Tensor Algebra

Ross Dempsey

March 2016

"I admire the elegance of your method of computation; it must be nice to ride through these fields upon the horse of true mathematics while the like of us have to make our way laboriously on foot."
–Albert Einstein, writing to Levi-Civita about tensors

1 Case Study: The Inertia Tensor

When you studied angular momentum for fixed-axis rotations, you encountered the formula

$$\boldsymbol{L} = \int \boldsymbol{r} \times d\boldsymbol{p} = \hat{\boldsymbol{\omega}} \int (r_\perp)(r_\perp \omega)\, dm = \boldsymbol{\omega} \int r_\perp^2\, dm = I\boldsymbol{\omega}.$$

This is a linear relationship between $\boldsymbol{L}$ and $\boldsymbol{\omega}$. However, it is simply a constant multiplier, which is not the most general linear relationship between two vectors. We know that the most general possible relationship would be given by a matrix multiplication.

In fact, this more general relationship occurs when we stop assuming a fixed axis. We are then forced to consider the integral

$$\boldsymbol{L} = \int \boldsymbol{r} \times d\boldsymbol{p} = \int \boldsymbol{r} \times (\boldsymbol{\omega} \times \boldsymbol{r})\, dm.$$

Using the identity $\boldsymbol{A} \times (\boldsymbol{B} \times \boldsymbol{C}) = \boldsymbol{B}(\boldsymbol{A} \cdot \boldsymbol{C}) - \boldsymbol{C}(\boldsymbol{A} \cdot \boldsymbol{B})$, this becomes

$$\begin{aligned}
\boldsymbol{L} &= \int r^2\boldsymbol{\omega} - \boldsymbol{r}(\boldsymbol{r} \cdot \boldsymbol{\omega})\, dm \\
&= \hat{\boldsymbol{\imath}} \int ((r^2 - x^2)\omega_x - xy\omega_y - xz\omega_z)\, dm + \\
&\quad \hat{\boldsymbol{\jmath}} \int (-xy\omega_x + (r^2 - y^2)\omega_y - yz\omega_z)\, dm + \\
&\quad \hat{\boldsymbol{k}} \int (-xz\omega_x - yz\omega_y + (r^2 - z^2)\omega_z)\, dm.
\end{aligned}$$

We see that each component of $\boldsymbol{L}$ is a linear combination of the components of $\boldsymbol{\omega}$, so we can rewrite this as

$$\begin{aligned}
\boldsymbol{L} &= \begin{pmatrix} \int (y^2 + z^2)\, dm & -\int xy\, dm & -\int xz\, dm \\ -\int xy\, dm & \int (x^2 + z^2)\, dm & -\int yz\, dm \\ -\int xz\, dm & -\int yz\, dm & \int (x^2 + y^2)\, dm \end{pmatrix} \begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix} \\
&\equiv \hat{I}\boldsymbol{\omega}.
\end{aligned}$$

We call the matrix $\hat{I}$ the *inertia tensor*.

Not all matrices are tensors. A matrix is simply an array of numbers; a tensor has physical significance. The physical significance of a tensor is evidenced by the way it transforms when the coordinate system is transformed. As an example, consider the inertia tensor in a rotated coordinate system. Let the new coordinate system be given by

$$\begin{aligned}\tilde{x} &= a_{11}x + a_{12}y + a_{13}z,\\ \tilde{y} &= a_{21}x + a_{22}y + a_{23}z,\\ \tilde{z} &= a_{31}x + a_{32}y + a_{33}z,\end{aligned}$$

where the a_{ij} form an orthogonal matrix. Then, for example, we have

$$\begin{aligned}-\int \tilde{x}\tilde{y}\,dm = -\int \Big(&a_{11}a_{21}x^2 + a_{12}a_{22}y^2 + a_{13}a_{23}z^2\\ &+ (a_{11}a_{22} + a_{12}a_{21})xy + (a_{11}a_{23} + a_{13}a_{21})xz + (a_{12}a_{23} + a_{13}a_{22})yz\Big)\,dm.\end{aligned}$$

