EOS 352 Continuum Dynamics
Scalars, vectors and tensors under coordinate transformations

April 30, 2014

Overview

These notes cover the following

- The concept of invariance
- Transforming vectors between coordinate systems
- Invariance in vector equations
- Scalars
- Tensors, contractions and derivatives

Physics and the choice of coordinate system

Having encountered vectors and tensors, it is worth at this point going a little deeper into what they are. The basic idea we will investigate below is that a set of equation describing a physical process should not depend on the choice of coordinate system. This is the concept of invariance, and is fundamental to physics. What we mean by 'not depending on the choice of coordinate system' is that we can define operations like taking a gradient or a divergence with respect to any given coordinate system. So we can write down, say, the heat equation in a Cartesian coordinate system $Ox_1x_2x_3$ as

$$\rho c\frac{\partial T}{\partial t} + \rho c u_i\frac{\partial T}{\partial x_i} - \frac{\partial}{\partial x_i}\left(k\frac{\partial T}{\partial x_i}\right) = a. \tag{1a}$$

Now, we expect that if we had chosen a different set of axes $Ox'_1x'_2x'_3$, we would have written

$$\rho c\frac{\partial T}{\partial t} + \rho c u'_i \frac{\partial T}{\partial x'_i} - \frac{\partial}{\partial x'_i}\left(k\frac{\partial T}{\partial x'_i}\right) = a. \tag{1b}$$

However, once we have written (1a), we specified a problem that will determine the function $T(x_1, x_2, x_3, t)$. If we now transform this function to (x'_1, x'_2, x'_3, t) as independent variables, it is not immediately clear that it will also satisfy (1b). If it doesn't, then this would imply that there was something special about the choice of coordinate axes — counter to our physical expectations.

To make further headway with this, we have to understand first how transformations between Cartesian coordinate systems work.

Transformations between different Cartesian coordinate systems

In really basic terms, if we go from one set of Cartesian coordinate axes[1] to another, the form of a vector as expressed in terms of its components parallel to the coordinate axes must also change, and this change has to take a certain predictable form. We illustrate this using the example of a rotation of the coordinate axes.

Take a vector $\mathbf{a}$ with components (a_1, a_2, a_3) expressed relative to a set of coordinate axes $Ox_1x_2x_3$. In the original, non-subscript vector notation, we might have written this as $\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$. More consistent with our idea that there is no intrinsic difference between the coordinate axes and with labelling with numerical indices (i.e.using (x_1, x_2, x_3) rather than (x, y, z)) is also to change the notation we use for unit vectors. This is usually done by writing the unit vector in the ith direction as $\mathbf{e}_i$, so $\mathbf{e}_1 = \mathbf{i}$, $\mathbf{e}_2 = \mathbf{j}$, $\mathbf{e}_3 = \mathbf{k}$. Then we can write $\mathbf{a} = a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + a_3\mathbf{e}_3$.

Now suppose we have a second coordinate system $Ox'_1x'_2x'_3$ that is rotated relative to $Ox_1x_2x_3$, and we'd like to know how to express the components (a'_1, a'_2, a'_3) of $\mathbf{a}$ relative to this second set of coordinate axes in terms of the original components (a_1, a_2, a_3). The answer is that we find the component of a vector relative to a particular coordinate axes by taking the dot product of the vector with the unit vector parallel to that axis. In other words, where we could previously have written $a_1 = \mathbf{a} \cdot \mathbf{e}_1$, we can now write

$$a'_1 = \mathbf{a} \cdot \mathbf{e}'_1$$

where $\mathbf{e}'_1$ is the unit vector in the x'_1-direction. But we have an expression for $\mathbf{a}$ in terms of the original components (a_1, a_2, a_3). Substituting this, we get

$$\begin{aligned} a'_1 &= (a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + a_3\mathbf{e}_3) \cdot \mathbf{e}'_1 \\ &= (\mathbf{e}_1 \cdot \mathbf{e}'_1)a_1 + (\mathbf{e}_2 \cdot \mathbf{e}'_1)a_2 + (\mathbf{e}_3 \cdot \mathbf{e}'_1)a_3, \end{aligned}$$

[1]The idea of invariance can also be developed for non-Cartesian coordinate systems, but this is considerably harder. A course in general relativity is often a useful starting point for this.

so a_1 is a linear combination of the original components (a_1, a_2, a_3). By the definition of the scalar product and the fact that all the unit vectors have magnitude one, the coefficients $\mathbf{e}_1 \cdot \mathbf{e}'_1$, $\mathbf{e}_2 \cdot \mathbf{e}'_1$ and $\mathbf{e}_3 \cdot \mathbf{e}'_1$ are simply the cosines of the angles between the original x_1-, x_2- and x_3-axes and the new x'_1-axis.

