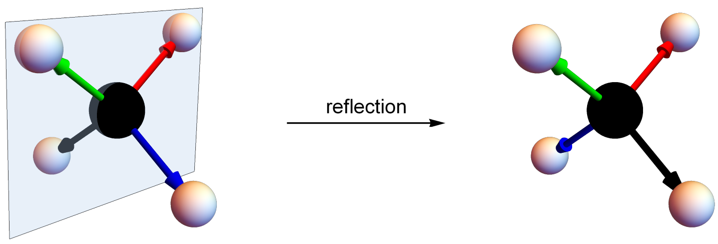
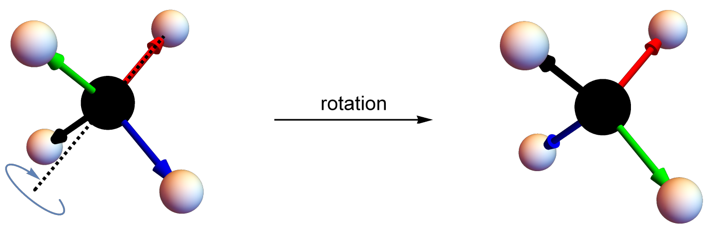
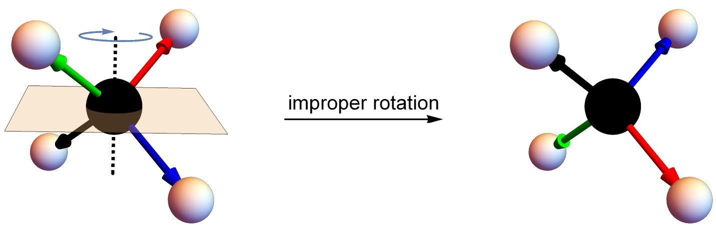
LECTURE 3

Orthogonal matrices

You've come across the ideas of molecular symmetry operations in your inorganic chemistry lectures. For example, methane is symmetric under certain reflections, rotations, and improper rotations (rotations folled by reflections). Some of these transformations are shown below. I've kept track of exactly how the molecule is transformed in each case, using coloured vectors to label the CH bonds.







In all cases, the lengths of the vectors are the same after the symmetry operation, as is the angle between any pair of vectors. Transformations with this property are known as *orthogonal transformations*.

It is easy to show that some linear transformations are not orthogonal. For example, consider the matrix

$$\mathsf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \tag{3.1}$$

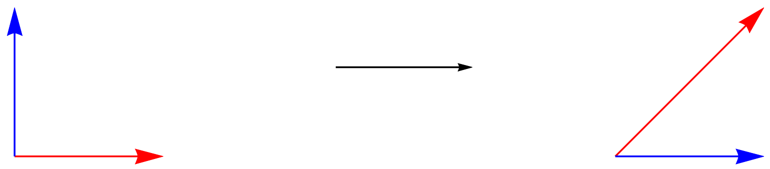
We have that

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

and

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

The vectors before and after the transformation are illustrated below. The red vector's length changes, and the angle between the two vectors changes from $\pi/2$ to $\pi/4$, so this transformation is not orthogonal.



Therefore, orthogonal transformations are special. Today, we will discover how orthogonal transformations are characterised, and the properties of the *orthogonal matrices* that describe them.

Orthogonal transformations are those that preserve the scalar product

Suppose two vectors $\mathbf{x}$ and $\mathbf{y}$ are transformed to new vectors by a matrix R:

$$\mathbf{x}' = \mathsf{R}\mathbf{x}$$
$$\mathbf{y}' = \mathsf{R}\mathbf{y}.$$

We shall show the following.

> If
>
> $$\mathbf{x}' \cdot \mathbf{y}' = \mathbf{x} \cdot \mathbf{y}, \tag{3.2}$$
>
> for any two vectors $\mathbf{x}$ and $\mathbf{y}$, the transformation must be orthogonal.

To see why, recall the definition of the scalar product,

$$\mathbf{x} \cdot \mathbf{y} = |\mathbf{x}||\mathbf{y}| \cos\theta$$

where θ is the angle betwen the vectors. In the special case of $\mathbf{y} = \mathbf{x}$, (3.2) becomes

$$\mathbf{x}' \cdot \mathbf{x}' = \mathbf{x} \cdot \mathbf{x}$$

which reduces to

$$|\mathbf{x}'|^2 = |\mathbf{x}|^2.$$

This shows that the length of $\mathbf{x}'$ is the same as the length of $\mathbf{x}$.