Making use of orthogonality, we can rewrite this as

$$\begin{aligned}-\int \tilde{x}\tilde{y}\,dm = \int \Big(&a_{11}a_{21}(y^2 + z^2) + a_{12}a_{22}(x^2 + z^2) + a_{13}a_{23}(x^2 + y^2)\\ &- (a_{11}a_{22} + a_{12}a_{21})xy - (a_{11}a_{23} + a_{13}a_{21})xz - (a_{12}a_{23} + a_{13}a_{22})yz\Big)\,dm.\end{aligned}$$

Now, observe that we can write this as

$$\tilde{\hat{I}}_{12} = \sum_{i=1}^{3}\sum_{j=1}^{3} a_{1i}a_{2j}\hat{I}_{ij}.$$

You can verify that this relation holds for each of the components:

$$\tilde{\hat{I}}_{mn} = \sum_{i=1}^{3}\sum_{j=1}^{3} a_{mi}a_{nj}\hat{I}_{ij} \qquad (1 \le m, n \le 3)$$

Note that this is equivalent to the familiar change-of-basis formula $\tilde{\hat{I}} = a\hat{I}a^{\mathrm{T}}$.

This is an example of a tensorial transformation law. We will see soon that the exact form of the transformation law will vary somewhat for different types of tensors, but the key point is that the components of a tensor under a coordinate transformation become a linear combination of the original components, with the coefficients of the combination related to the coordinate transformation.

2 Einstein Summation Notation

As we saw in the previous example, the transformation law of a tensor involves multiple sums. This is already irksome to write in two dimensions, and becomes even more so for tensors that are more complicated than the inertia tensor. To alleviate the difficulty of explicitly writing out the summation, we adopt a convention which allows almost all tensor expressions to be substantially compacted. We simply *remove the summation signs* and understand that repeated indices are summed from 1 to the understood dimension. For example, the transformation law of the inertia tensor would be written as

$$\tilde{\hat{I}}_{mn} = a_{mi}a_{nj}\hat{I}_{ij}.$$

Since i and j are repeated on the right hand side, they are summation indices. On the other hand, since m and n are not repeated on either side, they are *free indices.* Free indices indicate the presence of multiple equations contained within the single statement. In order to make the utility of index notation especially clear, we will write the transformation law as it would be written without this shorthand.

$$\begin{aligned}
\tilde{\hat{I}}_{11} &= \sum_{i=1}^{3}\sum_{j=1}^{3} a_{1i}a_{1j}\hat{I}_{ij} & \tilde{\hat{I}}_{12} &= \sum_{i=1}^{3}\sum_{j=1}^{3} a_{1i}a_{2j}\hat{I}_{ij} & \tilde{\hat{I}}_{13} &= \sum_{i=1}^{3}\sum_{j=1}^{3} a_{1i}a_{3j}\hat{I}_{ij} \\
\tilde{\hat{I}}_{21} &= \sum_{i=1}^{3}\sum_{j=1}^{3} a_{2i}a_{1j}\hat{I}_{ij} & \tilde{\hat{I}}_{22} &= \sum_{i=1}^{3}\sum_{j=1}^{3} a_{2i}a_{2j}\hat{I}_{ij} & \tilde{\hat{I}}_{23} &= \sum_{i=1}^{3}\sum_{j=1}^{3} a_{2i}a_{3j}\hat{I}_{ij} \\
\tilde{\hat{I}}_{31} &= \sum_{i=1}^{3}\sum_{j=1}^{3} a_{3i}a_{1j}\hat{I}_{ij} & \tilde{\hat{I}}_{32} &= \sum_{i=1}^{3}\sum_{j=1}^{3} a_{3i}a_{2j}\hat{I}_{ij} & \tilde{\hat{I}}_{33} &= \sum_{i=1}^{3}\sum_{j=1}^{3} a_{3i}a_{3j}\hat{I}_{ij}
\end{aligned}$$

Einstein notation requires a bit of practice to get used to, but it is very much worth the trouble. In addition to preventing carpal tunnel syndrome, it helps to strip away the unnecessary details and make algebraic manipulations clearer.

