Math 510 **Fall 2022**
Notes for Homework 1 **Drew Armstrong**

Let $\mathbb{R}$ be the set of real numbers. A *vector space over* $\mathbb{R}$ consists of a set V (of "vectors"), with two algebraic operations, called *addition* and *scalar multiplication*:

$$\mathbf{u},\mathbf{v}\in V \rightsquigarrow \mathbf{u}+\mathbf{v}\in V$$
$$a\in\mathbb{R},\mathbf{v}\in V \rightsquigarrow a\mathbf{v}\in V.$$

[Remark: We could also write scalar multiplication as $\mathbf{v}a$; the order doesn't matter.] These two operations are required to satisfy the following eight axioms:

(1) Axioms of Addition.

(a) $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$
(b) $\mathbf{u}+(\mathbf{v}+\mathbf{w})=(\mathbf{u}+\mathbf{v})+\mathbf{w}$
(c) There exists a vector $\mathbf{0}\in V$ such that $\mathbf{0}+\mathbf{v}=\mathbf{v}+\mathbf{0}=\mathbf{v}$ for all $\mathbf{v}\in V$.
(d) For every vector $\mathbf{v}\in V$ there exists a vector $\mathbf{u}\in V$ such that $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}=\mathbf{0}$.

Remarks: The vector $\mathbf{0}$ in axiom (1c) is unique. Indeed, if $\mathbf{0}$ and $\mathbf{0}'$ are two vectors satisfying (1c) then we must have

$$\mathbf{0}=\mathbf{0}+\mathbf{0}'=\mathbf{0}'.$$

We call the unique element $\mathbf{0}\in V$ satisfying (1c) the *zero vector*. The vector $\mathbf{u}$ in axiom (1d) is also unique. Indeed, suppose we have two vectors $\mathbf{u}$ and $\mathbf{u}'$ satisfying (1d). Then from axioms (1abc) we must have

$$\mathbf{u}=\mathbf{u}+\mathbf{0}=\mathbf{u}+(\mathbf{v}+\mathbf{u}')=(\mathbf{u}+\mathbf{v})+\mathbf{u}'=\mathbf{0}+\mathbf{u}'=\mathbf{u}'.$$

The unique element $\mathbf{u}$ satisfying (1d) is called the *additive inverse of* $\mathbf{v}$. We denote it by $-\mathbf{v}$. In other words, we have

$$\mathbf{v}+\mathbf{u}=\mathbf{0} \quad\Longleftrightarrow\quad \mathbf{u}=-\mathbf{v}.$$

Based on this, we define the operation of vector subtraction:

$$\mathbf{u}-\mathbf{v}:=\mathbf{u}+(-\mathbf{v}).$$

(2) Axioms of Scalar Multiplication.

(a) For the real number $1\in\mathbb{R}$ we have $1\mathbf{v}=\mathbf{v}$ for all $\mathbf{v}\in V$.
(b) For all real numbers $a,b\in\mathbb{R}$ and vectors $\mathbf{v}\in V$ we have $a(b\mathbf{v})=(ab)\mathbf{v}$.[1]
(c) For all $a,b\in\mathbb{R}$ and $\mathbf{v}\in V$ we have $(a+b)\mathbf{v}=a\mathbf{v}+b\mathbf{v}$.[2]
(d) For all $a\in\mathbb{R}$ and $\mathbf{u},\mathbf{v}\in V$ we have $a(\mathbf{u}+\mathbf{v})=a\mathbf{u}+a\mathbf{v}$.

Remarks: These eight axioms imply many other basic properties. For example, I claim that the real number $0\in\mathbb{R}$ satisfies $0\mathbf{v}=\mathbf{0}$ for all vectors $\mathbf{v}\in V$, where $\mathbf{0}$ is the zero vector.

[1] Note that this identity involves two **different** operations: multiplication of real numbers and scalar multiplication in V. This identity is the reason that we use the same notation for both operations.

[2] This identity is the reason that we use the same notation for addition in $\mathbb{R}$ and addition in V.

Indeed, since $0 + 0 = 0$ as real numbers, we have

$$\begin{aligned} 0 + 0 &= 0 \\ (0+0)\mathbf{v} &= 0\mathbf{v} \\ 0\mathbf{v} + 0\mathbf{v} &= 0\mathbf{v} \\ (0\mathbf{v} + 0\mathbf{v}) - 0\mathbf{v} &= 0\mathbf{v} - 0\mathbf{v} \\ 0\mathbf{v} + (0\mathbf{v} - 0\mathbf{v}) &= \mathbf{0} \\ 0\mathbf{v} + \mathbf{0} &= \mathbf{0} \\ 0\mathbf{v} &= \mathbf{0}. \end{aligned}$$