Now, if $\mathbf{x}$ and $\mathbf{y}$ are different vectors, (3.2) implies that

$$|\mathbf{x}'||\mathbf{y}'|\cos\theta' = |\mathbf{x}||\mathbf{y}|\cos\theta$$

(where θ' is the angle between $\mathbf{x}'$ and $\mathbf{y}'$). We have just shown that the transformation keeps $|\mathbf{x}'| = |\mathbf{x}|$ for all vectors, so this simplifies to

$$\cos\theta' = \cos\theta.$$

There is a special case to consider, if one or both vectors is the zero vector. I will leave this subtlety for you to think about.

Hence $\theta' = \theta$, i.e. the angle between the transformed vectors is unchanged.

The converse of (3.2) is also clearly true. If the transformation is orthogonal, i.e. lengths and relative angles are unchanged, then clearly the scalar product is also unchanged, by definition.

The general form of matrix multiplication

So, we need to find out which matrices preserve the scalar product between two vectors. To do this, we will need to introduce the idea of a rectangular matrix.

Up to now, we have been content to work only with *square matrices*, and have only needed two kinds of products involving matrices: the matrix-vector product, and the matrix-matrix product. Now, we will define an $m \times n$ matrix to have the following form, with m rows and n columns:

$$\begin{pmatrix} a_{11} & a_{12} \cdots\cdots a_{1n} \\ a_{21} & a_{22} \cdots\cdots a_{2n} \\ \vdots & \vdots \quad \ddots \quad \vdots \\ a_{m1} & a_{m2} \cdots\cdots a_{mn} \end{pmatrix}$$

When we defined the product AB of square ($n \times n$) matrices, we saw that this could be viewed as dot products of rows of A with columns of B. Let's now generalise this to rectangular matrices.

We shall say that an $m \times n$ matrix, A, can be multiplied by an $n \times p$ matrix, B, to give an $m \times p$ matrix $\mathsf{C} = \mathsf{AB}$. The elements of C are defined to be

$$\mathsf{c}_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}. \tag{3.3}$$

Notice that this is *exactly* the same formula as (6.13) from last Term. The only difference is that i now takes values from 1 to m, and j takes values from 1 to p.

As such, this definition is still just telling us to take dot products. It is easiest to see by an example.

Example 1. Calculate AB if

$$\mathsf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \qquad \mathsf{B} = \begin{pmatrix} 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 \end{pmatrix}$$

The product will be a 3×4 matrix. For example, the '2, 4' element is calculated from the dot product of the following row and column:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 \end{pmatrix} = \begin{pmatrix} \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & 3 \times 10 + 4 \times 14 \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

The whole thing comes out, after lots of tedious calculation, to be

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 \end{pmatrix} = \begin{pmatrix} 29 & 32 & 35 & 38 \\ 65 & 72 & 79 & 86 \\ 101 & 112 & 123 & 134 \end{pmatrix}$$

Example 2. Try to calculate BA using the matrices in Example 1.

Suppose we try to do the multiplication the other way round. We'd want to take dot products like the following

$$\begin{pmatrix} 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} = \begin{pmatrix} ? & \cdots \\ \cdots & \cdots \end{pmatrix}$$

But the row and columns have different numbers of elements. So it's not possible to multiply the matrices in this order. The product BA is undefined in this case.

Notice that if A has size $n \times n$, and B has size $n \times 1$, this definition encompasses the matrix-vector multiplication that motivated our whole study of matrix algebra

Example 3. Calculate AB if

$$\mathsf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \qquad \mathsf{B} = \begin{pmatrix} 5 \\ 6 \end{pmatrix}$$

We follow the same rule, taking the dot products of each row of A with the (only) column of B, e.g.

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 5 \\ 6 \end{pmatrix} = \begin{pmatrix} 1 \times 5 + 2 \times 6 \\ \cdots \end{pmatrix}$$

to get

$$\mathsf{AB} = \begin{pmatrix} 17 \\ 39 \end{pmatrix}$$

So column vectors are really just $n \times 1$ matrices! Likewise, we can work with row vectors, which are just $1 \times n$ matrices, e.g.

$$\begin{pmatrix} 7 & -2 \end{pmatrix}$$

And we can identify 1×1 matrices as scalars, e.g. (42) is just the number 42.

