Projections using orthogonal bases

We showed last time that if $\{\utilde{a}_1, \utilde{a}_2, \dots, \utilde{a}_k\}$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$ then the projection of $\utilde{v} \in \mathbb{R}^n$ onto W is given by

$$\utilde{p} = (\utilde{a}_1 \cdot \utilde{v})\utilde{a}_1 + (\utilde{a}_2 \cdot \utilde{v})\utilde{a}_2 + \cdots + (\utilde{a}_k \cdot \utilde{v})\utilde{a}_k. \tag{1}$$

We now obtain a similar formula for the case when we are given a basis of W that is merely orthogonal, rather than orthonormal.

Suppose that $\{\utilde{b}_1, \utilde{b}_2, \dots, \utilde{v}_k\}$ is an orthogonal basis for W. Normalizing, we have that $\{\frac{1}{\|\utilde{b}_1\|}\utilde{b}_1, \frac{1}{\|\utilde{b}_2\|}\utilde{b}_2, \dots, \frac{1}{\|\utilde{b}_k\|}\utilde{b}_k\}$ is an orthonormal basis for W. Now applying the formula (1) above (with $\utilde{a}_i = \frac{1}{\|\utilde{b}_i\|}\utilde{b}_i$) we see that the projection of $\utilde{v}$ onto W is given by

$$\utilde{p} = \sum_{i=1}^{k} \left(\frac{1}{\|\utilde{b}_i\|}\utilde{b}_i \cdot \utilde{v} \right) \frac{1}{\|\utilde{b}_i\|}\utilde{b}_i = \sum_{i=1}^{k} \frac{(\utilde{b}_i \cdot \utilde{v})}{\|\utilde{b}_i\|^2}\utilde{b}_i = \sum_{i=1}^{k} \frac{(\utilde{b}_i \cdot \utilde{v})}{(\utilde{b}_i \cdot \utilde{b}_i)}\utilde{b}_i. \tag{2}$$

Observe again that this is the sum of the projections onto the one-dimensional spaces spanned by the individual $\utilde{b}_i$'s. (See the discussion of projections onto one-dimensional subspaces at the start of the notes for Week 2.)

As an example, let $W = \mathrm{Span}(\utilde{b}_1, \utilde{b}_2)$, where $\utilde{b}_1 = (1,1,1)^T$ and $\utilde{b}_2 = (1,-3,2)^T$. Note that $\utilde{b}_1 \cdot \utilde{b}_2 = 1 - 3 + 2 = 0$, and so we do have an orthogonal basis for the subspace W. Observe also that $\utilde{b}_1 \cdot \utilde{b}_1 = 1^2 + 1^2 + 1^2 = 3$ and $\utilde{b}_2 \cdot \utilde{b}_2 = 1^2 + (-3)^2 + 2^2 = 14$. Now let $\utilde{v} = (-6, 8, 1)^T$. Substituting into the formula above we find that the projection $\utilde{p}$ is given by

$$\utilde{p} = \frac{(\utilde{v} \cdot \utilde{b}_1)}{3}\utilde{b}_1 + \frac{(\utilde{v} \cdot \utilde{b}_2)}{14}\utilde{b}_2 = \frac{3}{3}\utilde{b}_1 + \frac{(-28)}{14}\utilde{b}_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - 2\begin{pmatrix} 1 \\ -3 \\ 2 \end{pmatrix} = \begin{pmatrix} -1 \\ 7 \\ -3 \end{pmatrix}.$$

The Gram-Schmidt Orthogonalization Process

Since orthogonal bases are so convenient to work with, it is important to have a procedure for constructing them. There is indeed such a procedure, called the *Gram-Schmidt Orthogonalization Process*. The process itself is straightforward, despite its rather grand sounding name.

We suppose that we are given a basis $\{\utilde{a}_1, \utilde{a}_2, \dots, \utilde{a}_k\}$ for a subspace W of $\mathbb{R}^n$. The aim is to construct from it an orthogonal basis $\{\utilde{b}_1, \utilde{b}_2, \dots, \utilde{b}_k\}$ for the same space W. It is done as follows: define

$$\begin{aligned}
\utilde{b}_1 &= \utilde{a}_1 \\
\utilde{b}_2 &= \utilde{a}_2 - \text{the projection of } \utilde{a}_2 \text{ onto } \mathrm{Span}(\utilde{b}_1) \\
\utilde{b}_3 &= \utilde{a}_3 - \text{the projection of } \utilde{a}_3 \text{ onto } \mathrm{Span}(\utilde{b}_1, \utilde{b}_2) \\
\utilde{b}_4 &= \utilde{a}_4 - \text{the projection of } \utilde{a}_4 \text{ onto } \mathrm{Span}(\utilde{b}_1, \utilde{b}_2, \utilde{b}_3) \\
&\ \vdots \\
\utilde{b}_k &= \utilde{a}_k - \text{the projection of } \utilde{a}_k \text{ onto } \mathrm{Span}(\utilde{b}_1, \utilde{b}_2, \dots, \utilde{b}_{k-1}).
\end{aligned}$$

Recall that

$$\underset{\sim}{v} - \text{the projection of } \underset{\sim}{v} \text{ onto a subspace } X$$

is orthogonal to all elements of X. So the above formulas tell us immediately that

$$\begin{aligned}
&\underset{\sim}{b}_2 \text{ is orthogonal to all elements of } \mathrm{Span}(\underset{\sim}{b}_1) \\
&\underset{\sim}{b}_3 \text{ is orthogonal to all elements of } \mathrm{Span}(\underset{\sim}{b}_1, \underset{\sim}{b}_2) \\
&\underset{\sim}{b}_4 \text{ is orthogonal to all elements of } \mathrm{Span}(\underset{\sim}{b}_1, \underset{\sim}{b}_2, \underset{\sim}{b}_3)
\end{aligned}$$

and so on. So the vectors $\underset{\sim}{b}_j$ are orthogonal to each other. So we can use Eq. (2) above when computing the projections.