[We could have taken this as another axiom, but we didn't need to.] It follows from this that the additive inverse $-\mathbf{v}$ is the same as $(-1)\mathbf{v}$ for the real number $-1 \in \mathbb{R}$. Indeed, since $1 + (-1) = 0$ as real numbers, we have

$$\begin{aligned} 1 + (-1) &= 0 \\ (1 + (-1))\mathbf{v} &= 0\mathbf{v} \\ 1\mathbf{v} + (-1)\mathbf{v} &= 0\mathbf{v} \\ \mathbf{v} + (-1)\mathbf{v} &= \mathbf{0} \\ -\mathbf{v} + (\mathbf{v} + (-1)\mathbf{v}) &= -\mathbf{v} + \mathbf{0} \\ (-\mathbf{v} + \mathbf{v}) + (-1)\mathbf{v} &= -\mathbf{v} \\ \mathbf{0} + (-1)\mathbf{v} &= -\mathbf{v} \\ (-1)\mathbf{v} &= -\mathbf{v}. \end{aligned}$$

If this is too pedantic for you, feel free to take the properties $0\mathbf{v} = \mathbf{0}$ and $(-1)\mathbf{v} = -\mathbf{v}$ as axioms.

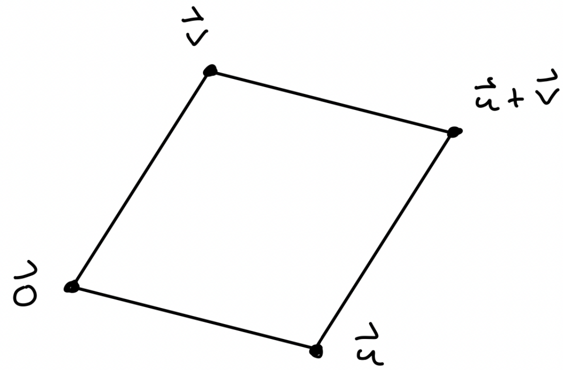
The Prorotype: Euclidean Space. Let $\mathbb{R}^n$ denote the set of ordered n-tupes of real numbers:

$$\mathbb{R}^n = \{\mathbf{v} = (v_1, v_2, \ldots, v_n) : v_1, v_2, \ldots, v_n \in \mathbb{R}\}.$$

It is easy to check that the following operations make $\mathbb{R}^n$ into a vector space over $\mathbb{R}$:

$$\begin{aligned} (u_1, \ldots, u_n) + (v_1, \ldots, v_n) &:= (u_1 + v_1, \ldots, u_n + v_n), \\ a(v_1, \ldots, v_n) &:= (av_1, \ldots, av_n). \end{aligned}$$

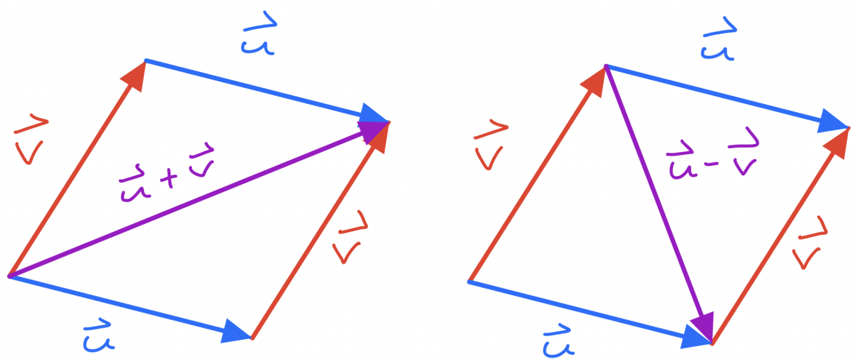
We can think of $\mathbf{v} = (v_1, \ldots, v_n)$ as the coordinates of a point in n-dimensional Euclidean space. In this case, the point $\mathbf{0} = (0, \ldots, 0)$ is called the *origin*. The Parallelogram Law says that for any points $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, the four points $\mathbf{0}$, $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{u} + \mathbf{v}$ are the vertices of a parallelogram. Picture:



We can also think of an n-tuple $\mathbf{v} = (v_1, \ldots, v_n)$ as a directed line segment (an "arrow") with head at the point $\mathbf{v}$ and tail at the origin $\mathbf{0}$. According to the Pythagorean Theorem, the length $\|\mathbf{v}\|$ of this line segment satisfies

$$\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2 \cdots + v_n^2}.$$

Geometrically, arrows add "head-to-tail" and subtract "tail-to-tail":



If we let θ denote the angle between arrows $\mathbf{u}$ and $\mathbf{v}$ then the Law of Cosines tells us that

$$\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\|\mathbf{u}\|\|\mathbf{v}\| \cos\theta.$$

On the other hand, the algebraic formula for the length of an arrow tells us that

$$\begin{aligned}
\|\mathbf{u} - \mathbf{v}\|^2 &= \|(u_1 - v_1, \ldots, u_n - v_n)\|^2 \\
&= (u_1 - v_1)^2 + \cdots + (u_n - v_n)^2 \\
&= (u_1^2 - 2u_1v_1 + v_1^2) + \cdots + (u_n^2 - 2u_nv_n + v_n^2) \\
&= (u_1^2 + \cdots + u_n)^2 + (v_1^2 + \cdots + v_n^2) - 2(u_1v_1 + \cdots + u_nv_n) \\
&= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2(u_1v_1 + \cdots + u_nv_n).
\end{aligned}$$