For clarity, I will continue to denote vectors using bold-faced letters (e.g. $\mathbf{r}$), even though they are just special cases of matrices.

By allowing for rectangular matrices, we are able to treat all of scalars, vectors and matrices under the same framework. This is key to understanding the results of today's lecture.

The transpose of a matrix

We can now define the transpose of a matrix. The transpose of an $m \times n$ matrix is obtained by turning all its columns into rows. This gives us an $n \times m$ matrix.

We saw a special case of the transpose, for a square matrix, in the formula for the matrix inverse in Lecture 1.

We shall use the superscript $^{\mathsf{T}}$ for transpose. So, for example, we write

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}^{\mathsf{T}} = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}.$$

Note that previous lectures and examiners on the course used a tilde, e.g. $\tilde{A}$, to denote the transpose

Notice, in particular, that the transpose of a column vector is a row vector, e.g.

$$\begin{pmatrix} 7 & -2 \end{pmatrix}^{\mathsf{T}} = \begin{pmatrix} 7 \\ -2 \end{pmatrix}.$$

Transpose of a product

We'll need an important result later. If $\mathsf{C} = \mathsf{AB}$, what is C^{T}? To answer this, we need to return to (3.3). Consider the ijth element of C. By definition of the transpose, this is the jith element of C^{T}:

$$(\mathsf{C})_{ij} = (\mathsf{C}^{\mathsf{T}})_{ji}.$$

Note that $(\mathsf{C})_{ij}$ is just what we called c_{ij} before: the ijth element of matrix C. It is normally convenient to use the simpler way of writing it, but sometimes (like now) we need the more explicit notation.

A common students' mistake here is to write $c_{ij} = c_{ji}^{\mathsf{T}}$. This is very wrong! The right-hand side of this equation would denote the transpose of the jith *element* of the matrix, not the jith element of the matrix transpose.

Combining this with the formula for the ijth element of C, (3.3), we can write

$$(\mathsf{C}^{\mathsf{T}})_{ji} = c_{ij} = \sum_k a_{ik} b_{kj}.$$

We can then say that $a_{ik} \equiv (\mathsf{A})_{ik} = (\mathsf{A}^{\mathsf{T}})_{ki}$, and likewise for B, to write

$$\begin{aligned}(\mathsf{C}^{\mathsf{T}})_{ji} &= \sum_k (\mathsf{A}^{\mathsf{T}})_{ki} (\mathsf{B}^{\mathsf{T}})_{jk} \\ &= \sum_k (\mathsf{B}^{\mathsf{T}})_{jk} (\mathsf{A}^{\mathsf{T}})_{ki}.\end{aligned}$$

(In the second step, we have used that the *elements* of the matrices A^{T} and B^{T} are just numbers, so they can be multiplied either way round.)

We can now use (3.3) again, to recognise that the RHS above is just the jith element of the product of two matrices. Specifically, we have that

$$(\mathsf{C}^\mathsf{T})_{ji} = (\mathsf{B}^\mathsf{T}\mathsf{A}^\mathsf{T})_{ji}.$$

This line in the proof is the main stumbling point for many students. Make sure you can see where it comes from. There are two ideas: we have changed which two matrices we are multiplying in (3.3), and we have adjusted the formula to calculate the jith element of the product, rather than the ijth element.

Since this holds for arbitrary i and j, we therefore have that the entire matrix C^T satisfies

$$\mathsf{C}^\mathsf{T} = \mathsf{B}^\mathsf{T}\mathsf{A}^\mathsf{T}.$$

In other words,

$$\boxed{(\mathsf{AB})^\mathsf{T} = \mathsf{B}^\mathsf{T}\mathsf{A}^\mathsf{T}. \qquad (3.4)}$$

Example 4. Verify (3.4) for matrices A and B of Example 3.

From Example 3,

$$(\mathsf{AB})^\mathsf{T} = \begin{pmatrix} 17 & 39 \end{pmatrix}.$$

On the other hand,

$$\mathsf{A}^\mathsf{T} = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} \qquad \text{and} \qquad \mathsf{B}^\mathsf{T} = \begin{pmatrix} 5 & 6 \end{pmatrix}.$$

Hence

$$\begin{aligned} \mathsf{B}^\mathsf{T}\mathsf{A}^\mathsf{T} &= \begin{pmatrix} 5 & 6 \end{pmatrix}\begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} \\ &= \begin{pmatrix} 17 & 39 \end{pmatrix} \\ &= (\mathsf{AB})^\mathsf{T} \end{aligned}$$

as required.

