Chapter 2. Linear transformations

This material is a reduced treatment of what is in Anton & Rorres chapter 4, chapter 6 (mostly for the case of 3 dimensions) and chapter 7. Rotations are mentioned in section 7.1 along with orthogonal matrices. See also section 4.9 which is a bit like the start of the notes below, or section 4.11 which is about two dimensions. Change of basis is dealt with (in a more general and abstact setting than here in section 4.6). Orthogonal matrices in 7.1. Gram-Schmidt is in section 6.3 (again in a more general setting). General linear transformations in section 8.1.

2.1 Introduction

Our aim first is to give a new way of looking at matrices. It is a development from ideas in MA1S11 (also mentioned in the Cramers rule section) where we expressed a system of linear equations as a single matrix equation

$$A\mathbf{x} = \mathbf{b}$$

Recall that a system of m linear equations in n unknowns

$$\begin{cases} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & & & & & & & \vdots & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{cases}$$

can be written as a single matrix equation $A\mathbf{x} = \mathbf{b}$ where A is an $m \times n$ matrix, $\mathbf{x}$ is an $n \times 1$ (column) matrix of unknowns and $\mathbf{b}$ is an $m \times 1$ column.

In single variable calculus we realise that it is convenient to use the language of functions to explain what happens when we are solving equations. The kind of examples we have there include

$$\begin{array}{rcll} 2x+3 & = & 5 & \text{(linear)} \\ x^2+2x+7 & = & 2 & \text{(quadratic)} \\ x^3+4x^2-2x+1 & = & 4 & \text{(cubic)} \\ \cos x & = & x & \text{(more complicated — can be} \\ & & & \text{written } \cos x - x = 0) \end{array}$$

We can think of all of these as $f(x) = 0$ or $f(x) = b$ for a suitable function $f(x)$.

Inspired by this experience we might be inclined to think of a system of linear equations

$$A\mathbf{x} = \mathbf{b}$$

in terms of functions.

The functions we need for this have to be functions of a vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

or alternatively can be viewed as functions of n variables $(x_1, x_2, \ldots, x_n)$. In fact the language of functions is already designed to cope with these more complex situations.

While we become used to having functions $f(x)$ where the independent variable x is a number, and where $f(x)$ is usually given by a formula, this is not the definition of a function. We may be used to saying "let $f(x) = x^3 + 4x^2 - 2x + 1$ be a function" (so given by the formula) but this is not the only kind of function that is allowed.

In general a function $f\colon S \to T$ is something that has a *domain* set S and a *target* set T (more often called a *codomain* set T) and the definition says that a function f is a rule that assigns one and only one element $f(s) \in T$ to each $s \in S$. The rule does not have to be a formula, but it does have to be an unambiguous rule that works for every s in the domain set S. We usually refer to $f(s)$ as the 'value' of the function at s.

The most familiar examples are functions $f\colon \mathbb{R} \to \mathbb{R}$ where the domain is $S = \mathbb{R}$ (the whole real line) and the values are also real numbers. So we have functions like $f(x) = x \cos x$ or $f(x) = x^2 + 2x + 3$ where the rule is given by a formula. Sometimes we also encounter cases where the domain S is not all of $\mathbb{R}$, but only a part of $\mathbb{R}$. For instance the square root function $f(x) = \sqrt{x}$ gets in trouble if we try to take square roots of negative numbers. We would usually take the domain $S = [0, \infty) = \{x \in \mathbb{R} : x \geq 0\}$ and $f\colon [0, \infty) \to \mathbb{R}$, $f(x) = \sqrt{x}$. (This might be a good place to remark that by the square root we mean the positive square root. So $\sqrt{4}$ means 2, not -2. If we did not do this the square root would not qualify as an ordinary function with values in $\mathbb{R}$. $f(x)$ to have just one value when we have a function.)

2.1.1 Remarks. We will be concentraing on linear functions, without a constant term. So in the case $f\colon \mathbb{R} \to \mathbb{R}$ we will look at examples or the form $f(x) = ax$ with a a constant.

So we might look at $f(x) = 3x$ of $f(x) = -2x$. No more compliucated functions!

We would be used to thinking of $f(x) = 3x$ via its graph, the staright line $y = 3x$ through the origin with slope 3. However, we can also look at what happens to an input x into this function, and picture where $f(x) = 3x$ is compared to x. The answer is that $3x$ is 3 times as far from the origin 0 as x was. We thing of the effect of f as 'transforming'or 'mapping' x in this way to a new point 3 times as far from the origin.

If we take the example $f(x) = -x$ instead, we can picture what happens in this case also. x gets mapped to $-x$, which is the same distance from the origin in $\mathbb{R}$ as x was, but on the other side of the origin. One can think of a mirror at the origin and $f(x) = -x$ is the reflection of x in the origin.

If we had say $f(x) = -5x$ we can think of a combination of a reflection and a stretching.

For functions $f\colon \mathbb{R}^2 \to \mathbb{R}$, there are examples we have more or less seen already. We could have for example

$$f(\mathbf{x}) = \|\mathbf{x}\| = \text{ distance from } \mathbf{x} \text{ to } \mathbf{0}$$

or we could write the same example

$$f(\mathbf{x}) = \sqrt{x_1^2 + x_2^2}$$

when we take $\mathbf{x} = x_1\mathbf{i} + x_2\mathbf{j}$ (or equivalently think of points $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$).

Another perfectly good example would be

$$f(x_1, x_2) = x_1^3 + x_2^4 - x_1$$

(which has no obvious geometrical interpretation, but makes sense as a rule that associates a value $f(x_1, x_2)$ to a point $(x_1, x_2) \in \mathbb{R}^2$).

An example that is a bit closer to the kinds of functions that will really interest us is

$$f(x_1, x_2) = 2x_1 - 3x_2$$

We can rewrite this using vectors as

$$f(x_1, x_2) = (2\mathbf{i} - 3\mathbf{j}) \cdot (x_1\mathbf{i} + x_2\mathbf{j}).$$

We'll be thinking of this example, or things very like it, in matrix language. The matrix product

$$\begin{bmatrix} 2 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

is a 1×1 matrix

$$[2x_1 - 3x_2]$$

and so has just one entry, the number $2x_1 - 3x_2$. If we are prepared to ignore the (arguably rather technical) difference between the number $2x_1 - 3x_2$ as a pure number and the 1×1 matrix $[2x_1 - 3x_2]$, then we can write the same example using matrix multiplication

$$f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 2 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Now that we've shown some examples in $\mathbb{R}^2$, we could show some examples in $\mathbb{R}^3$ without too much extra imagination. We could even look at functions $f\colon \mathbb{R}^n \to \mathbb{R}$ given by rules (or formulae) like

$$f(\mathbf{x}) = \|\mathbf{x}\|$$

which now looks like

$$f(x_1, x_2, \ldots, x_n) = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

Another example, generalising the one we just gave with $2\mathbf{i} - 3\mathbf{j}$, would be to fix $\mathbf{a} = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$ and define $f\colon \mathbb{R}^n \to \mathbb{R}$ by the rule

$$f(\mathbf{x}) = \mathbf{a} \cdot \mathbf{x}$$

We could think of this example in matrix terms also as

$$f\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right) = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

if we are still prepared to ignore the distinction between a scalar and a 1×1 matrix.

When A is an $m \times n$ matrix the formula

$$f(\mathbf{x}) = A\mathbf{x}$$

applied to an $n \times 1$ (column)

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

will produce an $m \times 1$ result. So the result can be interpreted as a point (or vector) in $\mathbb{R}^m$ (if we identify n-tuples $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ with $n \times 1$ column matrices, and similarly identify $m \times 1$ column matrices with points in $\mathbb{R}^m$). Our formula $f(\mathbf{x}) = A\mathbf{x}$ will then make sense to define a function

$$f \colon \mathbb{R}^n \to \mathbb{R}^m$$

Often people prefer to call these *transformation* rather than functions. Technically they are the same as functions (rules that work on some domain and produce values somewhere) but the idea of using the word 'transformation' instead of plain 'function' is partly due to the fact that the values are vectors rather than scalar values. There is also a geometrical way of thinking about what these transformations do, or look like, that we will try to explain in some examples soon. This viewpoint partly explains the 'transformation' word.

For now we will summarise what we have with a definition.

2.1.2 Definition. A *linear transformation* $f \colon \mathbb{R}^n \to \mathbb{R}^m$ is a function given by a rule

$$f(\mathbf{x}) = f\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right) = A \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = A\mathbf{x}$$

for some give $m \times n$ matrix A.

2.2 Examples of linear transformations

In the first examples we take $n = m = 2$, so that we have 2×2 matrices A and linear transformations $f \colon \mathbb{R}^2 \to \mathbb{R}^2$ of the plane.

(i) $A = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$,

$$A\mathbf{x} = A \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3x_1 \\ 3x_2 \end{bmatrix}$$

So we have $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ and $f(\mathbf{x}) = 3\mathbf{x} =$ the vector 3 times as long as $\mathbf{x}$ in the same direction. Or, if we think in terms of points in the plane $f(\mathbf{x})$ is the point 3 times further from the origin than $\mathbf{x}$, in the same direction.

Thinking of what happens to the points of $\mathbb{R}^2$ (picture the plane) after you apply the transformation f to them, you see everything expands by a factor 3 away from the central point at the origin. This is sometimes called 'dilation' by a factor 3 (like looking through a magnifying glass).

