5

Orthogonality

Life is like riding a bicycle. To keep your balance, you must keep moving.

Albert Einstein

5.1 Dot product

Given two n-vectors $u = (u_1, \ldots, u_n)$ and $v = (v_1, \ldots, v_n)$, their *dot product* is the following operation:

$$u \cdot v = \sum_{i=1}^{n} u_i v_i.$$

Is is worth noting that that dot product satisfies the following properties:

- $u \cdot v = v \cdot u$
- $u \cdot u \geq 0$

- If $u \neq \vec{0}$ then $u \cdot u > 0$
- $u \cdot \vec{0} = \vec{0} \cdot u = 0$
- $(u + v) \cdot w = (u \cdot w) + (v \cdot w)$
- $(\lambda u) \cdot v = u \cdot (\lambda v) = \lambda(u \cdot v)$

The *length* of a vector u, represented as $\|u\|$, can be defined via the dot product:

$$\|u\| = \sqrt{u \cdot u}$$

Note: Often times in the literature the dot product is represented as $u^t v$ because if we think of a vector as a single column matrix, this is exactly what comes out.

5.2 Angles between vectors

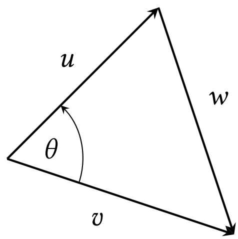
The dot product gives us an interesting way to rapidly assess what is the angle formed by the two vectors involved. In particular, we have the following result:

Proposition 16. *The angle θ formed by two n-vectors u, v satisfies the following:*

$$u \cdot v = \|u\|\|v\| \cos\theta$$

In particular, two non-zero vectors u and v are orthogonal if and only if $u \cdot v = 0$.

Let's adopt the arrow perspective and picture the vectors u, v and $w = v - u$:



Let's do a little computation,

$$w \cdot w = (v - u) \cdot (v - u) = (v \cdot v) - 2u \cdot v + (u \cdot u).$$

We can refactor this equation using lengths,

$$\|w\|^2 = \|v\|^2 + \|u\|^2 - 2u \cdot v.$$

But the *Law of Cosines* says that for a triangle with sides a, b, c with angle θ opposite to c, the following identity must hold:

$$c^2 = a^2 + b^2 - 2ab\cos\theta,$$

hence the conclusion follows.

We often denote that two non-zero vectors are orthogonal with the notation $u \perp v$. Whenever the zero vector may be involved, we will continue to use this notation implying that the dot product is zero.

5.3 Orthogonal spaces

Vectors can be orthogonal to one another, but also vector subspaces. We say that two vector subspaces U and V of $\mathbb{R}^n$ are orthogonal, denoted $U \perp V$, if any vector in U is orthogonal to any vector in V.

The set of all vectors that are othogonal to each vector in a given vector space U is referred to as the *orthogonal space* or *orthogonal complement* of U and denoted $U^\perp$:

$$U^\perp = \{v \in \mathbb{R}^n \mid v \cdot u = 0 \text{ for each } u \in U\}$$

We abuse a little this notation and refer to the orthogonal space of a vector u, denoted $u^\perp$, as the orthogonal space of the vector space spanned by u:

$$\mathbf{span}(u)^\perp = \{v \in \mathbb{R}^n \mid v \cdot u = 0\}.$$

5.3.1 Example

Given a matrix A, its null space $N(A)$ can be seen as $R(A)^\perp$, the orthogonal complement of the row space of A. Similarly, the null space of its transpose $N(A^t)$ can be seen as $C(A)^\perp$, the orthogonal complement of the column space of A.

Proposition 17. *Suppose that $u_1, \ldots, u_k$ is a linearly independent set from the vector space V and that $\mathbf{dim}(V) = n > k$. Then*

$$\mathbf{dim}(\mathbf{span}(u_1, \ldots, u_k)^\perp) = n - k$$

In other words, we can find a linearly independent set $u_{k+1}, \ldots, u_n$ that are orthogonal to $\mathbf{span}(u_1, \ldots, u_k)$. Moreover $\{u_1, \ldots, u_n\}$ is a basis of V.