Before analysing the theory, let us do an example.

Example 1. Suppose that

$$\underset{\sim}{a}_1 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \quad \underset{\sim}{a}_2 = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \quad \underset{\sim}{a}_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

We start by setting $\underset{\sim}{b}_1 = \underset{\sim}{a}_1$, and then use Eq. (2) to calculate $\underset{\sim}{b}_2$ and $\underset{\sim}{b}_3$.

$$\begin{aligned}
\underset{\sim}{b}_1 &= \underset{\sim}{a}_1 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \\
\underset{\sim}{b}_2 &= \underset{\sim}{a}_2 - \frac{(\underset{\sim}{a}_2 \cdot \underset{\sim}{b}_1)}{(\underset{\sim}{b}_1 \cdot \underset{\sim}{b}_1)} \underset{\sim}{b}_1 \\
&= \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} - \frac{2}{5} \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} -2/5 \\ 1/5 \\ 2 \end{pmatrix}, \\
\underset{\sim}{b}_3 &= \underset{\sim}{a}_3 - \frac{(\underset{\sim}{a}_3 \cdot \underset{\sim}{b}_1)}{(\underset{\sim}{b}_1 \cdot \underset{\sim}{b}_1)} \underset{\sim}{b}_1 - \frac{(\underset{\sim}{a}_3 \cdot \underset{\sim}{b}_2)}{(\underset{\sim}{b}_2 \cdot \underset{\sim}{b}_2)} \underset{\sim}{b}_2 \\
&= \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \frac{3}{5} \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} - \frac{9/5}{21/5} \begin{pmatrix} -2/5 \\ 1/5 \\ 2 \end{pmatrix} \\
&= \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 3/5 \\ 6/5 \\ 0 \end{pmatrix} + \begin{pmatrix} 6/35 \\ -3/35 \\ -6/7 \end{pmatrix} = \begin{pmatrix} 4/7 \\ -2/7 \\ 1/7 \end{pmatrix}.
\end{aligned}$$

Multiplying each vector by a nonzero scalar does not affect orthogonality of a set of vectors. So

$$\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} -2 \\ 1 \\ 10 \end{pmatrix}, \quad \begin{pmatrix} 4 \\ -2 \\ 1 \end{pmatrix}$$

also form an orthogonal set. And we can normalize the vectors (divide each by its length) to get three vectors that form an orthonormal basis of $\mathbb{R}^3$:

$$\frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \quad \frac{1}{\sqrt{105}} \begin{pmatrix} -2 \\ 1 \\ 10 \end{pmatrix}, \quad \frac{1}{\sqrt{21}} \begin{pmatrix} 4 \\ -2 \\ 1 \end{pmatrix}.$$

Let us now look at the process in more detail. To express the formulas above more compactly, let us use the following notation: if X is a subspace of $\mathbb{R}^n$ and $\underset{\sim}{a} \in \mathbb{R}^n$, let $\operatorname{Proj}(X, \underset{\sim}{a})$ be the orthogonal projection of the vector $\underset{\sim}{a}$ onto the subspace X. Recall that $\operatorname{Proj}(X, \underset{\sim}{a})$ is by definition the vector in X that is closest to $\underset{\sim}{a}$, and that $\underset{\sim}{a} - \operatorname{Proj}(X, \underset{\sim}{a})$ is orthogonal to all elements of X.

Recall that the subspace spanned by a collection of vectors is by definition the set of all linear combinations of those vectors. We shall write X_j for the subspace spanned by $\underset{\sim}{b}_1, \underset{\sim}{b}_2, \ldots, \underset{\sim}{b}_j$ (for each j from 1 to k):

$$X_j \stackrel{\text{def}}{=} \operatorname{Span}(\underset{\sim}{b}_1, \underset{\sim}{b}_2, \ldots, \underset{\sim}{b}_j).$$

For each $j > 1$ the vector $\underset{\sim}{b}_j$ is given by

$$\underset{\sim}{b}_j = \underset{\sim}{a}_j - \operatorname{Proj}(X_{j-1}, \underset{\sim}{a}_j) \tag{3}$$

and so all elements of X_{j-1} are orthogonal to $\underset{\sim}{b}_j$. Now if $i < j$ then $\underset{\sim}{b}_i$ is in the space X_{j-1}, and so it follows that $\underset{\sim}{b}_j$ is orthogonal to $\underset{\sim}{b}_i$. Similarly if $j < i$ then $\underset{\sim}{b}_j$ is in X_{i-1}), and so $\underset{\sim}{b}_i$ is orthogonal to $\underset{\sim}{b}_j$. Hence $\underset{\sim}{b}_i \cdot \underset{\sim}{b}_j = 0$ whenever $i \neq j$; so $\{\underset{\sim}{b}_1, \underset{\sim}{b}_2, \ldots, \underset{\sim}{b}_k\}$ is an orthogonal set of vectors.