(ii) $A = \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}$,

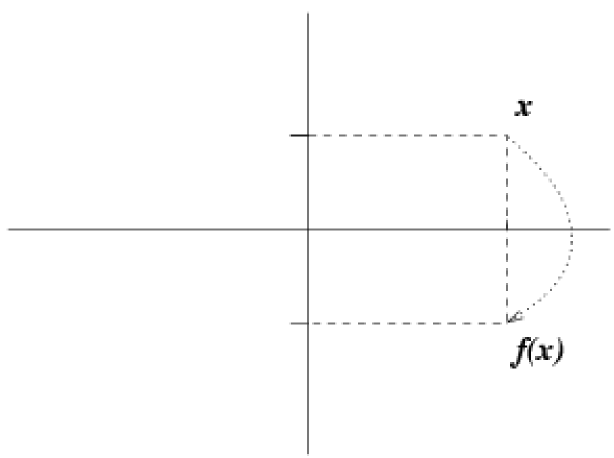
$$A\mathbf{x} = A\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 5x_1 \\ x_2 \end{bmatrix}$$

The picture here is that the plane is expanded or stretched horizontally (in the x_1- or x-direction) by a factor 5, while the vertical scale is not affected. The vertical axis does not move.

(iii) $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$,

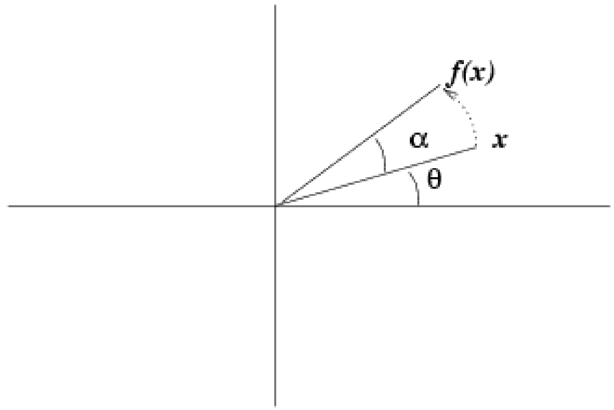
$$A\mathbf{x} = A\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ -x_2 \end{bmatrix}$$

The picture for this involves a mirror. The relation between $\mathbf{x}$ and $f(\mathbf{x})$ is that the point $f(\mathbf{x})$ arises by reflecting the original point $\mathbf{x}$ in the horizontal axis.



(iv) $A = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}$,

$$A\mathbf{x} = A\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} (\cos\alpha)x_1 - (\sin\alpha)x_2 \\ (\sin\alpha)x_1 + (\cos\alpha)x_2 \end{bmatrix}$$



In the picture the effect of this transformation is shown as a rotation by an angle α (radians) anticlockwise. To see why this is the case, it is easiest if we use polar coordinates. This means describing the position of a point $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$ via two other numbers (t, θ) where $r = \|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2} = \text{distance}\,(\mathbf{x}, \mathbf{0})$ and θ is the angle (in radians) from the positive x_1-axis (horizontal of x-axis) around anticlockwise to the radius from the origin to $\mathbf{x}$. In formulae, the relation is

$$(x_1, x_2) = (r\cos\theta, r\sin\theta)$$

If we apply the map, we get

$$\begin{aligned} f(\mathbf{x}) &= A\mathbf{x} = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} r\cos\theta \\ r\sin\theta \end{bmatrix} \\ &= \begin{bmatrix} (\cos\alpha)(r\cos\theta) - (\sin\alpha)(r\sin\theta) \\ (\sin\alpha)(r\cos\theta) + (\cos\alpha)(r\sin\theta) \end{bmatrix} \\ &= \begin{bmatrix} r(\cos\alpha\cos\theta - \sin\alpha\sin\theta) \\ r(\sin\alpha\cos\theta + \cos\alpha\sin\theta) \end{bmatrix} \\ &= \begin{bmatrix} r\cos(\alpha+\theta) \\ r\sin(\alpha+\theta) \end{bmatrix} \end{aligned}$$

At the last step above we used two trigonometric identities you may recall:

$$\begin{aligned} \cos(\alpha+\theta) &= \cos\alpha\cos\theta - \sin\alpha\sin\theta \\ \sin(\alpha+\theta) &= \sin\alpha\cos\theta + \cos\alpha\sin\theta \end{aligned}$$

The result of the above calculation is that the polar coordinates for the image point $f(\mathbf{x})$ have the same r as the original point $\mathbf{x}$, so same distance from the origin, but the polar coordinates angle for $f(\mathbf{x})$ is $\theta + \alpha$, or an increase by α on the angle θ for $\mathbf{x}$. This justifies the picture.

2.3 Orthogonal matrices

Before we look at more examples, we point out that the last two examples, the reflection and the rotation, both have the property that

$$A^{-1} = A^t$$

In fact for the reflection we have

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = A^{-1} = A^t$$

and for the rotation, you can check out that

$$A = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \text{ has } A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix} = \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix} = A^t$$

So in both cases $AA^t = I_2 = A^t A$.

2.3.1 Definition. An $n \times n$ matrix A is called an *orthogonal matrix* if $AA^t = I_n = A^t A$ (which means that A is invertible and $A^{-1} = A^t$.

If we write n-tuples as columns

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix},$$

then the dot product

$$\mathbf{v} \cdot \mathbf{w} = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n$$

can be expressed in matrix terms if we are prepared to overlook the distinction between a scalar and a 1×1 matrix.

$$(\mathbf{v}^t)\mathbf{w} = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = [v_1 w_1 + v_2 w_2 + \cdots + v_n w_n] = \mathbf{v} \cdot \mathbf{w}$$

Using this way of looking at things we can show that orthogonal matrices have some nice properties. (So these properties apply in particular to our 2×2 reflections and rotations.)

2.3.2 Proposition. *Let A be an $n \times n$ orthogonal matrix and $f \colon \mathbb{R}^n \to \mathbb{R}^n$ the associated linear transformation $f(\mathbf{x}) = A\mathbf{x}$. Then, for $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ we have*

(i) $(A\mathbf{v}) \cdot (A\mathbf{w}) = \mathbf{v} \cdot \mathbf{w}$

(ii) $f(\mathbf{v}) \cdot f(\mathbf{w}) = \mathbf{v} \cdot \mathbf{w}$ *(so that the orthogonal matrices preserve inner products, or the transformation associated with them preserves inner products)*

(iii) $\|A\mathbf{v}\| = \|\mathbf{v}\|$

(iv) $\|f(\mathbf{v})\| = \|\mathbf{v}\|$ *(so preserve distance from the origin)*

(v) $\|A\mathbf{v} - A\mathbf{w}\| = \|\mathbf{v} - \mathbf{w}\|$

(vi) $\|f(\mathbf{v}) - f(\mathbf{w})\| = \|\mathbf{v} - \mathbf{w}\|$ *(so preserve distances between points)*

(vii) *if θ denotes the angle between $\mathbf{v}$ and $\mathbf{w}$, then the angle between $f(\mathbf{v}) = A\mathbf{v}$ and $f(\mathbf{w}) = A\mathbf{w}$ is also θ (so preserves angles).*

Proof. Since we have mostly done the work already we can quickly show that all these properties are indeed correct.

(i) $(A\mathbf{v}) \cdot (A\mathbf{w}) = ((A\mathbf{v})^t)A\mathbf{w} = \mathbf{v}^t A^t A\mathbf{w} = \mathbf{v}^t I_n \mathbf{w} = (\mathbf{v}^t)\mathbf{w} = \mathbf{v} \cdot \mathbf{w}$

(We used here the rule that the transpose of a product is the product of the transposes in the reverse order.)

(ii) $f(\mathbf{v}) \cdot f(\mathbf{w}) = \mathbf{v} \cdot \mathbf{w}$ is just the same thing again in a different notation.

(iii) $\|A\mathbf{v}\|^2 = (A\mathbf{v}) \cdot (A\mathbf{v}) = \mathbf{v} \cdot \mathbf{v} = \|\mathbf{v}\|^2$ (by the first part with $\mathbf{v} = \mathbf{w}$). So $\|A\mathbf{v}\| = \|\mathbf{v}\|$.

(iv) $\|f(\mathbf{v})\| = \|\mathbf{v}\|$ is just the same statement again.

(v) $\|A\mathbf{v} - A\mathbf{w}\| = \|A(\mathbf{v} - \mathbf{w}\| = \|\mathbf{v} - \mathbf{w}\|$ (using properties of matrix multiplication and also the last statement).

(vi) $\|f(\mathbf{v}) - f(\mathbf{w})\| = \|\mathbf{v} - \mathbf{w}\|$ is just the same thing again in different notation.

(vii) The angle θ between $\mathbf{v}$ and $\mathbf{w}$ comes about (by definition when $n > 3$ or because we proved it for $n = 2$ and $n = 3$) from

$$\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\|\|\mathbf{w}\| \cos\theta,$$

But, using what we have just shown we can deduce

$$(A\mathbf{v}) \cdot (A\mathbf{w}) = \mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\|\|\mathbf{w}\| \cos\theta = \|A\mathbf{v}\|\|A\mathbf{w}\| \cos\theta,$$

which says that the angle between $f(\mathbf{v}) = A\mathbf{v}$ and $f(\mathbf{w}) = A\mathbf{w}$ is also θ.

$\square$

2.3.3 Proposition. *If A is an orthogonal matrix, the* $\det(A) = \pm 1$.

Proof. If A is orthogonal, $AA^t = I_n$. So $\det(AA^t) = \det(I_n) = 1$. So

$$\det(A)\det(A^t) = 1.$$

But $\det(A^t) = \det(A)$. Thus we get

$$(\det(A))^2 = 1$$

and that tells us $\det(A)$ has to be either 1 or -1. □

2.3.4 Proposition. *The 2×2 orthogonal matrices of determinant 1 are exactly the rotation matrices.*

$$A = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \qquad (\alpha \in \mathbb{R}, \textit{ or } 0 \le \alpha < 2\pi)$$

The 2×2 orthogonal matrices of determinant -1 are exactly the products

$$\begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

of a rotation and a reflection matrix.

Proof. We can see quite easily that the rotation matrices do have determinant 1, and they are all orthogonal as we already saw.

$$\det \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} = (\cos\alpha)^2 - (-\sin\alpha)(\sin\alpha) = \cos^2\alpha + \sin^2\alpha = 1$$

Now if we start with some orthogonal 2×2 matrix

$$B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix},$$

we want to show it is a rotation matrix, or a rotation matrix times a reflection.

