November 24, 2014

INNER PRODUCT SPACES

BASED ON NOTES BY RODICA D. COSTIN

Contents

1. Inner product

1.1. **Inner product.**

1.1.1. *Inner product on real spaces.* Vectors in $\mathbb{R}^3$ have more properties than the ones listed in the definition of vector spaces: we can define their length, and the angle between two vectors.

Recall that two vectors are orthogonal if and only if their dot product is zero, and, more generally, the cosine of the angle between two *unit* vectors in $\mathbb{R}^3$ is their dot product. The notion of *inner product* extracts the essential properties of the dot product, while allowing it to be defined on quite general vector spaces. We will first define it for real vector spaces, and then we will formulate it for complex ones.

Definition 1. *An* **inner product** *on vector space V over $F = \mathbb{R}$ is an operation which associate to two vectors $\mathbf{x}, \mathbf{y} \in V$ a scalar $\langle \mathbf{x}, \mathbf{y} \rangle \in \mathbb{R}$ that satisfies the following properties:*
(i) it is is positive definite: $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ and $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ if and only if $\mathbf{x} = 0$,
(ii) it is symmetric: $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$
(iii) it is linear in the second argument: $\langle \mathbf{x}, \mathbf{y} + \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{z} \rangle$ and $\langle \mathbf{x}, c\mathbf{y} \rangle = c\langle \mathbf{x}, \mathbf{y} \rangle$

Note that by symmetry it follows that an inner product is linear in the first argument as well: $\langle \mathbf{x} + \mathbf{z}, \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{z}, \mathbf{y} \rangle$ and $\langle c\mathbf{x}, \mathbf{y} \rangle = c\langle \mathbf{x}, \mathbf{y} \rangle$.

A function of two variables which is linear in one variable and linear in the other variable is called *bilinear*; hence, the inner product in a real vector space is bilinear.

Example 1. The dot product in $\mathbb{R}^2$ or $\mathbb{R}^3$ is clearly an inner product: if $\mathbf{x} = (x_1, x_2, x_3)$ and $\mathbf{y} = (y_1, y_2, y_3)$ then define

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 + x_3 y_3$$

Example 2. More generally, an inner product on $\mathbb{R}^n$ is

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x} \cdot \mathbf{y} = x_1 y_1 + \ldots + x_n y_n \tag{1}$$

Example 3. Here is another inner product on $\mathbb{R}^3$:

$$\langle \mathbf{x}, \mathbf{y} \rangle = 5x_1 y_1 + 10x_2 y_2 + 2x_3 y_3$$

(some directions are weighted more than others).

Example 4. On spaces of functions the most useful inner products use integration. For example, consider $C[a, b]$ be the linear space of functions continuous on $[a, b]$. Then

$$\langle f, g \rangle = \int_a^b f(t)g(t)\, dt$$

is an inner product on $C[a, b]$ (check!).

Example 5. Sometimes a weight is useful: let $w(t)$ be a positive function. Then

$$\langle f, g\rangle = \int_a^b w(t)f(t)g(t)\,dt$$

is also an inner product on $C[a,b]$ (check!).

1.1.2. *Inner product on complex spaces.* For complex vector spaces extra care is needed. The blueprint of the construction here can be seen on the simplest case, $\mathbb{C}$ as a one dimensional vector space over $\mathbb{C}$: the inner product of $\langle z, z\rangle$ needs to be a positive number! It makes sense to define $\langle z, z\rangle = \overline{z}z$.

Definition 2. *An* **inner product** *on vector space V over $F = \mathbb{C}$ is an operation which associate to two vectors $\mathbf{x}, \mathbf{y} \in V$ a scalar $\langle \mathbf{x}, \mathbf{y}\rangle \in \mathbb{C}$ that satisfies the following properties:*
(i) it is positive definite: $\langle \mathbf{x}, \mathbf{x}\rangle \geq 0$ and $\langle \mathbf{x}, \mathbf{x}\rangle = 0$ if and only if $\mathbf{x} = 0$,
(ii) it is linear in the second argument: $\langle \mathbf{x}, \mathbf{y} + \mathbf{z}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle + \langle \mathbf{x}, \mathbf{z}\rangle$ and $\langle \mathbf{x}, c\mathbf{y}\rangle = c\langle \mathbf{x}, \mathbf{y}\rangle$
(iii) it is conjugate symmetric*: $\langle \mathbf{x}, \mathbf{y}\rangle = \overline{\langle \mathbf{y}, \mathbf{x}\rangle}$.*

Exercise 1. Check that if $\langle \mathbf{x}, \mathbf{y}\rangle$ is an inner product, then so is $\overline{\langle \mathbf{x}, \mathbf{y}\rangle}$

Note that conjugate symmetry combined with linearity implies that $\langle .,. \rangle$ is *conjugate-linear in the first variable*: $\langle \mathbf{x} + \mathbf{z}, \mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle + \langle \mathbf{z}, \mathbf{y}\rangle$ and $\langle c\mathbf{x}, \mathbf{y}\rangle = \overline{c}\langle \mathbf{x}, \mathbf{y}\rangle$.

A function of two variables which is linear in one variable and conjugate-linear in the other variable is called *sesquilinear*; the inner product in a complex vector space is sesquilinear.

Please keep in mind that most mathematical books use inner product linear in the first variable, and conjugate linear in the second one. You should make sure you know the convention used by each author.

Example 1'. On the one dimensional complex vector space $\mathbb{C}$ an inner product is $\langle z, w\rangle = \overline{z}w$.

Example 2'. More generally, an inner product on $\mathbb{C}^n$ is

$$\langle \mathbf{x}, \mathbf{y}\rangle = \overline{x_1}y_1 + \ldots + \overline{x_n}y_n \tag{2}$$

Example 3'. Here is another inner product on $\mathbb{C}^3$:

$$\langle \mathbf{x}, \mathbf{y}\rangle = 5\overline{x_1}y_1 + 10\overline{x_2}y_2 + 2\overline{x_3}y_3$$

(some directions are weighted more than others).

Example 4'. Let $C([a,b],\mathbb{C})$ be the linear space of complex-valued functions which are continuous on $[a,b]$.[1] Then

$$\langle f,g\rangle = \int_a^b \overline{f(t)}\, g(t)\, dt$$

is an inner product on $C([a,b],\mathbb{C})$ (check!).

Example 5'. Weights need to be positive: for $w(t)$ a given positive function. Then

$$\langle f,g\rangle = \int_a^b w(t)\overline{f(t)}\, g(t)\, dt$$

is also an inner product on $C([a,b],\mathbb{C})$ (check!).

1.2. **Inner product spaces.**

Definition 3. *A vector space V equipped with an inner product $(V,\langle .,.\rangle)$ is called an* **inner product space**.

Examples 1.-5. before are examples of inner product spaces over $\mathbb{R}$, while Examples 1'.-5'. are inner product spaces over $\mathbb{C}$.

In an inner product space we can do geometry. First of all, we can define length of vectors:

Definition 4. *Let $(V,\langle .,.\rangle)$ be an inner product space. The quantity*

$$\|x\| = \sqrt{\langle x,x\rangle}$$

is called the **norm** *of the vector x. For $V=\mathbb{R}^3$ with the usual inner product (which is the dot product) the norm of a vector is its length.*

Vectors of norm one, $\mathbf{u}$ with $\|\mathbf{u}\|=1$, are called **unit vectors**.

In an inner product space we can define the angle between two vectors. Recall that in the usual Euclidian geometry in $\mathbb{R}^2$ or $\mathbb{R}^3$, the angle θ between two vectors $\mathbf{x},\mathbf{y}$ is calculated from $\mathbf{x}\cdot\mathbf{y} = \|\mathbf{x}\|\,\|\mathbf{y}\|\,\cos\theta$. The existence of an angle θ with this property in any (real) inner product space is guaranteed by the Cauchy-Schwartz inequality, one of the most useful, and deep, inequalities in mathematics, which holds in finite or infinite dimensional inner product spaces is:

Theorem 5. The Cauchy-Schwartz inequality

In an inner product vector space any two vectors $\mathbf{x},\mathbf{y}$ satisfy

$$|\langle \mathbf{x},\mathbf{y}\rangle| \le \|\mathbf{x}\|\,\|\mathbf{y}\| \tag{3}$$

with equality if and only if $\mathbf{x},\mathbf{y}$ are linearly dependent.

[1] A complex valued function $f(t)=u(t)+iv(t)$ is continuous if the $\mathbb{R}^2$-valued function $(u(t),v(t))$ is continuous.

