Structures Related to Inner Product Spaces

There are various generalizations of the notation of an inner product space. We will stick to real and complex inner product spaces. These are vector spaces equipped with an *inner product*. This gives us a way to compute things like lengths and angles. Once we have an inner product in hand, we can define related structures like a norm and metric.

But first we should recall some facts about the field of complex numbers: $\mathbb{C} = \{a + bi \mid a, b \in \mathbb{R}\}$ (i is the imaginary root, so $i^2 = -1$). Given $z = a + bi \in \mathbb{C}$, we call $\mathfrak{Re}(z) = a$ the real part of z and bi the imaginary part. Let $z = a + bi$ and $w = c + di$. We have $z + w = (a + c) + (b + d)i$ and $zw = (ac - bd) + (ad + bc)i$. The complex numbers also are equipped with the operation of **conjugation** defined by $\bar{z} = \overline{a + bi} = a - bi$ (we flip the sign between the real and imaginary parts). Conjugation is an involution on $\mathbb{C}$. In other words, $\overline{z + w} = \bar{z} + \bar{w}$, $\overline{zw} = \bar{z}\,\bar{w}$, and $\bar{\bar{z}} = z$. Notice also that $z = \bar{z}$ if and only if $z = \mathfrak{Re}(z) \in \mathbb{R}$ (i.e. z is self conjugate only when it is real). Notice $z\bar{z} = (a+bi)(a-bi) = a^2 + b^2 \in \mathbb{R}_{\geq 0}$. The modulus of a complex number is defined as $|z| = \sqrt{z\bar{z}} = \sqrt{a^2 + b^2}$. If we think of $z = a + bi$ as a point (a, b) in $\mathbb{R}^2$, then $|z|$ is the "length" of z (i.e., the distance from z to the origin).

Let us fix a particular a subfield of the complex numbers, say $\mathbb{F}$. Consequently $\mathbb{F}$ will contain (at least) all of the rational numbers, $\mathbb{Q}$, and be closed under addition, subtraction, multiplication, and division (not by zero). Usually, we have in mind $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$. Also, let V be a vector space over $\mathbb{F}$.

Definition: Let V be equipped with a mapping $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F}$ (called an *inner product*) such that...

(a) [Sesquilinear] (one and a half times linear) $\langle \cdot, \cdot \rangle$ is linear in the first slot and conjugate linear in the second slot.[1]

$$\langle \mathbf{u}+\mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle, \quad \langle \mathbf{u}, \mathbf{v}+\mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle, \quad \langle c\mathbf{v}, \mathbf{w} \rangle = c\langle \mathbf{v}, \mathbf{w} \rangle, \quad \text{and} \quad \langle \mathbf{v}, c\mathbf{w} \rangle = \bar{c}\langle \mathbf{v}, \mathbf{w} \rangle.$$

When $\mathbb{F}$ is a subfield of the real numbers (such as $\mathbb{R}$ itself), conjugation does not do anything. In this case, the inner product is linear in both slots (i.e. *bilinear*).

(b) [Conjugate Symmetric] $\langle \mathbf{v}, \mathbf{w} \rangle = \overline{\langle \mathbf{w}, \mathbf{v} \rangle}$.

Again, if $\mathbb{F}$ is a real field, conjugation does nothing. In such a case, we have plain old *symmetry*.

(c) [Positive Definite] Notice by conjugate symmetry, we have $\langle \mathbf{v}, \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \mathbf{v} \rangle}$ so $\langle \mathbf{v}, \mathbf{v} \rangle \in \mathbb{R}$. We require

$$\langle \mathbf{v}, \mathbf{v} \rangle \geq 0 \quad \text{and} \quad \langle \mathbf{v}, \mathbf{v} \rangle = 0 \text{ if and only if } \mathbf{v} = \mathbf{0}.$$

where $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $c \in \mathbb{F}$. If so, V is an **inner product space** equipped with inner product $\langle \cdot, \cdot \rangle$. In particular, when $\mathbb{F} = \mathbb{C}$ we say V is a complex inner product space and when $\mathbb{F} = \mathbb{R}$ we say V is a real inner product space.

Definition: Let V be equipped with a mapping $\| \cdot \| : V \to \mathbb{R}$ (called a *norm*) such that...

(a) [Positive Definite] $\|\mathbf{v}\| \geq 0$ and $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = \mathbf{0}$

(b) [Respects Scaling] $\|c\,\mathbf{v}\| = |c|\,\|\mathbf{v}\|$

(c) [Triangle Inequality] $\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$

where $\mathbf{v} \in V$ and $c \in \mathbb{F}$. If so, V is called a **normed space**.

Theorem: Every inner product space V has the structure of a normed space. Specifically, let V be an inner product space. Then $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$ for all $\mathbf{v} \in V$ is a norm on V.

Note: Given a normed space, its norm does not necessarily come from an inner product.

The first part of the definition of a norm follows from the inner product's positive definiteness. The second part is an easy consequence of the scalar multiple part of sesquilinearity. The triangle inequality quickly follows from the following (rather important) result:

Theorem: (Cauchy-Schwarz Inequality) Let V be an inner product space. For all $\mathbf{v}, \mathbf{w} \in V$, $|\langle \mathbf{v}, \mathbf{w} \rangle| \leq \|\mathbf{v}\|\,\|\mathbf{w}\|$.