We still have to prove that $\{\underset{\sim}{b}_1, \underset{\sim}{b}_2, \ldots, \underset{\sim}{b}_k\}$ is a basis for W. In fact, we shall prove more than this. We shall show that for each value of i in the range $1 \leq i \leq k$, the vectors $\underset{\sim}{b}_1, \underset{\sim}{b}_2, \ldots, \underset{\sim}{b}_i$ span the same subspace as the vectors $\underset{\sim}{a}_1, \underset{\sim}{a}_2, \ldots, \underset{\sim}{a}_i$:

$$\operatorname{Span}(\underset{\sim}{a}_1, \underset{\sim}{a}_2, \ldots, \underset{\sim}{a}_i) = \operatorname{Span}(\underset{\sim}{b}_1, \underset{\sim}{b}_2, \ldots, \underset{\sim}{b}_i) = X_i.$$

In other words, the vectors $\underset{\sim}{b}_1$, $\underset{\sim}{b}_2$, …, $\underset{\sim}{b}_i$ can all be expressed as linear combinations of the vectors $\underset{\sim}{a}_1$, $\underset{\sim}{a}_2$, …, $\underset{\sim}{a}_i$, and conversely.

The proof proceeds by induction on i. The case $i = 1$ is trivial: $\underset{\sim}{b}_1 = \underset{\sim}{a}_1$; so $\underset{\sim}{b}_1$ can be expressed in terms of $\underset{\sim}{a}_1$, and conversely. So let us suppose that $i > 1$ and that the result is true for $i - 1$. Thus our inductive assumption is that

$$\operatorname{Span}(\underset{\sim}{a}_1, \underset{\sim}{a}_2, \ldots, \underset{\sim}{a}_{i-1}) = \operatorname{Span}(\underset{\sim}{b}_1, \underset{\sim}{b}_2, \ldots, \underset{\sim}{b}_{i-1}) = X_{i-1}.$$

Since $\underset{\sim}{b}_1$, $\underset{\sim}{b}_2$, …, $\underset{\sim}{b}_{i-1}$ are all in X_{i-1}, they can be expressed as linear combinations of $\underset{\sim}{a}_1$, $\underset{\sim}{a}_2$, …, $\underset{\sim}{a}_{i-1}$; likewise, since $\underset{\sim}{a}_1$, $\underset{\sim}{a}_2$, …, $\underset{\sim}{a}_{i-1}$ are all in X_{i-1}, they can be expressed as linear combinations of $\underset{\sim}{b}_1$, $\underset{\sim}{b}_2$, …, $\underset{\sim}{b}_{i-1}$. To deduce the same result for i, all we have to do is show that $\underset{\sim}{b}_i$ can be expressed as a linear combination of $\underset{\sim}{a}_1$, $\underset{\sim}{a}_2$, …, $\underset{\sim}{a}_i$, and $\underset{\sim}{a}_i$ can be expressed as a linear combination of $\underset{\sim}{b}_1$, $\underset{\sim}{b}_2$, …, $\underset{\sim}{b}_i$. But this follows readily from the formula for $\underset{\sim}{b}_i$. By definition,

$$\begin{aligned} \underset{\sim}{b}_i &= \underset{\sim}{a}_i - \operatorname{Proj}(X_{i-1}, \underset{\sim}{a}_i) \\ &= \underset{\sim}{a}_i - \text{some linear combination of } \{\underset{\sim}{a}_1, \underset{\sim}{a}_2, \ldots, \underset{\sim}{a}_{i-1}\}, \end{aligned} \tag{4}$$

since the projection of $\underset{\sim}{a}_i$ onto X_{i-1} is an element of X_{i-1}, and $\underset{\sim}{a}_1, \underset{\sim}{a}_2, \ldots, \underset{\sim}{a}_{i-1}$ span X_{i-1}. The equation Eq. (4) shows that $\underset{\sim}{b}_i$ is a linear combination of $\underset{\sim}{a}_1, \underset{\sim}{a}_2, \ldots, \underset{\sim}{a}_i$, as required. Moreover, $\operatorname{Proj}(X_{i-1}, \underset{\sim}{a}_i)$ can also be expressed as a linear combination of $\underset{\sim}{b}_1, \underset{\sim}{b}_2, \ldots, \underset{\sim}{b}_{i-1}$, since these vectors also span X_{i-1}; so we see that

$$\begin{aligned} \underset{\sim}{a}_i &= \underset{\sim}{b}_i + \operatorname{Proj}(X_{i-1}, \underset{\sim}{a}_i) \\ &= \underset{\sim}{b}_i + \text{some linear combination of } \{\underset{\sim}{b}_1, \underset{\sim}{b}_2, \ldots, \underset{\sim}{b}_{i-1}\} \\ &= \text{some linear combination of } \{\underset{\sim}{b}_1, \underset{\sim}{b}_2, \ldots, \underset{\sim}{b}_i\}, \end{aligned}$$

Thus all of $\utilde{b}_1$, $\utilde{b}_2$, …, $\utilde{b}_i$ can be expressed in terms of $\utilde{a}_1$, $\utilde{a}_2$, …, $\utilde{a}_i$, and conversely, and this completes the induction.