Remarks: 1. Recall that $\mathbf{x}, \mathbf{y}$ are linearly dependent means that one of the two vectors is $\mathbf{0}$ or the two vectors are scalar multiples of each other.

2. It suffices to prove inequality (3) in $Sp(\mathbf{x}, \mathbf{y})$, a space which is at most two-dimensional.

Proof of (3).

If $\mathbf{y} = \mathbf{0}$ then the inequality is trivially true. Otherwise, we use the fact that the norm of a nonzero vector is nononzero. We have

$$0 \leq \langle \mathbf{x} - c\mathbf{y}, \mathbf{x} - c\mathbf{y} \rangle = \|\mathbf{x}\|^2 - 2\Re\left[c\langle \mathbf{x}, \mathbf{y} \rangle\right] + |c|^2\|\mathbf{y}\|^2$$

To illustrate the idea, think first that we are working in $\mathbb{R}$. Then we have

$$0 \leq \langle \mathbf{x} - c\mathbf{y}, \mathbf{x} - c\mathbf{y} \rangle = \|\mathbf{x}\|^2 + c\langle \mathbf{x}, \mathbf{y} \rangle + c^2\|\mathbf{y}\|^2 = P(c)$$

This is true for all c and we use the freedom to look at P as a function of c. $P(c)$ is a quadratic polynomial in c with $P \to +\infty$ as $c \to \pm\infty$. It must be nonnegative for all c. This means that it cannot have two *distinct* real roots (why?). But this is equivalent to the fact that the discriminant is *nonpositive*, which implies

$$4\langle \mathbf{x}, \mathbf{y} \rangle^2 - \|\mathbf{x}\|^2\|\mathbf{y}\|^2 \leq 0$$

exactly what we wanted to prove. If the discriminant *is* zero, then there is a real root c_0, for which $P(c_0) = \|\mathbf{x} - c_0\mathbf{y}\|^2 = 0 \Rightarrow \mathbf{x} = c_0\mathbf{y}$.

In the complex case, we get

$$0 \leq \langle \mathbf{x} - c\mathbf{y}, \mathbf{x} - c\mathbf{y} \rangle = \|\mathbf{x}\|^2 - 2|c||\langle \mathbf{x}, \mathbf{y} \rangle| \cos\theta + |c|^2\|\mathbf{y}\|^2$$

where θ is the angle between c and $\langle \mathbf{x}, \mathbf{y} \rangle$. We have the freedom to choose this angle at will. We choose two extreme values for $\cos\theta$, ± 1 meaning $\theta = 0, \pi$.

We get

$$0 \leq \langle \mathbf{x} - c\mathbf{y}, \mathbf{x} - c\mathbf{y} \rangle = \|\mathbf{x}\|^2 \pm 2|c||\langle \mathbf{x}, \mathbf{y} \rangle| + |c|^2\|\mathbf{y}\|^2$$

and we are back in the real case!

Straightforwardly, but without motivating the substitution choose $c = \overline{\langle \mathbf{x}, \mathbf{y} \rangle} / \langle \mathbf{y}, \mathbf{y} \rangle$. Ee obtain

$$0 \leq \langle \mathbf{x}, \mathbf{x} \rangle - \frac{|\langle \mathbf{x}, \mathbf{y} \rangle|^2}{\langle \mathbf{y}, \mathbf{y} \rangle|}$$

which gives (3).

The triangle inequality for vectors $\mathbf{x}, \mathbf{y}$ in $\mathbb{R}^3$ means that the sum of two sides of a triangle is at least as large as the third side (with equality only when the triangle is degenerated to a segment). More generally:

Theorem 6. *In an inner product space* **the triangle inequality** *holds: for any* $\mathbf{x}, \mathbf{y} \in V$

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$$

with equality if and only if $\mathbf{x}$, $\mathbf{y}$ *are linearly dependent.*

Proof.

Expanding, then using the Cauchy-Schwarts inequality:

$$\|\mathbf{x}+\mathbf{y}\|^2 = \langle \mathbf{x}+\mathbf{y}, \mathbf{x}+\mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{x}\rangle + \langle \mathbf{x}, \mathbf{y}\rangle + \langle \mathbf{y}, \mathbf{x}\rangle + \langle \mathbf{y}, \mathbf{y}\rangle$$

$$= \|\mathbf{x}\|^2 + 2\Re\langle \mathbf{x}, \mathbf{y}\rangle + \|\mathbf{y}\|^2 \le \|\mathbf{x}\|^2 + 2\|\mathbf{x}\|\,\|\mathbf{y}\| + \|\mathbf{y}\|^2 = (\|\mathbf{x}\| + \|\mathbf{y}\|)^2$$

with equality only when the Cauchy-Schwarts inequality has equality, which is if and only if $\mathbf{x}$, $\mathbf{y}$ are linearly dependent. $\square$

In usual Euclidian geometry, the lengths of the sides of triangle determine its angles. Similarly, in an inner product space, if we know the norm of vectors, then we know inner products. In other words, the inner product is completely recovered if we know the norm of every vector:

Theorem 7. The polarization identity*:*
In a real real inner space

$$\langle \mathbf{x}, \mathbf{y}\rangle = \frac{1}{4}\left(\|\mathbf{x}+\mathbf{y}\|^2 - \|\mathbf{x}-\mathbf{y}\|^2\right)$$

In a complex inner space

$$\langle \mathbf{x}, \mathbf{y}\rangle = \frac{1}{4}\left(\|\mathbf{x}+\mathbf{y}\|^2 + i\|i\mathbf{x}+\mathbf{y}\|^2 - \|\mathbf{x}-\mathbf{y}\|^2 - i\| - i\mathbf{x}+\mathbf{y}\|^2\right) = \frac{1}{4}\sum_{k=0}^{3} i^k\|i^k\mathbf{x}+\mathbf{y}\|^2$$

The proof is by a straightforward calculation.

Note. A norm does not always come from a scalar product. The condition above is a way to check if it does.

In an inner product space, the parallelogram law holds: "sum of the squares of the lengths of the four sides of a parallelogram equals the sum of the squares of the lengths of the two diagonals":

Proposition 8. *In an inner product space* the parallelogram law *holds: for any* $\mathbf{x}, \mathbf{y} \in V$

$$\|\mathbf{x}+\mathbf{y}\|^2 + \|\mathbf{x}-\mathbf{y}\|^2 = 2\|\mathbf{x}\|^2 + 2\|\mathbf{y}\|^2 \tag{4}$$

The proof of (4) is a simple exercise, left to the reader.

2. Orthogonal bases

Let $(V, \langle ., .\rangle)$ be an inner product space.

Definition 9. *Two vectors* $\mathbf{x}, \mathbf{y} \in V$ *are called* **orthogonal** *if* $\langle \mathbf{x}, \mathbf{y}\rangle = 0$.

In this case we denote for short $\mathbf{x} \perp \mathbf{y}$ (in this case, of course, also $\mathbf{y} \perp \mathbf{x}$).

Note that the zero vector is orthogonal to any vector:

$$\langle \mathbf{x}, \mathbf{0}\rangle = 0 \quad \text{for all } \mathbf{x} \in V$$

since $\langle \mathbf{x}, \mathbf{0}\rangle = \langle \mathbf{x}, 0\mathbf{y}\rangle = 0\langle \mathbf{x}, \mathbf{y}\rangle = 0$.

As in geometry:

Definition 10. *We say that a vector* $\mathbf{x} \in V$ *is orthogonal to a subset* S *of* V *if* $\mathbf{x}$ *is orthogonal to every vector in* S*:*

$$\mathbf{x} \perp S \quad \textit{if and only if} \quad \mathbf{x} \perp \mathbf{z} \textit{ for all } \mathbf{z} \in S$$

Proposition 11. *We have* $\mathbf{x} \perp Sp(\mathbf{y}_1, \ldots, \mathbf{y}_r)$ *if and only if* $\mathbf{x} \perp \mathbf{y}_1, \ldots, \mathbf{x} \perp \mathbf{y}_n$*.*

The proof is left as an exercise to the reader.

Definition 12. *A set* $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subset V$ *is called an* **orthogonal set** *if* $\langle \mathbf{v}_j, \mathbf{v}_i \rangle = 0$ *for all* $i \neq j$*.*

The set is called **orthonormal** *if it is orthogonal and all* $\mathbf{v}_j$ *are unit vectors.*

Note that $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is an orthonormal set is equivalent to

$$\langle \mathbf{u}_j, \mathbf{u}_k \rangle = \delta_{jk} \quad \text{for all } j, k = 1, \ldots, n$$

where δ_{jk} is the Kronecker symbol: $\delta_{jj} = 1$ and $\delta_{jk} = 0$ if $j \neq k$.