Proof: There are many proofs of this famous inequality. Here we present a quirky one. First, note if an inner product $\langle \mathbf{v}, \mathbf{w} \rangle = z$ is not a real number, we can multiply it by $\zeta = \bar{z}/|z|$ so that $\langle \zeta\mathbf{v}, \mathbf{w} \rangle = \zeta\langle \mathbf{v}, \mathbf{w} \rangle = \bar{z}/|z| \cdot z = |z|^2/|z| = |z| \in \mathbb{R}$. Moreover, $|\zeta| = |\bar{z}|/|z| = 1$ so that $|\langle \zeta\mathbf{v}, \mathbf{w} \rangle| = |\zeta| \cdot |\langle \mathbf{v}, \mathbf{w} \rangle| = |\langle \mathbf{v}, \mathbf{w} \rangle|$ and $\|\zeta\mathbf{v}\| = |\zeta| \cdot \|\mathbf{v}\| = \|\mathbf{v}\|$. Thus the inequality holds for $\mathbf{v}$ and $\mathbf{w}$ if and only if it holds for $\zeta\mathbf{v}$ and $\mathbf{w}$. So without loss of generality we may assume $\langle \mathbf{v}, \mathbf{w} \rangle$ is a real number.

[1] Some people follow the opposite convention and have their inner product conjugate linear in the first slot and linear in the second slot. This is common in matrix theory and especially in Physics.

Let x be any *real* number. Notice $0 \leq \|x\mathbf{v}+\mathbf{w}\|^2 = \langle x\mathbf{v}+\mathbf{w}, x\mathbf{v}+\mathbf{w}\rangle = x^2\langle \mathbf{v},\mathbf{v}\rangle + x\langle \mathbf{v},\mathbf{w}\rangle + x\langle \mathbf{w},\mathbf{v}\rangle + \langle \mathbf{w},\mathbf{w}\rangle$ where we used sequilinearity to pull apart the inner product. Keep in mind that x is real, so $\bar{x} = x$ (i.e., conjugation does nothing). Also, we have assumed $\langle \mathbf{v},\mathbf{w}\rangle$ is real, so $\langle \mathbf{w},\mathbf{v}\rangle = \overline{\langle \mathbf{v},\mathbf{w}\rangle} = \langle \mathbf{v},\mathbf{w}\rangle$. Thus our expression simplifies to $\|\mathbf{v}\|^2x^2 + 2\langle \mathbf{v},\mathbf{w}\rangle x + \|\mathbf{w}\|^2$.

If we let $A = \|\mathbf{v}\|^2$, $B = 2\langle \mathbf{v},\mathbf{w}\rangle$, and $C = \|\mathbf{w}\|^2$. Then $0 \leq \|x\mathbf{v}+\mathbf{w}\|^2 = Ax^2+Bx+C = y$ is a quadratic in x with real coefficients A, B, C. Since $y \geq 0$, our quadratic cannot have distinct real roots. This means that the discriminant for the quadratic $B^2 - 4AC = (2\langle \mathbf{v},\mathbf{w}\rangle)^2 - 4\|\mathbf{v}\|^2\|\mathbf{w}\|^2 \leq 0$ (recall that $B^2 - 4AC$ is the part of the quadratic formula inside the square root). Thus $(\langle \mathbf{v},\mathbf{w}\rangle)^2 \leq (\|\mathbf{v}\| \cdot \|\mathbf{w}\|)^2$. Square rooting, we have $|\langle \mathbf{v},\mathbf{w}\rangle| \leq \|\mathbf{v}\|\,\|\mathbf{w}\|$. ■

Now the norm's triangle inequality easily follows. Note: For real numbers a, b, we always have $a \leq |a| = \sqrt{a^2} \leq \sqrt{a^2+b^2} = |a+bi|$ (i.e., for a complex number z, $\mathfrak{Re}(z) \leq |z|$). Therefore,

$$\begin{aligned}\|\mathbf{v}+\mathbf{w}\|^2 = \langle \mathbf{v}+\mathbf{w},\mathbf{v}+\mathbf{w}\rangle = \langle \mathbf{v},\mathbf{v}\rangle + \langle \mathbf{v},\mathbf{w}\rangle + \langle \mathbf{w},\mathbf{v}\rangle + \langle \mathbf{w},\mathbf{w}\rangle &= \|\mathbf{v}\|^2 + \langle \mathbf{v},\mathbf{w}\rangle + \overline{\langle \mathbf{v},\mathbf{w}\rangle} + \|\mathbf{w}\|^2 \\ &= \|\mathbf{v}\|^2 + 2\,\mathfrak{Re}(\langle \mathbf{v},\mathbf{w}\rangle) + \|\mathbf{w}\|^2 \leq \|\mathbf{v}\|^2 + 2|\langle \mathbf{v},\mathbf{w}\rangle| + \|\mathbf{w}\|^2 \\ &\leq \|\mathbf{v}\|^2 + 2\|\mathbf{v}\|\|\mathbf{w}\| + \|\mathbf{w}\|^2 = (\|\mathbf{v}\| + \|\mathbf{w}\|)^2\end{aligned}$$

where our last inequality was the Cauchy-Schwarz inequality. Square-rooting yields $\|\mathbf{v}+\mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$.

An important consequence of the Cauchy-Schwarz inequality is that for nonzero $\mathbf{v}, \mathbf{w} \in V$ in a real inner product space, we can define the notion of the *angle* between $\mathbf{v}$ and $\mathbf{w}$. In particular, notice Cauchy-Schwarz says $\dfrac{|\langle \mathbf{v},\mathbf{w}\rangle|}{\|\mathbf{v}\|\,\|\mathbf{w}\|} \leq 1$. This means $-1 \leq \dfrac{\langle \mathbf{v},\mathbf{w}\rangle}{\|\mathbf{v}\|\,\|\mathbf{w}\|} \leq 1$ so that $\theta = \arccos\left(\dfrac{\langle \mathbf{v},\mathbf{w}\rangle}{\|\mathbf{v}\|\,\|\mathbf{w}\|}\right)$ makes sense. We note that $0 \leq \theta \leq \pi$ where $\theta = 0$ (respectively π) if $\mathbf{v}$ and $\mathbf{w}$ point in the same (respectively opposite) direction(s). We also have the familiar identity: $\langle \mathbf{v},\mathbf{w}\rangle = \|\mathbf{v}\|\,\|\mathbf{w}\|\cos(\theta)$ (even when $\mathbf{v}$ or $\mathbf{w}$ is $\mathbf{0}$ so that θ is ill defined).

