INNER PRODUCT SPACES

TIM SMITS

When we started our study of vector spaces, we had a goal in mind: find objects that generalized the algebraic structure on Euclidean space $\mathbb{R}^n$. However, if the ultimate goal of linear algebra is to fully generalize Euclidean space, there's something major that still hasn't been abstracted: the *geometry* of $\mathbb{R}^n$. The definition of an abstract vector space V does not include notions of length, distance, or angles, and therefore no concept of geometry. In order for a vector space to truly "act" Euclidean, we need to add more structure.

BASIC DEFINITIONS AND EXAMPLES

Throughout this document, we assume $F = \mathbb{R}$ or $F = \mathbb{C}$, and V is an inner product space over F.

Definition 0.1. An **inner product** $\langle -, - \rangle : V \times V \to F$ is a function that satisfies the following properties:

1. $\langle x + z, y \rangle = \langle x, y \rangle + \langle z, y \rangle$ for all $x, y, z \in V$
2. $\langle cx, y \rangle = c\langle x, y \rangle$ for all $x, y \in V$ and $c \in F$
3. $\overline{\langle x, y \rangle} = \langle y, x \rangle$ for all $x, y \in V$
4. $\langle x, x \rangle > 0$ for all $x \neq 0 \in V$.

An **inner product space** is a pair $(V, \langle -, - \rangle)$, i.e. a vector space V with a choice of inner product on V. From the conjugate symmetry of the inner product, we deduce the following basic properties:

Proposition 1. *Let V be an inner product space. Then the following hold:*

(a) $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$ *for all* $x, y, z \in V$.
(b) $\langle x, cy \rangle = \bar{c}\langle x, y \rangle$ *For all* $x, y \in V$ *and* $c \in F$.
(c) $\langle x, 0 \rangle = \langle 0, x \rangle = 0$ *for all* $x \in V$.
(d) $\langle x, x \rangle = 0$ *if and only if* $x = 0$.
(e) *If* $\langle x, y \rangle = \langle x, z \rangle$ *for all* $x \in V$, *then* $y = z$.

Proof. These are routine verifications and are omitted. □

The idea of an inner product is to generalize the dot product on $\mathbb{R}^n$ or $\mathbb{C}^n$. Having an inner product gives us a notion of length:

Definition 0.2. The **norm** of $x \in V$, denoted $\|x\|$, is defined by $\|x\| = \sqrt{\langle x, x \rangle}$.

Proposition 2. *Let V be an inner product space. Then the norm $\| \cdot \|$ satisfies the following properties:*

(a) $\|cx\| = |c|\|x\|$ *for all* $x \in V$, $c \in F$.
(b) $\|x\| = 0$ *if and only if* $x = 0$.
(c) *(Cauchy-Schwarz) For all* $x, y \in V$ $|\langle x, y \rangle| \leq \|x\|\|y\|$ *and equality holds if and only if* $x = cy$ *for some* $c \in F$.

(d) *(Triangle inequality) For all* $x, y \in V$, $\|x+y\| \leq \|x\| + \|y\|$ *and equality holds if and only if* $x = cy$ *for some* $c \in F$.

Proof.

(a) $\|cx\| = \langle cx, cx\rangle = c\bar{c}\langle x, x\rangle = |c|\|x\|$.
(b) This is (d) from the above proposition.
(c) If $y = 0$ this is obvious, so assume $y \neq 0$. For $c \in F$, we have $0 \leq \|x - cy\|^2 = \langle x - cy, x - cy\rangle = \langle x, x\rangle - \bar{c}\langle x, y\rangle - c\langle y, x\rangle + c\bar{c}\langle y, y\rangle$ by expanding out the inner product. In particular, setting $c = \frac{\langle x,y\rangle}{\langle y,y\rangle}$, this becomes $0 \leq \|x\|^2 - 2\frac{|\langle x,y\rangle|^2}{\|y\|^2} + \frac{|\langle x,y\rangle|^2}{\|y\|^2}$, so that $0 \leq \|x\|^2 - \frac{|\langle x,y\rangle|^2}{\|y\|^2}$. This then gives $|\langle x, y\rangle| \leq \|x\|\|y\|$ as desired. The case of equality is left as an exercise.
(d) We have $\|x + y\|^2 = \langle x + y, x + y\rangle = \langle x, x\rangle + \langle y, x\rangle + \langle x, y\rangle + \langle y, y\rangle = \|x\|^2 + 2\mathrm{Re}(\langle x, y\rangle) + \|y\|^2$ by expanding out the inner product and using conjugate symmetry. Since $2\mathrm{Re}(\langle x, y\rangle) \leq 2|\langle x, y\rangle| \leq 2\|x\|\|y\|$ by Cauchy-Schwarz, we then see $\|x + y\|^2 \leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2$, so take a square root to finish up. The equality case is again left as an exercise.

□

We now give some standard examples of inner product spaces.

Example 0.3. If V has an inner product $\langle -, -\rangle$ then for any subspace W of V, $\langle -, -\rangle$ is still an inner product on W.

Example 0.4. Set $V = \mathbb{R}^n$ and let $\cdot$ be the usual dot product, $(a_1, \ldots, a_n) \cdot (b_1, \ldots, b_n) = a_1b_1 + \ldots + a_nb_n$. This makes V a real inner product space. If instead $V = \mathbb{C}^n$, we define the dot product to be $(a_1, \ldots, a_n) \cdot (b_1, \ldots, b_n) = a_1\overline{b_1} + \ldots + a_n\overline{b_n}$. This makes V a complex inner product space. As an example in $\mathbb{C}^2$, we have $(1+2i, 3-i)\cdot(2, i) = 2(1+2i) - i(3-i) = 1 - i$.