Proposition 13. *An orthogonal set* $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ *where all* $\mathbf{v}_j \neq \mathbf{0}$ *is a linearly independent set.*

Indeed, if $c_1\mathbf{v}_1 + \ldots + c_k\mathbf{v}_k = \mathbf{0}$ then for any $j = 1, \ldots, k$

$$0 = \langle \mathbf{v}_j, c_1\mathbf{v}_1 + \ldots + c_k\mathbf{v}_k \rangle = c_1\langle \mathbf{v}_j, \mathbf{v}_1 \rangle + \ldots + c_j\langle \mathbf{v}_j, \mathbf{v}_j \rangle + \ldots + c_k\langle \mathbf{v}_j, \mathbf{v}_k \rangle$$

$$= c_j\langle \mathbf{v}_j, \mathbf{v}_j \rangle$$

and since $\mathbf{v}_j \neq \mathbf{0}$ then $\langle \mathbf{v}_j, \mathbf{v}_j \rangle \neq 0$ which implies $c_j = 0$. $\Box$.

Definition 14. *A basis for* V *with is an orthogonal set is called an* **orthogonal basis**.

An orthogonal basis made of unit vectors is called an **orthonormal basis***.*

Of course, if $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is an orthogonal basis, then $\{\frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1, \ldots, \frac{1}{\|\mathbf{v}_n\|}\mathbf{v}_n\}$ is an orthonormal basis.

For example, the standard basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is an orthonormal basis in $\mathbb{R}^n$ (when equipped this the inner product given by the dot product).

Orthonormal bases make formulas simpler and calculations easier. As a first example, here is how coordinates of vectors are found:

Theorem 15. *a) If* $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ *is an orthonormal basis for* V *then*

$$\mathbf{x} = x_1\mathbf{u}_1 + \ldots + x_n\mathbf{u}_n \qquad \textit{where } x_j = \langle \mathbf{u}_j, \mathbf{x} \rangle \tag{5}$$

b) If $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ *is an orthogonal basis for* V *then*

$$\mathbf{x} = c_1\mathbf{v}_1 + \ldots + c_n\mathbf{v}_n \qquad \textit{where } c_j = \frac{\langle \mathbf{v}_j, \mathbf{x} \rangle}{\|\mathbf{v}_j\|^2} \tag{6}$$

Proof.

a) Consider the expansion of x in the basis $\mathcal{B}$: $\mathbf{x} = x_1\mathbf{u}_1 + \ldots + x_n\mathbf{u}_n$. For each $j = 1, \ldots, n$

$$\langle \mathbf{u}_j, \mathbf{x}\rangle = \langle \mathbf{u}_j, x_1\mathbf{u}_1 + \ldots + x_n\mathbf{u}_n\rangle = x_1\langle \mathbf{u}_j, \mathbf{u}_1\rangle + \ldots + x_j\langle \mathbf{u}_j, \mathbf{u}_j\rangle + \ldots + x_n\langle \mathbf{u}_j, \mathbf{u}_n\rangle$$
$$= x_j\langle \mathbf{u}_j, \mathbf{u}_j\rangle = x_j$$

which gives the formula (5).

Part b) follows from a) applied to the orthonormal basis $\mathbf{u}_j = \frac{1}{\|\mathbf{v}_j\|}\mathbf{v}_j$, $j = 1, \ldots, n$. $\square$

When coordinates are given in an orthonormal basis, the inner product is the dot product of the coordinate vectors:

Proposition 16. *Let $\mathbf{u}_1, \ldots, \mathbf{u}_n$ be an orthonormal basis of V.*

$$\text{If } \mathbf{x} = \sum_{k=1}^{n} x_k\mathbf{u}_k, \quad \mathbf{y} = \sum_{k=1}^{n} y_k\mathbf{u}_k \quad \text{then} \quad \langle \mathbf{x}, \mathbf{y}\rangle = \sum_{k=1}^{n} \overline{x_k}\, y_k \tag{7}$$

The proof is an easy exercise left to the reader.

2.1. Existence of an orthogonal basis.

We know that any vector space has a basis. Moreover, any finite dimensional inner product space has an orthogonal basis, and here is how to find one:

Theorem 17. Gram-Schmidt Orthogonalization

Let $\{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$ be a basis for V. Then an orthonormal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ can be found as follows:

$$\mathbf{v}_1 = \mathbf{y}_1, \quad \mathbf{u}_1 = \frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1 \ \in Sp(\mathbf{y}_1) \tag{8}$$

$$\mathbf{v}_2 = \mathbf{y}_2 - \langle \mathbf{u}_1, \mathbf{y}_2\rangle\mathbf{u}_1, \quad \mathbf{u}_2 = \frac{1}{\|\mathbf{v}_2\|}\mathbf{v}_2 \ \in Sp(\mathbf{y}_1, \mathbf{y}_2) \tag{9}$$

$$\mathbf{v}_3 = \mathbf{y}_3 - \langle \mathbf{u}_1, \mathbf{y}_3\rangle\mathbf{u}_1 - \langle \mathbf{u}_2, \mathbf{y}_3\rangle\mathbf{u}_2, \quad \mathbf{u}_3 = \frac{1}{\|\mathbf{v}_3\|}\mathbf{v}_3 \ \in Sp(\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3) \tag{10}$$

$$\vdots$$

$$\mathbf{v}_k = \mathbf{y}_k - \sum_{j=1}^{k-1}\langle \mathbf{u}_j, \mathbf{y}_k\rangle\mathbf{u}_j, \quad \mathbf{u}_k = \frac{1}{\|\mathbf{v}_k\|}\mathbf{v}_k \ \in Sp(\mathbf{y}_1, \ldots, \mathbf{y}_k) \tag{11}$$

$$\vdots$$

$$\mathbf{v}_n = \mathbf{y}_n - \sum_{j=1}^{n-1}\langle \mathbf{u}_j, \mathbf{y}_n\rangle\mathbf{u}_j \quad \mathbf{u}_n = \frac{1}{\|\mathbf{v}_n\|}\mathbf{v}_n \ \in Sp(\mathbf{y}_1, \ldots, \mathbf{y}_n) \tag{12}$$

The Gram-Schmidt orthogonalization process is exactly as in $\mathbb{R}^n$. As a first step (8) we keep $\mathbf{y}_1$, only normalized.

Next, in (9), we replace $\mathbf{y}_2$ by a vector in $Sp(\mathbf{y}_1, \mathbf{y}_2)$ which is orthogonal to $\mathbf{u}_1$ (as it is easily shown by calculation). After normalization we have produced the orthonormal basis $\mathbf{u}_1, \mathbf{u}_2$ for $Sp(\mathbf{y}_1, \mathbf{y}_2)$.

In the next step, (10), we replace $\mathbf{y}_3$ by $\mathbf{v}_3$, a vector orthogonal to $Sp(\mathbf{u}_1, \mathbf{u}_2)$ (easy to check by calculation); then $\mathbf{v}_3$ is normalized, giving the orthonormal basis $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ for $Sp(\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3)$.

The procedure is continued.

To check that $\mathbf{u}_k \perp \mathbf{u}_i$ for all $i \neq k$ we note that at each step k we produce a vector $\mathbf{u}_k$ which is orthogonal to $Sp(\mathbf{u}_1, \ldots, \mathbf{u}_{k-1})$. $\Box$

Corollary 18. *Let U be a subspace of V. Any orthogonal basis of U can be completed to an orthogonal basis of V.*

Indeed, U has an orthogonal basis $\mathbf{u}_1, \ldots, \mathbf{u}_k$ by Theorem 17, which can be completed to a basis of V, $\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{y}_{k+1} \ldots \mathbf{y}_n$. Applying the Gram-Schmidt orthogonalization procedure to this basis, we obtain an orthogonal basis of V. It is easily seen that the procedure leaves $\mathbf{u}_1, \ldots, \mathbf{u}_k$ unchanged. $\Box$

2.2. **Orthogonal projections.** In the usual geometry of $\mathbb{R}^2$ or $\mathbb{R}^3$, the orthogonal projection of a vector $\mathbf{x}$ on a line ℓ through the origin is the unique vector $P_\ell \mathbf{x}$ in the direction ℓ with the property that $\mathbf{x} - P_\ell \mathbf{x}$ is orthogonal to ℓ. If $\mathbf{u}$ is the unit vector in the direction of ℓ (therefore $\ell = Sp(\mathbf{u})$) then $P_\ell \mathbf{x} = \langle \mathbf{u}, \mathbf{x} \rangle \mathbf{u}$.