Definition: Let X be a set equipped with a mapping $d : X \times X \to \mathbb{R}$ (called a *metric*) such that...

(a) [Positivity] $d(x,y) \geq 0$ and $d(x,y) = 0$ if and only if $x = y$

(b) [Symmetry] $d(x,y) = d(y,x)$

(c) [Triangle Inequality] $d(x,z) \leq d(x,y) + d(y,z)$

where $x, y, z \in X$. If so, X is called a **metric space**.

Theorem: Every normed space V has the structure of a metric space. Specifically, let V be a normed space. Then $d(\mathbf{v},\mathbf{w}) = \|\mathbf{v}-\mathbf{w}\|$ for all $\mathbf{v},\mathbf{w} \in V$ is a metric on X.

Note: However, given a metric space, its metric does not necessarily come from a norm. Also, if you study manifold theory, be aware that manifold theorists and differential geometers often use the term *metric* to mean something quite different. We are using *metric* in the topological sense.

Once again, the first part of the definition ultimately flows from positive definiteness: $d(\mathbf{v},\mathbf{w}) = \|\mathbf{v}-\mathbf{w}\| = 0$ if and only if $\mathbf{v}-\mathbf{w} = \mathbf{0}$ (i.e., $\mathbf{v} = \mathbf{w}$). Symmetry is also easy: $d(\mathbf{v},\mathbf{w}) = \|\mathbf{v}-\mathbf{w}\| = \|-(\mathbf{w}-\mathbf{v})\| = |-1|\cdot\|\mathbf{w}-\mathbf{v}\| = d(\mathbf{w},\mathbf{v})$. Finally, the metric version of the triangle inequality follows from the norm version:

$$d(\mathbf{u},\mathbf{w}) = \|\mathbf{u}-\mathbf{w}\| = \|\mathbf{u}-\mathbf{v}+\mathbf{v}-\mathbf{w}\| \leq \|\mathbf{u}-\mathbf{v}\| + \|\mathbf{v}-\mathbf{w}\| = d(\mathbf{u},\mathbf{v}) + d(\mathbf{v},\mathbf{w}).$$

In a linear algebra course, we are not typically concerned with issues of limits and continuity (i.e., analysis issues). However, it is worth (briefly) mentioning that a metric allows one to define a topological structure on our space X. In particular, $O \subseteq X$ is *open* if given any point $x_0 \in O$ we have that points arbitrarily close to x_0 also belong to O. More concretely, O is open if for every $x_0 \in O$, there exists some $\epsilon > 0$ such that whenever $d(x, x_0) < \epsilon$, we have $x \in O$. The collection of open subsets of X form a *topology* which allows one to explore notions such as continuity, limits, connectedness, and more. We will not pursue this further here.

Summing up: Every inner product space gives us a compatible normed space structure. Every normed space gives us a compatible metric space structure. Every metric space gives us a compatible topological space structure. None of these implications hold in reverse (in general). Inner product spaces have many layers of rich structure. In such a space, linear algebra, geometry, analysis, and topology all have something to tell us.[2]

[2] Functional analysis begins with a study of these structures (inner product, normed, and metric spaces). Such a space in which all Cauchy sequences converge is called complete. A complete normed space is called a Banach space. A complete inner product space is called a Hilbert space. These kinds of spaces play a big role in quantum mechanics and elsewhere.

Example: For $\mathbf{x} = (x_1, \ldots, x_n), \mathbf{y} = (y_1, \ldots, y_n) \in \mathbb{C}^n$ the mapping $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}\bar{\mathbf{y}}^t = \sum_{i=1}^{n} x_i \bar{y_i}$ is an inner product on $\mathbb{C}^n$. This is called the **dot product** or the standard inner product. If we consider $\mathbb{R}^n$ instead of $\mathbb{C}^n$, we have $\mathbf{x} \bullet \mathbf{y} = \sum_{i=1}^{n} x_i y_i$ for $\mathbf{x} = (x_1, \ldots, x_n), \mathbf{y} = (y_1, \ldots, y_n) \in \mathbb{R}^n$. In this case, the norm is $\|\mathbf{x}\| = \sqrt{x_1^2 + \cdots + x_n^2}$ and the corresponding metric (i.e. distance function) is $d(\mathbf{x}, \mathbf{y}) = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}$. These are the standard Euclidean norm and metric.

Example: (Related to the previous example.) Let $A = (a_{ij}), B = (b_{ij}) \in \mathbb{C}^{m \times n}$. We let $B^* = \bar{B}^t = \overline{B^t}$ denote the conjugate transpose of B. Define $\langle A, B \rangle = \mathrm{tr}(AB^*) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} \overline{b_{ij}}$. This is an inner product on $\mathbb{C}^{m \times n}$ called the Frobenius inner product. If we pull apart a matrix row by row (or column by column), we can think of $\mathbb{C}^{m \times n}$ as $\mathbb{C}^{mn}$. With such an identification, the Frobenius inner product on $\mathbb{C}^{m \times n}$ is really nothing more than the dot product on $\mathbb{C}^{mn}$. In the case of real matrices, $\mathbb{R}^{m \times n}$, this is just $\langle A, B \rangle = \mathrm{tr}(AB^T)$. In particular, for $A \in \mathbb{R}^{m \times n}$, $\|A\| = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} (a_{ij})^2}$.