Notice that

$$B\mathbf{i} = B\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} b_{11} \\ b_{21} \end{bmatrix}$$

is the first column of B. So it must be a unit vector (or the point (b_{11}, b_{21}) must be on the unit circle) since we know $\|B\mathbf{i}\| = \|\mathbf{i}\| = 1$ (because of Proposition 2.3.2 (iii)). So if we write this point in polar coordinates we can say

$$\begin{bmatrix} b_{11} \\ b_{21} \end{bmatrix} = \begin{bmatrix} \cos\beta \\ \sin\beta \end{bmatrix}$$

for some β.

Next notice that

$$B\mathbf{j} = B\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} b_{12} \\ b_{22} \end{bmatrix}$$

is the second column of B. It must be also be a unit vector and it has to be perpendicular to $B\mathbf{i} = \begin{bmatrix} \cos\beta \\ \sin\beta \end{bmatrix}$ (since $\mathbf{i} \perp \mathbf{j}$ and orthogonal matrices preserve angles — Proposition 2.3.2 (vii)).

But there are just two unit vectors in the plane perpendicular to $\begin{bmatrix} \cos\beta \\ \sin\beta \end{bmatrix}$. You can see that from a picture, and you can also check that those two vectors are

$$\begin{bmatrix} -\sin\beta \\ \cos\beta \end{bmatrix} \text{ and } -\begin{bmatrix} -\sin\beta \\ \cos\beta \end{bmatrix} = \begin{bmatrix} \sin\beta \\ -\cos\beta \end{bmatrix}$$

So these are the two choices for the second column of B and we have

$$B = \begin{bmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{bmatrix} \text{ or } B = \begin{bmatrix} \cos\beta & \sin\beta \\ \sin\beta & -\cos\beta \end{bmatrix}$$

In the first case we have a rotation matrix (and $\det(B) = 1$) while the second possibility can be written

$$B = \begin{bmatrix} \cos\beta & \sin\beta \\ \sin\beta & -\cos\beta \end{bmatrix} = \begin{bmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

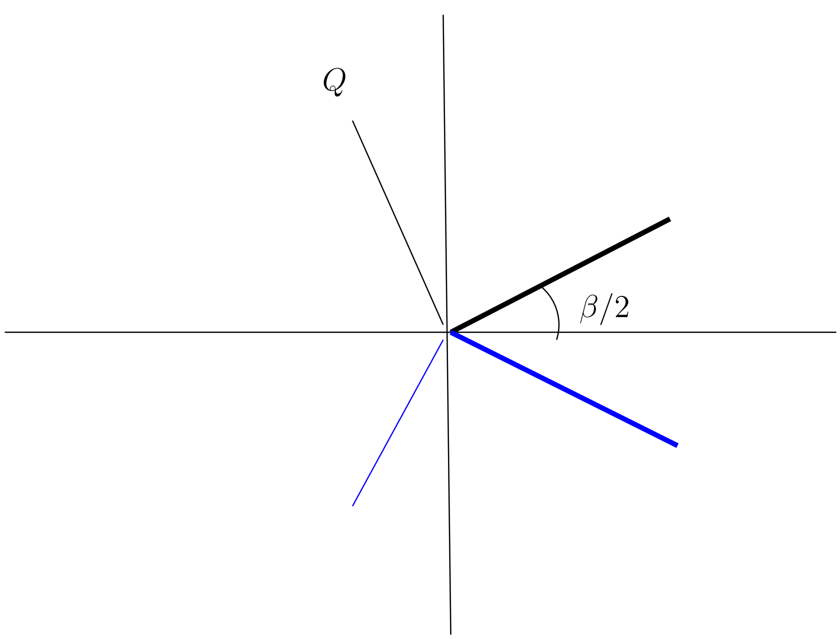
In this second case we have

$$\det(B) = \det\left(\begin{bmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{bmatrix}\right) \det\left(\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\right) = (1)(-1) = -1$$

So the sign of the determinant distinguished between the two cases. □

2.3.5 Remark. We can think of the transformation in the second case (when $\det(B) = -1$ above) as the result of first applying a reflection in the horizontal axis and then doing a rotation. So the transformation in this case is a reflection followed by a rotation.

In fact this combination of a reflection and a rotation is just a reflection in some line through the origin. The one inclined at angle $\beta/2$ in fact.



Looking at the picture you should see that if you replect the line at angle $\beta/2$ in the horizontal axis, it will go to the line at angle $-\beta/2$, and so if you rotate that by β you will be back at the original line.

For the perpendicular line, the one heading for the point marked Q in the diagram, the angle is $\beta/2 + \pi/2$. So if you reflect that it will go to angle $-(\beta/2 + \pi/2)$. And if you rotate that by β you will end up at angle $\beta/2 - \pi/2$, which is directly opposite the direction of Q (because the difference between $\beta/2 + \pi/2$ and $\beta/2 - \pi/2$ is π.

What you can see then by looking more carefully at the other directions is that the end result is a reflection in the axis at angle $\beta/2$.

But it also means that if we change our coordinate axes from the original x–y axes to ones tilted by $\beta/2$, then our reflection has a nice formula in the new coordinates. The first coordinate (along the $\beta/2$ direction) does not change while the other coordinate (in the axis heading for Q) gets flipped by a minus.

This idea that it might sometimes make life easier if we use nonstandard axes (for a particular situation) will arise again.

2.4 Rotations in space

Can we say anything about rotations in space? We described rotations in the plane $\mathbb{R}^2$ about the origin. What can we say in $\mathbb{R}^3$?

We need an axis to rotate around in space. If we choose a convenient axis, then it is quite easy to see what to do. For example, if we rotate around the z-axis, the z-coordinates (or altitudes above the horizontal x-y plane) will stay constant and the rotation will affect the x and y coordinates in the same way as a rotation of the plane about the origin. We can then use what we had above about rotations in $\mathbb{R}^2$ and show that the matrix

$$\begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

corresponds to rotation about the z-axis by the angle α.

To see this more clearly look at

$$\begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} (\cos\alpha)x - (\sin\alpha)y \\ (\sin\alpha)x + (\cos\alpha)y \\ z \end{bmatrix}$$

We see that (x, y) is rotated by α anticlockwise (from the x-axis towards the y-axis) to give

$$((\cos\alpha)x - (\sin\alpha)y, (\sin\alpha)x + (\cos\alpha)y)$$

and the height z stays fixed.

Here we used the fact that the axis was one of the coordinate axes. We can manage in a similar way if the axis is one of the other axes. For example rotation about the x-axis by an angle

α (from y towards the z-axis) is given by the matrix

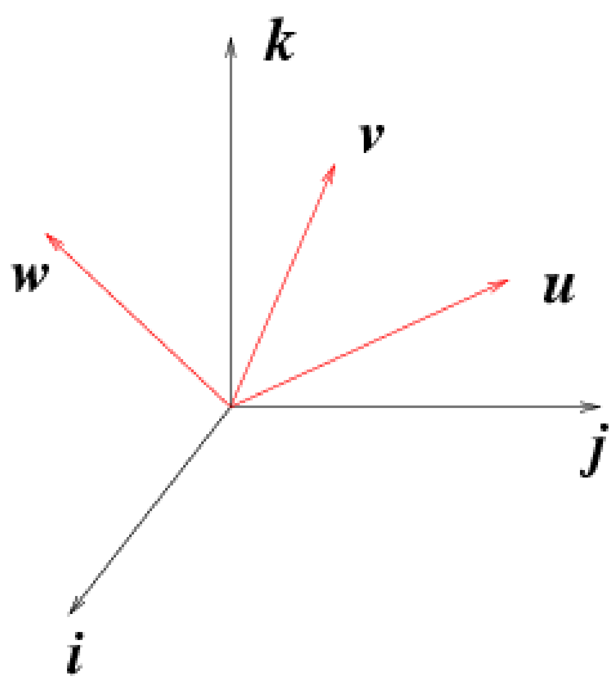
$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix}$$

But what about another axis? We would need a way to describe the direction of the axis and giving a nonzero vector $\mathbf{u}$ parallel to the axis seems like a handy way to describe that. Since the length of the vector does not seem to be useful, we can assume $\mathbf{u}$ is a unit vector. (If it is not, divide it by its length $\|\mathbf{u}\|$, which means replace the original $\mathbf{u}$ by $\mathbf{u}/\|\mathbf{u}\|$.)

One 'solution' is to choose 3 axes so that $\mathbf{u}$ is a unit vector in the direction of one of the axes. Then we have just seen how to write down a matrix to implement the rotation.

While this is an attractive solution, it is not really a complete solution. While in principle it is open to us to choose our coordinate axes to suit our problem (sometimes people call it choosing a frame of reference), mostly we are given the frame of reference earlier and we want to compute in terms of the given frame (or the given coordinates). What we really need is a way to go back and forth from the convenient coordinates to the original ones.

Suppose then we are given 3 perpendicular axes with the same origin as the original 3 axes. We will not discuss axes or frames with a different origin than the original. So, suppose we are thinking about describing the effect of a rotation about an axis through the origin in the direction of a unit vector $\mathbf{u}$. We assume we have 2 other axes perpendicular to each other and to $\mathbf{u}$. We can worry later about how to calculate them, but it is not hard to visualise $\mathbf{u}$ along with 2 other directions $\mathbf{v}$ and $\mathbf{w}$ so that $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ are all perpendicular unit vectors.



To make it more clear, imagine $\mathbf{u}$ is already known. Graphically, you can think of the plane through the origin perpendicular to $\mathbf{u}$ and visualise two unit vectors $\mathbf{v}$ and $\mathbf{w}$ in that plane and perpendicular to each other. We'll explain later how we might calculate $\mathbf{v}$ and $\mathbf{w}$ using components, starting from $\mathbf{u}$ given in components. (For the moment we will not worry about this aspect.)