Example 2.1. Use index notation to show that $(AB)^{\mathrm{T}} = B^{\mathrm{T}}A^{\mathrm{T}}$, where A is an $a \times b$ matrix and B is a $b \times c$ matrix.

Solution: When we multiply two matrices, we take the dot product of a row from the first with a column from the second. Thus, for example,

$$(AB)_{11} = A_{11}B_{11} + A_{12}B_{21} + \ldots + A_{1b}B_{b1}.$$

In index notation, this becomes

$$(AB)_{ij} = A_{ik}B_{kj}.$$

Note that we do not have to explicitly include the dimensions of the matrices when we state the multiplication law in this way.

Now, a component of $(AB)^{\mathrm{T}}$ is given by

$$(AB)^{\mathrm{T}}_{ij} = (AB)_{ji} = A_{jk}B_{ki} = B^{\mathrm{T}}_{ik}A^{\mathrm{T}}_{kj}.$$

We observe that this is identical to the multiplication law above, except we are multiplying B^{T} and A^{T} to obtain $(AB)^{\mathrm{T}}$. This completes the proof.

3 Tensor Transformation

We are now ready to look at the transformation law for the most general type of tensor. A tensor is characterized by three quantities: its *dimension*, its *rank*, and the types of its indices. The first two quantities are the easiest to understand. The dimension of a tensor is the number of distinct values that each of its indices can take. For example, the inertia tensor has dimension 3. The rank of a tensor is the number of indices. A matrix is a second-rank tensor, a vector is a first-rank tensor, and a scalar is a zero-rank tensor. It ia also possible to have tensors of higher rank – for example, continuum mechanics makes use of a fourth-rank elasticity tensor, and nonlinear optics involves an infinite sum of tensors of increasing rank – but in elementary applications, tensors with rank ≤ 2 are most common.

To understand the types of indices, we can examine two different vectors (or first-rank tensors): a displacement vector $\boldsymbol{r}$ and a gradient vector $\boldsymbol{g} = \boldsymbol{\nabla} f$, both located at a point P. *Intuitively, these transform in different ways.* For example, consider changing units from meters to kilometers. The components of a position vector are divided by 1000 (for example, 3000 m becomes 3 km). The components of a gradient, however, are multiplied by 1000 (for example, $0.02\,\mathrm{K\,m^{-1}}$ becomes $20\,\mathrm{K\,km^{-1}}$. We will now rigorously derive general transformation laws for both vectors, and show that they are in a sense reciprocal.

Let $\boldsymbol{r}$ be given by

$$\boldsymbol{r} = r_1\hat{\boldsymbol{x}}_1 + r_2\hat{\boldsymbol{x}}_2 + \boldsymbol{r}_3\hat{\boldsymbol{x}}_3.$$

Now consider an arbitrary curvilinear coordinate system, given by

$$x_1 = x_1(u, v, w), \qquad x_2 = x_2(u, v, w), \qquad x_3 = x_3(u, v, w).$$

Let $\boldsymbol{P}$ be the position vector of our point P, and let $\boldsymbol{P}_1 = \frac{\partial \boldsymbol{P}}{\partial u}$, $\boldsymbol{P}_2 = \frac{\partial \boldsymbol{P}}{\partial v}$, and $\boldsymbol{P}_3 = \frac{\partial \boldsymbol{P}}{\partial w}$. Define the curvilinear unit vectors to be

$$\hat{\boldsymbol{e}}_1 = \frac{\boldsymbol{P}_1}{|\boldsymbol{P}_1|} \qquad \hat{\boldsymbol{e}}_2 = \frac{\boldsymbol{P}_2}{|\boldsymbol{P}_2|} \qquad \hat{\boldsymbol{e}}_3 = \frac{\boldsymbol{P}_3}{|\boldsymbol{P}_3|}$$

Therefore,

$$d\boldsymbol{P} = dx_1\hat{\boldsymbol{x}}_1 + dx_2\hat{\boldsymbol{x}}_2 + dx_3\hat{\boldsymbol{x}}_3 = |\boldsymbol{P}_1|du\,\hat{\boldsymbol{e}}_{\boldsymbol{1}} + |\boldsymbol{P}_2|dv\,\hat{\boldsymbol{e}}_{\boldsymbol{2}} + |\boldsymbol{P}_3|dw\,\hat{\boldsymbol{e}}_{\boldsymbol{3}}.$$