We also have more interesting examples built from integration.

Example: For fixed real numbers $a < b$, let $C[a, b] = \{f : [a, b] \to \mathbb{R} \mid f \text{ is continuous}\}$ (i.e., continuous real valued functions defined on the closed interval $[a, b]$). This is a real vector space. Next, fix a positive continuous *weight* function $W(x) \in C[a, b]$ (so $W(x) > 0$ for $a \leq x \leq b$). For any $f(x), g(x) \in C[a, b]$, we define the inner product:

$$\langle f(x), g(x) \rangle = \int_a^b f(x)g(x)W(x)\, dx.$$

While bilinearity and symmetry are easily established, showing positive definiteness requires careful thought and referencing continuity. Note that for any $f(x) \in C[a, b]$, $f(x)^2 \geq 0$ and so $f(x)^2 W(x) \geq 0$ thus $\langle f(x), f(x) \rangle = \int_a^b f(x)^2 W(x)\, dx \geq 0$. Next, if $f(x_0) \neq 0$ for some $x_0 \in [a, b]$, then $f(x_0)^2 > 0$. Thus $f(x_0)^2 W(x_0) > 0$. Analysis tells us that a continuous function (such as $f(x)^2 W(x)$) that is positive at a point must also be positive on some open interval about that point. Eventually this allows us to conclude $\langle f(x), f(x) \rangle = \int_a^b f(x)^2 W(x)\, dx > 0$. Thus $\langle f(x), f(x) \rangle = 0$ only if $f(x) = 0$. For the space of polynomials[3], we can loosen our restriction on our weight function. In that context, we only need $W(x)$ to be a non-zero, non-negative continuous function.

The above example is a first step into the world of *orthogonal polynomials*. For example, choosing $[a, b] = [-1, 1]$ and $W(x) = 1$ so that $\langle f(x), g(x) \rangle = \int_{-1}^{1} f(x)g(x)\, dx$ and then running the Gram-Schmidt process on 1, x, x^2, $\ldots$, one gets: 1, x, $x^2 - 1/3$, $x^3 - 3/5x$, $\ldots$. These polynomials (up to a certain normalization) are Legendre polynomials which show up in numerous mathematical and scientific applications.

Coordinates:

Let V be a (finite dimensional) inner product space with basis $\beta = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. Given $\mathbf{v}, \mathbf{w} \in V$, we have $\mathbf{v} = \sum_{i=1}^{n} b_i \mathbf{v}_i$ and $\mathbf{w} = \sum_{i=1}^{n} c_i \mathbf{v}_i$ for some scalars $b_i, c_i \in \mathbb{F}$ (i.e., $[\mathbf{v}]_\beta = [b_1\ b_2\ \cdots\ b_n]^t$ and $[\mathbf{w}]_\beta = [c_1\ c_2\ \cdots\ c_n]^t$). We have:

$$\langle \mathbf{v}, \mathbf{w} \rangle = \left\langle \sum_{i=1}^{n} b_i \mathbf{v}_i, \sum_{j=1}^{n} c_j \mathbf{v}_j \right\rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} b_i \overline{c_j} \langle \mathbf{v}_i, \mathbf{v}_j \rangle.$$

If we define a square matrix $A = (a_{ij})$ by $a_{ij} = \langle \mathbf{v}_i, \mathbf{v}_j \rangle$, then $\langle \mathbf{v}, \mathbf{w} \rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} b_i a_{ij} \overline{c_j} = ([\mathbf{v}]_\beta)^T A\, \overline{[\mathbf{w}]_\beta}$. Here A is called the matrix of the inner product (relative to the basis β).

[3]Note that for $f(x) \in C[a, b]$, if there are real numbers $a_n, \ldots, a_0 \in \mathbb{R}$ such that $f(x) = a_n x^n + \cdots + a_0$, we say $f(x)$ is a polynomial function. Since real polynomials are determined by their values on any closed interval, it turns out that $\mathbb{R}[x]$ is isomorphic with the subspace of these polynomial functions on $[a, b]$. We often abuse such an isomorphism and use *polynomials* and *polynomial functions* interchangably. If we were considering polynomial functions on a finite field, this would no longer work. Over finite fields, there are infinitely many polynomials but only finitely many polynomial functions.

The matrix for the dot product (relative to the standard basis) is just the identity matrix. Also, not all matrices arise as matrices of inner products. For example, because of conjugate symmetry we must have $A = A^*$ (i.e., A is *Hermitian*). Over the reals, this amounts to requiring $A = A^T$ (i.e., A is *symmetric*). In addition to the symmetry requirement, the matrix must also satisfy: $\mathbf{x}^* A\,\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{F}^{n\times 1}$ and $\mathbf{x}^* A\,\mathbf{x} = 0$ only if $\mathbf{x} = \mathbf{0}$ (i.e., A is *positive definite*).

Definition: Let V be an inner product space. We say $\mathbf{v}, \mathbf{w} \in V$ are **orthogonal** (or **perpendicular**) if $\langle \mathbf{v}, \mathbf{w}\rangle = 0$. If $S \subseteq V$ and for all $\mathbf{v}, \mathbf{w} \in S$ and $\mathbf{v} \neq \mathbf{w}$ we have $\langle \mathbf{v}, \mathbf{w}\rangle = 0$, then we say S is an **orthogonal** set. If $\mathbf{u} \in V$ and $\|\mathbf{u}\| = 1$, we say $\mathbf{u}$ is a **unit vector**. Notice any non-zero vector $\mathbf{v} \in V$ can be **normalized** (i.e. turned into a unit vector) by rescaling: $\mathbf{u} = \mathbf{v}/\|\mathbf{v}\|$. If S is an orthogonal set of unit vectors, S is called an **orthonormal** set.