Since we are given that $\utilde{a}_1$, $\utilde{a}_2$, …, $\utilde{a}_k$ form a basis for W we know that they are linearly independent; hence $\utilde{a}_1$, $\utilde{a}_2$, …, $\utilde{a}_j$ are also linearly independent, and thus form a basis for the space they span, namely X_j. So X_j is j-dimensional, and since it is also spanned by $\utilde{b}_1, \utilde{b}_2, \ldots, \utilde{b}_j$, these vectors comprise another basis of X_j. Furthermore, we have also shown that the vectors $\utilde{b}_i$ have the property that $\utilde{b}_i \cdot \utilde{b}_l = 0$ whenever $i \neq l$; so $\{\utilde{b}_1, \utilde{b}_2, \ldots, \utilde{b}_j\}$ is an orthogonal basis for X_j. In view of this, and using the formula we derived for the projection in terms of an orthogonal basis, Eq (3) above gives

$$\utilde{b}_j = \utilde{a}_j - \sum_{i=1}^{j-1} \frac{(\utilde{a}_j \cdot \utilde{b}_i)}{(\utilde{b}_i \cdot \utilde{b}_i)} \utilde{b}_i. \tag{5}$$

This is the formula to use in practical calculations.

Gram-Schmidt process: summary

The input is a sequence of vectors $\utilde{a}_1$, $\utilde{a}_2$, …, $\utilde{a}_k$.
The output is another sequence of vectors, $\utilde{b}_1$, $\utilde{b}_2$, …, $\utilde{b}_k$, given by

$$\begin{aligned}
\utilde{b}_1 &= \utilde{a}_1 \\
\utilde{b}_2 &= \utilde{a}_2 - \frac{\utilde{a}_2 \cdot \utilde{b}_1}{\utilde{b}_1 \cdot \utilde{b}_1} \utilde{b}_1 \\
\utilde{b}_3 &= \utilde{a}_3 - \frac{\utilde{a}_3 \cdot \utilde{b}_1}{\utilde{b}_1 \cdot \utilde{b}_1} \utilde{b}_1 - \frac{\utilde{a}_3 \cdot \utilde{b}_2}{\utilde{b}_2 \cdot \utilde{b}_2} \utilde{b}_2 \\
&\vdots \\
\utilde{b}_k &= \utilde{a}_k - \frac{\utilde{a}_k \cdot \utilde{b}_1}{\utilde{b}_1 \cdot \utilde{b}_1} \utilde{b}_1 - \frac{\utilde{a}_k \cdot \utilde{b}_2}{\utilde{b}_2 \cdot \utilde{b}_2} \utilde{b}_2 - \cdots - \frac{\utilde{a}_k \cdot \utilde{b}_{k-1}}{\utilde{b}_{k-1} \cdot \utilde{b}_{k-1}} \utilde{b}_{k-1}.
\end{aligned}$$

For each i, the vectors $\utilde{b}_1$, $\utilde{b}_2$, …, $\utilde{b}_i$ form an orthogonal set spanning the same subspace as $\utilde{a}_1$, $\utilde{a}_2$, …, $\utilde{a}_i$.

Incidentally, we have assumed all along that the input vectors are linearly independent. If they are not then everything will still work in the same way, except that some of the output vectors $\utilde{b}_i$ will be zero.

Example 2. Although we usually use column vectors, the process of course works equally well for row vectors. Suppose that $\utilde{a}_1 = (1, 1, 1)$, $\utilde{a}_2 = (1, 1, 0)$ and $\utilde{a}_3 = (1, 0, 0)$. Then we obtain

$$\begin{aligned}
\utilde{b}_1 &= (1, 1, 1) \\
\utilde{b}_2 &= (1, 1, 0) - \frac{\utilde{b}_1 \cdot \utilde{a}_2}{\utilde{b}_1 \cdot \utilde{b}_1}(1, 1, 1) \\
&= (1, 1, 0) - \tfrac{2}{3}(1, 1, 1) \\
&= \left(\tfrac{1}{3}, \tfrac{1}{3}, -\tfrac{2}{3}\right)
\end{aligned}$$

$$
\begin{aligned}
\underset{\sim}{b}_3 &= (1,0,0) - \frac{\underset{\sim}{b}_1 \cdot \underset{\sim}{a}_3}{\underset{\sim}{b}_1 \cdot \underset{\sim}{b}_1}(1,1,1) - \frac{\underset{\sim}{b}_2 \cdot \underset{\sim}{a}_2}{\underset{\sim}{b}_2 \cdot \underset{\sim}{b}_2}\left(\tfrac{1}{3}, \tfrac{1}{3}, -\tfrac{2}{3}\right) \\
&= (1,0,0) - \tfrac{1}{3}(1,1,1) - \tfrac{1}{2}\left(\tfrac{1}{3}, \tfrac{1}{3}, -\tfrac{2}{3}\right) \\
&= \left(\tfrac{1}{2}, -\tfrac{1}{3}, 0\right)
\end{aligned}
$$

Thus we have an orthogonal set of vectors:

$$\underset{\sim}{b}_1 = (1,1,1), \quad \underset{\sim}{b}_2 = \left(\tfrac{1}{3}, \tfrac{1}{3}, -\tfrac{2}{3}\right), \quad \underset{\sim}{b}_3 = \left(\tfrac{1}{2}, -\tfrac{1}{3}, 0\right).$$