Similarly, in an inner product space we can define orthogonal projections onto subspaces:

Proposition 19. *Let W be a subspace of an inner product space $(V, \langle ., . \rangle)$.*

1. For any $\mathbf{x} \in W$ there exists a unique vector $P_W \mathbf{x}$ in W so that

$$\mathbf{x} - P_W \mathbf{x} \ \perp W$$

$P_W \mathbf{x}$ is called **the orthogonal projection of $\mathbf{x}$ on the subspace W.**

2. If $\mathbf{u}_1, \ldots \mathbf{u}_k$ is an orthonormal basis for W then

$$P_W \mathbf{x} = \langle \mathbf{u}_1, \mathbf{x} \rangle \mathbf{u}_1 + \ldots + \langle \mathbf{u}_k, \mathbf{x} \rangle \mathbf{u}_k \tag{13}$$

3. $P_W : V \to V$ is a linear transformation.

Proof. (This is true in infinite dimensional spaces too, but now we'll only prove it in finite dimensional ones.) By Theorem 17 the subspace W has an orthonormal basis, say $\mathbf{u}_1, \ldots \mathbf{u}_k$. Complete it to an orthonormal basis $\mathbf{u}_1, \ldots \mathbf{u}_k, \ldots, \mathbf{u}_n$ of V.

Writing $P_W \mathbf{x} = c_1 \mathbf{u}_1 + \ldots + c_k \mathbf{u}_k$ we search for scalars $c_1, \ldots, c_k$ so that $\mathbf{x} - (c_1 \mathbf{u}_1 + \ldots + c_k \mathbf{u}_k)$ is orthogonal to W. By Proposition 11 we only need to check orthogonality to all $\mathbf{u}_1, \ldots \mathbf{u}_k$. Since $\langle \mathbf{u}_j, \mathbf{x} - (c_1 \mathbf{u}_1 + \ldots + c_k \mathbf{u}_k) \rangle = \langle \mathbf{u}_j, \mathbf{x} \rangle - c_j$ formula (13) follows.

The fact that P_W is linear is seen from (13). $\Box$

Exercise. Show that $P_W^2 = P_W$. Find the range and the null space of P_W.

Note that in the Gram-Schmidt process, Theorem 17, at each step k, $\mathbf{v}_k$ is the difference between $\mathbf{y}_k$ and its orthogonal projection on $Sp(\mathbf{u}_1, \ldots, \mathbf{u}_{k-1})$, thus guaranteeing that $\mathbf{v}_k$ is orthogonal to all $\mathbf{u}_1, \ldots, \mathbf{u}_{k-1}$.

2.3. **Orthogonal subspaces, orthogonal complements.** Let $(V, \langle ., . \rangle)$ be an inner product space.

Definition 20. *Two subspaces $U, W \subset V$ are* **orthogonal** *($U \perp W$) if every $\mathbf{u} \in U$ is orthogonal to every $\mathbf{w} \in W$.*

Remark. If $U \perp W$ then $U \cap W = \{\mathbf{0}\}$.

Indeed, suppose that $\mathbf{x} \in U \cap W$. Then $\mathbf{x} \in U$ and $\mathbf{x} \in W$, and therefore we must have $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ which implies $\mathbf{x} = \mathbf{0}$.

Examples. Let $V = \mathbb{R}^3$.

(i) The (subspace consisting of all vectors on) the z-axis is orthogonal to a one dimensional subspace ℓ (a line through the origin) if and only if ℓ is included in the xy-plane.

(ii) In fact the z-axis is orthogonal to the xy-plane.

(iii) But the intuition coming from classical geometry stops here. As vector subspaces, the yz-plane is *not orthogonal* to the xy-plane (they have a subspace in common!).

Recall that any subspace $W \subset V$ as a complement in V (in fact, it has infinitely many). But exactly one of those is orthogonal to W:

Theorem 21. *Let W be a subspace of V. Denote*

$$W^\perp = \{\mathbf{x} \in V \,|\, \mathbf{x} \perp W\}$$

Then $W^\perp$ is a subspace, called the **orthogonal complement** *of W, and*

$$W \oplus W^\perp = V \tag{14}$$

In particular, $\dim W + \dim W^\perp = \dim V$.

Proof. Note that 0 is a vector space, orthogonal to any space. So $W^\perp$ is nonempty.

To show that $W^\perp$ is a subspace, let $\mathbf{x}, \mathbf{y} \in W^\perp$; then $\langle \mathbf{w}, \mathbf{y} \rangle = 0$ and $\langle \mathbf{w}, \mathbf{u} \rangle = 0$ for all $\mathbf{w} \in W$. Then for any c, d any scalars in F we have

$$\langle \mathbf{w}, c\mathbf{x} + d\mathbf{y} \rangle = c\langle \mathbf{w}, \mathbf{x} \rangle + d\langle \mathbf{w}, \mathbf{y} \rangle = 0$$

for all $\mathbf{w} \in W$, which shows that $c\mathbf{x} + d\mathbf{y} \in W^\perp$.

The sum is direct since if $\mathbf{w} \in W \cap W^\perp$ then $\mathbf{w} \in W$ and $\mathbf{w} \in W^\perp$, therefore $\langle \mathbf{w}, \mathbf{w} \rangle = 0$, implying $\mathbf{w} = \mathbf{0}$.

To show (14), let $\mathbf{u}_1, \ldots, \mathbf{u}_k$ be an orthogonal basis of W, and complete it to an orthogonal basis of V: $\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{v}_1, \ldots, \mathbf{v}_{n-k}$, by Corollary 18.

Then $W^\perp = Sp(\mathbf{v}_1, \ldots, \mathbf{v}_{n-k})$ since, one one hand, clearly $Sp(\mathbf{v}_1, \ldots, \mathbf{v}_{n-k}) \subset W^\perp$ and in the other hand it can be easily checked that if $\mathbf{x} \in W^\perp$ then $\mathbf{x} \in Sp(\mathbf{v}_1, \ldots, \mathbf{v}_{n-k})$. Therefore $W \oplus W^\perp = V$. $\square$

Remark. If V is *finite dimensional* then

$$\left(W^\perp\right)^\perp = W$$

The proof is left as an exercise.

2.4. **The Riesz Representation Theorem.** In an inner product space linear functionals are nothing more, and nothing less than inner products:

Theorem 22. The Riesz Representation Theorem

On an inner product space $(V, \langle\cdot,\cdot\rangle)$ any linear functional $\phi : V \to F$ has the form $(\phi, \mathbf{x}) = \langle z, \mathbf{x}\rangle$ for a unique $\mathbf{z} = \mathbf{z}_\phi \in V$.

Proof. Let $\mathbf{u}_1, \ldots, \mathbf{u}_n$ be an orthonormal basis of V. Then a linear functional ϕ acting on $\mathbf{x} \in V$ satisfies

$$(\phi, \mathbf{x}) = (\phi, \sum_{k=1}^n x_k \mathbf{u}_k) = \sum_{k=1}^n x_k (\phi, \mathbf{u}_k)$$

which, by (7), equals $\langle \mathbf{z}, \mathbf{x}\rangle$ if $\mathbf{z} = \sum_{k=1}^n \overline{(\phi, \mathbf{u}_k)}\mathbf{u}_k$. □

2.5. **The adjoint of a linear transformation.** To keep notations simple, we first consider linear transformations from a space to itself (endomorphisms); the general case is discussed after that (the differences are minor).

Definition 23. *Let $(V, \langle .,.\rangle)$ be an inner product space. Given a linear transformation $L : V \to V$, its adjoint L^* is a linear transformation $L^* : V \to V$ which satisfies*

$$\langle L\mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{x}, L^*\mathbf{y}\rangle \quad \textit{for all } \mathbf{x}, \mathbf{y} \in V \tag{15}$$

At this point it is not clear that such a transformation L^* exists, and it will require an argument. Before we do so, let us look at concrete examples.

Examples.

1. Let $V = \mathbb{R}^n$ with the usual inner product $\langle \mathbf{x}, \mathbf{y}\rangle = \mathbf{x}\cdot\mathbf{y} = \mathbf{x}^T\mathbf{y}$. Consider the linear transformation given by matrix multiplication: $L\mathbf{x} = M\mathbf{x}$ where M is an $n\times n$ real matrix. Equation (15) is in this case $(M\mathbf{x})^T\mathbf{y} = \mathbf{x}^T L^*\mathbf{y}$ or, using the familiar property of the transpose that $(AB)^T = B^T A^T$, the relation defining the adjoint becomes $\mathbf{x}^T M^T \mathbf{y} = \mathbf{x}^T L^*\mathbf{y}$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, which means that L^* is matrix multiplication by M^T.