The zero vector is weird. Since $\langle \mathbf{0}, \mathbf{v}\rangle = 0$ for any $\mathbf{v} \in V$, the zero vector is orthogonal to every vector – including itself! In fact, by positive definiteness, the zero vector is the *only* vector that is orthogonal to itself.

Lemma: Let S be an orthogonal subset of an inner product space V. Then S is linearly independent if and only if $\mathbf{0} \notin S$. In particular, orthonormal sets are always linearly independent.

Proof: If S is linearly independent, $\mathbf{0} \notin S$ since any set containing the zero vector is linearly dependent. Now suppose $\mathbf{0} \notin S$ and consider $\mathbf{v}_1, \dots \mathbf{v}_n \in S$ (distinct vectors) and $c_1, \dots, c_n \in \mathbb{F}$ such that $c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$. Since S is orthgonal, we have $\langle \mathbf{v}_i, \mathbf{v}_j\rangle = 0$ for $i \neq j$. Thus $0 = \langle \mathbf{0}, \mathbf{v}_j\rangle = \langle c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n, \mathbf{v}_j\rangle = c_1\langle \mathbf{v}_1, \mathbf{v}_j\rangle + \cdots + c_n\langle \mathbf{v}_n, \mathbf{v}_j\rangle = 0 + \cdots + 0 + c_j\langle \mathbf{v}_j, \mathbf{v}_j\rangle + 0 + \cdots + 0$. But $\mathbf{0} \notin S$ means $\mathbf{v}_j \neq \mathbf{0}$ so by positivity $\langle \mathbf{v}_j, \mathbf{v}_j\rangle \neq 0$. Therefore, $c_j = 0$ (for any $j = 1, \dots, n$). Consequently S is linearly independent.

Finally, if S is orthonormal, all its elements are unit vectors (i.e., length 1). This means $\mathbf{0} \notin S$ (and orthonormal implies orthogonal), so S is linearly independent. ■

The above proof indicates that coordinates should be particularly nice when working with orthogonal bases. Let $\beta = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be an orthogonal basis for a finite dimensional inner product space V. Let $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$ for $c_1, \dots, c_n \in \mathbb{F}$ (i.e., the coordinates of $\mathbf{v}$ are $[c_1 \ \dots \ c_n]^T$). Then, just like the calculation in our proof above, $\langle \mathbf{v}, \mathbf{v}_i\rangle = c_1\langle \mathbf{v}_1, \mathbf{v}_i\rangle + \cdots + c_n\langle \mathbf{v}_n, \mathbf{v}_i\rangle = 0 + \cdots + 0 + c_i\langle \mathbf{v}_i, \mathbf{v}_i\rangle + 0 + \cdots + 0$. This means $c_i = \frac{\langle \mathbf{v}, \mathbf{v}_i\rangle}{\langle \mathbf{v}_i, \mathbf{v}_i\rangle}$ so that $\mathbf{v} = \sum_{i=1}^{n} \frac{\langle \mathbf{v}, \mathbf{v}_i\rangle}{\langle \mathbf{v}_i, \mathbf{v}_i\rangle}\mathbf{v}_i$ (i.e., $\mathbf{v}$'s i^{th} β-coordinate is $\frac{\langle \mathbf{v}, \mathbf{v}_i\rangle}{\langle \mathbf{v}_i, \mathbf{v}_i\rangle}$).

When β is orthonormal (i.e., $\langle \mathbf{v}_i, \mathbf{v}_j\rangle = \delta_{ij}$ which is 1 if $i = j$ and 0 otherwise), we simply have $\mathbf{v} = \sum_{i=1}^{n} \langle \mathbf{v}, \mathbf{v}_i\rangle \mathbf{v}_i$ (i.e., $\mathbf{v}$'s i^{th} β-coordinate is $\langle \mathbf{v}, \mathbf{v}_i\rangle$ – that is – just the inner product with the appropriate basis element). For example, the (standard) coordinates of $\mathbf{v} = [1\ 2\ 3] \in \mathbb{R}^{1\times 3}$ are $\mathbf{v} \bullet \mathbf{i} = 1$, $\mathbf{v} \bullet \mathbf{j} = 2$, and $\mathbf{v} \bullet \mathbf{k} = 3$. Also, if β is orthonormal, notice that the matrix of the inner product relative to β is $A = [a_{ij}]$ where $a_{ij} = \langle \mathbf{v}_i, \mathbf{v}_j\rangle = \delta_{ij}$. In other words, the matrix of an inner product relative to an orthonormal basis is just the identity matrix.

Lengths also work nicely with orthogonal sets. Suppose $S = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$ is orthogonal. Then

$$\|\mathbf{w}_1 + \cdots + \mathbf{w}_n\|^2 = \left\langle \sum_{i=1}^{n} \mathbf{w}_i, \sum_{j=1}^{n} \mathbf{w}_j \right\rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} \langle \mathbf{v}_i, \mathbf{v}_j\rangle = \sum_{i=1}^{n} \langle \mathbf{w}_i, \mathbf{w}_i\rangle = \|\mathbf{w}_1\|^2 + \cdots + \|\mathbf{w}_n\|^2$$

where one of the sums drops out because $\langle \mathbf{w}_i, \mathbf{w}_j\rangle = 0$ if $i \neq j$. This then implies that if $\beta = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is an orthonormal basis, then $\|\mathbf{v}\|^2 = \sum_{i=1}^{n} \|\langle \mathbf{v}, \mathbf{v}_i\rangle \mathbf{v}_i\|^2 = \sum_{i=1}^{n} |\langle \mathbf{v}, \mathbf{v}_i\rangle|^2$ (i.e. the length of $\mathbf{v}$ is the sum of the squares of the lengths of its coordinates.