Now that we have 3 perpendicular unit vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$, we will write them in components

$$\begin{aligned}\mathbf{u} &= u_1\mathbf{i} + u_2\mathbf{j} + u_3\mathbf{k},\\ \mathbf{v} &= v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k},\\ \mathbf{w} &= w_1\mathbf{i} + w_2\mathbf{j} + w_3\mathbf{k}.\end{aligned}$$

We claim that the matrix P made by using these vectors as the columns of P is an orthogonal matrix. That is

$$P = \begin{bmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ u_3 & v_3 & w_3 \end{bmatrix}$$

is orthogonal. The reason is that when we compute

$$P^tP = \begin{bmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{bmatrix} \begin{bmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ u_3 & v_3 & w_3 \end{bmatrix}$$

we multiply rows of P^t into columns of P and that means we end up taking dot products between columns of P. We get

$$P^tP = \begin{bmatrix} \mathbf{u}\cdot\mathbf{u} & \mathbf{u}\cdot\mathbf{v} & \mathbf{u}\cdot\mathbf{w} \\ \mathbf{v}\cdot\mathbf{u} & \mathbf{v}\cdot\mathbf{v} & \mathbf{v}\cdot\mathbf{w} \\ \mathbf{w}\cdot\mathbf{u} & \mathbf{w}\cdot\mathbf{v} & \mathbf{w}\cdot\mathbf{w} \end{bmatrix} = \begin{bmatrix} \|\mathbf{u}\|^2 & 0 & 0 \\ 0 & \|\mathbf{v}\|^2 & 0 \\ 0 & 0 & \|\mathbf{w}\|^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I_3$$

because the 3 vectors are perpendicular to one another and are unit vectors (length 1).

Since P is a square matrix and $P^tP = I_3$, then $PP^t = I_3$ is automatically true also and so P^t is P^{-1}.

We will show soon how to make use of the observation that P is orthogonal. One thing we can see is that

$$P\mathbf{i} = \begin{bmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ u_3 & v_3 & w_3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \mathbf{u}$$

and similarly

$$P\mathbf{j} = \mathbf{v}, \quad P\mathbf{k} = \mathbf{w}.$$

If we multiply the 3 equations

$$\boxed{\mathbf{u} = P\mathbf{i}, \quad \mathbf{v} = P\mathbf{j} \text{ and } \mathbf{w} = P\mathbf{k}.} \tag{2.4.1}$$

by $P^t = P^{-1}$ on the left we get

$$\boxed{P^t\mathbf{u} = \mathbf{i}, \quad P^t\mathbf{v} = \mathbf{j} \text{ and } P^t\mathbf{w} = \mathbf{k}} \tag{2.4.2}$$

In words (2.4.1) says that multiplication by P sends the standard basis vectors $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$, to the new basis vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$, while (2.4.2) says that multiplication by P^t does the opposite.

What we need though is to relate coordinates of any $\mathbf{x} \in \mathbb{R}^3$ with respect to the new $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ axes to coordinates of $\mathbf{x}$ is the standard axes.

Notice first that it is possible to express any vector $\mathbf{x}$ is $\mathbb{R}^3$ as a combination

$$\mathbf{x} = \xi_1 \mathbf{u} + \xi_2 \mathbf{v} + \xi_3 \mathbf{w}$$

with components ξ_1, ξ_2, ξ_3 (scalars). In MA1S11 we explained every vector $\mathbf{x}$ must be able to be expressed as

$$\mathbf{x} = x_1 \mathbf{i} + x_2 \mathbf{j} + x_3 \mathbf{k}$$

for some $x_1, x_2, x_3 \in \mathbb{R}$. We argued then in a geometrical way, based on the given axes. If we argue now in the same way, using the new $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ axes, we find that there must be scalars ξ_1, ξ_2, ξ_3.

That is a geometrical argument, which says something can be done. But in practice we need a way to calculate ξ_1, ξ_2, ξ_3. There is a nice trick for this using dot products. Remember, we know there are numbers ξ_1, ξ_2, ξ_3 but we want a way to calculate them. Starting from

$$\mathbf{x} = \xi_1 \mathbf{u} + \xi_2 \mathbf{v} + \xi_3 \mathbf{w}$$

we can compute

$$\begin{aligned}
\mathbf{x} \cdot \mathbf{u} &= (\xi_1 \mathbf{u} + \xi_2 \mathbf{v} + \xi_3 \mathbf{w}) \cdot \mathbf{u} \\
&= \xi_1 \mathbf{u} \cdot \mathbf{u} + \xi_2 \mathbf{v} \cdot \mathbf{u} + \xi_3 \mathbf{w} \cdot \mathbf{u} \\
&= \xi_1 + 0 + 0 \\
&= \xi_1
\end{aligned}$$

and similarly we can check

$$\mathbf{x} \cdot \mathbf{v} = \xi_2 \text{ and } \mathbf{x} \cdot \mathbf{w} = \xi_3.$$

In matrix terms, we have

$$\begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix} = \begin{bmatrix} \mathbf{x} \cdot \mathbf{u} \\ \mathbf{x} \cdot \mathbf{v} \\ \mathbf{x} \cdot \mathbf{w} \end{bmatrix} = \begin{bmatrix} \mathbf{u} \cdot \mathbf{x} \\ \mathbf{v} \cdot \mathbf{x} \\ \mathbf{w} \cdot \mathbf{x} \end{bmatrix} = \begin{bmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = P^t \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

We can thus add to our earlier observations about P and P^t the following.

2.4.1 Proposition. *If* $\mathbf{x} = x_1 \mathbf{i} + x_2 \mathbf{j} + x_3 \mathbf{k} = \xi_1 \mathbf{u} + \xi_2 \mathbf{v} + \xi_3 \mathbf{w}$*, where* $\mathbf{u}$*,* $\mathbf{v}$*,* $\mathbf{w}$ *are three mutually perpendicular unit vectors, and if*

$$P = \begin{bmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ u_3 & v_3 & w_3 \end{bmatrix}$$

(the matrix with $\mathbf{u}$*,* $\mathbf{v}$ *and* $\mathbf{w}$ *as its columns) then*

$$\begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix} = P^t \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \tag{2.4.3}$$

and

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = P \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix} \tag{2.4.4}$$

The second statement (2.4.4) follows by multiplying both sides of (2.4.3) by $P = (P^t)^{-1}$.

2.4.2 Remark. Note that (2.4.1) said multiplication by P sends the original basis vectors **i**, **j** and $\mathbf{k}$ to the 'new' basis vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$. But (2.4.3) says that $P^t = P^{-1}$ maps coordinates in the standard basis to coordinates in the new basis.

We will refer to P as the change of basis matrix, but there is also a case for giving that name to P^t (which has $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ as its rows).

2.4.3 Rotations in 3 dimensions. We can now write down a rotation around an axis in the direction $\mathbf{u}$ by an angle θ (a rotation by θ from $\mathbf{v}$ towards $\mathbf{w}$ in fact). In the coordinates of the new axes, we can use what we did earlier. Rotation sends the point with coordinates $\begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix}$ to the point with coordinates

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix}$$

Let us write

$$\begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix} \tag{2.4.5}$$

for the coordinates of the rotated point in the $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ axes. Let us also name this rotated point $\mathbf{y}$.

We are saying then that when we apply the rotation to

$$\mathbf{x} = \xi_1\mathbf{u} + \xi_2\mathbf{v} + \xi_3\mathbf{w}$$

the result is

$$\mathbf{y} = \eta_1\mathbf{u} + \eta_2\mathbf{v} + \eta_3\mathbf{w}.$$

What we want is to be able to do all this in terms of coordinates with respect to the original basis **i**, **j** and $\mathbf{k}$. We can figure this out using what we did before about the matrix P.

If $\mathbf{x} = x_1\mathbf{i} + x_2\mathbf{j} + x_3\mathbf{k}$ in standard coordinates, we know from (2.4.3)

$$\begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix} = P^t \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

On the other hand if $\mathbf{y} = y_1\mathbf{i} + y_2\mathbf{j} + y_3\mathbf{k}$ is the rotated point expressed in standard coordinates, then (2.4.4) tells us

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = P \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix}$$

If we put these last two equations together with (2.4.5) we get

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = P \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix} = P \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix} = P \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} P^t \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

The summary then is this

2.4.4 Proposition. *The matrix for a rotation around the axis through the origin in the direction* $\mathbf{u}$*, has the form*

$$R = P \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} P^t \tag{2.4.6}$$

where $\mathbf{u}$ *is assumed to be a unit vector,* $\mathbf{v}$ *and* $\mathbf{w}$ *are two other unit vectors perpendicular to each other and to* $\mathbf{u}$*, and the angle* θ *is measured from* $\mathbf{v}$ *towards* $\mathbf{w}$*. Here*

$$P = \begin{bmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ u_3 & v_3 & w_3 \end{bmatrix}$$

where $\mathbf{u} = u_1\mathbf{i} + u_2\mathbf{j} + u_3\mathbf{k}$*,* $\mathbf{v} = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}$ *and* $\mathbf{w} = w_1\mathbf{i} + w_2\mathbf{j} + w_3\mathbf{k}$*.*

2.4.5 Remarks. If we apply the rotation R from (2.4.6) to $\mathbf{u}$ we get

$$\begin{aligned}
R\mathbf{u} &= P \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} P^t\mathbf{u} \\
&= P \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \mathbf{i} \\
&\qquad \text{(using } P^t\mathbf{u} = \mathbf{i} \text{ from (2.4.2))} \\
&= P \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \\
&= P \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \\
&= P\mathbf{i} \\
&= \mathbf{u}
\end{aligned}$$

(using $P\mathbf{i} = \mathbf{u}$ from (2.4.1) at the last step).