Example 0.5. If V is a finite dimensional F-vector space, we can always give V the structure of an inner product space as follows. Say that $\dim_F(V) = n$, and fix an isomorphism $\varphi : V \to F^n$. Define an inner product $\langle -, -\rangle$ on V by $\langle v, w\rangle = \varphi(v) \cdot \varphi(w)$, where the dot product on the right hand side happens in F^n.

Example 0.6. Set $V = C([a, b], \mathbb{C})$ and define $\langle f, g\rangle = \int_a^b f(t)\overline{g(t)}\, dt$. Then calculus says this makes V an inner product space. In $C([-\pi, \pi], \mathbb{C})$ with $f = 1 + 2x$ and $g = \cos(x)$, one can check that $\|f\| = \sqrt{2\pi + \frac{8\pi^3}{3}}$, $\|g\| = \sqrt{\pi}$, and that $\langle f, g\rangle = 0$.

Example 0.7. Let $V = M_n(\mathbb{C})$ and define $\langle A, B\rangle = \mathrm{tr}(B^*A)$, where $(A^*)_{ij} = \overline{A_{ji}}$ is the **conjugate transpose** of A. Then by linearity of tr and definition of B^*, one sees that this defines an inner product. For any $A \in M_n(\mathbb{C})$, we see that the ij-th entry of A^*A is simply the i-th row of A^* dotted with j-column of A. In particular, if v_i is the i-th column of A, then $(A^*A)_{ii} = \|v_i\|^2$, so that $\|A\| = \sqrt{\|v_1\|^2 + \ldots + \|v_n\|^2}$ where $v_1, \ldots, v_n$ are the columns of A. In $M_2(\mathbb{C})$, set $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ and $B = \begin{pmatrix} i & 0 \\ 1+i & -i \end{pmatrix}$. Then we see $\|A\| = \sqrt{30}$, $\|B\| = \sqrt{2}$, and $\langle A, B\rangle = 3$.

Example 0.8. Consider the sequence space $\mathbb{C}^\infty$. Let $\ell^2 = \{(a_n) \in \mathbb{C}^\infty : \sum_{n=1}^\infty |a_n|^2 < \infty\}$. For sequences $(a_n), (b_n)$, note that for any n, we have $2|a_n||b_n| \leq |a_n|^2 + |b_n|^2$. Combined with the triangle inequality, we then have $|a_n + b_n|^2 \leq (|a_n| + |b_n|)^2 \leq 2|a_n|^2 + 2|b_n|^2$, which then immediately tells us that ℓ^2 is a $\mathbb{C}$-vector space. Define $\langle (a_n),(b_n)\rangle = \sum_{n=1}^\infty a_n\overline{b_n}$. If this expression is finite, then it's clear that this makes ℓ^2 a complex inner product space, since it's just the usual dot product extended to vectors with infinitely many coordinates. To see this is the case, for any N, we have $\sum_{n=1}^N |a_n||\overline{b_n}| \le \sqrt{\sum_{n=1}^N |a_n|^2 \sum_{n=1}^N |b_n|^2}$ by applying Cauchy-Schwarz to the vector space $\mathbb{R}^N$ with vectors $(|a_1|,\ldots,|a_N|)$ and $(|b_1|,\ldots,|b_N|)$ (here we use that for complex numbers, $|\bar{z}| = |z|$). Since $(a_n),(b_n)\in\ell^2$, taking $N\to\infty$ says the right hand side of the above inequality converges to something finite, so that $\lim_{N\to\infty}\sum_{n=1}^N |a_n||\overline{b_n}| = \sum_{n=1}^\infty |a_n||\overline{b_n}| < \infty$. This says $\sum_{n=1}^\infty a_n\overline{b_n}$ converges absolutely, so that $\sum_{n=1}^\infty a_n\overline{b_n}$ converges. We then have proved that this is indeed an inner product. This space is important in functional analysis.

Orthogonality

In $\mathbb{R}^n$, one of the most important properties of the dot product was that it was able to measure the angle between two vectors: this was detected by the quantity $\frac{x\cdot y}{\|x\|\|y\|}$. For a general inner product space V, it doesn't make sense to define general angles, since the expression $\frac{\langle x,y\rangle}{\|x\|\|y\|}$ may be a complex number. However, we may still make sense of orthogonality.

Definition 0.9. Vectors $x,y\in V$ are called **orthogonal** if $\langle x,y\rangle = 0$. A subset S of V is called orthogonal if any two vectors in S are orthogonal, and S is called **orthonormal** if S is orthogonal and $\|x\| = 1$ for all $x\in S$.

Since inner products have a notion of orthogonality, the Pythagorean theorem is still true:

Theorem 0.10 (Pythagorean theorem). *Let $x,y\in V$ be orthogonal. Then $\|x+y\|^2 = \|x\|^2+\|y\|^2$.*

Proof. $\|x+y\|^2 = \langle x+y,x+y\rangle = \langle x,x\rangle+\langle x,y\rangle+\langle y,x\rangle+\langle y,y\rangle = \langle x,x\rangle+\langle y,y\rangle = \|x\|^2+\|y\|^2$ because $\langle x,y\rangle = \langle y,x\rangle = 0$ by assumption. □

One of the reasons that we impose the extra structure of an inner product is that they become very nice to work with: orthogonality makes linear independence easy to check, as well as finding the coordinates of a vector with respect to some basis.