The corresponding orthonormal basis of $(\mathbb{R}^3)'$ is

$$\underset{\sim}{v}_1 = \left(\tfrac{1}{\sqrt{3}}, \tfrac{1}{\sqrt{3}}, \tfrac{1}{\sqrt{3}}\right), \quad \underset{\sim}{v}_2 = \left(\tfrac{1}{\sqrt{6}}, \tfrac{1}{\sqrt{6}}, -\tfrac{2}{\sqrt{6}}\right), \quad \underset{\sim}{v}_3 = \left(\tfrac{1}{\sqrt{2}}, -\tfrac{1}{\sqrt{2}}, 0\right).$$

Using MAGMA, it can be done as follows. (Note that, in a MAGMA command, // indicates the beginning of a comment: MAGMA ignores everything following // up to the end of the line).

```
> R:=RealField();
> V:=VectorSpace(R,3);
> // We now enter the three input vectors as the terms of a
> // sequence of vectors
> a := [V![1,1,1], V![1,1,0], V![1,0,0]];
> // Print it, just as a check
> a;
[
    (1,1,1),
    (1,1,0),
    (1,0,0),
]
> // Now calculate a sequence of vectors whose terms are the
> // Gram-Schmidt output.  Initially, the sequence is empty; we then
> // add the terms one at a time.
> b:=[];
> b[1]:=a[1];
> // If a sequence has n terms, then magma permits us to increase
> // its length by 1 by simply defining term number n + 1.
> b;
[
    (1,1,1)
]
> b[2]:=a[2];
> b[2]:=b[2]-(InnerProduct(b[2],b[1])/InnerProduct(b[1],b[1]))*b[1];
> b[3]:=a[3];
> b[3]:=b[3]-(InnerProduct(b[3],b[1])/InnerProduct(b[1],b[1]))*b[1];
> b[3]:=b[3]-(InnerProduct(b[3],b[2])/InnerProduct(b[2],b[2]))*b[2];
> b;
[
    (1,1,1)
    (1/3 1/3 1/3)
    (1/2 -1/2 0)
]
```

We can save ourselves some typing by using some “for” loops. We define a function `GramSchmidt` that takes a sequence `a` of vectors as its input, and outputs the orthogonalized sequence of vectors `b`. First, we put `b` equal to the empty sequence `[]`, then for `i` running from 1 to the number of terms in the sequence `a`, we define `b[i]` by means of another loop: put `b[i]` equal to `a[i]` initially, then for `j` running from 1 to `i - 1`, redefine `b[i]` via `b[i]:=b[i]-(InnerProduct(b[i],b[j])/InnerProduct(b[j],b[j]))*b[j]`.

```
> //    We assume that a is a sequence of vectors.  magma denotes
> //    the number of terms in a sequence s by #s.
> GramSchmidt:=function(a);
function> b:=[];
function> for i in [1..#a] do
function|for> b[i]:=a[i]
function|for> for j in [1..i-1] do
function|for|for> b[i]:=b[i]-(InnerProduct(b[i],b[j])
function|for|for> /InnerProduct(b[j],b[j]))*b[j];
function|for|for> end for;
function|for> end for;
function> return(b);
> //    Having defined this function, it is now easy to do
> //    Example 1 above.
> R:=RealField();
> V:=VectorSpace(R,2);
> a:=[V![1,2,0],V![0,1,2],V![1,1,1]];
> G:=GramSchmidt(a);
> G;
[
(1 2 0),
(-2/5 1/5 2),
( 4/7 -2/7 1/7)
]
```

Orthogonal matrices

Definition. A matrix is said to be *orthogonal* if it is square and satisfies $A^T A = I$, the identity matrix.

Recall that $A^T A = I$ if and only if the columns of A form an orthonormal set of vectors. This is because the (i, j)-entry of $A^T A$ is the dot product of the i-th and j-th columns of A; so $A^T A = I$ if and only if this dot product is 0 for $i \neq j$ and 1 for $i = j$. So an $n \times n$ matrix is orthogonal if and only if its columns form an orthonormal basis of $\mathbb{R}^n$. (In view of this, it is a pity that such matrices are called orthogonal matrices rather than orthonormal matrices. However, the terminology has been fixed for so long that it cannot be changed now.)

It is a fact that you should know from earlier courses that if A is a square matrix and $BA = I$ then necessarily $AB = I$ also, and so A is invertible and $B = A^{-1}$. So it follows from the definition above that a square matrix is orthogonal if and only if its transpose is also its inverse. Of course, if the transpose of A is the inverse of A then the transpose of A^T (namely, A) is the inverse of A^T. In other words, the condition “transpose equals inverse", holds for A^T if it holds for A. We conclude that if A is orthogonal then so is A^T. And so if the columns of A form an orthonormal basis of $\mathbb{R}^n$ then so do the columns of A^T. But the columns of A^T are the transposes of the rows of A, and so the columns of A^T form an orthonormal basis of $\mathbb{R}^n$ if and only if the rows of A form an orthonormal basis of $(R^n)'$. Thus we have shown that if an $n \times n$ matrix has orthonormal columns then it also has orthonormal rows. The converse of this also holds, by similar reasoning.