If we arrange $u_1, \ldots, u_k$ as columns in a $n \times k$ matrix A, applying Theorem 1 to A then we know that

$$n = \mathbf{dim}(N(A^t)) + \mathbf{rank}(A^t) \Rightarrow \mathbf{dim}(N(A^t)) = n - k$$

where all vectors in $N(A^t)$ are orthogonal to $u_1, \dots, u_k$, the columns of A. So we can pick $u_{k+1}, \dots, u_n$ some basis of $N(A^t)$. If $u_1, \dots, u_n$ were not linearly independent then there would be a linear combination

$$\sum_{i=1}^{k} \lambda_i u_i + \sum_{j=k+1}^{n} \lambda_j u_j = \vec{\mathbf{0}} .$$

Not all $\lambda_{k+1}, \dots, \lambda_n$ can be zero, otherwise $u_1, \dots, u_k$ would not be linearly independent, therefore

$$\sum_{i=1}^{k} \lambda_i u_i + w = \vec{\mathbf{0}} \quad \text{with } w \in N(A^t), w \neq \vec{\mathbf{0}} .$$

But since w is orthogonal to the $u_1, \dots, u_k$ we can apply the dot-product-trick:

$$w \cdot \left(\sum_{i=1}^{k} \lambda_i u_i \right) + w \cdot w = 0 \Rightarrow \|w\|^2 = 0,$$

which contradicts the fact that $\{u_1, \dots, u_n\}$ are linearly dependent.

5.4 Orthogonal sets of vectors

A set of non-zero n-vectors $\{v_1, \dots, v_k\}$ is said to be an *orthogonal set* if every vector is orthogonal to every other vector in the set:

$$v_i \perp v_j \text{ whenever } i \neq j.$$

Orthogonal sets are a particularly easy type of set to handle. We have previously discussed linear independence, a property whereby every element of the set provides some filling-up of the vector space that none of the other elements can provide either on their own or combined. One would intuitively think that orthogonality would bear some relationship with linear independence because, in the end of the day, what filling-up can be more Orthogonal sets are linearlyefficient than adding orthogonal vectors to the collection?

In fact we can prove that for orthogonal sets linear independence is guaranteed.

Proposition 18. *Orthogonal sets are linearly independent.*

Suppose that $u_1, \ldots, u_k$ are non-zero and orthogonal to each other. If they were linearly dependent, there would be scalars $\lambda_1, \ldots, \lambda_k$, not all zero, such that

$$\lambda_1 v_1 + \ldots + \lambda_k v_k = \overrightarrow{\mathbf{0}} \ .$$

Relabeling if necessary, suppose that $\lambda_1 \neq 0$. Doing dot product of the linear combination with v_1 we get:

$$0 = (\lambda_1 v_1 + \ldots + \lambda_k v_k) \cdot v_1 = \lambda_1 (v_1 \cdot v_1) + \ldots + \lambda_k (v_k \cdot v_1) = \lambda_1 \|v_1\|^2$$

since $v_i \cdot v_1 = 0$ for all $i \neq 1$, hence the only surviving term is the $v_1 \cdot v_1$ term. But this identity contradicts the fact that all the vectors in the set were non-zero.

It follows that any generating set that is orthogonal is a basis: an *orthogonal basis*. Moreover, if all the elements of a basis has unit length, the basis can be referred to as an *orthonormal basis*.

From an orthogonal basis, say $v_1, \ldots, v_n$, we can always an orthonormal one, simply by rescaling each element according to their length:

$$u_i = \frac{1}{\|v_i\|} v_i$$

In this case $u_1, \ldots, u_n$ would be an orthonormal basis.

Orthonormal bases can be really handy. For example, the coordinates of a vector in an orthonormal basis can be recovered by doing dot product against the elements of the basis themselves.

Proposition 19. *Given an n-vector v and an orthonormal basis $\mathcal{V} = \{v_1, \ldots, v_n\}$, the coordinates of v in the basis $\mathcal{V}$ are given by $(v_1 \cdot v, \ldots, v_n \cdot v)$.*

To prove it, suppose that $v = \lambda_1 v_1 + \ldots + \lambda_n v_n$. Then

$$v_i \cdot v = \lambda_i v_i \cdot v_i = \lambda_i \|v_i\|^2 = \lambda_i.$$

5.5 Orthogonal matrices

Orthonormal sets of vectors give rise to a special type of matrix known as *orthogonal matrices*: these are square matrices whose column vectors are mutually orthogonal and have unit length. This makes inverting orthogonal matrices an easy operation:

Proposition 20. *If Q is an orthogonal matrix, then $Q^{-1} = Q^t$. In particular, $\textbf{det}(Q) = \pm 1$*

If q_i are the column vectors of Q, multiplying Q^tQ yields a matrix that has $q_i \cdot q_i = 1$ as elements in the diagonal and $q_j \cdot q_i = 0$ off the diagonal, hence the identity matrix. Note that this implies that the determinant must be either 1 or -1:

$$1 = \mathbf{det}(\mathbf{I}) = \mathbf{det}(Q^tQ) = \mathbf{det}(Q^t)\mathbf{det}(Q) = (\mathbf{det}(Q))^2$$

In $\mathbb{R}^2$ the repertoire of orthogonal transformations is relatively limited: they can be either of the form

$$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \text{ or } \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