Then comparing the two equations gives the amazing formula

$$u_1v_1 + \cdots + u_nv_n = \|\mathbf{u}\|\|\mathbf{v}\| \cos\theta.$$

This formula allows us to express angles simply in terms of the coordinates. To be precise, we define the *dot product* of two arrows:

$$\mathbf{u} \bullet \mathbf{v} := u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

Observe that

$$\mathbf{v} \bullet \mathbf{v} = v_1 v_1 + \cdots + v_n v_n = v_1^2 + \cdots + v_n^2 = \|\mathbf{v}\|^2.$$

Hence we have

$$\cos\theta = \frac{\mathbf{u} \bullet \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|} = \frac{\mathbf{u} \bullet \mathbf{v}}{\sqrt{\mathbf{u} \bullet \mathbf{u}}\sqrt{\mathbf{v} \bullet \mathbf{v}}}.$$

Note that θ is a right angle if and only if $\mathbf{u} \bullet \mathbf{v} = 0$.

More generally, an *inner product space over* $\mathbb{R}$ consists of a vector space V over $\mathbb{R}$ together with another algebraic operation

$$\mathbf{u}, \mathbf{v} \in V \rightsquigarrow \langle \mathbf{u}, \mathbf{v} \rangle \in \mathbb{R},$$

which must satisfy the following axioms:

(3) Axioms of Inner Products.

(a) $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
(b) $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$[3]
(c) For all $a \in \mathbb{R}$ and $\mathbf{u}, \mathbf{v} \in V$ we have $\langle a\mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, a\mathbf{v} \rangle = a\langle \mathbf{u}, \mathbf{v} \rangle$.
(d) For all $\mathbf{v} \in V$ we have $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$, with $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ if and only if $\mathbf{v} = \mathbf{0}$.

The following important inequality is a direct consequence of the axioms, but its proof is just a little bit tricky. I'll give you a hint and have you prove it on the homework.

Cauchy-Schwarz Inequality. For any vectors $\mathbf{u}, \mathbf{v} \in V$ in an inner product space we have

$$|\langle \mathbf{u}, \mathbf{v} \rangle|^2 \leq \langle \mathbf{u}, \mathbf{u} \rangle \langle \mathbf{v}, \mathbf{v} \rangle.$$

Why should we bother with this level of abstraction? There are two reasons.

First of all, there exist important examples of abstract inner product spaces that have nothing to do with arrows or points in Euclidean space.

Example: L^2 Space. Let $L^2[0,1]$ denote the set of real-valued functions $f(x)$ on the interval $[0,1]$ such that the integral of $f(x)$ converges:[4]

$$L^2[0,1] = \{f : [0,1] \to \mathbb{R}, \int_0^1 f(x)^2\, dx < \infty\}.$$

Given functions $f, g \in L^2[0,1]$ and scalar $a \in \mathbb{R}$ we define the new functions $f + g \in L^2[0,1]$ and $af \in L^2[0,1]$ by adding and multiplying their values, as one does in Calculus:

$$(f+g)(x) := f(x) + g(x),$$
$$(af)(x) := af(x).$$

[3]By combining (3ab) we also have $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$.

[4]Any statement about integrals has some very technical conditions, but we will proceed intuitively, just as a physicist would.

One can check that these operations make $L^2[0,1]$ into a vector space over $\mathbb{R}$.[5] Furthermore, one can check that the following operation satisfies the inner product axioms:

$$\langle f(x), g(x)\rangle := \int_0^1 f(x)g(x)\,dx.$$

Such inner product spaces of functions are extremely important in applied mathematics. We will say more below.

Another reason for abstraction in linear algebra has to do with "subspaces".

(4) Axioms of Subspaces. Given a vector space V over $\mathbb{R}$ and a subset $U \subseteq V$, we say that U is a *subspace* when it satisfies the following axioms:

(a) $\mathbf{0} \in U$
(b) If $\mathbf{u}, \mathbf{v} \in U$ then $\mathbf{u} + \mathbf{v} \in U$.
(c) If $a \in \mathbb{R}$ and $\mathbf{v} \in U$ then $a\mathbf{v} \in U$.

For example, any line or plane through the origin in Euclidean space is a subspace.[6] We note that Euclidean spaces comes with a collection of *standard basis vectors*:

$$\begin{aligned}
\mathbf{e}_1 &= (1,0,0,\ldots,0,0),\\
\mathbf{e}_2 &= (0,1,0,\ldots,0,0),\\
&\vdots\\
\mathbf{e}_n &= (0,0,0,\ldots,0,1).
\end{aligned}$$

By definition, every vector $\mathbf{v} = (v_1,\ldots,v_n) \in \mathbb{R}^n$ has a unique expression as a *linear combination* of these basis vectors:

$$\begin{aligned}
\mathbf{v} &= (v_1,\ldots,v_n)\\
&= v_1(1,0,\ldots,0,0) + \cdots + v_n(0,0,\ldots,0,1)\\
&= v_1\mathbf{e}_1 + \cdots + v_n\mathbf{e}_n.
\end{aligned}$$