Proposition 3. *Let $\{v_1,\ldots,v_k\}$ be an orthogonal subset of non-zero vectors. Then $\{v_1,\ldots,v_k\}$ is linearly independent.*

Proof. If $c_1v_1+\ldots+c_kv_k = 0$ is a linear dependence relation among the v_i's , then $\langle c_1v_1+\ldots+c_kv_k, v_i\rangle = \langle 0,v_i\rangle = 0$. On the other hand, $\langle c_1v_1+\ldots+c_kv_k,v_i\rangle = c_1\langle v_1,v_i\rangle+\ldots+c_k\langle v_k,v_i\rangle = c_i\|v_i\|^2$ by orthogonality, so $c_i = 0$. This says that $\{v_1,\ldots,v_k\}$ is linearly independent. □

Proposition 4. *Let $S = \{v_1,\ldots,v_k\}$ be an orthogonal subset of non-zero vectors. Then if $x = c_1v_1+\ldots+c_kv_k$, $c_i = \frac{\langle x,v_i\rangle}{\|v_i\|^2}$.*

Proof. Taking an inner product with v_i says $\langle x,v_i\rangle = c_i\langle v_i,v_i\rangle = c_i\|v_i\|^2$ by orthogonality. □

In particular, the above says that if we have a basis β for V consisting of orthogonal vectors, then finding the coordinates $[x]_\beta$ is reduced to an inner product computation. If V is finite dimensional, is it always possible to find an orthonormal basis? The answer is yes, and follows from a more general result.

Theorem 0.11 (Gram-Schmidt process)**.** *Let $S = \{w_1, \ldots, w_m\}$ be a linearly independent subset of V. Define $S' = \{v_1, \ldots, v_m\}$ where $v_1 = w_1$ and $v_k = w_k - \sum_{j=1}^{k-1} \frac{\langle w_k, v_j \rangle}{\|v_j\|^2} v_j$ for $2 \leq j \leq m$. Then S' is an orthogonal subset of non-zero vectors and Span$(S) =$ Span(S').*

Proof. The proof is by induction on m. If $m = 1$, then the result is immediate. Now suppose for every linearly independent set of size $m-1$ the theorem is true. Define $S_k = \{w_1, \ldots, w_k\}$ and $S'_k = \{v_1, \ldots, v_k\}$, so that in particular the set $S'_{m-1} = \{v_1, \ldots, v_{m-1}\}$ is orthogonal. We will check that the theorem is true for $S' = S'_{m-1} \cup \{v_m\}$ where v_m is defined as in the statement of the theorem. If $v_m = 0$, this says $w_m \in \text{Span}(S'_{m-1})$, but $\text{Span}(S'_{m-1}) = \text{Span}(S_{m-1})$ by induction hypothesis, which contradicts that S is linearly independent. Therefore $v_m \neq 0$. We then see $\langle v_m, v_i \rangle = \langle w_m, v_i \rangle - \langle w_m, v_i \rangle = 0$ since by assumption S'_{m-1} is orthogonal, so that S' is therefore orthogonal. As $w_i \in \text{Span}(S'_{m-1})$ for all $1 \leq i \leq m-1$ by assumption, combined with the definition of v_m we get $w_i \in \text{Span}(S')$ for all i so that $\text{Span}(S) = \text{Span}(S')$ as desired. □

As an immediate corollary to the Gram-Schmidt process, we get the following:

Corollary 0.12. *If V is a finite dimensional inner product space, then V has an orthonormal basis.*

Proof. Apply the Gram-Schmidt process to a basis of V to get a basis of orthogonal vectors. Then normalize. □

Example 0.13. Set $V = \mathbb{R}^3$ and $\beta = \{(1,1,1), (0,1,1), (0,0,1)\} = \{w_1, w_2, w_3\}$, which is a basis of $\mathbb{R}^3$. To construct an orthogonal basis, set $v_1 = (1,1,1)$. Then $v_2 = w_2 - \frac{\langle w_2, v_1 \rangle}{\|v_1\|^2} v_1 = w_2 - \frac{2}{3} v_1 = (-2/3, 1/3, 1/3)$, and $v_3 = w_3 - \frac{\langle w_3, v_1 \rangle}{\|v_1\|^2} v_1 - \frac{\langle w_3, v_2 \rangle}{\|v_2\|^2} v_2 = w_3 - \frac{1}{3} v_1 - \frac{1}{2} v_2 = (0, -1/2, 1/2)$. This produces an orthogonal basis, so normalizing each vector with give an orthonormal basis. We see $\|v_1\| = \sqrt{3}$, $\|v_2\| = \sqrt{\frac{2}{3}}$, and $\|v_3\| = \frac{1}{\sqrt{2}}$. Then $\{(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}), (-\sqrt{\frac{2}{3}}, \sqrt{\frac{1}{6}}, \sqrt{\frac{1}{6}}), (0, -\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})\}$ is an orthonormal basis of $\mathbb{R}^3$.

Example 0.14. Set $V = P_2(\mathbb{R})$,which may be viewed as a subspace of $C([-1,1])$ with the inner product $\langle f, g \rangle = \int_{-1}^{1} f(x)g(x)\,dt$. Let $\beta = \{1, x, x^2\} = \{w_1, w_2, w_3\}$ be the standard basis of V. To produce an orthonormal basis, we use Gram-Schmidt. The vectors 1 and x are already orthogonal, so we do not need to compute v_1 and v_2. Then $v_3 = w_3 - \frac{\langle w_3, v_1 \rangle}{\|v_1\|^2} v_1 - \frac{\langle w_3, v_2 \rangle}{\|v_2\|^2} v_2 = x^2 - \frac{1}{3}$, so $\{1, x, x^2 - \frac{1}{3}\}$ is an orthogonal basis. We compute $\|1\| = \sqrt{2}$, $\|x\| = \sqrt{\frac{2}{3}}$, and $\|x^2 - \frac{1}{3}\| = \sqrt{\frac{8}{45}}$. This produces an orthonormal basis $\{\frac{1}{\sqrt{2}}, \sqrt{\frac{3}{2}} x, \sqrt{\frac{5}{8}}(3x^2 - 1)\}$. These are the first three *Legendre polynomials*, which have applications in physics. Repeating this process with the basis $\beta = \{1, x, \ldots, x^n\}$ of $P_n(\mathbb{R})$ allows one to compute the n-th Legendre polynomial.