Orthogonal matrices produce orthogonal transformations

Now, we are finally ready to put all the ingredients together, to see which matrices lead to orthogonal transformations. Suppose that $\mathbf{x}$ and $\mathbf{y}$ are two column vectors:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \qquad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

We saw last Term that their scalar product can be calculated from

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

Notice that this is exactly what we get from the matrix product:

Technically, this product gives a 1×1 matrix, but we said earlier that such a matrix is the same as the corresponding scalar.

$$\begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = x_1 y_1 + x_2 y_2 + x_3 y_3.$$

The row vector in the above is the transpose of $\mathbf{x}$, so we can write the scalar product of two vectors as

$$\mathbf{x}\cdot\mathbf{y} \equiv \mathbf{x}^{\mathsf{T}}\mathbf{y}$$

when $\mathbf{x}$ and $\mathbf{y}$ are column vectors.

Now, suppose that

$$\begin{aligned}\mathbf{x}' &= \mathsf{R}\mathbf{x}\\ \mathbf{y}' &= \mathsf{R}\mathbf{y}.\end{aligned}$$

such that

$$\mathbf{x}'\cdot\mathbf{y}' = (\mathbf{x}')^{\mathsf{T}}\mathbf{y}'.$$

Substituting in the definitions of the transformed vectors, and then using (3.4), we see that

$$\begin{aligned}\mathbf{x}'\cdot\mathbf{y}' &= (\mathsf{R}\mathbf{x})^{\mathsf{T}}(\mathsf{R}\mathbf{y})\\ &= \mathbf{x}^{\mathsf{T}}\mathsf{R}^{\mathsf{T}}\mathsf{R}\mathbf{y}.\end{aligned}$$

If $\mathsf{R}^{\mathsf{T}}\mathsf{R} = \mathrm{I}$, then this becomes

$$\begin{aligned}\mathbf{x}'\cdot\mathbf{y}' &= \mathbf{x}^{\mathsf{T}}\mathbf{y}\\ &= \mathbf{x}\cdot\mathbf{y}.\end{aligned}$$

Therefore, any matrix satisfying $\mathsf{R}^{\mathsf{T}}\mathsf{R} = \mathrm{I}$ will produce an orthogonal transformation. We call such matrices *orthogonal matrices*.

The converse can be proven: *only* orthogonal matrices lead to orthogonal transformations. I will leave this proof to an appendix for those who are interested.

$$\boxed{\mathsf{R}^{\mathsf{T}}\mathsf{R} = \mathrm{I} \qquad \mathsf{R} \text{ is an orthogonal matrix} \qquad (3.5)}$$

Determinant and inverse of an orthogonal matrix

From (2.4), we have that

$$\det(\mathsf{R}^{\mathsf{T}})\det\mathsf{R} = \det\mathrm{I}.$$

The RHS is clearly 1, and we know that taking the transpose of a matrix does not affect its determinant. Hence, the equation above simplifies to

$$(\det\mathsf{R})^2 = 1$$

and so

$$\det\mathsf{R} = \pm 1$$

for any orthogonal matrix. This means that every orthogonal matrix is invertible and, in particular, we can see from (3.5) that

$$\mathsf{R}^{-1} = \mathsf{R}^{\mathsf{T}} \qquad \text{for an orthogonal matrix.}$$

So another way to think about an orthogonal matrix is that its transpose equals its inverse. From the properties of the inverse matrix, discussed in Lecture 2, we see that this also implies that $\mathsf{R}\mathsf{R}^{\mathsf{T}} = \mathrm{I}$.

If you've been thinking that we've been overly pedantic in showing that the converse of certain true statements is also true, here is a good opportunity to point out why that's important. We see that *if* R is orthogonal, its determinant is ± 1. However, if a matrix has determinant (say) -1, it is not necessarily orthogonal. A simple counterexample is the one I showed you earlier in (3.1):

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

The columns of orthogonal matrices are orthonormal. So are the rows.

The next property of orthogonal matrices will be important when we turn to *diagonalisation* of matrices: one of the main things that makes them useful in chemistry.