However, **subspaces** of $\mathbb{R}^n$ do not come with standard basis vectors. For example, consider the plane $V \subseteq \mathbb{R}^3$ defined by the equation $x - 2y + z = 0$.[7] I claim that every vector $\mathbf{v} \in V$ this plane has a unique expression of the form

$$\mathbf{v} = a_1(1,1,1) + a_2(1,2,3).$$

Hence we say that $B = \{\mathbf{b}_1, \mathbf{b}_2\}$ with $\mathbf{b}_1 = (1,1,1)$ and $\mathbf{b}_2 = (1,2,3)$ is a *basis* for the vector space V, and if $\mathbf{v} = a_1\mathbf{b}_1 + a_2\mathbf{b}_2$ we say that $\mathbf{v} = (a_1,a_2)_B$ are the *coordinates of* $\mathbf{v}$ *in the* B*-basis*. For example, the vector $\mathbf{v} = (1,-1,-3) \in \mathbb{R}^3$ is in the plane V. It has coordinates $(1,-1,-3)$ as an element of $\mathbb{R}^3$ but it has coordinates $(3,-2)_B$ as an element of V, with respect to the B-basis. Here is why we need the concept of an abstract vector space:

> *Subspaces of $\mathbb{R}^n$ do not come with a standard basis. Therefore we must study them via the axioms of abstract vector spaces.*

[5]The hardest part of the proof is to show that the sum of square integrable functions is square integrable. This can be shown with the Cauchy-Schwarz inequality.

[6]A line or plane not through the origin is not a subspace because it doesn't satisfy (3a). The concept of "subspace" is not immediately intuitive but it is vital to the theory.

[7]Check that this is a subspace.

Here is the technical definition of a basis in an abstract vector space.

Definition of Basis. Let V be a vector space over $\mathbb{R}$ and consider a finite subset $B = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ of vectors in V.

- We say that B is a *spanning set* if for all $\mathbf{v} \in V$ **there exists at least one choice** of scalars $a_1, \ldots, a_n \in \mathbb{R}$ such that
$$\mathbf{v} = a_1\mathbf{b}_1 + \cdots + a_n\mathbf{b}_n.$$
- We say that B is an *independent set*[8] if for all $\mathbf{v} \in V$ **there exists at most one choice** of scalars $a_1, \ldots, a_n \in \mathbb{R}$ such that
$$\mathbf{v} = a_1\mathbf{b}_1 + \cdots + a_n\mathbf{b}_n.$$
- We say that B is a *basis* if it is spanning and independent; that is, if for all $\mathbf{v} \in V$ **there exists a unique choice** of scalars $a_1, \ldots, a_n \in \mathbb{R}$ such that
$$\mathbf{v} = a_1\mathbf{b}_1 + \cdots + a_n\mathbf{b}_n.$$
  In this case we say that $a_1, \ldots, a_n \in \mathbb{R}$ are the *B-coordinates* of $\mathbf{v}$, and we write
$$\mathbf{v} = (a_1, \ldots, a_n)_B.$$
  After we have chosen a basis, we can work with coordinates and pretend that V is $\mathbb{R}^n$.

The following point is fundamental, but its proof is more subtle than you would think.

Definition of Dimension. If a vector space V has a basis with n vectors, then any basis of V must have n vectors. In this case we say that V has *dimension* n, and we write
$$\dim V = n.$$

Proof. This uses a famous trick called "Steinitz Exchange". See the homework.

Example: Euclidean Space. The vector space $\mathbb{R}^n$ has a standard basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ consisting of n vectors. It follows from Steinitz Exchange that **any** basis for $\mathbb{R}^n$ must have n vectors, and hence $\dim \mathbb{R}^n = n$, as it should be. It is relatively easy to find a basis: any sufficiently random collection in n vectors in $\mathbb{R}^n$ will do. For example:

$$(1, 4, 3, 2), (3, -7, 4, 1), (100, 89, -72, 36), (23, 24, 25, 26) \text{ is almost certainly a basis of } \mathbb{R}^4.$$

Not every vector space has a finite basis.

Example: Polynomials. Let $\mathbb{R}[x]$ denote the set of polynomials in x with real coefficients. This set is a vector space over $\mathbb{R}$. It does not have a finite basis, but it does have a fairly obvious infinite basis B consisting of the elements
$$B = \{1(= x^0), x, x^2, \ldots\}.$$
For infinite bases we need to modify slightly the definitions of independence and spanning. In this case, the key fact is that each polynomial $f(x) \in \mathbb{R}[x]$ **has a unique expression**
$$f(x) = \sum_{k \geq 0} a_k x^k,$$

[8] In proofs it is often convenient to use a different form of the definition. Say that B is *independent* if for any scalars $a_1, \ldots, a_n$ we have
$$a_1\mathbf{b}_1 + \cdots + a_n\mathbf{b}_n = \mathbf{0} \text{ implies } a_i = 0 \text{ for all } i.$$
Exercise: Check that the two definitions are equivalent.

where only finitely many of the coefficients $a_0, a_1, a_2, \ldots$ **are nonzero**. If we allow infinitely many nonzero coefficients then we obtain *power series*, instead of polynomials.

In order to say anything about convergence of infinite series in a vector space, one needs a way to measure "distance" between vectors.