Definition 0.15. Let $S \subset V$ be a subset. The **orthogonal complement** of S, denoted $S^\perp$ is defined by $S^\perp = \{v \in V : \langle v, x \rangle = 0 \text{ for all } x \in S\}$.

It's an easy verification that $S^\perp$ is always a subspace of V. When S itself is a subspace, we have the following decomposition:

Theorem 0.16. *Let $W \subset V$ be a finite dimensional subspace. Then $V = W \oplus W^\perp$.*

Proof. Let $\{w_1, \ldots, w_k\}$ be an orthonormal basis of W. We will try to find a vector $w \in W$ such that $x = w + (x - w)$ with $x - w \in W^\perp$. Write $w = c_1w_1 + \ldots + c_kw_k$. If $x - w \in W^\perp$, then necessarily, $0 = \langle x - w, w_i\rangle = \langle x - c_1w_1 - \ldots - c_kw_k, w_i\rangle = \langle x, w_i\rangle - c_i\|w_i\|^2$. Since $\|w_i\| = 1$, this says $c_i = \langle x, w_i\rangle$, so this choice of coefficients gives us the vector w that works. This says $V = W + W^\perp$. If $w \in W \cap W^\perp$, then $\langle w, w\rangle = 0$ so that $w = 0$ says the sum is direct. □

An immediate consequence is the following dimension formula:

Corollary 0.17. *If V is finite dimensional,* $\dim(V) = \dim(U) + \dim(U^\perp)$.

Using this, we easily get the following:

Proposition 5. *Let V be finite dimensional. Then $(W^\perp)^\perp = W$.*

Proof. Set $n = \dim(V)$. Since $W \subset (W^\perp)^\perp$, we get $\dim(W) \leq \dim(W^\perp)^\perp$. This then says $n = \dim(W) + \dim(W^\perp) \leq \dim((W^\perp)^\perp) + \dim(W^\perp) = n$ so that $\dim(W) = \dim((W^\perp)^\perp)$ gives $W = (W^\perp)^\perp$. □

Example 0.18. Let $V = \mathbb{R}^3$ and $W = \text{Span}\{v_1\}$ where $v_1 = (1, 1, 1)$. Then $W^\perp = \{(x, y, z) : (x, y, z) \cdot (1, 1, 1) = 0\}$, i.e. $W^\perp$ is simply the plane $x + y + z = 0$.

Example 0.19. Let $V = \mathbb{R}^4$ and $W = \text{Span}\{v_1, v_2\}$ where $v_1 = (1, 2, 3, -4)$ and $v_2 = (-5, 4, 3, 2)\}$. If $x = (x, y, z, t)$ is in $U^\perp$, we see that $Ax = 0$, where $A = \begin{pmatrix} 1 & 2 & 3 & -4 \\ -5 & 4 & 3 & 2 \end{pmatrix}$. Using row reduction, one can easily compute $U^\perp = \ker(A) = \text{Span}\{(-3, -9, 7, 0), (-10, 9, 0, 7)\}$.

Definition 0.20. Let W be a subspace with an orthonormal basis $\{w_1, \ldots, w_k\}$. Define the **orthogonal projection** onto W, P_W, by $P_W(x) = \langle x, w_1\rangle w_1 + \ldots + \langle x, w_k\rangle w_k$.

Given $x \in V$, The orthogonal projection $P_W(x)$ has the property that it is the vector in W that is closest to x:

Theorem 0.21. *Let $x \in V$. Then $\|x - y\| \geq \|x - P_W(x)\|$ for all $y \in W$.*

Proof. Write $x = P_W(x) + z$ where $z \in W^\perp$. Then for any $y \in W$, we have $x - y = (P_W(x)-y)+z$. Then we see z is orthogonal to $P_W(x)-y$, so $\|x-y\|^2 = \|P_W(x)-y\|^2+\|z\|^2 \geq \|P_W(x) - y\|^2$. □

Example 0.22. Let $V = \mathbb{R}^3$, and set $v = (1, 2, 3)$. What's the minimal distance from v to a point on the plane $W : x + 2y + z = 0$? A basis of W can be easily computed as $\{(-2, 1, 0), (-1, 0, 1)\} = \{w_1, w_2\}$. Running Gram-Schmidt gives an orthogonal basis of $\{v_1, v_2\} = \{(-2, 1, 0), (-5, -2, 5)\}$. The minimal distance the the plane is given by the quantity $\|v - P_W(v)\|$. One can check that $P_W(v) = \frac{5}{3}v_2$, so $v - P_W(v) = (4/3, 8/3, 4/3)$ which has length $\frac{4\sqrt{6}}{3}$.