2. Let V be the complex vector space $\mathbb{C}^n$ with the usual inner product $\langle \mathbf{x}, \mathbf{y}\rangle = \overline{\mathbf{x}}\cdot\mathbf{y} = \overline{\mathbf{x}}^T\mathbf{y}$. Consider the linear transformation given by matrix multiplication: $L\mathbf{x} = M\mathbf{x}$ where M is an $n \times n$ complex matrix. Equation (15) is in this case $(\overline{M\mathbf{x}})^T\mathbf{y} = \overline{\mathbf{x}}^T L^*\mathbf{y}$ or, $\overline{\mathbf{x}}^T \overline{M}^T \mathbf{y} = \overline{\mathbf{x}}^T L^*\mathbf{y}$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$, which means that L^* is multiplication by $\overline{M}^T$.

Proof of existence of the adjoint.

Since any linear transformation of a finite dimensional vector space is, essentially, matrix multiplication (once a bases are chosen), and since the inner product is the usual dot product if the basis is orthonormal, the examples above show that the adjoint exists, and that the matrix of L^* (in an orthonormal basis) is the conjugate transpose of the matrix of L (in the same basis).

An alternative argument is as follows. For every fixed $\mathbf{y}$, the formula $(\phi, \mathbf{x}) = \langle \mathbf{y}, L\mathbf{x}\rangle$ defines a linear functional on V. By the Riesz representation theorem, there exists a unique vector $\mathbf{z} \in V$ (which depends on $\mathbf{y}$) so that $(\phi, \mathbf{y}) = \langle \mathbf{z}, \mathbf{x} \rangle$. Define $L^*\mathbf{y} = \mathbf{z}$. It only remains to check that L^* thus defined is linear, which follows using the linearity of the inner product and by the uniqueness of each $\mathbf{z}$. $\Box$

More generally, if $(V, \langle ., . \rangle_V)$, $(U, \langle ., . \rangle_U)$ are two inner product spaces over the same scalar field F, and L is a linear transformation between them, L^* is defined similarly:

Definition 24. *If $L : U \to V$ is a linear transformation, its adjoint L^* is the linear transformation $L^* : V \to U$ which satisfies*

$$\langle L\mathbf{x}, \mathbf{y} \rangle_V = \langle \mathbf{x}, L^*\mathbf{y} \rangle_U \quad \text{for all } \mathbf{x} \in U, \ \mathbf{y} \in V$$

Arguments similar to the ones before show that that the adjoint exists, and that the matrix of L^* (corresponding to orthonormal bases) is the conjugate transpose of the matrix of L (corresponding to the same bases).

Note that

$$(L^*)^* = L \tag{16}$$

Other notations used in the literature: In the case of real inner spaces the adjoint is sometimes simply called the transpose, and denoted L^T (which is the same as L', the transpose operator acting on the dual space V' if we identify V' to V via the Riesz Representation Theorem). In the case of complex inner spaces the adjoint L^* is sometime called the *Hermitian*, and denoted by L^H or $L^\dagger$. The notation for matrices is similar: the hermitian of a matrix, M^H, is the complex conjugate of the transpose of M.

2.6. **Projections.** The prototype of an oblique projection is the shadow cast by objects on the ground (when the sun is not directly vertical). Mathematically, we have a plane (the ground), a direction (direction of the rays of light, which are approximately parallel, since the sun is very far), and the projection of a person along the light rays is its shadow on the ground.

More generally, we can have projections parallel to higher dimensional subspaces (they are quite useful in image processing among other applications).

The simplest way to define a projection mathematically is

Definition 25. *Given a linear space V, a* **projection** *is a linear transformation $P : V \to V$ so that $P^2 = P$.*

Exercise. Show that the eigenvalues of a projection can be only 0 or 1.

The subspace $R = \mathcal{R}(P)$ is the subspace on which P projects, and the projection is parallel to (or, along) the subspace $N = \mathcal{N}(P)$.

If $\mathbf{y} \in R$ then $P\mathbf{y} = \mathbf{y}$ since, for any $\mathbf{y} \in R$, then $\mathbf{y} = P\mathbf{x}$, hence $P\mathbf{y} = P(P\mathbf{x}) = P^2\mathbf{x} = P\mathbf{x} = \mathbf{y}$.

Since $\dim \mathcal{R}(P) + \dim \mathcal{N}(P) = n$ it follows that $R \oplus N = V$. Therefore any $\mathbf{x} \in V$ has the form $\mathbf{x} = \mathbf{x}_R + \mathbf{x}_N$ where $\mathbf{x}_R \in R$ and $\mathbf{x}_N \in N$ and $P\mathbf{x} = \mathbf{x}_{\mathbf{R}}$. This is a true projection!

2.6.1. *Geometrical characterization of projections.* Conversely, given two subspaces R and N so that $R \oplus N = V$, a projection onto R, parallel to N is defined geometrically as follows.

Proposition 26. *If $R \oplus N = V$, then for any $\mathbf{x} \in V$, there is a unique vector $P\mathbf{x} \in R$ so that $\mathbf{x} - P\mathbf{x} \in N$.*

P is a projection (the projection onto R along N*).*

The *proof* is immediate, since any $\mathbf{x} \in V$ can be uniquely written as $\mathbf{x} = \mathbf{x}_R + \mathbf{x}_N$ with $\mathbf{x}_R \in R$ and $\mathbf{x}_N \in N$, hence $P\mathbf{x} = \mathbf{x}_R$. $\square$

Of course, the projection is orthogonal if $N = R^\perp$.

Example. In $V = \mathbb{R}^3$ consider the projection on $R = Sp(\mathbf{e}_1, \mathbf{e}_2)$ (the x_1x_2-plane) along a line $N = Sp(\mathbf{n})$ where $\mathbf{n}$ is a vector not in R, $\mathbf{n} = (n_1, n_2, n_3)$ with $n_3 \neq 0$. We construct the projection geometrically. Given a point $\mathbf{x} = (x_1, x_2, x_3)$, the line parallel to $\mathbf{n}$ and passing through $\mathbf{x}$ is $\mathbf{x} + t\mathbf{n}$, $t \in \mathbb{R}$. The line intersects the x_1x_2-plane for $t = -\frac{x_3}{n_3}$. Hence $P\mathbf{x} = \mathbf{x} - \frac{x_3}{n_3}\mathbf{n}$, or

$$P\mathbf{x} = \mathbf{x} - \frac{\langle \mathbf{x}, \mathbf{u} \rangle}{\langle \mathbf{n}, \mathbf{u} \rangle}\mathbf{n}, \quad \text{where } \mathbf{u} \perp R$$

2.6.2. *The matrix of a projection.* Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ be a basis for R, and $\{\mathbf{u}_{r+1}, \ldots, \mathbf{u}_n\}$ a basis for S; then $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is a basis for V, and in this basis the matrix of the sought-for projection is

$$M_P = \begin{bmatrix} I & O \\ O & O \end{bmatrix}$$

To find the matrix in another basis, say, for example, $V = \mathbb{R}^n$ or $\mathbb{C}^n$, and the vectors $\mathbf{u}_j$ are given by their coordinates in the standard basis, then the transition matrix (from old basis to standard basis) is $S = [\mathbf{u}_1, \ldots, \mathbf{u}_n]$ and then the matrix of P in the standard basis is $S^{-1} M_P S$.

An alternative formula is[2]

Proposition 27. *Let R and N be subspaces of $V = \mathbb{R}^n$ or $\mathbb{C}^n$ such that $R \oplus N = V$.*

Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ be a basis for R and $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ be a basis for $N^\perp$. Let $A = [\mathbf{u}_1, \ldots, \mathbf{u}_r]$ and $B = [\mathbf{v}_1, \ldots, \mathbf{v}_r]$ (where the vectors are expressed as their coordinates in the standard basis).

Then the matrix (in the standard basis) of the projection on R parallel to N is

$$A(B^*A)^{-1}B^*$$

Proof.