(5) Axioms of Norms. Let V be a vector space with a function

$$\mathbf{v} \in V \rightsquigarrow \|\mathbf{v}\| \in \mathbb{R}.$$

We call this function a *norm* when it satisfies the following axioms:

(a) $\|\mathbf{v}\| \geq 0$ for all $\mathbf{v} \in V$ with $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = \mathbf{0}$.
(b) For all $a \in \mathbb{R}$ and $\mathbf{v} \in V$ we have $\|a\mathbf{v}\| = |a|\|\mathbf{v}\|$.
(c) For all $\mathbf{u}, \mathbf{v} \in V$ we have $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$.

(6) Axioms of Metrics. Let V be a vector space with a function

$$\mathbf{u}, \mathbf{v} \in V \rightsquigarrow \text{dist}(\mathbf{u}, \mathbf{v}) \in \mathbb{R}.$$

We call this function a *metric* when it satisfies the following axioms:

(a) $\text{dist}(\mathbf{u}, \mathbf{v}) = \text{dist}(\mathbf{v}, \mathbf{u})$
(b) $\text{dist}(\mathbf{u}, \mathbf{v}) \geq 0$ for all $\mathbf{u}, \mathbf{v} \in V$ with $\text{dist}(\mathbf{u}, \mathbf{v}) = 0$ if and only if $\mathbf{u} = \mathbf{v}$.
(c) $\text{dist}(\mathbf{u}, \mathbf{v}) \leq \text{dist}(\mathbf{u}, \mathbf{w}) + \text{dist}(\mathbf{w}, \mathbf{v})$ for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$.

Every inner product space becomes a normed space[9] by taking $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$, and every normed space becomes a metric space by taking $\text{dist}(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$.

Concept of Orthonormal Sets. Let V be an inner product space. A collection of vectors $\mathbf{b}_1, \mathbf{b}_2, \ldots$ is called *orthonormal* if

- $\langle \mathbf{b}_i, \mathbf{b}_j \rangle = 0$ for all i, j with $i \neq j$
- $\langle \mathbf{b}_i, \mathbf{b}_i \rangle = 1$ for all i

The first statement says that any two vectors in the set are *orthogonal*,[10] and the second statement says that each vector has length 1:

$$\|\mathbf{b}_i\| = \sqrt{\langle \mathbf{b}_i, \mathbf{b}_i \rangle} = \sqrt{1} = 1.$$

Orthonormal sets are very easy to work with. You will show on the homework that if $\mathbf{b}_i$ are orthonormal and $\mathbf{v} = a_1\mathbf{b}_1 + \cdots + a_n\mathbf{b}_n$ then we must have

$$a_i = \langle \mathbf{v}, \mathbf{b}_i \rangle \quad \text{and} \quad \|\mathbf{v}\|^2 = a_1^2 + \cdots + a_n^n.$$

If the orthonormal set spans V then it is called an *orthonormal basis.* Orthonormal bases are analogous to the standard basis in Euclidean space.

Example: Fourier Series. The inner product space $L^2[0,1]$ of square integrable functions $[0,1] \to \mathbb{R}$ contains a particularly famous orthonormal set of functions. If we define

$$s_n(x) := \sqrt{2}\sin(2\pi n x),$$
$$c_n(x) := \sqrt{2}\cos(2\pi n x),$$

then you will show on the homework that the following set of functions is orthonormal:

$$B = \{1, s_1(x), s_2(x), \ldots, c_1(x), c_2(x), \ldots\}.$$

[9] You will prove this on the homework.
[10] In Euclidean space this corresponds to perpendicular vectors

That is, you will show

- $\langle 1, s_n(x)\rangle = \langle 1, c_n(x)\rangle = 0$ for all $n \geq 1$,
- $\langle s_m(x), c_n(x)\rangle = 0$ for all $m, n \geq 1$,
- $\langle s_m(x), s_n(x)\rangle = 0$ for $m \neq n$ and 1 for $m = n$,
- $\langle c_m(x), c_n(x)\rangle = 0$ for $m \neq n$ and 1 for $m = n$.

It follows from this that the set is independent. Is it also a spanning set? For a given function $f(x) \in L^2[0,1]$, the problem is to find scalars $a_0, a_1, a_2, \ldots, b_1, b_2, \ldots \in \mathbb{R}$ such that

$$(*) \qquad f(x) = a_0 + \sum_{n=1}^{\infty} a_n s_n(x) + \sum_{n=1}^{\infty} b_n c_n(x).$$

In Fourier's paper on the analytic theory of heat (1822) he gave a clever formula to find the coefficients. For us this formula is an immediate consequence of the fact that B is orthonormal:

$$\begin{aligned} a_0 &= \langle f(x), 1\rangle = \int_0^1 f(x)\,dx, \\ a_n &= \langle f(x), s_n(x)\rangle = \sqrt{2}\int_0^1 f(x)\sin(2\pi n x)\,dx, \\ b_n &= \langle f(x), c_n(x)\rangle = \sqrt{2}\int_0^1 f(x)\cos(2\pi n x)\,dx. \end{aligned}$$

So the coefficients are easy to find. The hard question is whether, and in what sense, the infinite series $(*)$ converges. This is an important problem in the history of mathematics; controversies surrounding its solution led to many of the concepts of modern analysis.