Gram-Schmidt Orthgonalization:

Seeing that orthogonal (or even better orthonormal) bases are so nice. One might ask if these actually exist? Perhaps surprisingly, for finite dimensional inner product spaces, they always do. This is a consequence of the Gram-Schmidt orthgonalization procedure. First, we need to be reminded about orthogonal projections.

Definition: Let V be an inner product space and $\mathbf{v}, \mathbf{w} \in V$ with $\mathbf{v} \neq \mathbf{0}$. Let $\text{proj}_{\mathbf{v}}(\mathbf{w}) = \dfrac{\langle \mathbf{w}, \mathbf{v}\rangle}{\|\mathbf{v}\|^2}\mathbf{v}$. This is the **orthogonal projection** of $\mathbf{w}$ onto $\mathbf{v}$.

If we let $\mathbf{p} = \text{proj}_{\mathbf{v}}(\mathbf{w})$, then $\langle \mathbf{w} - \mathbf{p}, \mathbf{v}\rangle = \langle \mathbf{w}, \mathbf{v}\rangle - \langle \mathbf{p}, \mathbf{v}\rangle = \langle \mathbf{w}, \mathbf{v}\rangle - \frac{\langle \mathbf{w}, \mathbf{v}\rangle}{\|\mathbf{v}\|^2}\langle \mathbf{v}, \mathbf{v}\rangle = \langle \mathbf{w}, \mathbf{v}\rangle - \langle \mathbf{w}, \mathbf{v}\rangle = 0$. In other words, we have $\mathbf{w} = (\mathbf{w} - \mathbf{p}) + \mathbf{p}$ where $\mathbf{w} - \mathbf{p}$ is orthogonal to $\mathbf{v}$ and $\mathbf{p}$ is parallel to (i.e., a scalar multiple of) $\mathbf{v}$. Thus projections give us a way to pull apart vectors into parallel and perpendicular parts.

Gram-Schmidt is amounts to successively removing parallel parts of vectors leaving only the orthogonal bits. Let $\alpha = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$ be a linearly independent set. Recursively define a set $\beta = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ as follows:

$$\mathbf{v}_1 = \mathbf{w}_1 \qquad \mathbf{v}_{j+1} = \mathbf{w}_{j+1} - \text{proj}_{\mathbf{v}_1}(\mathbf{w}_{j+1}) - \cdots - -\text{proj}_{\mathbf{v}_n}(\mathbf{w}_{j+1})$$

This means each new $\mathbf{v}_i$ is $\mathbf{w}_i$ with its orthogonal projections onto previous stuff removed. Notice span$\{\mathbf{w}_1, \ldots, \mathbf{w}_j\}$ = span$\{\mathbf{v}_1, \ldots, \mathbf{v}_j\}$ for each $j = 1, \ldots, n$. We can see this since each new vector $\mathbf{v}_{j+1}$ is equal to $\mathbf{w}_{j+1}$ modulo the span of previous vectors. Thus swapping out $\mathbf{w}_{j+1}$ with $\mathbf{v}_{j+1}$ does not effect the span. Also, we have:

$$\langle \mathbf{v}_{j+1}, \mathbf{v}_k \rangle = \langle \mathbf{w}_{j+1}, \mathbf{v}_k \rangle - \textstyle\sum_{\ell=1}^{j} \langle \text{proj}_{\mathbf{v}_\ell}(\mathbf{w}_{j+1}), \mathbf{v}_k \rangle = \langle \mathbf{w}_{j+1}, \mathbf{v}_k \rangle - \sum_{\ell=1}^{j} \dfrac{\langle \mathbf{w}_{j+1}, \mathbf{v}_\ell \rangle}{\|\mathbf{v}_\ell\|^2} \langle \mathbf{v}_\ell, \mathbf{v}_k \rangle$$

$$= \langle \mathbf{w}_{j+1}, \mathbf{v}_k \rangle - \frac{\langle \mathbf{w}_{j+1}, \mathbf{v}_k \rangle}{\|\mathbf{v}_k\|^2} \langle \mathbf{v}_k, \mathbf{v}_k \rangle = \langle \mathbf{w}_{j+1}, \mathbf{v}_k \rangle - \langle \mathbf{w}_{j+1}, \mathbf{v}_k \rangle = 0$$

where the sum collapses since orthogonality of previous vectors implies $\langle \mathbf{v}_\ell, \mathbf{v}_k \rangle = 0$ for $\ell \neq k$. Thus each new $\mathbf{v}_{j+1}$ is orthogonal to the previous $\mathbf{v}_k$'s.

Theorem: (Gram-Schmidt) Let $\alpha = \{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ be a basis for a finite dimensional inner product space V. Define $\mathbf{v}_1 = \mathbf{w}_1$ and then $\mathbf{v}_{j+1} = \mathbf{w}_{j+1} - \sum_{\ell=1}^{j} \text{proj}_{\mathbf{v}_\ell}(\mathbf{w}_{j+1})$. Then for each $1 \leq j \leq n$, span$\{\mathbf{w}_1, \ldots, \mathbf{w}_j\}$ = span$\{\mathbf{v}_1, \ldots, \mathbf{v}_j\}$. Also, $\beta = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is an orthogonal basis for V. Moreover, normalizing the vectors in β yields an orthonormal basis for V. In particular, every finite dimensional inner product space has an orthonormal basis.