Example 0.23. Set $W = P_2(\mathbb{R})$ viewed as a subspace of $V = C([-1, 1])$ with the inner product $\langle f, g\rangle = \int_{-1}^{1} f(x)g(x)\,dx$. With $f(x) = e^x$, which polynomial $p(x)$ of degree at most 2 minimizes the quantity $\int_{-1}^{1}(e^x - p(x))^2\,dt$, and what is this value? Equivalently, what is the minimizer of $\|e^x - p(x)\|$? We saw before than an orthonormal basis of $P_2(\mathbb{R})$ with respect to this inner product is given by the Legendre polynomials, with basis $\{\frac{1}{\sqrt{2}}, \sqrt{\frac{3}{2}}x, \sqrt{\frac{5}{8}}(3x^2 - 1)\}$, so the minimizer is just the orthogonal projection of e^x onto W. This is given by $p(x) =$

$\langle e^x, \frac{1}{\sqrt{2}}\rangle\frac{1}{\sqrt{2}} + \langle e^x, \sqrt{\frac{3}{2}}x\rangle\sqrt{\frac{3}{2}}x + \langle e^x, \sqrt{\frac{5}{8}}(3x^2-1)\rangle\sqrt{\frac{5}{8}}(3x^2-1) = (\frac{15e}{4} - \frac{105}{4e})x^2 + \frac{3}{e}x + (\frac{33}{4e} - \frac{3e}{4})$. Numerically, the actual minimal value of the integral is $\approx .00144$.

The adjoint of a linear operator

Definition 0.24. The **dual space** of V, denoted V^* is defined as $V^* = \text{Hom}_F(V, F)$. An element $\varphi \in V^*$ is called a **linear functional**.

If V is finite dimensional, then we have seen that $V \cong V^*$. However, this isomorphism is not "natural" in the sense that it requires picking a basis if V. However, when V is an inner product space, the isomorphism *is* natural:

Theorem 0.25 (Riesz Representation Theorem)**.** *Let V be a finite dimensional inner product space. Then the map $\Phi : V \to V^*$ given by $\Phi(v) = \varphi_v$ is an isomorphism, where $\varphi_v(x) = \langle x, v\rangle$.*

Proof. First, we show that Φ is linear. For $x, y \in V$, We have $\Phi(x + y) = \varphi_{x+y}$. For any $z \in V$, we have $\varphi_{x+y}(z) = \langle z, x+y\rangle = \langle z, x\rangle + \langle z, y\rangle = \varphi_x(z) + \varphi_y(z)$, so that $\varphi_{x+y} = \varphi_x + \varphi_y$. This then says that $\Phi(x + y) = \Phi(x) + \Phi(y)$. Similarly, we conclude that for any $c \in F$, $\Phi(cx) = c\Phi(x)$, so that Φ is linear. Now suppose that $\Phi(x) = 0$. This says that $\varphi_x(z) = 0$ for all $z \in V$, i.e. $\langle z, x\rangle = 0$ for all $z \in V$. Picking $z = x$, we get $\|x\|^2 = 0$, so that $x = 0$. This says that Φ is injective, and since $\dim V = \dim V^*$, we conclude that Φ is an isomorphism. □

The Riesz Representation Theorem says the structure of the dual space of a finite dimensional inner product space is very rigid: for any linear functional $\varphi \in V^*$, the surjectivity of the map Φ in the above proof says there is a vector $v \in V$ such that $\varphi = \langle -, v\rangle$. This is very important in functional analysis (where it holds in a more general setting), but for our purposes, we will only need it for the following:

Definition 0.26. Let V be a finite dimensional inner product space. The **adjoint** of a linear operator $T : V \to V$, denoted T^* is defined via the relation $\langle T(x), y\rangle = \langle x, T^*(y)\rangle$ for all $x, y \in V$.

Proposition 6. *The adjoint T^* of a linear operator T exists and is unique, and furthermore $T^* \in Hom_F(V, V)$.*

Proof. Define $\varphi_y(x) = \langle T(x), y\rangle$. Then $\varphi_y(x + z) = \langle T(x + z), y\rangle = \langle T(x) + T(z), y\rangle = \langle T(x), y\rangle + \langle T(z), y\rangle = \varphi_y(x) + \varphi_y(z)$. Similarly, $\varphi_y(cx) = c\varphi_y(x)$ for $x, z \in V$ and $c \in F$, so $\varphi_y(x)$ is a linear functional. By the Riesz Representation Theorem, $\varphi_y(x) = \langle x, y'\rangle$ for some $y' \in V$. Define a map $T^* : V \to V$ by $T^*(y) = y'$. By definition T^* satisfies the desired property. If there is another function $S : V \to V$ such that $\langle T(x), y\rangle = \langle x, S(y)\rangle$ for all x, y, this says $\langle x, T^*(y)\rangle = \langle x, S(y)\rangle$ for all x, y so that $T^* = S$. Finally, it remains to show linearity. We see $\langle x, T^*(y + z)\rangle = \langle T(x), y + z\rangle = \langle T(x), y\rangle + \langle T(x), z\rangle = \langle x, T^*(y)\rangle + \langle x, T^*(z)\rangle = \langle x, T^*(y) + T^*(z)\rangle$ for all $x, y, z \in V$. This says $T^*(y + z) = T^*(y) + T^*(z)$. Similarly one can check $T^*(cy) = cT^*(y)$, so that $T^* \in \text{Hom}_F(V, V)$.