I will just state the simplest form of the answer; the proof is well beyond the scope of this course. Consider the distance function induced by the inner product on $L^2[0,1]$. That is, for any functions $f(x), g(x) \in L^2[0,1]$ we define the "distance" between then by

$$\text{dist}(f(x), g(x))^2 = \|f(x) - g(x)\|^2 = \langle f(x) - g(x), f(x) - g(x)\rangle = \int_0^1 (f(x) - g(x))^2\,dx.$$

Now consider any function $f(x) \in L^2[0,1]$ and let a_n, b_n be the corresponding Fourier coefficients. Then we have the following theorems.

- **Convergence of Fourier Series.** The series $(*)$ converges in L^2. That is, we have

$$\text{dist}\left(f(x), a_0 + \sum_{n=1}^{N} a_n s_n(x) + \sum_{n=1}^{N} b_n c_n(x)\right) \to 0 \qquad \text{as } N \to \infty.$$

- **Parseval's Identity.** Computing the "length" of each side of $(*)$ gives a convergent series of real numbers:

$$\int_0^1 f(x)^2\,dx = \langle f(x), f(x)\rangle = a_0^2 + a_1^2 + b_1^2 + a_2^2 + b_2^2 + \cdots.$$

For example, consider the square wave function

$$f(x) = \begin{cases} 1 & 0 \leq x < 1/2, \\ 0 & 1/2 \leq x \leq 1. \end{cases}$$

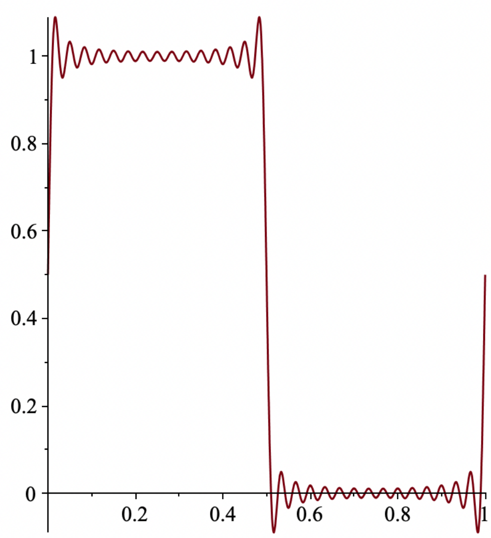
It is easy to check that $a_0 = \langle f(x), 1\rangle = 1/2$ and $b_n = \langle f(x), c_n(x)\rangle = 0$ for all $n \geq 1$. Next, we compute

$$\begin{aligned} a_n &= \langle f(x), s_n(x)\rangle \\ &= \sqrt{2}\int_0^1 f(x)\sin(2\pi n x)\,dx \\ &= \sqrt{2}\int_0^{1/2} \sin(2\pi n x)\,dx \\ &= \frac{\sqrt{2}}{2\pi n}\left[-\cos(2\pi n x)\right]_0^{1/2} \\ &= \frac{\sqrt{2}}{2\pi n}\left[-\cos(\pi n) + 1\right] \\ &= \frac{\sqrt{2}}{2\pi n}\left[-(-1)^n + 1\right] \\ &= \begin{cases} 0 & n \text{ even}, \\ \frac{\sqrt{2}}{\pi n} & n \text{ odd}. \end{cases} \end{aligned}$$

It follows that

$$f(x) = \frac{1}{2} + \frac{\sqrt{2}}{\pi}\sin(2\pi x) + \frac{\sqrt{2}}{3\pi}\sin(6\pi x) + \frac{\sqrt{2}}{5\pi}\sin(10\pi x) + \cdots.$$

Here is a picture of the first 30 terms of this sequence:



Finally, Parseval's Identity gives the following interesting identity:

$$\begin{aligned}
\int_0^1 f(x)^2\,dx &= \left(\frac{1}{2}\right)^2 + \left(\frac{\sqrt{2}}{\pi}\right)^2 + \left(\frac{\sqrt{2}}{3\pi}\right)^2 + \left(\frac{\sqrt{2}}{5\pi}\right)^2 + \cdots \\
\frac{1}{2} &= \frac{1}{4} + \frac{2}{\pi^2} + \frac{2}{3^2\pi^2} + \frac{2}{5^2\pi^2} + \cdots \\
\frac{1}{4} &= \frac{2}{\pi^2} + \frac{2}{3^2\pi^2} + \frac{2}{5^2\pi^2} + \cdots \\
\frac{1}{4} &= \frac{2}{\pi^2}\left(\frac{1}{1^2} + \frac{1}{3^2} + \frac{1}{5^2} + \cdots\right) \\
\frac{\pi^2}{8} &= \frac{1}{1^2} + \frac{1}{3^2} + \frac{1}{5^2} + \cdots
\end{aligned}$$

That's weird.[11]

Now seems like a good time to bring in complex numbers. I have a beef with the American educational system, in that there is no course that reliably introduces complex numbers. The system is able to sleep at night because complex numbers are in the pre-Calculus curriculum, but the treatment is inadequate, and most math majors don't take pre-Calculus. Indeed, I believe it possible for a student to graduate with a math major having never seen a good introduction to complex numbers. As is traditional, I will give a quick review and pretend that you have seen this before, even if you haven't.