Similarly, we can write the remaining two components a'_2 and a'_3 relative to the new coordinate system in the form

$$a'_2 = (\mathbf{e}_1 \cdot \mathbf{e}'_2)a_1 + (\mathbf{e}_2 \cdot \mathbf{e}'_2)a_2 + (\mathbf{e}_3 \cdot \mathbf{e}'_2)a_3$$
$$a'_3 = (\mathbf{e}_1 \cdot \mathbf{e}'_3)a_1 + (\mathbf{e}_2 \cdot \mathbf{e}'_3)a_2 + (\mathbf{e}_3 \cdot \mathbf{e}'_3)a_3$$

The above can be written more succinctly as

$$a'_i = \sum_{j=1}^{3} (\mathbf{e}'_i \cdot \mathbf{e}_j)a_j$$

Defining a transformation matrix R_{ij} through

$$R_{ij} = \mathbf{e}'_i \cdot \mathbf{e}_j,$$

this becomes

$$a'_i = \sum_{j=1}^{3} R_{ij}a_j \tag{2}$$

Employing the summation convention, to which we will return shortly, this would be written as $a'_i = R_{ij}a_j$.[2]

Suppose we now have the components (a'_1, a'_2, a'_3) in the new coordinate system and want to go back to the old (a_1, a_2, a_3). By the same process as above, we would write using $\mathbf{a} = a'_1\mathbf{e}_1 + a'_2\mathbf{e}_2 + a'_3\mathbf{e}_3$ that

$$a_1 = \mathbf{a} \cdot \mathbf{e}_1 = (\mathbf{e}'_1 \cdot \mathbf{e}_1)a'_1 + (\mathbf{e}'_2 \cdot \mathbf{e}_1)a'_2 + (\mathbf{e}_3 \cdot \mathbf{e}_1)a'_3$$

and similarly for a_2 and a_3. In short form, these would be

$$a_i = \sum_{j=1}^{3} (\mathbf{e}_i \cdot \mathbf{e}'_j)a'_j = \sum_{j=1}^{3} R'_{ij}a'_j$$

with $R'_{ij} = \mathbf{e}_i \cdot \mathbf{e}'_j$. But reference to the definition of R_{ij} above shows that

$$R'_{ij} = R_{ji}$$

[2]In perhaps more familiar matrix notation

$$\begin{pmatrix} a'_1 \\ a'_2 \\ a'_3 \end{pmatrix} = \begin{pmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{32} \\ R_{31} & R_{32} & R_{33} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}.$$

so that[3]

$$a_i = \sum_{j=1}^{3} R_{ji} a'_j. \tag{3}$$

We also know that if we first transform from the $Ox_1x_2x_3$ coordinate system to the $Ox'_1x'_2x'_3$ coordinate system and then back again, we must get the original components back. In other words, we must always have

$$a_i = \sum_{j=1}^{3} R_{ji} a'_j = \sum_{j=1}^{3} R_{ji} \left(\sum_{k=1}^{3} R_{jk} a_k \right) = \sum_{j=1}^{3} \sum_{k=1}^{3} R_{ji} R_{jk} a_k \tag{4}$$

as $a'_j = \sum_{k=1}^{3} R_{jk} a_k$. This must hold for any vector $\mathbf{a}$. Hence we can demand that it should hold for the vector with $a_1 = 1$, $a_2 = a_3 = 0$. But then (4) becomes

$$\sum_{j=1}^{3} R_{j1} R_{j1} = 1, \qquad \sum_{j=1}^{3} R_{j2} R_{j1} = \sum_{j=1}^{3} R_{j3} R_{j1} = 0.$$

Similarly picking $a_1 = a_3 = 0$, $a_2 = 1$ gives

$$\sum_{j=1}^{3} R_{j1} R_{j2} = \sum_{j=1}^{3} R_{j3} R_{j2} = 0, \qquad \sum_{j=1}^{3} R_{j2} R_{j2} = 0,$$

while $a_1 = a_2 = 0$, $a_3 = 1$ leads to

$$\sum_{j=1}^{3} R_{j1} R_{j3} = \sum_{j=1}^{3} R_{j2} R_{j3} = 0, \qquad \sum_{j=1}^{3} R_{j3} R_{j3} = 0,$$