Let $\mathbf{x} \in V$. We search for $P\mathbf{x} \in R$, therefore, for some scalars $y_1, \ldots, y_r$

$$P\mathbf{x} = \sum_{j=1}^{r} y_j \mathbf{u}_j = [\mathbf{u}_1, \ldots, \mathbf{u}_r] \begin{bmatrix} y_1 \\ \vdots \\ y_r \end{bmatrix} = A\mathbf{y}$$

Since $\mathbf{x} - P\mathbf{x} \in N = (N^\perp)^\perp$ this means that $\mathbf{x} - P\mathbf{x} \perp N^\perp$, therefore $\langle \mathbf{v}_k, \mathbf{x} - P\mathbf{x} \rangle = 0$ for all $k = 1, \ldots, r$. Hence

$$0 = \langle \mathbf{v}_k, \mathbf{x} - P\mathbf{x} \rangle = \langle \mathbf{v}_k, \mathbf{x} \rangle - \langle \mathbf{v}_k, P\mathbf{x} \rangle = \langle \mathbf{v}_k, \mathbf{x} \rangle - \langle \mathbf{v}_k, \sum_{j=1}^{r} y_j \mathbf{u}_j \rangle$$

so

$$0 = \langle \mathbf{v}_k, \mathbf{x} \rangle - \sum_{j=1}^{r} y_j \langle \mathbf{v}_k, \mathbf{u}_j \rangle, \quad k = 1, \ldots, r \tag{17}$$

Note that element (j,k) of the matrix B^*A is precisely $\langle \mathbf{v}_k, \mathbf{u}_j \rangle = \mathbf{v}_k^* \mathbf{u}_j = \overline{\mathbf{v}_k}^T \mathbf{u}_j$ hence the system (17) is

$$B^* A\mathbf{y} = B^*\mathbf{x}$$

which we need to solve for $\mathbf{y}$.

Claim: The $r \times r$ matrix B^*A is invertible.

[2]Reference: http://en.wikipedia.org/wiki/Projection_(linear_algebra)

Its justification will have to wait for a little while. Assuming for the moment the claim is true, then it follows that $\mathbf{y} = (B^*A)^{-1}B^*\mathbf{x}$. Therefore $P\mathbf{x} = PA\mathbf{y} = A(B^*A)^{-1}B^*\mathbf{x}$. $\square$

2.7. **More about orthogonal projections.** We have defined projections in §2.6 as endomorphisms $P : V \to V$ satisfying $P^2 = P$ and we saw that in the particular case when $\mathcal{N}(P) = \mathcal{R}(P)^\perp$ we have orthogonal projections, characterized in Theorem 19.

The first theorem shows that a projection is orthogonal if and only if it is self-adjoint:

Theorem 28. *Let V be a finite dimensional inner product space. A linear transformation $P : V \to V$ is an orthogonal projection if and only if $P^2 = P$ and $P = P^*$.*

Proof.

We already showed that $P^2 = P$ means that P is a projection. We only need to check that $P = P^*$ if and only if P is an orthogonal projection, that is, if and only if $\mathcal{N}(P) = \mathcal{R}(P)^\perp$.

Assume $P = P^*$. Let $\mathbf{x} \in \mathcal{N}(P)$ and $P\mathbf{y} \in \mathcal{R}(P)$. Then $\langle \mathbf{x}, P\mathbf{y}\rangle = \langle P^*\mathbf{x}, \mathbf{y}\rangle = \langle P\mathbf{x}, \mathbf{y}\rangle = 0$ therefore $\mathcal{N}(P) \subset \mathcal{R}(P)^\perp$ and since $\mathcal{N}(P) \oplus \mathcal{R}(P) = V$ then $\mathcal{N}(P) = \mathcal{R}(P)^\perp$.

The converse implication is similar. $\square$

2.8. **Fundamental subspaces of matrices.**

2.8.1. *Fundamental subspaces of linear transformations.*

Theorem 29. *Let $L : U \to V$ be a linear transformation between two inner product spaces $(V, \langle .,. \rangle_V)$, $(U, \langle .,. \rangle_U)$. Then*
(i) $\mathcal{N}(L^*) = \mathcal{R}(L)^\perp$
(ii) $\mathcal{R}(L^*)^\perp = \mathcal{N}(L)$

Note: The theorem is true in finite or infinite dimensions. Of course, in finite dimensions also: $\mathcal{N}(L^*)^\perp = \mathcal{R}(L)$ and $\mathcal{R}(L^*) = \mathcal{N}(L)^\perp$ (but in infinite dimensions extra care is needed).

Remark 30. *Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. We have:*
$\langle \mathbf{x}, \mathbf{v} \rangle = 0$ *for all* $\mathbf{v} \in V$ *if and only if* $\mathbf{x} = \mathbf{0}$.

This is easy to see, since if we take in particular $\mathbf{v} = \mathbf{x}$ then $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ which implies $\mathbf{x} = \mathbf{0}$.

Proof of Theorem 29.
(i) We have $\mathbf{y} \in \mathcal{R}(L)^\perp$ if and only if $\langle \mathbf{y}, L\mathbf{x} \rangle = 0$ for all $\mathbf{x} \in U$ if and only if $\langle L^*\mathbf{y}, \mathbf{x} \rangle = 0$ for all $\mathbf{x} \in U$ if and only if (by Remark 30) $L^*\mathbf{y} = \mathbf{0}$ hence $\mathbf{y} \in \mathcal{N}(L^*)$.
(ii) follows from (i) after replacing L by L^* and using (16). □

Exercise. Show that if P is a projection onto R along N then P^* is a projection onto $N^\perp$ along $R^\perp$.

2.8.2. *The four fundamental subspaces of a matrix.*

Recall that if M is the matrix associated L in two bases which are orthonormal, then the matrix associated to L^* (in the same bases) is the conjugate transpose of M, denoted $M^* = \overline{M}^T = \overline{M^T}$.

To transcribe Theorem 29 in language of matrices, once orthonormal bases are chosen L becomes matrix multiplication taking (coordinate) vectors from F^n to (coordinate) vectors $M\mathbf{x} \in F^m$, where M is an $m \times n$ matrix with elements in F and the inner products on the coordinate spaces are the dot products if $F = \mathbb{R}$ and conjugate-dot products in $F = \mathbb{C}$.

Recall that $\mathcal{R}(M)$ is the column space of M, and $\mathcal{N}(M)$ is the right null space of M: the space of all $\mathbf{x}$ so that $M\mathbf{x} = \mathbf{0}$.

Recall that the left null space of M is defined as the space of all vectors so that $\mathbf{y}^T M = \mathbf{0}$.

In the real case, when $F = \mathbb{R}$, $M^* = M^T$. Then $\mathbf{y}$ belongs to the left null space of M means that $\mathbf{y} \in \mathcal{N}(M^T)$. Also, the row space of M is $\mathcal{R}(M^T)$. Theorem 29 states that:

Theorem 31. *Let M be an $m \times n$ real matrix. Then:*

1) $\mathcal{R}(M)$ (the column space of M) and $\mathcal{N}(M^T)$ (the left null space of M) are orthogonal complements to each other in $\mathbb{R}^n$, and

2) $\mathcal{R}(M^T)$ (the row space of M) and $\mathcal{N}(M)$ (the right null space of M) are orthogonal complements to each other in $\mathbb{R}^m$.

As a corrolary: *The linear system $M\mathbf{x} = \mathbf{b}$ has solutions if and only if $\mathbf{b} \in \mathcal{R}(M)$ if and only if $\mathbf{b}$ is orthogonal to all solutions of $\mathbf{y}^T M = \mathbf{0}$.*

2.9. **Decomposing linear transformations.** Let $L : U \to V$ be a linear transformation between two inner product spaces $(V, \langle .,. \rangle_V)$, $(U, \langle .,. \rangle_U)$. It makes sense to split the spaces U and V into subspaces which carry information about L, and their orthogonal complements, which are redundant.

For example, only the subspace $\mathcal{R}(L)$ of V is "necessary" to L, so we could decompose $V = \mathcal{R}(L) \oplus \mathcal{R}(L)^\perp$, which, by Theorem 29, is the same as $V = \mathcal{R}(L) \oplus \mathcal{N}(L^*)$.

Also, the subspace $\mathcal{N}(L)$ of U is taken to zero through L, and we may wish to decompose $U = \mathcal{N}(L) \oplus \mathcal{N}(L)^\perp = \mathcal{N}(L) \oplus \mathcal{R}(L^*)$.