So $\mathbf{u}$ is fixed by R. But that is right because $\mathbf{u}$ is the axis.

Let's look now at $R\mathbf{v}$. Following steps that are similar till near the end, we get

$$\begin{aligned}
R\mathbf{v} &= P\begin{bmatrix}1 & 0 & 0\\ 0 & \cos\theta & -\sin\theta\\ 0 & \sin\theta & \cos\theta\end{bmatrix} P^t\mathbf{v}\\
&= P\begin{bmatrix}1 & 0 & 0\\ 0 & \cos\theta & -\sin\theta\\ 0 & \sin\theta & \cos\theta\end{bmatrix}\mathbf{j}\\
&= P\begin{bmatrix}1 & 0 & 0\\ 0 & \cos\theta & -\sin\theta\\ 0 & \sin\theta & \cos\theta\end{bmatrix}\begin{bmatrix}0\\1\\0\end{bmatrix}\\
&= P\begin{bmatrix}0\\ \cos\theta\\ -\sin\theta\end{bmatrix}\\
&= P((\cos\theta)\mathbf{j} - (\sin\theta)\mathbf{k})\\
&= (\cos\theta)P\mathbf{j} - (\sin\theta)P\mathbf{k}\\
&= (\cos\theta)\mathbf{v} - (\sin\theta)\mathbf{w}
\end{aligned}$$

Again, this is what we should expect to get when we think in terms of the axes $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$.

We'll give some terminology now. It is standard terminology and it deals with concepts that arose above.

2.4.6 Definition. Vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^3$ are called *orthogonal* if they are all perpendicular to one another. That is if $\mathbf{u} \perp \mathbf{v}$, $\mathbf{u} \perp \mathbf{w}$ and $\mathbf{v} \perp \mathbf{w}$. (Using dot products we can express this $\mathbf{u}\cdot\mathbf{v} = 0$, $\mathbf{u}\cdot\mathbf{w} = 0$ and $\mathbf{v}\cdot\mathbf{w} = 0$.)

Vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^3$ are called *orthonormal* if they are orthogonal and they are also all unit vectors. (So $\mathbf{u}\cdot\mathbf{u} = \mathbf{v}\cdot\mathbf{v} = \mathbf{w}\cdot\mathbf{w} = 1$ as well as orthogonality.)

This terminology (orthogonal and also orthonormal) is used for different numbers of vectors, not just for 3 vectors. However there is not room for more than 3 orthonormal vectors in $\mathbb{R}^3$ and room for only 2 in $\mathbb{R}^2$. So if we are in space, we can only have 1, 2 or 3 orthonormal vectors. In $\mathbb{R}^n$ we can have more.

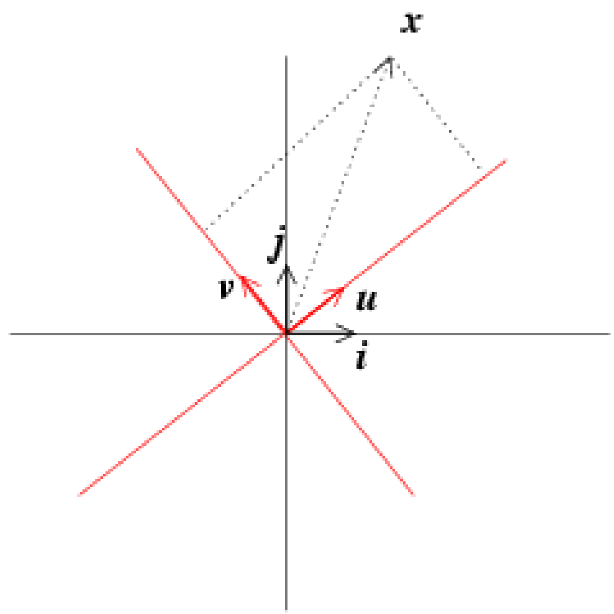
The definition of orthogonality is that each vector listed should be perpendicular to every other one. For orthonormal, each one should be a unit vector in addition.

2.4.7 Remark. We saw above that if $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ are 3 orthonormal vectors in $\mathbb{R}^3$ then we can write every $\mathbf{x} \in \mathbb{R}^3$ as a combination

$$\mathbf{x} = (\mathbf{x}\cdot\mathbf{u})\mathbf{u} + (\mathbf{x}\cdot\mathbf{v})\mathbf{v} + (\mathbf{x}\cdot\mathbf{w})\mathbf{w}$$

In two dimensions we have a similar fact. If $\mathbf{u}$ and $\mathbf{v}$ are orthonormal vectors in $\mathbb{R}^2$ then we can write every $\mathbf{x} \in \mathbb{R}^2$ as a combination

$$\mathbf{x} = (\mathbf{x}\cdot\mathbf{u})\mathbf{u} + (\mathbf{x}\cdot\mathbf{v})\mathbf{v}$$



In this 2-dimensional setting, if we write $\mathbf{u} = u_1\mathbf{i} + u_2\mathbf{j}$ and $\mathbf{v} = v_1\mathbf{i} + v_2\mathbf{j}$ then the matrix P made by using the coordinates of $\mathbf{u}$ and $\mathbf{v}$ as its columns

$$P = \begin{bmatrix} u_1 & v_1 \\ u_2 & v_2 \end{bmatrix}$$

has properties like we had in the 3×3 case earlier. These are

(i) P is an orthogonal matrix

(ii) $P\mathbf{i} = \mathbf{u}$ and $P\mathbf{j} = \mathbf{v}$.

(iii) $P^t\mathbf{u} = \mathbf{i}$ and $P^t\mathbf{v} = \mathbf{j}$.

(iv) For $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2$,

$$P^t\mathbf{x} = \begin{bmatrix} u_1 & u_2 \\ v_1 & v_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \mathbf{x} \cdot \mathbf{u} \\ \mathbf{x} \cdot \mathbf{v} \end{bmatrix}$$

gives the components of $\mathbf{x}$ in the $\mathbf{u}$-$\mathbf{v}$ basis (or frame).

Our next goal is to be able to complete the programme we started of writing down the matrix R from (2.4.6) starting with the axis $\mathbf{u}$ and the angle θ. What we did in (2.4.6) is a solution to that but it requires us to know the other two vectors $\mathbf{v}$ and $\mathbf{w}$ (so that the 3 together make up an orthonormal triple of vectors). As explained before, it is not hard to visualise $\mathbf{v}$ and $\mathbf{w}$ but we need a way to calculate them if we want to write down R.

The Gram-Schmidt method, which we now explain, is a useful method for other things. It starts with a number of vectors and 'straightens them out' so as to make them orthonormal to one another. In our case we will have 3 vectors.

In order to use it we need to have 3 vectors to start with, let's call them $\mathbf{u}$, $\mathbf{r}$ and $\mathbf{s}$. The vector $\mathbf{u}$ will be the one we know, the direction of the axis of rotation. We need $\mathbf{r}$ and $\mathbf{s}$ so that none of the 3 vectors lies in the same plane as the other two. This might sound hard, but it will work to take $\mathbf{r} = \mathbf{i}$ and $\mathbf{s} = \mathbf{j}$ unless $\mathbf{u}$ is a combination of $\mathbf{i}$ and $\mathbf{j}$ (with no $\mathbf{k}$ component).

If the axis $\mathbf{u} = (1/\sqrt{5})(\mathbf{i} + 2\mathbf{j})$ for example, we can take $\mathbf{r} = \mathbf{i}$ and $\mathbf{s} = \mathbf{k}$. So finding $\mathbf{r}$ and $\mathbf{s}$ is not a big deal.

2.4.8 Gram-Schmidt procedure. Starting with 3 vectors $\mathbf{u}$, $\mathbf{r}$ and $\mathbf{s}$, the procedure makes 3 orthonormal vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$. It is a procedure with 3 steps (or really 2).

Step 1: If $\mathbf{u}$ is not a unit vector already, replace it by $(1/\|\mathbf{u}\|)\mathbf{u}$ (the unit vector with the same direction).

Step 2: Take
$$\mathbf{v} = \frac{\mathbf{r} - (\mathbf{r} \cdot \mathbf{u})\mathbf{u}}{\|\mathbf{r} - (\mathbf{r} \cdot \mathbf{u})\mathbf{u}\|}$$

Step 3: Take
$$\mathbf{w} = \frac{\mathbf{s} - (\mathbf{s} \cdot \mathbf{u})\mathbf{u} - (\mathbf{s} \cdot \mathbf{v})\mathbf{v}}{\|\mathbf{s} - (\mathbf{s} \cdot \mathbf{u})\mathbf{u} - (\mathbf{s} \cdot \mathbf{v})\mathbf{v}\|}$$

The process is straightforward enough, though the calculations can take a bit of time to do. Here is an explanation of what is going on.

Step 1 is clear enough, perhaps. In our situation we will probably have a unit vector $\mathbf{u}$ anyhow and so step 1 will not be needed. But the rest of the steps need $\mathbf{u}$ to be a unit vector.

The numerator $\mathbf{r} - (\mathbf{r} \cdot \mathbf{u})\mathbf{u}$ in step 2 works out as the difference
$$\mathbf{r} - \mathrm{proj}_{\mathbf{u}}(\mathbf{r})$$
and so is perpendicular to $\mathbf{u}$. (Recall how projections work out to see why it is perpendicular.) Without thinking in terms of projections we can calculate
$$(\mathbf{r} - (\mathbf{r} \cdot \mathbf{u})\mathbf{u}) \cdot \mathbf{u} = \mathbf{r} \cdot \mathbf{u} - (\mathbf{r} \cdot \mathbf{u})\mathbf{u} \cdot \mathbf{u} = \mathbf{r} \cdot \mathbf{u} - (\mathbf{r} \cdot \mathbf{u})1 = 0$$
to see that they are indeed perpendicular. The numerator $\mathbf{r} - (\mathbf{r} \cdot \mathbf{u})\mathbf{u}$ can't be zero because of the assumption that $\mathbf{r}$ is not in the same plane as $\mathbf{u}$ and $\mathbf{s}$. So, when we divide the numerator vector by its length we get a unit vector $\mathbf{v}$ which is perpendicular to $\mathbf{u}$.