As an addendum to this process, we note that if we apply the Gram-Schmidt process to a linearly *dependent* set, it still works – after a minor modification. Suppose $\mathbf{w}_{j+1}$ depends linearly on previous vectors. Then when we attempt to compute $\mathbf{v}_{j+1}$, its formula will yield $\mathbf{0}$. Ultimately this means we can still apply the process to a linearly dependent set. Modification for linearly dependent sets: If we get $\mathbf{v}_j = \mathbf{0}$, then we should toss out this vector (and $\mathbf{w}_j$) and continue as usual. Thus we detect and then ignore the dependent stuff. We also note that some people, calculate each new vector and immediately normalize (to be a unit vector). If one does this, $\|\mathbf{v}_\ell\|^2 = 1$ in the above calculations and our output is already an orthonormal basis.

In the end there is not a lot of variety in the world of finite dimensional inner product spaces. First, we define the notion of isometry (what it means to be essentially the "same" inner product space).

Definition: A mapping $T : V \to W$ is **morphism** of inner product spaces if T is linear and it preserves inner products. Specifically, for all $\mathbf{v}_1, \mathbf{v}_2 \in V$, we have $\langle T(\mathbf{v}_1), T(\mathbf{v}_2) \rangle = \langle \mathbf{v}_1, \mathbf{v}_2 \rangle$ (the first inner product is the inner product in W and the second is the inner product in V). When T is an isomorphism (i.e., an invertible morphism), we say T is an **isometry**.

The Gram-Schmidt process guarantees that every finite dimensional inner product space V has an orthonormal basis β. The coordinate map $[\cdot]_\beta : V \to \mathbb{F}^{n\times 1}$ is easily seen to be an isometry when $\mathbb{F}^{n\times 1}$ is given the standard inner product (i.e., the dot product). Thus *every n-dimensional real inner product space is essentially just $\mathbb{R}^n$* (with its standard dot product) in disguise. Likewise, every n-dimensional complex inner product space is essentially just $\mathbb{C}^n$ (with its standard dot product). Just as with finite dimensional vector spaces, when dealing with finite dimensional inner product spaces, every computation and question can be realized in terms of column vectors (with the usual addition, scalar multiplication, and dot product).

Example: Let $S = \{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\} = \{(4, -1, 1), (8, -1, 3), (5, -2, -4)\} \subseteq \mathbb{R}^3$.

- $\mathbf{v}_1 = \mathbf{w}_1 = (4, -1, 1)$
- $\mathbf{v}_2 = \mathbf{w}_2 - \dfrac{\mathbf{w}_2 \bullet \mathbf{v}_1}{\|\mathbf{v}_1\|^2} \mathbf{v}_1 = (8, -1, 3) - \dfrac{36}{18}(4, -1, 1) = (0, 1, 1)$
- $\mathbf{v}_3 = \mathbf{w}_3 - \dfrac{\mathbf{w}_3 \bullet \mathbf{v}_1}{\|\mathbf{v}_1\|^2} \mathbf{v}_1 - \dfrac{\mathbf{w}_3 \bullet \mathbf{v}_2}{\|\mathbf{v}_2\|^2} \mathbf{v}_2 = (5, -2, -4) - \dfrac{18}{18}(4, -1, 1) - \dfrac{-6}{2}(0, 1, 1) = (1, 2, -2)$

We have span(S) = span$\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ but our new spanning set is an orthogonal set (without zero vector). This implies that both S and our new set are linearly independent sets. Since $\dim(\mathbb{R}^3) = 3$, these are bases for $\mathbb{R}^3$. Specifically, $\{(4, -1, 1), (0, 1, 1), (1, 2, -2)\}$ is an orthogonal basis for $\mathbb{R}^3$. We can normalize our set and have that $\left\{ \frac{1}{3\sqrt{2}}(4, -1, 1), \frac{1}{\sqrt{2}}(0, 1, 1), \frac{1}{3}(1, 2, -2) \right\}$ is an orthonormal basis for $\mathbb{R}^3$.

Example: Consider $S = \{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \mathbf{w}_4\} = \{(1, 0, 0, i), (1+i, 1, 1, 1+i), (3+i, 1, 1, 1+3i), (0, 1+2i, 0, 1)\} \subseteq \mathbb{C}^4$.

- $\mathbf{v}_1 = \mathbf{w}_1 = (1, 0, 0, i)$
- $\mathbf{v}_2 = \mathbf{w}_2 - \dfrac{\mathbf{w}_2 \bullet \mathbf{v}_1}{\|\mathbf{v}_1\|^2} \mathbf{v}_1 = (1+i, 1, 1, 1+i) - \dfrac{2}{2}(1, 0, 0, i) = (i, 1, 1, 1)$

- Notice that $\mathbf{w}_3 - \dfrac{\mathbf{w}_3 \bullet \mathbf{v}_1}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \dfrac{\mathbf{w}_3 \bullet \mathbf{v}_2}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 = (3+i,1,1,1+3i) - \dfrac{6}{2}(1,0,0,i) - \dfrac{4}{4}(i,1,1,1) = (0,0,0,0)$.

  This means that $\mathbf{w}_3 \in \text{span}\{\mathbf{v}_1, \mathbf{v}_2\} = \text{span}\{\mathbf{w}_1, \mathbf{w}_2\}$. So we can toss out $\mathbf{w}_3$ (and $\mathbf{0}$).