The upshot of all this is that, for an $n \times n$ matrix A, the following conditions are equivalent:

- A is orthogonal;
- $A^T = A^{-1}$;
- the columns of A form an orthonormal basis of $\mathbb{R}^n$;
- the rows of A form an orthonormal basis of $(\mathbb{R}^n)'$.

Example 3. Since the set of vectors

$$\left\{ \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$$

is clearly an orthonormal basis of $\mathbb{R}^3$, it follows that the matrix

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

is orthogonal. So its transpose must be its inverse. Let us check this:

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Example 4. We saw in our discussion of the Gram-Schmidt process that the following set of vectors is an orthonormal basis of $\mathbb{R}^3$:

$$\frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \quad \frac{1}{\sqrt{105}} \begin{pmatrix} -2 \\ 1 \\ 10 \end{pmatrix}, \quad \frac{1}{\sqrt{21}} \begin{pmatrix} 4 \\ -2 \\ 1 \end{pmatrix}.$$

So it follows that the matrix

$$\begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{105}} & \frac{4}{\sqrt{21}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{105}} & \frac{-2}{\sqrt{21}} \\ 0 & \frac{10}{\sqrt{105}} & \frac{1}{\sqrt{21}} \end{pmatrix}$$

is orthogonal. So its rows must form an orthonormal basis of $(\mathbb{R}^3)'$. Let us check that the dot product of the first two rows is indeed zero:

$$\begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{105}} & \frac{4}{\sqrt{21}} \end{pmatrix} \cdot \begin{pmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{105}} & \frac{-2}{\sqrt{21}} \end{pmatrix} = \frac{2}{5} - \frac{2}{105} - \frac{8}{21} = \frac{42-2-40}{105} = 0.$$

Example 3. The following orthogonal matrix is one of my favourites:

$$\begin{pmatrix} \frac{-1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{-1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{-1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{-1}{2} \end{pmatrix}.$$

This matrix is also symmetric; so its transpose equals itself as well as its inverse. So it is self-inverse: $A^2 = I$.

Orthogonal matrices of degree 2

It is not hard to completely determine all 2×2 orthogonal matrices. If

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

then expanding the left hand side gives the three equations

$$\begin{aligned} a^2 + c^2 &= 1 \\ ba + dc &= 0 \\ b^2 + d^2 &= 1. \end{aligned}$$

From the first of these equations we can deduce that there exists a number $\psi \in \mathbb{R}$ such that $a = \cos\psi$ and $c = \sin\psi$. Putting this into the second equation gives $b\cos\psi + d\sin\psi = 0$, and from this we deduce

$$\begin{aligned} b &= \lambda \sin\psi \\ d &= -\lambda\cos\psi \end{aligned}$$

for some scalar λ. (In fact, if $\cos\psi \neq 0$ then it is clear that these equations hold with $\lambda = \frac{-d}{\cos\psi}$. If $\cos\psi = 0$ then $\sin\psi \neq 0$, and the alternative formula $\lambda = \frac{b}{\sin\psi}$ can be used.) Now $b^2 + d^2 = 1$ gives $\lambda^2 \sin^2\psi + \lambda^2\cos^2\psi = 1$, and it follows that $\lambda^2 = 1$. So $\lambda = \pm 1$, and we have shown that

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix} = \begin{pmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} \cos\psi & \sin\psi \\ \sin\psi & -\cos\psi \end{pmatrix}. \tag{6}$$

Conversely, it is trivial to check that any 2×2 matrix of either of these forms is orthogonal. Note that in one case the matrix has determinant $\cos^2\psi + \sin^2\psi = 1$, and in the other case it has determinant $-cos^2\psi - \sin^2\psi = -1$.

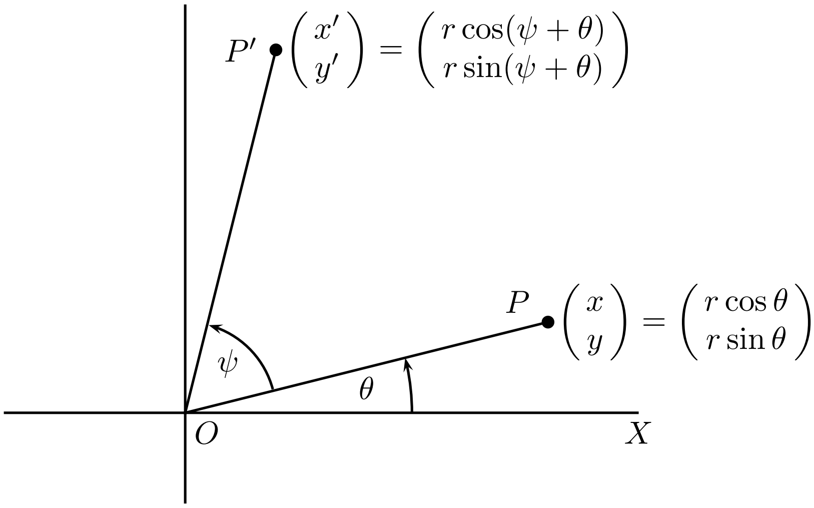
Geometrically, the first of the two alternatives in Eq. (6)—the one with positive determinant—corresponds to an anticlockwise rotation through ψ. More precisely, the linear transformation