An important little fact is that the plane of $\mathbf{u}$ and $\mathbf{v}$ is the same as the plane of $\mathbf{u}$ and $\mathbf{s}$, so that $\mathbf{r}$ is not in that plane.

At step 3, the numerator vector turns out to be perpendicular to both $\mathbf{u}$ and to $\mathbf{v}$. That can be checked using dot products. For example
$$\begin{aligned}
\big(\mathbf{s} - (\mathbf{s} \cdot \mathbf{u})\mathbf{u} - (\mathbf{s} \cdot \mathbf{v})\mathbf{v}\big) \cdot \mathbf{u} &= \mathbf{s} \cdot \mathbf{u} - (\mathbf{s} \cdot \mathbf{u})\mathbf{u} \cdot \mathbf{u} - (\mathbf{s} \cdot \mathbf{v})\mathbf{v} \cdot \mathbf{u} \\
&= \mathbf{s} \cdot \mathbf{u} - (\mathbf{s} \cdot \mathbf{u})1 - 0 \\
&= 0
\end{aligned}$$

Similarly the dot product with $\mathbf{v}$ works out as 0.

Dividing by the length gives a unit vector. We will not be dividing by zero since that would mean that $\mathbf{s}$ would be the same as $(\mathbf{s}\cdot\mathbf{u})\mathbf{u}+(\mathbf{s}\cdot\mathbf{v})\mathbf{v}$ and that would say $\mathbf{s}$ would be in the plane of $\mathbf{u}$ and $\mathbf{v}$ (which it is not).

2.4.9 Right handed and left handed. So now $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ are orthonormal vectors and one of them is $\mathbf{u}$. So

$$P=\begin{bmatrix} u_1 & v_1 & w_1\\ u_2 & v_2 & w_2\\ u_3 & v_3 & w_3\end{bmatrix}$$

is an orthogonal matrix.

But there is another (apparently equally good) triple of vectors $\mathbf{u}$, $\mathbf{w}$ and $\mathbf{v}$ and another orthonormal change of basis matrix

$$Q=\begin{bmatrix} u_1 & v_1 & w_1\\ u_3 & v_3 & w_3\\ u_2 & v_2 & w_2\end{bmatrix}$$

where the order of $\mathbf{v}$ and $\mathbf{w}$ is reversed.

Which is better? Or is there any difference?

We know from Proposition 2.3.3 that orthogonal matrices have determinant ±1. [The argument applied to P is that $PP^t=I_3$ implies $\det(PP^t)=\det(I_3)=1$. But $\det(PP^t)=\det(P)\det(P^t)$ by the rule about determinants of a product and also we know $\det(P^t)=\det(P)$ (see §6.4). So, combining these things we find $1=\det(P)\det(P^t)=\det(P)^2$ and so $\det(P)=\pm1$.]

Now

$$\det(Q)=-\det(P)$$

because we get from one to the other by swapping two rows. It turns out that the one with determinant $+1$ is the one that corresponds to a right handed choice of axes, and so that one is perhaps better to use.

The link with right or left handed axes is as follows. We worked out in section 1.8 that

$$\det(P)=\det(P^t)=\det\begin{bmatrix} u_1 & u_2 & u_3\\ v_1 & v_2 & v_3\\ w_1 & w_2 & w_3\end{bmatrix}=\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})$$

If we think geometrically for a while, we see that, since $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ are orthonormal, $\mathbf{v}\times\mathbf{w}$ is a vector in either the same or the opposite direction to $\mathbf{u}$. (Recall $\mathbf{v}\times\mathbf{w}$ is perpendicular to the plane of $\mathbf{v}$ and $\mathbf{w}$, but so also is $\mathbf{u}$.) In fact as the length of $\mathbf{v}\times\mathbf{w}$ is

$$\|\mathbf{v}\times\mathbf{w}\|=\|\mathbf{v}\|\|\mathbf{w}\|\sin(\pi/2)=1=\|\mathbf{u}\|,$$

we must have $\mathbf{v}\times\mathbf{w}=\pm\mathbf{u}$. For $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ to be a right handed frame, we should have $\mathbf{v}\times\mathbf{w}=\mathbf{u}$ and so

$$\det(P)=\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})=\mathbf{u}\cdot(\mathbf{u})=\mathbf{u}\cdot\mathbf{u}=1.$$

On the other hand if $\mathbf{v} \times \mathbf{w} = -\mathbf{u}$, then we can reverse the order to get $\mathbf{w} \times \mathbf{v} = \mathbf{u}$ and so $\det(Q) = 1$ in this case.

2.4.10 Example. Find the matrix for the rotation about the axis $\mathbf{u} = (1/\sqrt{3})(\mathbf{i} + \mathbf{j} + \mathbf{k})$ by an angle $\theta = \pi/4$, the rotation to be in the direction so that a right-handed screw placed along the axis $\mathbf{u}$ will travel in the direction of the vector $\mathbf{u}$ (rather than the opposite direction).
Solution: The idea is to come up with a right-handed frame $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$, to take

$$P = \begin{bmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ u_3 & v_3 & w_3 \end{bmatrix}$$

and then (using (2.4.6)) compute

$$R = P \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\pi/4) & -\sin(\pi/4) \\ 0 & \sin(\pi/4) & \cos(\pi/4) \end{bmatrix} P^t$$

First we should use Gram-Schmidt to find an orthonormal frame including an axis along the direction $\mathbf{u}$. We can use $\mathbf{r} = \mathbf{i}$ and $\mathbf{s} = \mathbf{j}$. Now the steps of Gram-Schmidt.

Step 1: $\mathbf{u}$ is a unit vector already and so this step is not needed.

Just to check

$$\|\mathbf{u}\| = \frac{1}{\sqrt{3}}\|\mathbf{i} + \mathbf{j} + \mathbf{k}\| \frac{1}{\sqrt{3}}\sqrt{1^2 + 1^2 + 1^2} = 1$$

Step 2: Calculate

$$\begin{aligned}
\mathbf{r} - (\mathbf{r} \cdot \mathbf{u})\mathbf{u} &= \mathbf{i} - \left(\frac{1}{\sqrt{3}}\right)\mathbf{u} \\
&= \mathbf{i} - \left(\frac{1}{3}\right)(\mathbf{i} + \mathbf{j} + \mathbf{k}) \\
&= \frac{2}{3}\mathbf{i} - \frac{1}{3}\mathbf{j} - \frac{1}{3}\mathbf{k} \\
&= \frac{1}{3}(2\mathbf{i} - \mathbf{j} - \mathbf{k}) \\
\|\mathbf{r} - (\mathbf{r} \cdot \mathbf{u})\mathbf{u}\| &= \frac{1}{3}\|2\mathbf{i} - \mathbf{j} - \mathbf{k}\| \\
&= \frac{1}{3}\sqrt{2^2 + (-1)^2 + (-1)^2} \\
&= \frac{\sqrt{6}}{3} \\
\mathbf{v} &= \frac{\mathbf{r} - (\mathbf{r} \cdot \mathbf{u})\mathbf{u}}{\|\mathbf{r} - (\mathbf{r} \cdot \mathbf{u})\mathbf{u}\|} \\
&= \frac{1}{\sqrt{6}}(2\mathbf{i} - \mathbf{j} - \mathbf{k})
\end{aligned}$$

Step 3: Next compute

$$
\begin{aligned}
\mathbf{s} - (\mathbf{s}\cdot\mathbf{u})\mathbf{u} - (\mathbf{s}\cdot\mathbf{v})\mathbf{v} &= \mathbf{j} - \left(\frac{1}{\sqrt{3}}\right)\mathbf{u} - \left(\frac{1}{\sqrt{6}}\right)\mathbf{v} \\
&= \mathbf{j} - \left(\frac{1}{3}\right)(\mathbf{i}+\mathbf{j}+\mathbf{k}) - \left(-\frac{1}{6}\right)(2\mathbf{i}-\mathbf{j}-\mathbf{k}) \\
&= \mathbf{j} - \left(\frac{1}{3}\right)(\mathbf{i}+\mathbf{j}+\mathbf{k}) + \left(\frac{1}{6}\right)(2\mathbf{i}-\mathbf{j}-\mathbf{k}) \\
&= \left(\frac{2}{6}-\frac{1}{3}\right)\mathbf{i} + \left(1-\frac{1}{3}-\frac{1}{6}\right)\mathbf{j} - \left(\frac{1}{3}+\frac{1}{6}\right)\mathbf{k} \\
&= \frac{1}{2}\mathbf{j} - \frac{1}{2}\mathbf{k} \\
&= \frac{1}{2}(\mathbf{j}-\mathbf{k}) \\
\|\mathbf{s} - (\mathbf{s}\cdot\mathbf{u})\mathbf{u} - (\mathbf{s}\cdot\mathbf{v})\mathbf{v}\| &= \frac{1}{2}\|\mathbf{j}-\mathbf{k}\| \\
&= \frac{\sqrt{2}}{2} \\
\mathbf{w} &= \frac{\mathbf{s} - (\mathbf{s}\cdot\mathbf{u})\mathbf{u} - (\mathbf{s}\cdot\mathbf{v})\mathbf{v}}{\|\mathbf{s} - (\mathbf{s}\cdot\mathbf{u})\mathbf{u} - (\mathbf{s}\cdot\mathbf{v})\mathbf{v}\|} \\
&= \frac{1}{\sqrt{2}}((\mathbf{j}-\mathbf{k})
\end{aligned}
$$