Then

$$L : U = \mathcal{R}(L^*) \oplus \mathcal{N}(L) \to V = \mathcal{R}(L) \oplus \mathcal{N}(L^*)$$

Recall that the rank of a matrix equals the one of its transpose. It also equals the rank of its complex conjugate, since $\det(\overline{M}) = \overline{\det M}$. Therefore

$$\operatorname{rank} M = \dim \mathcal{R}(L) = \dim \mathcal{R}(L^*) = \operatorname{rank} M^*$$

Let $B_1 = \{\mathbf{u}_1, \ldots \mathbf{u}_r\}$ be an orthonormal basis for $\mathcal{R}(L^*)$ and $\mathbf{u}_{r+1}, \ldots, \mathbf{u}_n$ an orthonormal basis for $\mathcal{N}(T)$. Then $B_U = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is a basis for U. Similarly, let $B_2 = \{\mathbf{v}_1, \ldots \mathbf{v}_r\}$ be an orthonormal basis for $\mathcal{R}(L)$ and $\mathbf{v}_{r+1}, \ldots, \mathbf{v}_m$ an orthonormal basis for $\mathcal{N}(L^*)$. Then $B_V = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is an orthonormal basis for V. It is easily checked that the matrix associated to L in the bases B_U, B_V has the block form

$$\begin{bmatrix} M_0 & 0 \\ 0 & 0 \end{bmatrix}$$

where M_0 is an $k \times k$ invertible matrix associated to the restriction of L:

$$L_0 : \mathcal{R}(L^*) \to \mathcal{R}(L) \text{ given by } L_0\mathbf{x} = L\mathbf{x}$$

which is onto, hence also one-to-one: it is invertible!

L_0, and its associated matrix M_0 constitute an invertible "core" of any linear transformation, respectively, matrix.

Exercise the **Pythagorean theorem** Show that

$$\text{if } \mathbf{y} \perp \mathbf{z} \text{ then } \|\mathbf{y}\|^2 + \|\mathbf{z}\|^2 = \|\mathbf{y} + \mathbf{z}\|^2 \tag{18}$$

3. Least squares approximations

The **distance** *between two vectors* $\mathbf{x}, \mathbf{y}$ *in an inner product space is defined as* $\|\mathbf{x} - \mathbf{y}\|$.

The following characterization of orthogonal projections is very useful in approximations. As in Euclidian geometry, the shortest distance between a point and a line (or a plane) is the one measured on a perpendicular line:

Theorem 32. *Let* W *be a subspace of the inner product space* $(V, \langle\, , \rangle)$.

For any $\mathbf{x} \in V$ *the point in* W *which is at minimal distance to* $\mathbf{x}$ *is* $P\mathbf{x}$*, the orthogonal projection of* $\mathbf{x}$ *onto* W*:*

$$\|x - P\mathbf{x}\| = \min\{\|\mathbf{x} - \mathbf{w}\| \mid \mathbf{w} \in W\}$$

Proof.

We use the Pythagorean theorem for $\mathbf{y} = P\mathbf{x} - \mathbf{w}$ and $\mathbf{z} = \mathbf{x} - P\mathbf{x}$ (orthogonal to W, hence to $\mathbf{y}$) gives

$$\|\mathbf{x} - \mathbf{w}\|^2 = \|\mathbf{w} - P\mathbf{x}\|^2 + \|\mathbf{x} - P\mathbf{x}\|^2, \quad \text{for any } \mathbf{w} \in W$$

implying that $\|\mathbf{x} - \mathbf{w}\| \geq \|\mathbf{x} - P\mathbf{x}\|$ with equality only for $\mathbf{w} = P\mathbf{x}$. $\square$

3.1. **Overdetermined systems: best fit solution.** Let M be an $m \times n$ matrix with entries in $\mathbb{R}$ (one could also work in $\mathbb{C}$). By abuse of notation we will speak of the matrix M both as a matrix and as the linear transformation $\mathbb{R}^n \to \mathbb{R}^m$ which takes $\mathbf{x}$ to $M\mathbf{x}$, denoting by $\mathcal{R}(M)$ the range of the transformation (the column space of M).

The linear system $M\mathbf{x} = \mathbf{b}$ has solutions if and only if $\mathbf{b} \in \mathcal{R}(M)$. If $\mathbf{b} \notin \mathcal{R}(M)$ then the system has no solutions, and it is called *overdetermined*.

In practice overdetermined systems are not uncommon, usual sources being that linear systems are only models for more complicated phenomena, and the collected data is subject to errors and fluctuations. For practical problems it is important to produce a best fit solution: an $\mathbf{x}$ for which the error $M\mathbf{x} - \mathbf{b}$ is as small as possible.

There are many ways of measuring such an error, often this is *the least squares*: find $\mathbf{x}$ which minimizes the square error:

$$S = r_1^2 + \ldots + r_m^2 \quad \text{where } r_j = (M\mathbf{x})_j - b_j$$

Of course, this is the same as minimizing $\|M\mathbf{x} - \mathbf{b}\|$ where the inner product is the usual dot product on $\mathbb{R}^m$. By Theorem 32 it follows that $M\mathbf{x}$ must equal $P\mathbf{b}$, the orthogonal projection of $\mathbf{b}$ onto the subspace $\mathcal{R}(M)$.

We now need to solve the system $M\overline{\mathbf{x}} = P\mathbf{b}$ (solvable since $P\mathbf{b} \in \mathcal{R}(M)$). **The notation $\overline{\mathbf{x}}$ is standard in statistics for the best fit. It can be confused with the complex conjugate!**

Remark: (a) If M is one to one, then so is M^*M (since if $M^*M\mathbf{x} = \mathbf{0}$ then $0 = \langle M^*M\mathbf{x}, \mathbf{x}\rangle = \langle M\mathbf{x}, M\mathbf{x}\rangle$ therefore $M\mathbf{x} = \mathbf{0}$).

(b) Note that M^*M and MM^* are always square matrices.

Case I: If M is one to one, then there is a unique solution $\overline{\mathbf{x}} = M^{-1}P\mathbf{b}$.

An easier to implement formula can be found as follows. Since P is an orthogonal projection on $\mathcal{R}(M)$, we have $(\mathbf{b}-P\mathbf{b}) \perp \mathcal{R}(M)$ then $(\mathbf{b}-P\mathbf{b}) \in \mathcal{N}(M^*)$ (by Theorem 29 (i)) therfeore $M^*\mathbf{b} = M^*P\mathbf{b}$, so

$$M^*\mathbf{b} = M^*M\overline{\mathbf{x}}$$

which is also called *the normal equation* in statistics.

Since we assumed M is one to one, then M^*M is one to one, therefore it is invertible (being a square matrix), and we can solve

$$\overline{\mathbf{x}} = (M^*M)^{-1}M^*\mathbf{b}$$

Since $M\overline{\mathbf{x}} = P\mathbf{b}$ we also found a formula for the projection

$$P\mathbf{b} = M(M^*M)^{-1}M^*\mathbf{b} \tag{19}$$

Case II: If M is not one to one, then, given one solution $\overline{\mathbf{x}}$ then any vector in the space $\overline{\mathbf{x}} + \mathcal{N}(M)$ is a solution as well. Choosing the vector $\overline{\overline{\mathbf{x}}}$ with the smallest norm in $\overline{\mathbf{x}} + \mathcal{N}(M)$, this gives $\overline{\overline{\mathbf{x}}} = M^+\mathbf{b}$ where M^+ is called the *pseudoinverse* of M. The notion of pseudoinverse will be studied in more detail later on. By Theorem 32 $\overline{\overline{\mathbf{x}}} = \mathbf{x} - P_{\mathcal{N}(M)}\mathbf{x}$.

3.2. **Another formula for orthogonal projections.** Formula (19) is another useful way of writing projections. Suppose that W is a subspace in $\mathbb{R}^n$ (or $\mathbb{C}^n$) and $\mathbf{x}_1, \ldots, \mathbf{x}_r$ is a basis for W (not necessarily orthonormal). The matrix $M = [\mathbf{x}_1, \ldots, \mathbf{x}_r]$ has its column space equal to W and has linearly independent columns, therefore a zero null space. Then (19) is the orthogonal projection to W.

4. Orthogonal and unitary matrices, QR factorization

4.1. **Unitary and orthogonal matrices.** Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space. The following type of linear transformation are what isomorphisms of inner product spaces should be: linear, invertible, and they preserve the inner product (therefore angles and lengths):

Definition 33. *A linear transformation $U : V \to V$ is called a* **unitary transformation** *if $U^*U = UU^* = I$.*

Remark. U is unitary if and only if U^* is unitary.

In the language of matrices:

Definition 34. *A* **unitary matrix** *is an $n\times n$ matrix with $U^*U = UU^* = I$ and therefore $U^{-1} = U^*$.*

Definition 35. *A unitary matrix with real entries is called an* **orthogonal matrix***.*

In other words, Q is an orthogonal matrix means that Q has real elements and $Q^{-1} = Q^T$.