5.6 Orthogonal projection onto a vector

Consider two n-vectors u and v. The *orthogonal projection of v onto $\mathbf{span}(u)$* is another vector $p(v)$ that satisfies the following two conditions:

- $p(v) \in \mathbf{span}(u)$
- $v - p(v) \perp u$

How can we compute $p(v)$? If $v \in \mathbf{span}(u)$ then $p(v) = v$ is the trivial solution to the problem. So we can assume from now on that $v \notin \mathbf{span}(u)$. By the first condition, we know that there is some scaling factor λ such that $p(v) = \lambda u$, so we need to find λ. Using this piece of information with the second condition, we also know that $(v - \lambda u) \cdot u = 0$, implying that

$$\lambda = \frac{v \cdot u}{u \cdot u},$$

leading to one possible form of the solution:

$$p(v) = \frac{v \cdot u}{u \cdot u}\, u.$$

What is the length of $p(v)$?

$$\|p(v)\| = \frac{v \cdot u}{u \cdot u}\|u\| = \frac{v \cdot u}{\|u\|} = \|v\| \cos\theta,$$

where θ is the angle formed by u and v.

We can also see that the projection onto u is in fact a linear transformation:

$$p(v+w) = \left(\frac{v \cdot u}{u \cdot u} + \frac{w \cdot u}{u \cdot u}\right) u = p(v) + p(w)$$
$$p(\alpha v) = \frac{\alpha v \cdot u}{u \cdot u}\, u = \alpha \frac{v \cdot u}{u \cdot u}\, u = \alpha p(v)$$

So there must be a $n \times n$ matrix P that gets the job done. What matrix P accomplishes it?

$$p(v) = \frac{v \cdot u}{u \cdot u}\, u = \frac{1}{\|u\|^2} u u^t v$$

Therefore,

$$P = \frac{1}{\|u\|^2} u u^t = \left(\frac{u}{\|u\|}\right)\left(\frac{u}{\|u\|}\right)^t$$

works as intented.

Note that the expression of P simplifies quite a bit if u has unit length $\|u\| = 1$:

$$P = u u^t$$

The projection $p(v)$ of v onto $\mathbf{span}(u)$ is what v needs to get rid of to become orthgonal to u, i.e. $p(v)$ is what we need to substract from v to render it orthogonal to u.

5.7 Gram-Schmidt method

Given a vector space of interest V, we are now concerned about finding an orthonormal basis of V. Our raw materials for this task will be some given basis $\mathscr{V} = v_1, \ldots, v_k$ of V, not necessarily orthogonal. Our strategy will be to go about modifying the vectors of the given basis so that we end up with another basis that is orthogonal.

5.7.1 Prunning

Before going to a general procedure, we first investigate how to proceed in the simple case of two vectors of the plane $\mathbb{R}^2$. Suppose that u and v are a basis of $\mathbb{R}^2$ and that u and v are not orthogonal. How can we change one of the vectors, say v, to obtain a new vector v' in such a way that u and v' are orthogonal?

We remove from v what separates it from being orthogonal to u. From the last section we now that this is precisely the projection $p(v)$ of v onto $\mathbf{span}(u)$. Therefore we can conclude that:

$$v' = v - p(v) = v - \frac{v \cdot u}{u \cdot u}\, u$$

is orthogonal to u. We have managed to prune the part of v that was impeding orthogonality with u. This simple case illuminates the way to transform any basis into an orthogonal basis.

5.7.2 The algorithm

Starting from a basis $v_1, \ldots, v_k$ of V we will conduct a series of steps to produce an orthonormal basis of V: $w_1, \ldots, w_n$.

Take

$$w_1 = \frac{v_1}{\|v_1\|}.$$

This will be the first element of the basis.

Now suppose that we have already computed the first i elements of our orthonormal basis: $w_1, \ldots, w_i$. The next element will be computed from v_{i+1} by prunning everything that impedes it to be orthogonal to each of the preceding members of the basis:

$$w'_{i+1} = v_{i+1} - (v_{i+1} \cdot w_1)w_1 - \ldots - (v_{i+1} \cdot w_i)w_i$$

then we convert it to a unit length vector:

$$w_{i+1} = \frac{w'_{i+1}}{\|w'_{i+1}\|}.$$

It is easy to check that at every step we are adding a vector that is orthogonal to all the previously added vectors. Let's compute that $w'_{i+1} \cdot w_j = 0$ for all $1 \leq j \leq i$:

$$w'_{i+1} \cdot w_j = v_{i+1} \cdot w_j - (v_{i+1} \cdot w_1)(w_1 \cdot w_j) - \ldots - (v_{i+1} \cdot w_i)(w_i \cdot w_j).$$

Observe that all the terms with $w_k \cdot w_j$ with $k \neq j$ must be zero. The only remaining terms are

$$v_{i+1} \cdot w_j - (v_{i+1} \cdot w_j)(w_j \cdot w_j) = v_{i+1} \cdot w_j - v_{i+1} \cdot w_j = 0.$$

Therefore the algorithm works as intended.