- $\mathbf{v}_3 = \mathbf{w}_4 - \dfrac{\mathbf{w}_4 \bullet \mathbf{v}_1}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \dfrac{\mathbf{w}_4 \bullet \mathbf{v}_2}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 = (0,1+2i,0,1) - \dfrac{-i}{2}(1,0,0,i) - \dfrac{2+2i}{4}(i,1,1,1) = \dfrac{1}{2}(1,1+3i,-1-i,-i)$

We have $\text{span}(S) = \text{span}\left\{(1,0,0,i),(i,1,1,1),\frac{1}{2}(1,1+3i,-1-i,-i)\right\}$ where our new set is an orthogonal basis for $\text{span}(S)$. If we normalize our new vectors, we have $\alpha = \left\{\frac{1}{\sqrt{2}}(1,0,0,i),\frac{1}{2}(i,1,1,1),\frac{1}{\sqrt{14}}(1,1+3i,-1-i,-i)\right\}$ which is an orthonormal basis for $\text{span}(S)$.

Example: Consider the inner product $\langle f, g\rangle = \displaystyle\int_{-1}^{1} f(x)g(x)\,dx$ defined on $\mathbb{R}[x]$. If we run the Gram-Schmidt algorithm (unnormalized) on $\{1, x, x^2, \dots\}$ and then rescale them so that $\|P_n(x)\|^2 = \dfrac{2}{2n+1}$, we will get a collection of polynomials known as *Legendre polynomials*. Let us denote theses by $P_0(x), P_1(x)$, etc. We calculate the first few Legendre polynomials:

- $f_0(x) = 1$. We need $2(?)^2 = (?)^2 \int_{-1}^{1} 1^2\,dx = (?)^2\|f_0(x)\|^2 = \dfrac{2}{2(0)+1}$, so $? = 1$ and thus $P_0(x) = f_0(x) = 1$.

- $f_1(x) = x - \dfrac{\langle x, 1\rangle}{\|1\|^2}1 = x - \dfrac{\int_{-1}^{1} x \cdot 1\,dx}{\int_{-1}^{1} 1^2\,dx}1 = x - \dfrac{0}{2}1 = x$. Using the Legendre normalization, we need

  $\dfrac{2}{3}(?)^2 = \|(?)^2 f_1(x)\|^2 = \dfrac{2}{2(1)+1}$ so $? = 1$ and thus $P_1(x) = f_1(x) = x$.

- $f_2(x) = x^2 - \dfrac{\langle x^2, 1\rangle}{\|1\|^2}1 - \dfrac{\langle x^2, x\rangle}{\|x\|^2}x = x^2 - \dfrac{\int_{-1}^{1} x^2 \cdot 1\,dx}{\int_{-1}^{1} 1^2\,dx}1 - \dfrac{\int_{-1}^{1} x^2 \cdot x\,dx}{\int_{-1}^{1} x^2\,dx}x = x^2 - \dfrac{2/3}{2}1 - \dfrac{0}{2/3}x = x^2 - \dfrac{1}{3}$.

  Again, we need $\dfrac{8}{45}(?)^2 = (?)^2 \int_{-1}^{1}(x^2 - 1/3)^2\,dx = \|?\, f_2(x)\|^2 = \dfrac{2}{2(2)+1} = \dfrac{2}{5}$ so $(?)^2 = 9/4$ and thus $? = 3/2$.

  Therefore, $P_2(x) = (3/2)f_2(x) = \dfrac{3}{2}x^2 - \dfrac{1}{2}$.

Connection with Dual Spaces:

Notice that since the inner product is linear in its first slot, for any $\mathbf{w} \in W$ we have $\langle \cdot, \mathbf{w}\rangle : W \to \mathbb{F}$ is a linear map. In other words, the mapping f defined by $f(\mathbf{v}) = \langle \mathbf{v}, \mathbf{w}\rangle$ is a dual vector (i.e., $\langle \cdot, \mathbf{w}\rangle \in W^*$). In fact, when $\beta = \{\mathbf{w}_1, \dots, \mathbf{w}_m\}$ is an orthonormal basis, $\beta^* = \{\langle \cdot, \mathbf{w}_1\rangle, \dots, \langle \cdot, \mathbf{w}_m\rangle\}$ is its dual basis. Why? Denoting $\mathbf{w}_j^* = \langle \cdot, \mathbf{w}_j\rangle$, we have $\mathbf{w}_j^*(\mathbf{w}_i) = \langle \mathbf{w}_i, \mathbf{w}_j\rangle = \delta_{ij}$.

Suppose $T : V \to W$ is a linear transformation of finite dimensional vector spaces where $\alpha = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a basis for V and $\beta = \{\mathbf{w}_1, \dots, \mathbf{w}_m\}$ is an orthogonal basis for W. Then the (i,j)-entry of the coordinate matrix $[T]_\alpha^\beta$ is $\mathbf{w}_i^*(T(\mathbf{v}_j)) = \langle T(\mathbf{v}_j), \mathbf{w}_i\rangle$.

Working over a real field, we can make this cleaner. If $\mathbb{F}$ is a subfield of the real numbers, the inner product is bilinear and symmetric. Thus the (i,j)-entry of the coordinate matrix $[T]_\alpha^\beta$ (still assuming β is orthonormal) could be written as $\langle \mathbf{w}_i, T(\mathbf{v}_j)\rangle$. Also, in this case, the association of $\mathbf{w} \in W$ with $\langle \mathbf{w}, \cdot\rangle = \langle \cdot, \mathbf{w}\rangle \in W^*$ is actually an isomorphism. For any finite dimensional vector space, W, we have $W \cong W^*$. But this isomorphism is not "natural" in a some technical sense I won't get into (the isomorphism is basis dependent – we usually show $W \cong W^*$ by choosing a basis for W and relating it to a dual basis for W^*). It turns out that picking an isomorphism between W and W^* is almost the same thing as choosing an inner product on W. For a finite dimensional real inner product space, its inner product can allow us (with some care) to treat vectors in our space and its dual as interchangeable.