It makes sense for the curvilinear components of this displacement to be (du, dv, dw), but we see that in this expression, these components are multiplied by scale factors. Thus, we define the components of our vector in the new coordinate system at a point P by the relation

$$\boldsymbol{r} = \tilde{r}_1|\boldsymbol{P}_1|\hat{\boldsymbol{e}}_1 + \tilde{r}_2|\boldsymbol{P}_2|\hat{\boldsymbol{e}}_2 + \tilde{r}_3|\boldsymbol{P}_3|\hat{\boldsymbol{e}}_3.$$

From here, we can solve for the components $\tilde{r}_i$. This is accomplished most easily by writing the curvilinear basis vectors in terms of the original basis vectors, as

$$\begin{aligned}
\hat{\boldsymbol{e}}_{\boldsymbol{1}} &= \frac{1}{|\boldsymbol{P}_1|}\left(\frac{\partial x_1}{\partial u}\hat{\boldsymbol{x}}_1 + \frac{\partial x_2}{\partial u}\hat{\boldsymbol{x}}_2 + \frac{\partial x_3}{\partial u}\hat{\boldsymbol{x}}_3\right), \\
\hat{\boldsymbol{e}}_{\boldsymbol{2}} &= \frac{1}{|\boldsymbol{P}_2|}\left(\frac{\partial x_1}{\partial v}\hat{\boldsymbol{x}}_1 + \frac{\partial x_2}{\partial v}\hat{\boldsymbol{x}}_2 + \frac{\partial x_3}{\partial v}\hat{\boldsymbol{x}}_3\right), \\
\hat{\boldsymbol{e}}_{\boldsymbol{3}} &= \frac{1}{|\boldsymbol{P}_3|}\left(\frac{\partial x_1}{\partial w}\hat{\boldsymbol{x}}_1 + \frac{\partial x_2}{\partial w}\hat{\boldsymbol{x}}_2 + \frac{\partial x_3}{\partial w}\hat{\boldsymbol{x}}_3\right).
\end{aligned}$$

This gives the system of equations

$$\begin{aligned}
r_1 &= \tilde{r}_1\frac{\partial x_1}{\partial u} + \tilde{r}_2\frac{\partial x_1}{\partial v} + \tilde{r}_3\frac{\partial x_1}{\partial w}, \\
r_2 &= \tilde{r}_1\frac{\partial x_2}{\partial u} + \tilde{r}_2\frac{\partial x_2}{\partial v} + \tilde{r}_3\frac{\partial x_2}{\partial w}, \\
r_3 &= \tilde{r}_1\frac{\partial x_3}{\partial u} + \tilde{r}_2\frac{\partial x_3}{\partial v} + \tilde{r}_3\frac{\partial x_3}{\partial w}.
\end{aligned}$$

To solve this system, we need to invert the matrix of partial derivatives

$$\begin{pmatrix}
\frac{\partial x_1}{\partial u} & \frac{\partial x_1}{\partial v} & \frac{\partial x_1}{\partial w} \\
\frac{\partial x_2}{\partial u} & \frac{\partial x_2}{\partial v} & \frac{\partial x_2}{\partial w} \\
\frac{\partial x_3}{\partial u} & \frac{\partial x_3}{\partial v} & \frac{\partial x_3}{\partial w}
\end{pmatrix}.$$

This can be accomplished by remembering the chain rule. If we multiply by the matrix

$$\begin{pmatrix} \frac{\partial u}{\partial x_1} & \frac{\partial u}{\partial x_2} & \frac{\partial u}{\partial x_3} \\ \frac{\partial v}{\partial x_1} & \frac{\partial v}{\partial x_2} & \frac{\partial v}{\partial x_3} \\ \frac{\partial w}{\partial x_1} & \frac{\partial w}{\partial x_2} & \frac{\partial w}{\partial x_3} \end{pmatrix},$$