Complex Numbers. The complex numbers are defined as

$$\mathbb{C} = \{a + ib : a, b \in \mathbb{R}\},$$

where i is an abstract symbol satisfying $i^2 = -1$. Given a complex number $\alpha = a + ib$, we define its *absolute value* and *complex conjugate*:[12]

$$\begin{aligned}
|\alpha| &:= \sqrt{a^2 + b^2}, \\
\alpha^* &:= a - ib.
\end{aligned}$$

These satisfying the following properties.

(7) Properties of Complex Numbers. For all $a, b \in \mathbb{R}$ and $\alpha, \beta \in \mathbb{C}$ we have

(a) $(a\alpha + b\beta)^* = a\alpha^* + b\beta^*$
(b) $(\alpha\beta)^* = \alpha^*\beta^*$
(c) $\alpha = \alpha^* \iff \alpha \in \mathbb{R}$
(d) $|\alpha| \geq 0$ with $|\alpha| = 0$ if and only if $\alpha = 0$.
(d) $|\alpha| = \alpha^*\alpha$
(e) $|\alpha\beta| = |\alpha||\beta|$.
(f) If $\alpha \neq 0$ then $\alpha^{-1} = \alpha^*/|\alpha|^2$.

---

[11] This series is related to the famous *Basel problem*. It is easy to see that the infinite series $1/1^2 + 1/2^2 + 1/3^2 + \cdots$ converges, but is not at all clear how to find a formula for the sum. This problem was posed by Pietro Mengoli 1650 and finally solved by Leonhard Euler in 1734, who showed that the limit is exactly $\pi^2/6$. The appearance of π in the answer was a big surprise.

[12] I will use α^* instead of the traditional $\overline{\alpha}$ to avoid conflict with the whiteboard notation for vectors: $\vec{v}$.

Many applications of linear algebra use complex instead of real scalars. Almost all of the axioms are the same, but there is a key change in the definition of inner product.

(8) Axioms of Hermitian Inner Products. Let V be a vector space over $\mathbb{C}$, together with an algebraic operation

$$\mathbf{u}, \mathbf{v} \in V \rightsquigarrow \langle \mathbf{u}, \mathbf{v} \rangle \in \mathbb{C}.$$

We call this a *Hermitian inner product* if it satisfies the following axioms:

(a) $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle^*$
(b) $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$[13]
(c) For all $\alpha \in \mathbb{C}$ we have $\langle \mathbf{u}, \alpha\mathbf{v} \rangle = \alpha \langle \mathbf{u}, \mathbf{v} \rangle$.[14]
(d) For all $\mathbf{v} \in V$, part (a) tells us that $\langle \mathbf{v}, \mathbf{v} \rangle \in \mathbb{R}$. Furthermore, we must have $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ with $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ if and only if $\mathbf{v} = \mathbf{0}$.

Jargon: A Hermitian inner product is sometimes called *sesquilinear* (one and a half times linear) because it is linear in the second coordinate:

$$\langle \mathbf{u}, \alpha\mathbf{v} + \beta\mathbf{w} \rangle = \alpha \langle \mathbf{u}, \mathbf{v} \rangle + \beta \langle \mathbf{u}, \mathbf{w} \rangle,$$
$$\langle \alpha\mathbf{u} + \beta\mathbf{v}, \mathbf{w} \rangle = \alpha^* \langle \mathbf{u}, \mathbf{w} \rangle + \beta^* \langle \mathbf{v}, \mathbf{w} \rangle.$$

Beware, some books switch these.

Example: The standard Hermitian product on $\mathbb{C}^n$. Consider the set

$$\mathbb{C}^n = \{\mathbf{v} = (v_1, \ldots, v_n) : v_1, \ldots, v_n \in \mathbb{C}\}.$$

This is naturally a vector space over $\mathbb{C}$ with the usual operations of addition and scalar multiplication. We can still define the usual dot product

$$\mathbf{u} \bullet \mathbf{v} = u_1 v_1 + \cdots + u_n v_n,$$

but this turns out to have bad properties. For example, we might have $\mathbf{v} \bullet \mathbf{v} < 0$, as with the vector $\mathbf{v} = (i, i)$. To fix this, we instead consider the following operation:

$$\langle \mathbf{u}, \mathbf{v} \rangle = u_1^* v_1 + \cdots + u_n^* v_n.$$

One can check that this satisfies the axioms of a Hermitian inner product. Most importantly, we have $\langle \mathbf{v}, \mathbf{v} \rangle = v_1^* v_1 + \cdots + v_n^* v_n = |v_1|^2 + \cdots + |v_n|^2 \geq 0$, with $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ if and only if $\mathbf{v} = \mathbf{0}$, which allows us to define a norm and a metric:

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle},$$
$$\text{dist}(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|.$$

Quantum mechanics is the big reason for using complex Hermitian spaces, but the complex numbers also allow us to simplify some classical problems.