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

maps the point P in the plane (identified with $\mathbb{R}^2$) to the point P' the same distance from the origin O as P, and such that $\angle P'OX = \psi + \angle POX$ (where X is on the positive x-axis). To see this, write the coordinates of P as $x = r\cos\theta$ and $y = r\sin\theta$. Thus P has distance r from O, and $\angle OPX = \theta$. Since

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{pmatrix} \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix}$$
$$= \begin{pmatrix} r\cos\psi\cos\theta - r\sin\psi\sin\theta \\ r\sin\psi\cos\theta + r\cos\psi\sin\theta \end{pmatrix} = \begin{pmatrix} r\cos(\psi+\theta) \\ r\sin(\psi+\theta) \end{pmatrix},$$

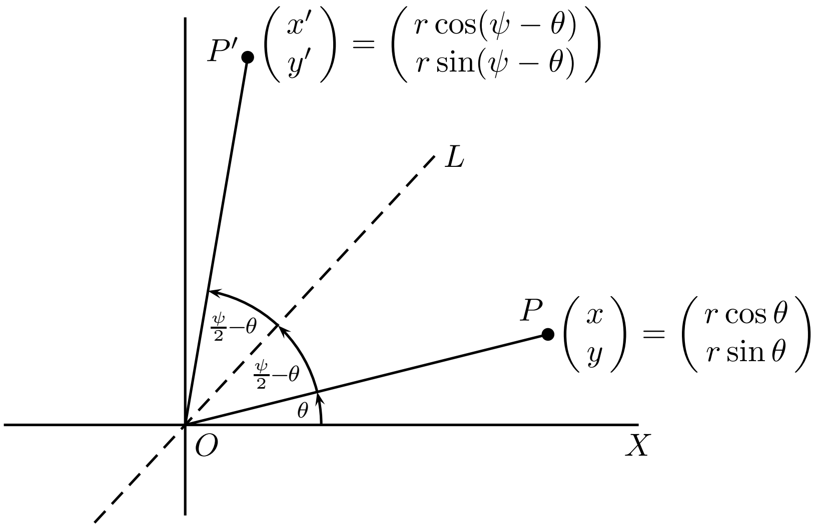
we see that the distance from P' to O is also r, and $\angle P'OX = \psi + \theta$, as required.



In a similar fashion it can be shown that the other matrix in Eq. (6) above corresponds to the reflection in the line OL such that $\angle LOX = \psi/2$. This time the calculation goes as follows:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\psi & \sin\psi \\ \sin\psi & -\cos\psi \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\psi & \sin\psi \\ \sin\psi & -\cos\psi \end{pmatrix} \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix}$$
$$= \begin{pmatrix} r\cos\psi\cos\theta + r\sin\psi\sin\theta \\ r\sin\psi\cos\theta - r\cos\psi\sin\theta \end{pmatrix} = \begin{pmatrix} r\cos(\psi-\theta) \\ r\sin(\psi-\theta) \end{pmatrix}$$

As can be seen from the diagram, this ensures that $\angle P'OL = \angle POL = \frac{\psi}{2} - \theta$, making P' the mirror image of P relative to OL.



Example 5. Let $A = \begin{pmatrix} 3/5 & 4/5 \\ -4/5 & 3/5 \end{pmatrix}$. Then

$$A^T A = \begin{pmatrix} 3/5 & -4/5 \\ 4/5 & 3/5 \end{pmatrix} \begin{pmatrix} 3/5 & 4/5 \\ -4/5 & 3/5 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

so that A is orthogonal. And $\det A = (3/5)^2 + (4/5)^2 = 1$; so A is a rotation. Precisely, it is an anticlockwise rotation through an angle θ such that $\cos\theta = 3/4$ and $\sin\theta = -4/5$. Since $\sin\theta$ is negative and $\cos\theta$ positive, θ is in the fourth quadrant. And since $\arcsin(4/5) \approx 63.13°$, we deduce that $\theta \approx -63.13°$. Alternatively put, A corresponds to a clockwise rotation through $63.13°$ (approximately).

Here is some MAGMA.

```
> R := RealField();
> root2 := Sqrt(2);
> root3 := Sqrt(3);
> root6 := Sqrt(6);
> M := KMatrixSpace(R,3,3);
> A := M![1/root3,1/root3,1/root3,1/root6,1/root6,-2/root6,
>     1/root2,-1/root2,0];
> A * Transpose(A);
[0.9999999999999999999999999996 0.E-29 0.E-29]
[ 0.E-29 0.9999999999999999999999999996 0.E-29]
[ 0.E-29 0.E-29 0.9999999999999999999999999997]
```

The answer MAGMA obtains in the above calculation is the identity matrix, to 28 decimal places. Of course, if the calculations had been exact, the answer would have been exactly the identity matrix.

In fact, MAGMA can do this kind of calculation exactly. Here is how.

```
> Q := RationalField();
> P<x> := PolynomialRing(Q);
> N< root2, root3 > := NumberField([x∧2-2,x∧2-3]);
> // These commands have declared N to be a field with elements
> // root2 and root3 satisfying root2² = 2 and root3² = 3 (exactly)
> root6 := root2*root3
> M := KMatrixSpace(N,3,3);
> A := M![1/root3,1/root3,1/root3,1/root6,1/root6,-2/root6,
>     1/root2,-1/root2,0];
> A * Transpose(A);
[1 0 0]
[0 1 0]
[0 0 1]
```

Abstract inner product spaces

So far in this course the only vector spaces we have been concerned with have been $\mathbb{R}^n$ and $(\mathbb{R}^n)'$, and their subspaces. However, there are innumerable other examples of vector spaces, and it is quite straightforward to adapt the ideas we have been using so that they can be applied in a wider context. In particular, we shall be interested in vector spaces whose elements are continuous real-valued functions defined on subsets of $\mathbb{R}$, and subspaces of these, such as spaces of polynomial functions.