We should immediately state and prove the following properties of unitary matrices, each one can be used as a definition for a unitary matrix:

Theorem 36. *Let U be $n \times n$ a matrix.*

The following statements are equivalent:
(i) U is unitary.
(ii) The columns of U form an orthonormal set of vectors (therefore an orthonormal basis of $\mathbb{C}^n$).
(iii) U preserves inner products: $\langle U\mathbf{x}, U\mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$.
(iv) U is an isometry: $\|U\mathbf{x}\| = \|\mathbf{x}\|$ for all $\mathbf{x} \in \mathbb{C}^n$.

Remark. If U is unitary, also *the rows of U from an orthonormal set.*

Examples. Rotation matrices and reflexion matrices in $\mathbb{R}^n$, and their products, are orthogonal (they preserve the length of vectors).

Remark. An isometry is necessarily one to one, and therefore, it is also onto (in finite dimensional spaces).

Remark. The equivalence between (i) and (iv) is not true in infinite dimensions (unless the isometry is assumed onto).

Proof of Theorem 36.

(i)$\Longleftrightarrow$(ii) is obvious by matrix multiplication (line j of U^* multiplying, place by place, column i of U is exactly the dot product of column j complex conjugated and column i).

(i)$\Longleftrightarrow$(iii) is obvious because $U^*U = I$ is equivalent to $\langle U^*U\mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, which is equivalent to (iii).

(iii)$\Rightarrow$(iv) by taking $\mathbf{y} = \mathbf{x}$.

(iv)$\Rightarrow$(iii) follows from the polarization identity. $\square$

4.2. Rectangular matrices with orthonormal columns. A simple formula for orthogonal projections.

Let $M = [\mathbf{u}_1, \ldots, \mathbf{u}_k]$ be an $m \times k$ matrix whose columns $\mathbf{u}_1, \ldots, \mathbf{u}_k$ are an orthonormal set of vectors. Then necessarily $m \geq k$. If $m = k$ then the matrix M is unitary, but assume here that $m > k$.

Note that $\mathcal{N}(M) = \{\mathbf{0}\}$ since the columns are independent.

Note also that

$$M^*M = I$$

(since line j of M^* multiplying, place by place, column i of M is exactly $\overline{\mathbf{u}_j} \cdot \mathbf{u}_i$ which equals 1 for $i = j$ and 0 otherwise). Then the least squares minimization formula (19) takes the simple form $P = MM^*$:

Theorem 37. *Let $\mathbf{u}_1, \ldots, \mathbf{u}_k$ be an orthonormal set, and $M = [\mathbf{u}_1, \ldots, \mathbf{u}_k]$.*

The orthogonal projection onto $Sp(\mathbf{u}_1, \ldots, \mathbf{u}_k)$ (cf. §3.2) is

$$P = MM^*$$

4.3. QR factorization.

The following decomposition of matrices has countless applications, and extends to infinite dimensions.

If an $m \times k$ matrix $M = [\mathbf{y}_1, \ldots, \mathbf{y}_k]$ has linearly independent columns (hence $m \geq k$ and $\mathcal{N}(M) = \{\mathbf{0}\}$) then applying the Gram-Schmidt process on the columns $\mathbf{y}_1, \ldots, \mathbf{y}_k$ amounts to factoring $M = QR$ as described below.

Theorem 38. QR factorization of matrices
Any $m \times k$ matrix $M = [\mathbf{y}_1, \ldots, \mathbf{y}_k]$ with linearly independent columns can be factored as $M = QR$ where Q is an $m \times k$ matrix whose columns form an orthonormal basis for $\mathcal{R}(M)$ (hence $Q^*Q = I$) and R is an $k \times k$ upper triangular matrix with positive entries on its diagonal (hence R is invertible).

In the case of a square matrix M then Q is also square, and it is a unitary matrix.

If the matrix M has real entries, then Q and R have real entries, and if $k = n$ then Q is an orthogonal matrix.

Remark 39. *A similar factorization can be written $A = Q_1R_1$ with Q_1 an $m \times m$ unitary matrix and R_1 an $m \times k$ rectangular matrix whose first k rows are the upper triangular matrix R and the last $m - k$ rows are zero:*

$$A = Q_1R_1 = [Q\ Q_2] \begin{bmatrix} R \\ 0 \end{bmatrix} = QR$$

Proof of Theorem 38.

Every step of the Gram-Schmidt process (8), (9), … (11) is completed by a special normalization: after obtaining an orthonormal set $\mathbf{u}_1, \ldots, \mathbf{u}_k$ let

$$\mathbf{q}_j = \gamma_j\, \mathbf{u}_j \quad \text{for all } j = 1, \ldots k, \quad \text{where } \gamma_j \in \mathbb{C},\ |\gamma_j| = 1 \tag{20}$$

where the numbers γ_j will be suitably determined to obtain a special orthonormal set $\mathbf{q}_1, \ldots, \mathbf{q}_k$.

First, replace $\mathbf{u}_1, \ldots, \mathbf{u}_k$ in (8), (9),…,(11) by the orthonormal set $\mathbf{q}_1, \ldots, \mathbf{q}_k$.

Then invert: write $\mathbf{y}_j$'s in terms of $\mathbf{q}_j$'s. Since $\mathbf{q}_1 \in Sp(\mathbf{y}_1)$, $\mathbf{q}_2 \in Sp(\mathbf{y}_1, \mathbf{y}_2)$,…, $\mathbf{q}_j \in Sp(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_j), \ldots$ then $\mathbf{y}_1 \in Sp(\mathbf{q}_1)$, $\mathbf{y}_2 \in Sp(\mathbf{q}_1, \mathbf{q}_2)$,…, $\mathbf{y}_j \in Sp(\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_j), \ldots$ and therefore there are scalars c_{ij} so that

$$\mathbf{y}_j = c_{1j}\mathbf{q}_1 + c_{2j}\mathbf{q}_2 + \ldots + c_{jj}\mathbf{q}_j \quad \text{for each } j = 1, \ldots, k \tag{21}$$

and since $\mathbf{q}_1, \ldots, \mathbf{q}_k$ are orthonormal, then

$$c_{ij} = \langle \mathbf{q}_i, \mathbf{y}_j \rangle \tag{22}$$

Relations (21), (22) can be written in matrix form as

$$M = QR$$

with

$$Q = [\mathbf{q}_1, \ldots, \mathbf{q}_k], \quad R = \begin{bmatrix} \langle \mathbf{q}_1, \mathbf{y}_1 \rangle & \langle \mathbf{q}_1, \mathbf{y}_2 \rangle & \ldots & \langle \mathbf{q}_1, \mathbf{y}_k \rangle \\ 0 & \langle \mathbf{q}_2, \mathbf{y}_2 \rangle & \ldots & \langle \mathbf{q}_2, \mathbf{y}_k \rangle \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & \langle \mathbf{q}_k, \mathbf{y}_k \rangle \end{bmatrix}$$

We have the freedom of choosing the constants γ_j of modulus 1, and they can be chosen so that all diagonal elements of R are positive (since $\langle \mathbf{q}_j, \mathbf{y}_j \rangle = \overline{\gamma_j} \langle \mathbf{u}_j, \mathbf{y}_j \rangle$ choose $\gamma_j = \langle \mathbf{u}_j, \mathbf{y}_j \rangle / |\langle \mathbf{u}_j, \mathbf{y}_j \rangle|$). $\square$

For numerical calculations the Gram-Schmidt process described above accumulates round-off errors. For large m and k other more efficient, numerically stable, algorithms exist, and should be used.

Applications of the QR factorization to solving linear systems.

1. Suppose M is an invertible square matrix. To solve $M\mathbf{x} = \mathbf{b}$, factoring $M = QR$, the system is $QR\mathbf{x} = \mathbf{b}$, or $R\mathbf{x} = Q^*\mathbf{b}$ which can be easily solved since R is triangular.

2. Suppose M is an $m \times k$ rectangular matrix, of full rank k. Since $m > k$ the linear system $M\mathbf{x} = \mathbf{b}$ may be overdetermined. Using the QR factorization in Remark 39, the system is $Q_1 R_1 \mathbf{x} = \mathbf{b}$, or $R_1 \mathbf{x} = Q_1^* \mathbf{b}$, which is easy to see if it has solutions: the last $m-k$ rows of $Q_1^*\mathbf{b}$ must be zero. If this is the case, the system can be easily solved since R is upper triangular.