In short, this can be written as[4]

$$\sum_{j=1}^{3} R_{ji} R_{jk} = \delta_{ik}$$

Similarly, it is possible to show by demanding that $a'_i = \sum_{j=1}^{3} R_{ij} a_j = \sum_{j=1}^{3} \sum_{k=1}^{3} R_{ij} R_{kj} a'_k$ for all vectors (a'_1, a'_2, a'_3) that[5]

$$\sum_{j=1}^{3} R_{ij} R_{kj} = \delta_{ik}$$

[3] In standard matrix a language, the transformation matrix for going from a'_i to a_i is the transpose of the transformation matrix used to go from a_i to a'_i.

[4] In standard matrix notation, with $\mathbf{R}$ being used to denote the matrix with entries R_{ij} and $\mathbf{I}$ being the identity matrix, this would become $\mathbf{R}^{\mathrm{T}}\mathbf{R} = \mathbf{I}$, and the matrix $\mathbf{R}$ would be described as being an *orthonormal matrix*, or a rotation matrix.

[5] Better still, if you know something about matrices, you will know that pre- and post-multiplicative inverses are the same, so that if $\mathbf{R}^{\mathrm{T}}\mathbf{R} = \mathbf{I}$, then we also have $\mathbf{R}\mathbf{R}^{\mathrm{T}} = \mathbf{I}$

Reverting to standard summation notation from now on (i.e., dropping the explicit summation signs above), we can now say that a vector **a** is not just an object with 3 components (a_1, a_2, a_3), but also that these components transform to the new frame as

$$a'_i = R_{ij} a_j \tag{5}$$

where the transformation matrix R_{ij} satisfies

$$R_{ji} R_{jk} = R_{ij} R_{kj} = \delta_{ik} \tag{6}$$

Exercise 1 *Consider a counterclockwise rotation through an angle θ about the x_3 axis. Show that*

$$\begin{pmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{32} \\ R_{31} & R_{32} & R_{33} \end{pmatrix} = \begin{pmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and that $R_{ij} R_{ik} = R_{ji} R_{ki} = \delta_{jk}$.

Invariance in vector equations: gradients as vectors

If we have an equation describing some physics involving vectors, we have to be sure that both sides of such an equation transform in the same way from one coordinate system to another. Take for instance Fourier's law, which states that $\mathbf{q} = -k\nabla T$, or

$$q_i = -k \frac{\partial T}{\partial x_i}.$$

Now, if heat flux **q** is a vector, then we should have

$$q'_i = R_{ij} q_j.$$

But we would also like Fourier's law to hold in the rotated coordinate frame, meaning that the flux q'_i should be given by the gradient with respect to the rotated coordinates,

$$q'_i = -k \frac{\partial T}{\partial x'_i}.$$

This would require that

$$\frac{\partial T}{\partial x'_i} = R_{ij} \frac{\partial T}{\partial x_j}.$$

In other words, the question is whether the gradient of a scalar function transforms as a vector. This is not immediately obvious. What we need to do is to apply the chain rule, which states that[6]

$$\frac{\partial T}{\partial x'_i} = \frac{\partial T}{\partial x_j}\frac{\partial x_j}{\partial x'_i}.$$

But $x_j = R_{kj} x'_k$ from (3), so

$$\frac{\partial x_j}{\partial x'_i} = R_{kj}\frac{\partial x'_k}{\partial x'_i} = R_{kj}\delta_{ki} = R_{ij}$$

and hence

$$\frac{\partial T}{\partial x'_i} = R_{ij}\frac{\partial T}{\partial x_j}, \tag{7}$$

and the gradient of a scalar really does transform as a vector.

Scalars: dot products and divergences

Having characterized a vector as something that has three components that transform in a particular way, we can also refine our understanding of a scalar. A scalar is a quantity that has a single value *and that does not change under a change in coordinate systems*. For instance, the component q_1 of the vector $\mathbf{q}$ is not a scalar, because $q'_1 \neq q$, but temperature T at a given point is a scalar because it does not matter which coordinate system we use to express position.