From the Matrix Applications course, you should know that P_n, the set of polynomials in the variable X of degree at most n, is a vector space over $\mathbb{R}$. By definition,

$$P_n = \{\, a_0 + a_1X + a_2X^2 + \cdots + a_nX^n \mid a_0,\ a_1,\ \ldots,\ a_n \in \mathbb{R}\,\}$$

and it is clear that P_n has dimension $n+1$, the set $\{1, X, X^2, \ldots, X^n\}$ being a basis.

If $a,\ b \in \mathbb{R}$ with $a \le b$ then the set $\mathcal{F}[a,b]$ consisting of all functions $[a,b] \to \mathbb{R}$ is a vector space over $\mathbb{R}$ under the operations of addition and scalar multiplication defined as follows:

$$\begin{aligned}(f+g)(t) &= f(t) + g(t) \quad \text{for all } t \in \mathbb{R}\\ (\lambda f)(t) &= \lambda(f(t)) \quad \text{for all } t \in \mathbb{R}\end{aligned}$$

for all $f,\ g \in \mathcal{F}[a,b]$ and $\lambda \in \mathbb{R}$. From basic calculus we know that sums and scalar multiples of continuous functions are continuous, and it follows readily that the set $\mathcal{C}[a,b]$ consisting of all continuous real-valued functions on $[a,b]$ is a subspace of $\mathcal{F}[a,b]$. And $\mathcal{P}[a,b]$, the set of polynomial functions on $[a,b]$, is a subspace of $\mathcal{C}[a,b]$.

Let V be any vector space over $\mathbb{R}$. Suppose that we have a rule for multiplying vectors, such that the product of two vectors is a scalar. We adopt the notation $(\underset{\sim}{u}, \underset{\sim}{v})$ for the product of $\underset{\sim}{u}$ and $\underset{\sim}{v}$. We are interested in products that satisfy the following properties:

IP1) $(\underset{\sim}{u}, \underset{\sim}{v}) = (\underset{\sim}{v}, \underset{\sim}{u})$ for all $\underset{\sim}{u},\ \underset{\sim}{v} \in V$;

IP2) $(\underset{\sim}{u} + \underset{\sim}{v}, \underset{\sim}{w}) = (\underset{\sim}{u}, \underset{\sim}{w}) + (\underset{\sim}{v}, \underset{\sim}{w})$ for all $\underset{\sim}{u},\ \underset{\sim}{v}\,\underset{\sim}{w} \in V$;

IP3) $(k\underset{\sim}{u}, \underset{\sim}{v}) = k(\underset{\sim}{u}, \underset{\sim}{v})$ for all $\underset{\sim}{u},\ \underset{\sim}{v} \in V$ and all $k \in \mathbb{R}$;

IP4) $(\underset{\sim}{u}, \underset{\sim}{u}) \ge 0$ for all $\underset{\sim}{u} \in V$, and if $(\underset{\sim}{u}, \underset{\sim}{u}) = 0$ then $\underset{\sim}{v} = \underset{\sim}{0}$.

We noted in Week 1 that the dot product on $\mathbb{R}^n$ has these properties; the proofs are given in the solutions to Tutorial 1. So the dot product on $\mathbb{R}^n$ is an example of an inner product. It is called the *standard inner product* on $\mathbb{R}^n$, and it is the motivating example for the definition. Every other inner product, on any vector space, should be thought of as a generalization of the standard inner product. We can expect the properties of the standard inner product to be mirrored by other inner products.

If A is any $n \times n$ symmetric matrix with positive eigenvalues, then the rule

$$(\underset{\sim}{u}, \underset{\sim}{v}) = \underset{\sim}{u}^T A \underset{\sim}{v}$$

defines an inner product on $\mathbb{R}^n$. For example, let $A = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$. This is symmetric, and its eigenvalues are 2 and 3; so it should determine an inner product on $\mathbb{R}^2$. Let us express the formula in terms of the entries of $\underset{\sim}{u}$ and $\underset{\sim}{v}$:

$$\begin{aligned}\left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}\right) &= \begin{pmatrix} x_1 & y_1 \end{pmatrix} A \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 & y_1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}\\ &= \begin{pmatrix} 2x_1 & 3y_1 \end{pmatrix} \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = 2x_1x_2 + 3y_1y_2\end{aligned}$$

It is easy to check that (IP1)–(IP4) are satisfied; the proofs are very similar to those for the standard inner product. In particular, (IP4) holds since

$$\left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}\right) = 2x_1^2 + 3y_1^2 \geq 0,$$

with equality only if $x_1 = y_1 = 0$. It is clear that positivity of the eigenvalues is crucial for this: if we changed the diagonal entries of A from 2 and 3 to 2 and -3 it would no longer work.

Note that using the identity matrix in the above construction gives the standard inner product.