□

Although it may not be clear from the above defintion, the point of the adjoint is that it's a analogous operation on linear operators to taking a conjugate transpose. The following properties make this more clear:

Proposition 7. *Let $S, T \in Hom_F(V, V)$. The following hold:*

(a) $(S+T)^* = S^* + T^*$
(b) $(cT)^* = \overline{c}T^*$
(c) $(T^*)^* = T$
(d) $I^* = I$
(e) $(ST)^* = T^*S^*$

Proof. All the above properties can be proved using a similar approach to the one in the proposition above by pulling the adjoint through the inner product. We omit the proofs. □

Proposition 8. *Let V be a finite dimensional inner product space, and let β be an orthonormal basis of V. Then $[T^*]_\beta = [T]^*_\beta$.*

Proof. Let $\beta = \{v_1, \ldots, v_n\}$ be an orthonormal basis for V. Set $[T]_\beta = [a_{ij}]$. Then $T(v_i) = a_{1i}v_1 + \ldots + a_{ni}v_n$, so $a_{ji} = \langle T(v_i), v_j\rangle$. This says $([T]^*_\beta)_{ij} = \overline{a_{ji}} = \overline{\langle T(v_i), v_j\rangle} = \langle v_j, T(v_i)\rangle = \langle T^*(v_j), v_i\rangle = ([T^*]_\beta)_{ij}$, so that $[T^*]_\beta = [T]^*_\beta$. □

Geometrically, the relationship between T^* and T is as follows:

Theorem 0.27. *Let V be a finite dimensional inner product space, and let $T : V \to V$ be a linear operator. Then* $\ker(T^*) = \operatorname{Im}(T)^\perp$ *and* $\operatorname{Im}(T^*) = \ker(T)^\perp$.

Proof. Let $x \in \ker(T^*)$, so that $T^*(x) = 0$. Then for any $y \in V$, $\langle y, T^*(x)\rangle = 0$. Pulling the adjoint through the inner product says $\langle T(y), x\rangle = 0$ for all y, so that $\ker(T^*) \subset \operatorname{Im}(T)^\perp$. Similarly, if $x \in \operatorname{Im}(T)^\perp$ this says $\langle x, T(y)\rangle = 0$ for all $y \in V$ so that $\langle T^*(x), y\rangle = 0$ for all $y \in V$. This says $T^*(x) = 0$, so that $\operatorname{Im}(T)^\perp \subset \ker(T^*)$ says $\ker(T^*) = \operatorname{Im}(T)^\perp$. Setting $T = T^*$ and taking orthogonal complements of both sides gives the second statement. □

Example 0.28. Let $T : \mathbb{C}^2 \to \mathbb{C}^2$ be given by $T(z_1, z_2) = (z_1 - 2iz_2, 3z_1 + iz_2)$, where $\mathbb{C}^2$ is equipped with the usual dot product. Then the standard basis $\{e_1, e_2\}$ is orthonormal. We see $[T]_\beta = \begin{pmatrix} 1 & -2i \\ 3 & i \end{pmatrix}$, so that $[T^*]_\beta = [T]^*_\beta = \begin{pmatrix} 1 & 3 \\ 2i & -i \end{pmatrix}$.

Example 0.29. Let $T : M_2(\mathbb{R}) \to M_2(\mathbb{R})$ be the transpose map, $T(A) = A^t$. Equip $M_2(\mathbb{R})$ with the inner product $\langle A, B\rangle = \operatorname{tr}(B^tA)$. With respect to this inner product, the standard basis $\{E_{11}, E_{12}, E_{21}, E_{22}\}$ is orthonormal. Then $[T^*]_\beta = [T]^*_\beta = [T]^t_\beta$. We see that $[T]_\beta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$. This matrix is symmetric, so $T = T^*$.

Example 0.30. Let $V \subset C^\infty(\mathbb{R})$ be the vector space of infinitely differentiable functions that are 1-periodic, i.e. $f(x+1) = f(x)$ for all $x \in \mathbb{R}$. Give V an inner product structure by $\langle f, g\rangle = \int_0^1 f(t)g(t)\,dt$. Let $D : V \to V$ be the derivative map. To compute the adjoint of D, we use the definition. For $f, g \in V$, $\langle D(f), g\rangle = \int_0^1 f'(t)g(t)\,dt$. Integrating by parts and using $f(1) = f(0)$, the latter integral equals $-\int_0^1 f(t)g'(t)\,dt = \langle f(t), -D(g)\rangle$. This says $D^* = -D$.

The spectral theorem

We will now return to diagonalizability. We previously saw what conditions are necessary for a linear operator on V to be diagonalizable, i.e. for V to have a basis of eigenvectors for T. If V is an inner product space, a natural question is when can we find an *orthonormal* basis of eigenvectors? The Spectral theorem gives a precise answer.

Definition 0.31. A linear operator $T : V \to V$ is called **normal** if $TT^* = T^*T$. T is called **self-adjoint** if $T = T^*$.

Proposition 9. *Suppose that $T : V \to V$ is normal. Then if v is an eigenvector of T with eigenvalue λ, then v is an eigenvector of T^* with eigenvalue $\overline{\lambda}$.*

Proof. It's easy to check that since T is normal, then so is $T - cI_V$ for any $c \in F$. Since $T(v) = \lambda v$, this says $0 = \|(T-\lambda I_V)(v)\|^2 = \langle (T-\lambda I_V)(v), (T-\lambda I_V)(v)\rangle = \langle v, (T^*-\overline{\lambda} I_V)(T-\lambda I_V)(v)\rangle = \langle v, (T-\lambda I_V)(T^*-\overline{\lambda} I_V)(v)\rangle = \langle (T^*-\overline{\lambda} I_V)(v), (T^*-\overline{\lambda} I_V)(v) = \|(T^*-\overline{\lambda} I_V)(v)\|^2$. This says $T^*(v) = \overline{\lambda} v$ as desired. ☐

Proposition 10. *Suppose that $T : V \to V$ is normal. Then if λ_1, λ_2 are distinct eigenvalues of T with eigenvectors v_1 and v_2 respectively, then v_1 and v_2 are orthogonal.*