Let us check now if we have a right-handed frame from $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$. The matrix P is

$$
\begin{bmatrix}
\frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} & 0 \\
\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\
\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}}
\end{bmatrix}
$$

and we can calculate its determinant. Using the rules for determinants it will make our lives easier if we factor out $1/\sqrt{3}$ from the first column, $2/\sqrt{6}$ from the second and $1/\sqrt{2}$ from the third.

$$
\begin{aligned}
\det(P) &= \left(\frac{1}{\sqrt{3}}\right)\left(\frac{2}{\sqrt{6}}\right)\left(\frac{1}{\sqrt{2}}\right)\det\begin{bmatrix} 1 & 2 & 0 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{bmatrix} \\
&= \frac{1}{\sqrt{36}}\det\begin{bmatrix} 1 & 2 & 0 \\ 0 & -3 & 1 \\ 0 & -3 & -1 \end{bmatrix} = \left(\frac{1}{6}\right)\det\begin{bmatrix} -3 & 1 \\ -3 & -1 \end{bmatrix} \\
&\quad \text{(expanding along column 1)} \\
&= \frac{1}{6}(3+3) = 1
\end{aligned}
$$

Now to calculate R

$$
\begin{aligned}
R &= P \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\pi/4) & -\sin(\pi/4) \\ 0 & \sin(\pi/4) & \cos(\pi/4) \end{bmatrix} P^t \\
&= P \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \\
&= P \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{2}{2\sqrt{3}} & -\frac{1}{2\sqrt{3}} - \frac{1}{2} & -\frac{1}{2\sqrt{3}} + \frac{1}{2} \\ \frac{2}{2\sqrt{3}} & -\frac{1}{2\sqrt{3}} + \frac{1}{2} & \frac{1}{2\sqrt{3}} - \frac{1}{2} \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} & 0 \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{2\sqrt{3}} - \frac{1}{2} & -\frac{1}{2\sqrt{3}} + \frac{1}{2} \\ \frac{1}{\sqrt{3}} & -\frac{1}{2\sqrt{3}} + \frac{1}{2} & \frac{1}{2\sqrt{3}} - \frac{1}{2} \end{bmatrix}
\end{aligned}
$$

Finally we get the answer

$$
R = \begin{bmatrix} \frac{1}{3}(1+\sqrt{2}) & \frac{1}{6}(2-\sqrt{2}-\sqrt{6}) & \frac{1}{6}(2-\sqrt{2}+\sqrt{6}) \\ \frac{1}{6}(2-\sqrt{2}+\sqrt{6}) & \frac{1}{3}(1+\sqrt{2}) & \frac{1}{6}(2-\sqrt{2}-\sqrt{6}) \\ \frac{1}{6}(2-\sqrt{2}-\sqrt{6}) & \frac{1}{6}(2-\sqrt{2}+\sqrt{6}) & \frac{1}{3}(1+\sqrt{2}) \end{bmatrix}
$$

2.4.11 Theorem. *The 3×3 rotation matrices are exactly all the 3×3 orthogonal matrices of determinant 1.*

(This statement is also true about 2×2 matrices — see Proposition 2.3.4.)

Steps required for a proof. Perhaps we don't need to prove this in every detail, but some parts are easy enough on the basis of what we know and others are illuminating for what we will do later. The main steps are

1. Rotation matrices are orthogonal.
2. Rotation matrices have determinant 1.
3. Orthogonal matrices with determinant 1 are all rotations.

The last step is the hardest. We won't do it but we will show in an example some of the ingredients that go into a proof. □

2.4.12 Example. Show that the matrix

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

is a rotation matrix and find the axis of rotation (up to an ambiguity of $\pm$) and the angle of rotation (up to an ambiguity between θ and $2\pi - \theta$).

Solution: Assuming we know the theorem is correct, what we have to do to establish that A is a rotation matrix is to show $AA^t = I_3$ (so A is orthogonal) and $\det(A) = 1$. Neither of these is difficult to check, but then finding the axis and the angle will be the things that involve some new techniques.

To check the two things we notice

$$AA^t = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and (computing by cofactor expansion along the first row)

$$\det(A) = 0 - 1 \det \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + 0 = -1(-1) = 1$$

(In fact A is an example of a permutation matrix — a single nonzero entry in each row and each column, and the nonzero entries are all $= 1$. We mentioned these permutation matrices briefly before in §6.3. In fact all of them are orthogonal matrices, but some have determinant -1.)

Now, if we want to find out which rotation has matrix A (or if we were going to try and prove the theorem) we would have to come up with a way to find out the axis for the rotation. Recall that the axis for a rotation matrix R is in a direction $\mathbf{u}$ which has $R\mathbf{u} = \mathbf{u}$. That means that the axis itself does not move. So if there is an axis for the rotation A, then we must have $A\mathbf{u} = \mathbf{u}$. We can rewrite this as

$$\begin{aligned} A\mathbf{u} &= I_3\mathbf{u} \\ (A - I_3)\mathbf{u} &= \mathbf{0} \\ \left(\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) \mathbf{u} &= \mathbf{0} \\ \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{bmatrix} \mathbf{u} &= \mathbf{0} \end{aligned}$$

We need a solution $\mathbf{u}$ to this which is a unit vector, or if we just find any nonzero solution $\mathbf{u}$ we can take $(1/\|\mathbf{u}\|)\mathbf{u}$ to get a unit vector. We know how to solve equations like this. What we have is a matrix form of an equation and we can solve it by row-reducing the augmented matrix

$$\left[\begin{array}{rrrl} -1 & 1 & 0 & :0 \\ 0 & -1 & 1 & :0 \\ 1 & 0 & -1 & :0 \end{array} \right]$$

First Gauss-Jordan step: multiply row 1 by -1 (to get 1 in top left).

$$\left[\begin{array}{rrrl} 1 & -1 & 0 & :0 \\ 0 & -1 & 1 & :0 \\ 1 & 0 & -1 & :0 \end{array} \right]$$

Next Gauss-Jordan step: subtract first row from third (to get new third row and zeros below leading 1 in top left).

$$\begin{bmatrix} 1 & -1 & 0 & :0 \\ 0 & -1 & 1 & :0 \\ 0 & 1 & -1 & :0 \end{bmatrix}$$

Next Gauss-Jordan step: multiply row 2 by -1 (to get leading 1 in row 2)

$$\begin{bmatrix} 1 & -1 & 0 & :0 \\ 0 & 1 & -1 & :0 \\ 0 & 1 & -1 & :0 \end{bmatrix}$$

Next Gauss-Jordan step: subtract second row from third (to get new third row and zeros below leading 1 in row 2).

$$\begin{bmatrix} 1 & -1 & 0 & :0 \\ 0 & 1 & -1 & :0 \\ 0 & 0 & 0 & :0 \end{bmatrix}$$

Next (final) Gauss-Jordan step: add second row to first (to get zeros above the last leading 1, the leading 1 in row 2).

$$\begin{bmatrix} 1 & 0 & -1 & :0 \\ 0 & 1 & -1 & :0 \\ 0 & 0 & 0 & :0 \end{bmatrix}$$

Rewrite as equations:

$$\begin{cases} u_1 & & - & u_3 & = & 0 \\ & u_2 & - & u_3 & = & 0 \end{cases}$$

So we get

$$\begin{cases} u_1 = u_3 \\ u_2 = u_3 \\ \phantom{u_2 = {}} u_3 \quad \text{free} \end{cases}$$

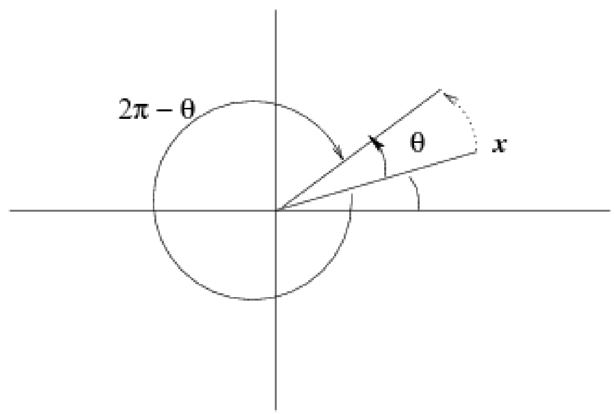
By taking a nonzero value for u_3 we get a nonzero solution. For example we could take $u_3 = 1$ which gives $u_1 = u_2 = u_3 = 1$ and the vector $\mathbf{i}+\mathbf{j}+\mathbf{k}$ parallel to the axis of rotation. If we want a unit vector we should divide this by its length and get

$$\mathbf{u} = \frac{1}{\sqrt{3}}(\mathbf{i}+\mathbf{j}+\mathbf{k})$$

Is there any other possible answer for $\mathbf{u}$? Well, yes there is. If we take any positive value for u_3 we will end up with the same unit vector $\mathbf{u}$ but if we take a negative value for u_3 we end up with the unit vector in the opposite direction. So we are left with

$$\mathbf{u} = \pm\frac{1}{\sqrt{3}}(\mathbf{i}+\mathbf{j}+\mathbf{k})$$

Which should it be? The answer is that both are possible because a rotation by an angle θ becomes an rotation by $2\pi - \theta$ when viewed upside down.



The idea of the picture is to convince you that a rotation of θ radians anticlockwise looked at from above has the same effect as a rotation by $2\pi - \theta$ radians clockwise, and the clockwise rotation of $2\pi - \theta$ will look like an anticlockwise rotation if you look at it from below (or from behind the page).

So both the plus and minus signs are possible in the vector $\mathbf{u}$ but the choice will affect the correct angle θ for the rotation.

There is a shortcut to working out the angle θ based on the trace of a matrix. We discussed the trace in §5.17 (it is the sum of the diagonal entries of a square matrix) and most of its properties are pretty easy. The only curious one was that $\mathrm{trace}(AB) = \mathrm{trace}(BA)$ even though AB and BA need not be equal at all.