We've said before that matrix multiplication is all about dot products. Suppose that R is an orthogonal matrix. We can extract from R a set of column vectors $\mathbf{c}_i$, corresponding to the individual columns of the matrix.

$$\mathsf{R} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad \rightarrow \quad \mathbf{c}_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{c}_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{c}_3 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

Now think about R^T. The *rows* of this matrix are the same vectors as above, but expressed as row vectors.

$$\mathsf{R}^\mathsf{T} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

To multiply the two matrices together, we just take dot products of the rows of R^T with the columns of R. Hence,

$$\mathsf{R}^\mathsf{T}\mathsf{R} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} \mathbf{c}_1 \cdot \mathbf{c}_1 & \mathbf{c}_1 \cdot \mathbf{c}_2 & \mathbf{c}_1 \cdot \mathbf{c}_3 \\ \mathbf{c}_2 \cdot \mathbf{c}_1 & \mathbf{c}_2 \cdot \mathbf{c}_2 & \mathbf{c}_2 \cdot \mathbf{c}_3 \\ \mathbf{c}_3 \cdot \mathbf{c}_1 & \mathbf{c}_3 \cdot \mathbf{c}_2 & \mathbf{c}_3 \cdot \mathbf{c}_3 \end{pmatrix}.$$

But we know that $\mathsf{R}^\mathsf{T}\mathsf{R} = \mathtt{I}$. Therefore,

$$\mathbf{c}_i \cdot \mathbf{c}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}. \tag{3.6}$$

What does this mean? Well, consider the $i = j$ case first. As emphasised last Term, the dot product of the same vector with itself gives the length squared:

$$\mathbf{c}_i \cdot \mathbf{c}_i = |\mathbf{c}_i|^2.$$

So if $\mathbf{c}_i \cdot \mathbf{c}_i = 1$, all the column vectors must be unit vectors. We might say that their lengths are 'normalised to 1'.

On the other hand, we see that the dot product of two different column vectors ($i \neq j$) is zero. Therefore, they must be perpendicular, or 'orthogonal'.

It's slightly unfortunate that the word 'orthogonal' means different things when applied to vectors and matrices. Orthogonal *vectors* are perpendicular, orthogonal *matrices* satisfy (3.5).

Hence, the column vectors are individually normalised, and mutually orthogonal. We say that they are *orthonormal*.

It is fiddly to write out (3.6) every time we want to say that vectors are orthonormal. So we use a shorthand notation, called the *Kronecker delta*, δ_{ij}. We shall *define*

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{3.7}$$

For example, in three dimensions, $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ are a set of orthonormal vectors.

So, for example, $\delta_{77} = 1$, and $\delta_{93} = 0$. Then, (3.6) becomes

$$\mathbf{c}_i \cdot \mathbf{c}_j = \delta_{ij} \qquad \text{orthonormal set of vectors} \qquad (3.8)$$

It is easy to show, from $\mathsf{RR}^\mathsf{T} = \mathtt{I}$, that the rows of an orthogonal matrix are also orthonormal vectors.

Let's take a look at some examples, to see this in action. All the following matrices are orthogonal: you can confirm for yourself that their columns are orthonormal *and* their rows are also orthonormal.

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \text{(identity)}$$

$$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \qquad \text{(inversion of coordinates)}$$

$$\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \text{(rotation around } z \text{ axis)}$$

$$\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & -1 \end{pmatrix} \qquad \text{(improper rotation around } z \text{ axis)}$$

$$\begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix} \qquad \text{(reflection in line at angle } \theta \text{ to } x \text{ axis)}$$

Notice, as we said at the beginning, that these are all transformations corresponding to molecular symmetry elements. This is not a coincedence. Following this rabbit hole would lead us into the mathematical field of *group theory*.

Summary / take-home message

We've seen a lot of derivations today. This was to give us more practice in using some of the earlier results, and to see where the properties of orthogonal matrices actually come from. The key things to remember about orthogonal matrices are as follows.

- ☆ Orthogonal transformations preserve the lengths of, and relative angles between, all vectors.

- ☆ They are described by orthogonal matrices, which satisfy

  $$\mathsf{R}^\mathsf{T}\mathsf{R} = \mathsf{RR}^\mathsf{T} = \mathtt{I}$$

  i.e. the transpose of an orthogonal matrix is its inverse.

- ☆ The rows of an orthogonal matrix are orthonormal.

- ☆ The columns of an orthogonal matrix are orthonormal.