Where this gets interesting is when we have a scalar quantity derived from a vector. For instance, we might say that the kinetic energy of a point particle of mass m and velocity $\mathbf{u}$ is $m|\mathbf{u}|^2/2$. We would like this to be a scalar, so that kinetic energy does not change when we rotate our coordinate system. But is it a scalar?

Take a vector $\mathbf{a} = (a_1, a_2, a_3)$. Then the square of the length of the vector is

$$|\mathbf{a}|^2 = a_i a_i$$

In a rotated coordinate system, we would compute $a'_i a'_i$. From (5) and (6), we can show that this gives the same value for $|\mathbf{a}|^2$ as $a_i a_i$. We have

$$a'_i a'_i = (R_{ij} a_j)(R_{ik} a_k)$$

[6]Recall that we are using the summation convention, so

$$\frac{\partial T}{\partial x'_i} = \frac{\partial T}{\partial x_j}\frac{\partial x_j}{\partial x'_i} = \sum_{j=1}^{3}\frac{\partial T}{\partial x_j}\frac{\partial x_j}{\partial x'_i}.$$

In more classical notation, this would have been for instance

$$\frac{\partial T}{\partial x'} = \frac{\partial T}{\partial x}\frac{\partial x}{\partial x'} + \frac{\partial T}{\partial y}\frac{\partial y}{\partial x'} + \frac{\partial T}{\partial z}\frac{\partial z}{\partial x'},$$

and similarly for $\partial T/\partial y'$ and $\partial T/\partial z'$.

making sure not to repeat an index more than once. But

$$(R_{ij}a_j)(R_{ik}a_k) = R_{ij}R_{ik}a_ja_k = \delta_{jk}a_ja_k = a_ka_k$$

from (6). Hence $a'_ia'_i = a_ka_k = a_ia_i$, and the length of a vector really is a scalar.

More generally, we can show that a dot product is a scalar. As $|\mathbf{a}|^2 = \mathbf{a}\cdot\mathbf{a}$, the result above is then simply a special case. We have

$$\mathbf{a}\cdot\mathbf{b} = a_ib_i.$$

In a rotated frame, we would have calculated

$$a'_ib'_i = (R_{ij}a_j)(R_{ik}b_k) = R_{ij}R_{ik}a_jb_k = \delta_{jk}a_jb_k = a_jb_j = a_ib_i,$$

and we obtain the same answer for $\mathbf{a}\cdot\mathbf{b}$ in the rotated frame. Knowing that a gradient is a vector, it follows that $\rho c\mathbf{u}\cdot\nabla T = \rho cu_i\partial T/\partial x_i$ is a scalar.[7]

Many of the equations in continuum mechanics contain divergences. For instance, we now know that $\mathbf{q} = -k\nabla T$ is a vector. But the heat equation (1a) really reads

$$\rho c\frac{\partial T}{\partial t} + \rho c\mathbf{u}\cdot\nabla T + \nabla\cdot\mathbf{q} = a,$$

and we know from the above that $\rho c\mathbf{u}\cdot\nabla T$ is a scalar (just as $\rho c\partial T/\partial t$ and a are). In order for this equation not to depend on the choice of coordinate system — in other words, for (1b) to hold if (1a) holds — we have to be sure that $\nabla\cdot\mathbf{q}$ is a scalar, i.e., that

$$\frac{\partial q_i}{\partial x_i} = \frac{\partial q'_i}{\partial x'_i}.$$

But if $\mathbf{q}$ is a vector, then $q'_i = R_{ij}q_j$. By the same steps as those leading up to (7), we also know

$$\frac{\partial}{\partial x'_i} = R_{ik}\frac{\partial}{\partial x_k} \tag{8}$$

where we have used a different dummy variable for convenience. But then

$$\frac{\partial q'_i}{\partial x'_i} = R_{ik}\frac{\partial(R_{ij}q_j)}{\partial x_k} = R_{ik}R_{ij}\frac{\partial q_j}{\partial x_k} = \delta_{kj}\frac{\partial q_j}{\partial x_k} = \frac{\partial q_k}{\partial x_k} = \frac{\partial q_i}{\partial x_i}$$

as required, and the divergence $\nabla\cdot\mathbf{q}$ is a scalar. It follows that, if we transform (1a) to (x'_1, x'_2, x'_3) and t is independent variables, we obtain (1b).