Proof. Suppose $T(v_1) = \lambda_1 v_1$ and $T(v_2) = \lambda_2 v_2$. Then $\langle T(v_1), v_2\rangle = \lambda_1\langle v_1, v_2\rangle$. On the other hand, $\langle T(v_1), v_2\rangle = \langle v_1, T^*(v_2)\rangle = \langle v_1, \overline{\lambda_2} v_2\rangle = \lambda_2\langle v_1, v_2\rangle$ by the above proposition. Since $\lambda_1 \neq \lambda_2$, this forces $\langle v_1, v_2\rangle = 0$. ☐

Theorem 0.32 (Complex Spectral Theorem)**.** *Let V be a finite dimensional complex inner product space. Then a linear operator $T : V \to V$ is normal if and only if there is an orthonormal basis for V consisting of eigenvectors of T.*

Proof. First suppose that T is normal. We prove that T is orthogonally diagonalizable by induction on the dimension of V. If $\dim(V) = 1$ then this is obvious, because any non-zero vector is an eigenvector, so just normalize. Now suppose that any normal operator on an $n-1$ dimensional complex inner product space is orthogonally diagonalizable. If $\dim(V) = n$ and $T : V \to V$ is a normal operator, because $\mathbb{C}$ is algebraically closed T has an eigenvector, say v. Set $U = \text{Span}(\{v\})$ and write $V = U \oplus U^\perp$. Note that because T is normal, both T, T^* are U-invariant. If $x \in U^\perp$, then for $y \in U$, we have $\langle y, T(x)\rangle = \langle T^*(y), x\rangle = 0$ because $T^*(y) \in U$. This says $T(x) \in U^\perp$ so that T is $U^\perp$-invariant. Similarly, T^* is $U^\perp$-invariant. Then we may write $T(x) = T|_U(u) + T|_{U^\perp}(u')$ for $x = u + u'$ with $u \in U$ and $u' \in U^\perp$. We now show that $T|_{U^\perp}$ is a normal operator on $U^\perp$.

By definition, for $x, y \in U^\perp$, $\langle T|_{U^\perp}(x), y\rangle = \langle x, (T|_{U^\perp})^*(y)\rangle$. However, by definition $T|_{U^\perp}$ and T agree on $U^\perp$, so $\langle T|_{U^\perp}(x), y\rangle = \langle T(x), y\rangle = \langle x, T^*(y)\rangle = \langle x, (T^*)|_{U^\perp}(y)\rangle$. This says $(T|_{U^\perp})^* = (T^*)_{U^\perp}$. Then $(T|_{U^\perp})(T|_{U^\perp})^* = T|_{U^\perp}(T^*)|_{U^\perp} = TT^* = T^*T = (T^*)|_{U^\perp}T|_{U^\perp} = (T|_{U^\perp})^*(T|_{U^\perp})$, which proves $T|_{U^\perp}$ is normal. By induction, there is an orthonormal basis $\{v_2, \ldots, v_n\}$ of $U^\perp$ consisting of eigenvectors for $T|_{U^\perp}$. Then $\{v, v_2, \ldots, v_n\}$ is an orthogonal basis of V consisting of eigenvectors of T. Normalizing v makes this orthonormal, so we are done.

Conversely, suppose that T is orthogonally diagonalizable. Let $\beta = \{v_1, \ldots, v_n\}$ be a basis of eigenvectors with eigenvalues λ_i. Then $(T^*T)(v_i) = T^*(\lambda_i v_i) = \lambda_i T^*(v_i) = |\lambda_i|^2 v_i$. On the other hand, $(TT^*)(v_i) = T(\overline{\lambda_i} v_i) = |\lambda_i|^2 v_i$. Then T^*T and TT^* agree on a basis of V, so they are equal which shows T is normal as desired. □

We now move onto the Spectral Theorem for operators on real inner product spaces. In the complex case, we were able to make the argument work because the fundamental theorem of algebra says every linear operator over a complex vector space has an eigenvalue, which led to a decomposition $V = U \oplus U^\perp$. The key part of the proof is the normality of T said that it restricted to *normal* operators on U and $U^\perp$, allowing the induction to kick in. If V is a real inner product space, this no longer remains true, as we have seen that a rotation by some angle in $\mathbb{R}^2$ has no *real* eigenvalue. If we can find a class of normal operators that are guaranteed to have a real eigenvalue, then the same argument as above goes through. As it turns out, the key to this is self-adjointness:

Proposition 11. *Suppose that $T : V \to V$ is self-adjoint. Then if λ is an eigenvalue of T, then λ is real.*

Proof. Write $T(v) = \lambda v$. Then $\langle T(v), v\rangle = \lambda\langle v, v\rangle = \lambda\|v\|^2$. On the other hand, because T is self-adjoint we can write $\langle T(v), v\rangle = \langle v, T(v)\rangle = \overline{\lambda}\|v\|^2$. Since v is non-zero, this says $\lambda = \overline{\lambda}$ so that λ is real. □

Theorem 0.33. *(Real Spectral Theorem) Let V be a finite dimensional real inner product space. Then a linear operator $T : V \to V$ is self-adjoint if and only if there is an orthonormal basis for V consisting of eigenvectors of T.*

Proof. The characteristic polynomial p_T of T has a complex root by the fundamental theorem of algebra. Since T is self-adjoint, the above says this root is real, so that T has an eigenvector. Since a self-adjoint operator is normal, we can run the same argument in the complex case and the proof still goes through, so that T is orthogonally diagonalizable.