Example: Complex Fourier Series. Recall Euler's identities:

$$e^{i\theta} = \cos\theta + i \sin\theta,$$
$$e^{-i\theta} = \cos\theta - i \sin\theta,$$
$$\cos\theta = (e^{i\theta} + e^{-i\theta})/2,$$
$$\sin\theta = (e^{i\theta} - e^{-i\theta})/(2i).$$

[13] By combining (8ab) we also have $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v} + \mathbf{w} \rangle$.
[14] By combining (8ac) we also have $\langle \alpha\mathbf{u}, \mathbf{v} \rangle = \alpha^* \langle \mathbf{u}, \mathbf{v} \rangle$.

Suppose that we have a *real Fourier series*[15]

$$f(x) = a_0 + \sum_{n\geq 1} a_n \sin(2\pi nx) + \sum_{n\geq 1} b_n \cos(2\pi nx).$$

We can write this as a *complex Fourier series*

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{i2\pi nx}$$

by defining the complex coefficients:

$$c_n := \begin{cases} a_0 & n = 0, \\ (b_n - ia_n)/2 & n \geq 1, \\ (b_{-n} + ia_{-n})/2 & n \leq -1. \end{cases}$$

Why would we do this? Because the functions $e^{i2\imath nx}$ are easier to work with! Let's define the *complex L^2 space* $L^2[0,1]$ as the set of functions $[0,1] \to \mathbb{C}$ satisfying

$$\int_0^1 |f(x)|^2 \, dx < \infty.$$

This space has a standard Hermitian inner product:

$$\langle f(x), g(x) \rangle = \int_0^1 f(x)^* g(x) \, dx.$$

And the functions $e^{i2\pi nx}$ for $n \in \mathbb{Z}$ are an orthonormal set:

$$\begin{aligned} \langle e^{i2\pi mx}, e^{i2\pi nx} \rangle &= \int_0^1 (e^{i2\pi mx})^* e^{2\pi inx} \, dx \\ &= \int_0^1 e^{-i2\pi mx} e^{2\pi inx} \, dx \\ &= \int_0^1 e^{i2\pi(n-m)x} \, dx. \end{aligned}$$

If $m = n$ then this gives

$$\langle e^{i2\pi nx}, e^{i2\pi nx} \rangle = \int_0^1 1 \, dx = 1,$$

and if $m \neq n$ then we get

$$\begin{aligned} \langle e^{i2\pi mx}, e^{i2\pi nx} \rangle &= \int_0^1 e^{i2\pi(n-m)x} \, dx \\ &= \frac{1}{i2\pi(n-m)} \left[e^{i2\pi(n-m)x} \right]_0^1 \\ &= \frac{1}{i2\pi(n-m)} [1 - 1] \\ &= 0. \end{aligned}$$

[15] I'll absorb the $\sqrt{2}$ factors into the coefficients this time.

Note: That was much easier than messing around with trigonometric identities. It also means that we have a single formula for the complex Fourier coefficients:[16]

$$c_n = \langle e^{i2\pi nx}, f(x)\rangle = \int_0^1 e^{-i2\pi nx} f(x)\,dx.$$

Then we can convert back to real coefficients if desired.

Fourier Transform. For the physicists among you, I should mention what happens for functions on the whole real line. Let $L^2(\mathbb{R})$ denote the set of functions $f : \mathbb{R} \to \mathbb{C}$ that are square integrable:

$$\int_{-\infty}^{\infty} |f(x)|^2\,dx < \infty.$$

As with $L^2[0,1]$, this is a Hermitian space with Hermitian product

$$\langle f(x), g(x)\rangle = \int_{-\infty}^{\infty} f(x)^* g(x)\,dx.$$

This space is more complicated than $L^2[0,1]$ because it does not have a countable basis.[17] However, the situation is not hopeless because we can generalize the *Fourier series* to the *Fourier transform*:

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{i2\pi nx} \quad \rightsquigarrow \quad f(x) = \int_{-\infty}^{\infty} c(\omega) e^{i2\pi\omega x}\,d\omega.$$

We can view the function $c : \mathbb{R} \to \mathbb{C}$ as a generalization of the sequence of coefficients c_n for $n \in \mathbb{Z}$. This function $c(\omega)$ is called the *Fourier transform* of $f(x)$ and it is sometimes denoted $\hat{f}(\omega)$. In some sense we can view the set

$$\{e^{i2\pi\omega x} : \omega \in \mathbb{R}\}$$

as an uncountably infinite basis for the space $L^2(\mathbb{R})$. There is just one issue; the functions $e^{i2\pi\omega x}$ are not square integrable:

$$\int_{-\infty}^{\infty} |e^{i2\pi\omega x}|^2\,dx = \int_{-\infty}^{\infty} 1\,dx = \infty.$$

This is a typical problem in physics. It can be surmounted by generalizing the concept of function to that of "distribution", but the rigorous mathematical definitions make the subject less understandable. Dirac showed that the intuitive point of view is a powerful tool for studying quantum mechanics.

[16] Taking the inner product in the other direction gives $\langle f(x), e^{i2\pi nx}\rangle = c_n^*$.

[17] The issue is that $[0,1]$ is a compact infinite set, while $\mathbb{R}$ is not compact.