[7] You might wonder about velocity $\mathbf{u}$ being a vector but this is easy to show: if $(x_1(t), x_2(t), x_3(t))$ denotes the position of a particular bit of matter, then the velocity field at that point is $(u_1, u_2, u_3) = \mathrm{d}(x_1(t), x_2(t), x_3(t))/\mathrm{d}t$ or $u_i = \mathrm{d}x_i(t)/\mathrm{d}t$. But then, in the $Ox'_1x'_2x'_3$ frame, we have $u'_i = \mathrm{d}x'_i(t)/\mathrm{d}t = \mathrm{d}R_{ij}x_j(t)/\mathrm{d}t$ from (5), so that $u'_i = R_{ij}\,\mathrm{d}x_j(t)/\mathrm{d}t = R_{ij}u_j$.

Tensors

While we had a straightforward geometrical guide to how scalars and vector should transform, this is not as simple for tensors. The rule here is that a tensor A_{ij} should transform as

$$A'_{ij} = R_{ik} R_{jl} A_{kl}.$$

Higher order tensors with more than two indices (such as $\rho(\sigma_{jk} x_i - \sigma_{ik} x_j)$, which can be identified as the conductive flux of angular momentum $\mathcal{L}_{ij}$ in the k-direection) should transform correspondingly as

$$A'_{ijk\ldots} = R_{ip} R_{jq} R_{kr} \ldots A_{pqr\ldots}.$$

Note that this is consistent with how a vector transforms, if we treat a vector as a tensor with only on index.

This definition of how a tensor should transform has several essential properties. First of all, so-called *direct products* (or *outer products*) of vectors are naturally tensors. For instance, if $A_{ij} = a_i b_j$ with $\mathbf{a}$ and $\mathbf{b}$ tensors, then

$$A'_{ij} = R_{ik} R_{jl} A_{kl} = (R_{ik} a_k)(R_{jl} b_l) = a'_i b'_j$$

as expected. Hence angular momentum $m(x_i u_j - x_j u_i)$ transforms as a tensor as required.

Also, we frequently multiply a tensor by a vector and sum over an index to form a new vector or tensor, for instance as in $a_i = A_{ij} b_j$. This is called an *inner product* or a *contraction* over the repeated index. The result is then indeed a vector:

$$a'_i = A'_{ij} b'_j = (R_{ik} R_{jl} A_{kl})(R_{jm} b_m) = R_{ik} (R_{jl} R_{jm}) A_{kl} b_m = R_{ik} R_{lm} A_{kl} b_m = R_{ik} A_{kl} b_l = R_{ik} a_k$$

as required. This ensures for instance that an element of surface force $f_i = \sigma_{ij} n_j \Delta S$ is indeed a vector of σ_{ij} is a tensor. Alternatively, we can take a tensor and sum over a repeated index within it to form a new tensor, vector or scalar.[8] This is also called a *contraction*. For instance, if A_{ijk} is a tensor, then $a_i = A_{ijj}$ is a vector:

$$a'_i = A'_{ijj} = R_{ik} R_{jl} R_{jm} A_{klm} = R_{ik} \delta_{lm} A_{klm} = R_{ik} A_{kll} = R_{ik} a_k.$$

Lastly, derivatives of vectors or tensors are tensors. Take for instance $A_{ij} = \partial a_i / \partial x_j$. From (8) above, we have

$$\frac{\partial}{\partial x'_j} = R_{jl} \frac{\partial}{\partial x_l}$$

and hence

$$A'_{ij} = \frac{\partial a'_i}{\partial x'_j} = R_{jl} \frac{\partial (R_{ik} a_k)}{\partial x_l} = R_{ik} R_{jl} \frac{\partial a_k}{\partial x_l} = R_{ik} R_{jl} A_{kl}$$

[8] This distinction actually becomes unnecessary: a vector is simply a tensor with only one index, and a scalar is a tensor with no indices at all.

as required. Together with the results on contracting over repeated indices above, this ensures for instance that

$$u_j \frac{\partial u_i}{\partial x_j} \qquad \text{and} \qquad \frac{\partial \sigma_{ij}}{\partial x_j}$$

are vectors, while the so-called strain rate

$$D_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)$$

behaves as a tensor.

Exercise 2 *Demonstrate explicitly that, if*

$$a_i = u_j \frac{\partial u_i}{\partial x_j}$$

then $a'_i = R_{ij} a_j$. *Repeat for* $a_i = \partial \sigma_{ij} / \partial x_j$.