Conversely, suppose that T is orthogonally diagonalizable. The argument from before shows that T is normal. Let $\beta = \{v_1, \ldots, v_n\}$ be an eigenbasis for V with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. Then $T(v_i) = \lambda_i v_i$, and $T^*(v_i) = \overline{\lambda_i} v_i$. However, λ_i are *real*, which says that $T = T^*$ so that T is self-adjoint. □

We then immediately get the corresponding statements for matrices:

Corollary 0.34. *Let $T : V \to V$ be a normal operator over a finite dimensional complex inner product space, or a self-adjoint operator on a real inner product space. Then there is an orthonormal basis γ of V such that $[T]_\gamma = PDP^*$ where P is orthogonal, i.e. $PP^* = P^*P = I$ and D is diagonal.*

Proof. Fix an orthonormal basis β of V. By the Spectral Theorem, there is a basis γ of V consisting of orthonormal eigenvectors of T. Then $S^{\beta}_{\beta'}$ is orthogonal, so the change of basis formula gives the result with $P = S^{\beta}_{\beta'}$ and D the diagonal matrix of eigenvalues of T. □

The proof of the Spectral Theorem tells us how to orthogonally diagonalize an operator when it is possible. If $V = U \oplus U^\perp$, running Gram-Schmidt on bases of U and $U^\perp$ give orthogonal bases of these spaces, and then the union is an orthogonal basis of V, so after normalizing, an orthonormal basis. Suppose T is normal with distinct eigenvalues $\lambda_1, \ldots, \lambda_k$. In the proof of the Spectral Theorem, we may instead run the argument with $U = E_{\lambda_1}$ (the invariance condition is still true). Then since $E_{\lambda_i} \perp E_{\lambda_1}$ for $i \neq 1$, this says $E_{\lambda_2} \oplus \ldots \oplus E_{\lambda_k} \subset$

$U^\perp$ so that $E_{\lambda_2} \oplus \ldots \oplus E_{\lambda_k} = U^\perp$ for dimensional reasons. By inductively applying the above obeservation, this says running Gram-Schmidt on each eigenspace E_{λ_i} and taking the union of these orthogonal basis is then an orthogonal basis for V consisting of eigenvalues of T, and then normalizing gives an orthonormal basis.

Example 0.35. The operator $T : \mathbb{C}^2 \to \mathbb{C}^2$ given by $T(z_1, z_2) = (z_2, 0)$ is not normal, because it is not diagonalizable.

Example 0.36. The operator $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by $T(x, y) = (-y, x)$ is normal. With respect the the standard basis, $[T]_\beta = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, so $[T^*]_\beta = [T]^t_\beta = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = -[T]_\beta$. However, T is not self-adjoint, because $[T]_\beta$ is not a symmetric matrix. T has no real eigenvalues so it is not diagonalizable over $\mathbb{R}$, but over $\mathbb{C}$ has eigenvalues $i, -i$. To orthogonally diagonalize T over $\mathbb{C}$, a basis of eigenvectors is given by $\{(i, 1), (-i, 1)\}$, which we see is orthogonal. Normalizing says an orthonormal basis of eigenvectors is $\beta' = \{(\frac{i}{\sqrt{2}}, \frac{1}{\sqrt{2}}), (\frac{-i}{\sqrt{2}}, \frac{1}{\sqrt{2}})\}$. Since β' is orthonormal, the change of basis matrix $S^\beta_{\beta'}$ satisfies the relation $(S^\beta_{\beta'})^{-1} = S^{\beta'}_\beta = (S^\beta_{\beta'})^*$ This gives the matrix factorization $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \frac{i}{\sqrt{2}} & \frac{-i}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \begin{pmatrix} \frac{-i}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$.

Example 0.37. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $T(x, y, z) = (-2z, -x + 2y - z, x + 3z)$, so that $[T]_\beta = \begin{pmatrix} 0 & 0 & -2 \\ -1 & 2 & -1 \\ 1 & 0 & 3 \end{pmatrix}$ with β the standard basis. We see that T is diagonalizable with eigenvalues $1, 2$ and basis of the eigenspaces E_1 and E_2 are given by $\{(2, 1, -1)\}$ and $\{(0, 1, 0), (-1, -1, 1)\}$ respectively. However, T is not self-adjoint because $[T]_\beta$ is not symmetric, so the Spectral Theorem says that T is not orthogonally diagonalizable. What goes wrong? An orthogonal basis of E_2 is given by $\{(0, 1, 0), (1, 0, -1)\}$. However, $(2, 1, -1) \cdot (0, 1, 0) = 1 \neq 0$. Since any eigenvector $v \in E_2$ is of the form $(c_2, c_1, -c_2)$ for $c_1, c_2 \in \mathbb{R}$, we see that $(2, 1, -1) \cdot (c_2, c_1, -c_2) = 2c_1 + 2c_2$ is 0 only when $c_2 = -c_1$, i.e. the eigenvector is of the form $(-c_1, c_1, c_1)$. Therefore it's impossible to find two eigenvectors orthogonal to $(2, 1, -1)$, so that T cannot be orthogonally diagonalizable. Explicitly, with $U = E_1$, we see that $[T^*]_\beta = [T]^t_\beta$. T^* is not U-invariant, because $T^*(2, 1, -1) = (-2, 2, -8) \notin U$, so that T^* is not U-invariant and the argument cannot continue. Since all the eigenvalues of T are real, we see that even viewed as an operator on $\mathbb{C}^3$, the only eigenvector in E_2 that is orthogonal to $(2, 1, -1)$ is in the $\mathbb{C}$-span of $(-1, 1, 1)$, so again it is not possible to find two eigenvectors orthogonal to $(2, 1, -1)$. This then says that T is not normal when viewed as an operator on $\mathbb{C}^3$, and therefore not as an operator on $\mathbb{R}^3$ because the matrix of T^* is the same in either case.