Recall from (2.4.6) that a rotation has a matrix of the form

$$R = P \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} P^t$$

where P is orthogonal matrix. Using the remark about traces

$$\begin{aligned}
\operatorname{trace}(R) &= \operatorname{trace}\left(P\left(\begin{bmatrix}1 & 0 & 0\\ 0 & \cos\theta & -\sin\theta\\ 0 & \sin\theta & \cos\theta\end{bmatrix}P^t\right)\right)\\
&= \operatorname{trace}\left(\left(\begin{bmatrix}1 & 0 & 0\\ 0 & \cos\theta & -\sin\theta\\ 0 & \sin\theta & \cos\theta\end{bmatrix}P^t\right)P\right)\\
&= \operatorname{trace}\left(\begin{bmatrix}1 & 0 & 0\\ 0 & \cos\theta & -\sin\theta\\ 0 & \sin\theta & \cos\theta\end{bmatrix}P^tP\right)\\
&= \operatorname{trace}\left(\begin{bmatrix}1 & 0 & 0\\ 0 & \cos\theta & -\sin\theta\\ 0 & \sin\theta & \cos\theta\end{bmatrix}I_3\right)\\
&= \operatorname{trace}\left(\begin{bmatrix}1 & 0 & 0\\ 0 & \cos\theta & -\sin\theta\\ 0 & \sin\theta & \cos\theta\end{bmatrix}\right)\\
&= 1+\cos\theta+\cos\theta = 1+2\cos\theta
\end{aligned}$$

Our matrix A has

$$\operatorname{trace}(A) = \operatorname{trace}\left(\begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ 1 & 0 & 0\end{bmatrix}\right) = 0$$

and so the angle for our rotation has to satisfy

$$1+2\cos\theta = 0$$

or $\cos\theta = -1/2$. This gives two possible angles

$$\theta = \frac{2\pi}{3} \text{ or } \theta = \frac{4\pi}{3} = 2\pi - \frac{2\pi}{3}.$$

As we explained before, either choice could be right depending on whether we choose one sign or the other for the vector $\mathbf{u}$.

Unfortunately, we are still left with an ambiguity

$$\mathbf{u} = \frac{1}{\sqrt{3}}(\mathbf{i}+\mathbf{j}+\mathbf{k}) \text{ and } \theta = \frac{2\pi}{3} \text{ or } \theta = \frac{4\pi}{3}$$

We could work out the matrix for both of these rotations and one of them must turn out to be A. But that is quite a bit more work.

2.5 Abstract approach to linear transformations

We defined linear transformations using matrices. Given an $m \times n$ matrix A we consider $f\colon \mathbb{R}^n \to \mathbb{R}^m$ given by $f(\mathbf{x}) = A\mathbf{x}$. Recall that we found it handy for this to write $\mathbf{x} \in \mathbb{R}^n$ as an $n \times 1$ column matrix $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ (and the same for elements of $\mathbb{R}^m$).

We have looked at some examples and we considered the examples of rotation matrices (2×2 and 3×3) is some detail. An important aspect in the 3×3 case was the desirability of changing from the original **i**, **j**, **k** frame of reference to a different orthonormal frame (or basis) $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$.

It is useful in association with this kind of idea to have a way of describing linear transformations without reference to any particular frame of reference.

2.5.1 Theorem. *Linear transformations $f\colon \mathbb{R}^n \to \mathbb{R}^m$ are exactly those functions (or transformations) that satisfy the two properties*

(i) $f(\mathbf{x} + \mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y})$ *(for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$)*

(ii) $f(k\mathbf{x}) = kf(\mathbf{x})$ *(for $\mathbf{x} \in \mathbb{R}^n$, k a scalar)*

Proof. There are two things to prove here:

1. Every linear transformation $f(\mathbf{x}) = A\mathbf{x}$ satisfies (i) and (ii).

2. If we have a map that satisfies (i) and (ii) then there is some $m \times n$ matrix that gives rise to the map f.

Here is how we do them

1. This is quite easy. If $f(\mathbf{x}) = A\mathbf{x}$ then we can see

$$\begin{aligned} f(\mathbf{x}+\mathbf{y}) &= A(\mathbf{x}+\mathbf{y}) \\ &= A\mathbf{x} + A\mathbf{y} \\ &\quad \text{(by properties of matrix multiplication)} \\ &= f(\mathbf{x}) + f(\mathbf{y}) \end{aligned}$$

and

$$\begin{aligned} f(k\mathbf{x}) &= A(k\mathbf{x}) \\ &= kA\mathbf{x} \\ &\quad \text{(by properties of matrix multiplication again)} \\ &= kf(\mathbf{x}) \end{aligned}$$

2. Suppose $f\colon \mathbb{R}^n \to \mathbb{R}^m$ satisfies (i) and (ii). Then we need to find a matrix A and show it works out.

   To explain how to do this we need to introduce vectors (or points) in $\mathbb{R}^n$ which are the obvious extension of the standard basis vectors $\mathbf{i}, \mathbf{j} \in \mathbb{R}^2$ and $\mathbf{i}, \mathbf{j}, \mathbf{k} \in \mathbb{R}^3$. To avoid running out of letters we use the same letter $\mathbf{e}$ and a subscript.

   We let
$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \cdots \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix},$$
   where they are all $n \times 1$ matrices (or n-tuples) and the j^{th} one $\mathbf{e}_j$ has 1 in the j^{th} position, zeros elsewhere.

   The idea of these (called the 'standard basis vectors' in $\mathbb{R}^n$) is that we can write every $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ as a combination
$$\begin{aligned} \mathbf{x} &= x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \cdots + x_n \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \\ &= x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n \end{aligned}$$

   To find the columns of the matrix A we use
$$f(\mathbf{e}_1), f(\mathbf{e}_2), \ldots f(\mathbf{e}_n)$$

   That is we write
$$f(\mathbf{e}_1) = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}, f(\mathbf{e}_2) = \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix}, \cdots, f(\mathbf{e}_n) = \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix},$$

and take
$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

We claim now that this choice of A works out. That is we claim that it must be true that $f(\mathbf{x}) = A\mathbf{x}$ for every $\mathbf{x} \in \mathbb{R}^n$.

To verify this claim we first do it when $\mathbf{x}$ is any of the standard basis vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$. That is quite easy because of the way matrix multiplication works

$$\begin{aligned} A\mathbf{e}_1 &= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} \\ & \quad = \text{the first column of } A \\ &= f(\mathbf{e}_1) \end{aligned}$$

Similarly

$$A\mathbf{e}_2 = \text{ second column of } A \ = f(\mathbf{e}_2)$$

and we get $f(\mathbf{e}_j) = A\mathbf{e}_j$ for $j = 1, 2, \ldots, n$ this way.

Finally, for any arbitrary $\mathbf{x} \in \mathbb{R}^n$ we can write

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n$$

and use (i) and (ii) to get

$$\begin{aligned} f(\mathbf{x}) &= f(x_1\mathbf{e}_1) + f(x_2\mathbf{e}_2) + \cdots + f(x_n\mathbf{e}_n) \\ &= x_1 f(\mathbf{e}_1) + x_2 f(\mathbf{e}_2) + \cdots + x_n f(\mathbf{e}_n) \\ &= x_1 A\mathbf{e}_1 + x_2 A\mathbf{e}_2 + \cdots + x_n A\mathbf{e}_n \\ &= A(x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n) \\ &= A\mathbf{x} \end{aligned}$$

$\square$

2.5.2 Remark. The idea then of (i) and (ii) in the Theorem above is to give an abstract interpretation of what we mean by the word "linear" in "linear transformation".

One consequence (i) and (ii) is that we always have $f(\mathbf{0}) = \mathbf{0}$ for a linear transformation f. That is really obvious from the matrix approach — if $f(\mathbf{x}) = A\mathbf{x}$ always, then $f(\mathbf{0}) = \mathbf{0}$.

We can also get it directly from (i) like this. From (i) we have

$$\begin{aligned} f(\mathbf{0}) &= f(\mathbf{0} + \mathbf{0}) \\ &= f(\mathbf{0}) + f(\mathbf{0}) \end{aligned}$$

Then add $(-1)f(\mathbf{0})$ to both sides of this equation to get

$$\mathbf{0} = f(\mathbf{0})$$

2.6 Rigid motions

We have looked above at rotations (fixing the origin) in $\mathbb{R}^2$ and $\mathbb{R}^3$ and we've seen that they are implemented by multiplication by orthogonal matrices of determinant equal to 1.

It is possible to show (we will not do it) that the linear transformations $f\colon \mathbb{R}^n \to \mathbb{R}^n$ that arise from orthogonal matrices are exactly those that have the distance preserving property

$$\mathrm{distance}(f(\mathbf{x}), f(\mathbf{y})) = \mathrm{distance}(\mathbf{x}, \mathbf{y})$$

or

$$\|f(\mathbf{x}) - f(\mathbf{y})\| = \|\mathbf{x} - \mathbf{y}\|$$

(for every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$).

There is a point of view (which we will not pursue) that the rotations are the result of 'rigid motions' of the plane $\mathbb{R}^2$ or space $\mathbb{R}^3$. The idea is that a rigid motion is a continuous change starting from the identity (which leaves every point $\mathbf{x}$ unchanged) to end up at some final position, but we have to do it in such a way that at every intermediate time distances are preserved. Points of the plane or of space should be moved along gradually so that distances are maintained. We insist also that the origin never moves. Then it is the case that we must end up with some rotation applied to all the points.

It is not so hard to see that rotations are in this category, as we can start with a rotation by angle zero (which moves nothing) and gradually change the angle until we end up at the angle we want to reach.

Reflections of the plane cannot be arrived at by a continuous change from the identity (while also insisting that distances are always preserved). You could flip the plane gradually around the axis to end up with a rotation of π (or 180 degrees) but that involves moving the plane out of itself, and so is not considered a rigid motion of the plane. It would be a rigid motion in space.

We will not try to justify any of these statements.