Remark: The primary difference between this problem and the previous one using Euclidean vectors is that $G^\perp$ is infinite in dimension.

— / —

So we went from $$d_j = \int_a^b g_j(x)\, m(x)\, dx$$ Fredholm eq^n of 1st kind.

Recognize that this is a definition of an inner product. Work in a Hilbert space where $\|m\| = (m,m)^{1/2}$. Then with the decomposition th$^{\underline{m}}$

$$m = m^\parallel + m^\perp$$

Note: Our data could also be $\vec{d}_j = \vec{g}_j^{\,T} \vec{M} \qquad \in \mathbb{R}^N$
We note that this is also an inner product.

And, the smallest model $m^\parallel = \sum_{j=1}^{N} \alpha_j g_j$

$$f + g \quad \text{and} \quad \alpha f$$

where $f, g \in \mathrm{V}$ and $\alpha \in \mathbb{R}$. For any $f, g, h \in \mathrm{V}$ and any scalars α and β, the following set of nine relations must be valid:

$$f + g \in \mathrm{V} \tag{V1}$$

$$\alpha f \in \mathrm{V} \tag{V2}$$

$$f + g = g + f \tag{V3}$$

$$f + (g + h) = (f + g) + h \tag{V4}$$

$$f + g = f + h, \text{ if and only if } g = h \tag{V5}$$

$$\alpha(f + g) = \alpha f + \alpha g \tag{V6}$$

$$(\alpha + \beta)f = \alpha f + \beta f \tag{V7}$$

$$\alpha(\beta f) = (\alpha\beta)f \tag{V8}$$

$$1f = f \ . \tag{V9}$$

Table 1.01A: Some Linear Vector Spaces

Symbol	Description	Remarks
$\mathbb{R}^N$	The set of ordered real N-tuples $(x_1, x_2, \cdots x_N)$	The flagship of the finite-dimensional linear vector spaces
E^N	$\mathbb{R}^N$ equipped with any norm	–
$\mathrm{C}^n[a, b]$	The set of functions, continuously differentiable to order n on the real interval $[a, b]$	Not a normed space
$\mathrm{C}[a, b]$	$\mathrm{C}^0[a, b]$ equipped with the uniform norm $\lVert f \rVert_\infty = \max_{a \le x \le b} \lvert f(x) \rvert$	A Banach (complete normed) space
$\mathrm{CL}_1[a, b]$	$\mathrm{C}^0[a, b]$ equipped with the L_1-norm $\lVert f \rVert_1 = \int_a^b \lvert f(x) \rvert \, dx$	An incomplete normed space
$\mathrm{CL}_2[a, b]$	$\mathrm{C}^0[a, b]$ equipped with the L_2-norm $\lVert f \rVert_2 = [\int_a^b f(x)^2 dx\ ]^{1/2}$; implied inner product $(f, g) = \int_a^b f(x) g(x) dx$	An inner product or pre-Hilbert space; this is an incomplete space
$\mathrm{C}^n \mathrm{L}_2[a, b]$	$\mathrm{C}^n[a, b]$ equipped with a 2-norm that penalizes $d^n f / dx^n$	Another pre-Hilbert space
$\mathrm{L}_2[a, b]$	The completion of $\mathrm{CL}_2[a, b]$; each element is an equivalence class	The flagship Hilbert (complete inner product) space
l_2	The set of infinite ordered real sequences $(x_1, x_2, x_3 \cdots)$ normed by $\lVert x \rVert = [\sum_j x_j^2]^{1/2}$	Another Hilbert space
$\mathrm{W}_2^n[a, b]$	The completion of $\mathrm{C}^n \mathrm{L}_2[a, b]$	A Sobolev space (the norm acts on a derivative)