Since we can apply this method to any given basis of any vector space, it is guaranteed that:

Proposition 21. *Any vector space has an orthonormal basis.*

5.8 Orthogonal projection: general case

Consider now an orthonormal basis $\mathscr{U} = \{u_1, \ldots, u_k\}$ of a subspace of V and a vector $v \in V$. The *orthogonal projection of v onto* $\boldsymbol{span}(\mathscr{U})$ is another vector $p(v)$ that satisfies the following conditions:

- $p(v) \in \mathbf{span}(\mathscr{U})$
- $v - p(v) \perp u$ for all $u \in \mathbf{span}(\mathscr{U})$

We can define an orthonormal basis of V by *orthogonal completion* of $\mathscr{U}$, i.e. adding new unit length vectors $u_{k+1}, \ldots, u_n$ that are all orthogonal to $\mathscr{U}$. The resulting set $\mathscr{U}' = \mathscr{U} \cup \{u_{k+1}, \ldots, u_n\}$ must be an orthonormal basis of V. What would be the effect of projecting the vectors of $\mathscr{U}'$ onto $\mathbf{span}(\mathscr{U})$? The matrix of the resulting linear transformation in coordinates of $\mathscr{U}'$ would be the following $n \times n$ matrix:

$$\tilde{P} = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \end{bmatrix}$$

implying that $\tilde{P}u_i = u_i$ as long as $1 \leq i \leq k$ and $\tilde{P}u_i = \overrightarrow{\mathbf{0}}$ otherwise.

If Q is the orthogonal matrix that has the u_j's as columns, we can recover the matrix of the projection in the canonical basis, as we saw previously, as follows:

$$P = Q\tilde{P}Q^t$$

Note that since the last $n-k$ columns and rows of $\tilde{P}$ are zero, the last $n-k$ columns of Q have no relevance on the results and we can get rid of them.

Proposition 22. *The matrix P of the orthogonal projection onto a vector subspace S of $\mathbb{R}^n$ is given by*

$$P = UU^t$$

where U is the matrix with an orthonormal basis of S as column vectors.

We can also recover a general expression when we are given just an ordinary basis of S, under no orthonormality assumption whatsoever, say $w_1, \dots, w_k$. Suppose that A is the matrix with w_j as columns. The key observation to come up with it is that there is an invertible $k \times k$ matrix Λ such that $U = A\Lambda$ has an orthonormal set as column vectors: we can think of the columns of Λ as linear combination recipes telling us how to combine the columns of A to generate an orthonormal basis of S.

Hence the projection matrix must be:

$$P = UU^t = (A\Lambda)(A\Lambda)^t = A\Lambda\Lambda^t A^t$$

Amazingly, the matrix $\Lambda\Lambda^t$ can be recovered entirely from A!

$$\Lambda\Lambda^t = (A^t A)^{-1}$$

Since $A = U\Lambda^{-1}$ we can see that

$$(A^t A)^{-1} = ((U\Lambda^{-1})^t (U\Lambda^{-1}))^{-1} = ((\Lambda^{-1})^t U^t U \Lambda^{-1})^{-1}$$

Using the fact that $U^t U = \mathbf{I}$, it follows that:

$$(A^t A)^{-1} = ((\Lambda^{-1})^t \Lambda^{-1})^{-1} = \Lambda((\Lambda^{-1})^t)^{-1} = \Lambda\Lambda^t$$

Proposition 23. *The matrix P of the orthogonal projection onto a vector subspace S of $\mathbb{R}^n$ is given by*

$$P = A(A^t A)^{-1} A^t$$

where A is the matrix with any basis of S as column vectors.

5.9 TL;DR

When we represent vectors as arrows, we can study their geometric properties, like the length or the angles formed between vectors. Here we introduce the dot product, yet another operation that takes a pair of vectors as input and returns a number: when this number is zero it signals that the vectors are orthogonal to each other. With this simple computational tool we can then go about studying orthgonal sets of vectors and orthogonal matrices. Orthgonality is a particularly strong condition that enforces linear independence. On the other hand, the changes of coordinates between orthogonal bases turn out to be particularly simple to calculate. We describe an algorithm, known as Gram-Schmidt, that takes as input a basis and returns an orthgonal basis spanning the same vector space. Using everything we learned, we end up by figuring out a completely general expression for the matrix of the orthogonal projection onto any given vector subspace.