we obtain

$$\begin{pmatrix} \frac{dx_1}{dx_1} & \frac{dx_1}{dx_2} & \frac{dx_1}{dx_3} \\ \frac{dx_2}{dx_1} & \frac{dx_2}{dx_2} & \frac{dx_2}{dx_3} \\ \frac{dx_3}{dx_1} & \frac{dx_3}{dx_2} & \frac{dx_3}{dx_3} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Using our inverse matrix to solve the system, we obtain finally the transformation law:

$$\begin{aligned} \tilde{r}_1 &= \frac{\partial u}{\partial x_1} r_1 + \frac{\partial u}{\partial x_2} r_2 + \frac{\partial u}{\partial x_3} r_3, \\ \tilde{r}_2 &= \frac{\partial v}{\partial x_1} r_1 + \frac{\partial v}{\partial x_2} r_2 + \frac{\partial v}{\partial x_3} r_3, \\ \tilde{r}_3 &= \frac{\partial w}{\partial x_1} r_1 + \frac{\partial w}{\partial x_2} r_2 + \frac{\partial w}{\partial x_3} r_3. \end{aligned}$$

As an exercise, you can repeat this derivation using index notation, by defining $\tilde{x}_1 = u$, $\tilde{x}_2 = v$, and $\tilde{x}_3 = w$. The final result becomes

$$\tilde{r}_i = \frac{\partial \tilde{x}_i}{\partial x_j} r_j.$$

Now, we will determine the transformation law of $\boldsymbol{g} = \boldsymbol{\nabla} f$, this time using index notation. Recall that the gradient is the vector that satisfies $df = (\boldsymbol{\nabla} f) \cdot d\boldsymbol{P}$. Clearly this is satisfied in the original Cartesian coordinate system by $g_i = \frac{\partial f}{\partial x_i}$. In the curvilinear system, we can substitute for $d\boldsymbol{P}$ and obtain

$$df = \frac{\partial f}{\partial \tilde{x}_i} d\tilde{x}_i = (\boldsymbol{\nabla} f) \cdot (|\boldsymbol{P}_i| d\tilde{x}_i \, \hat{\boldsymbol{e}}_{\boldsymbol{i}}).$$

This implies that

$$\boldsymbol{\nabla} f = \frac{1}{|\boldsymbol{P}_i|} \frac{\partial f}{\partial \tilde{x}_i} \hat{\boldsymbol{e}}_i.$$

Now, by analogy with the position vector where we did not include the scale factors $|\boldsymbol{P}_i|$ in the components, here we will not include the scale factors $\frac{1}{|\boldsymbol{P}_i|}$ in the components. Thus, we have

$$\tilde{g}_i = \frac{\partial f}{\partial \tilde{x}_i} = \frac{\partial x_j}{\partial \tilde{x}_i} \frac{\partial f}{\partial x_j} = \frac{\partial x_j}{\partial \tilde{x}_i} g_j.$$

Comparing this with the transformation law for $\boldsymbol{r}$, we see that it is indeed reciprocal in a formal sense.

With this analysis, we are ready to define the two types of vectors. We call vectors satisfying a transformation law like the position vector *contravariant*. The components of contravariant vectors are written with raised indices. Likewise, since the coordinates themselves can be considered components of contravariant position vectors, we write them with raised indices. Thus, the contravariant transformation law is

$$\tilde{A}^i = \frac{\partial \tilde{x}^i}{\partial x^j} A^j.$$

Vectors that transform like the gradient are called *covariant*. Their components are written with lowered indices. The covariant transformation law is

$$\tilde{A}_i = \frac{\partial x^j}{\partial \tilde{x}^i} A_j.$$

To form tensors of different rank, we can combine vectors. First, consider forming the dot product between a contravariant and covariant vector, $A^i B_i$. The transformation law of this quantity is

$$\tilde{A}^i \tilde{B}_i = \left(\frac{\partial \tilde{x}^i}{\partial x^j} A^j \right) \left(\frac{\partial x^j}{\partial \tilde{x}^i} B_j \right) = \left[\frac{\partial \tilde{x}^i}{\partial x^j} \frac{\partial x^j}{\partial \tilde{x}^i} \right] A^j B_j.$$

The quantity in brackets is simply 1. Replacing j with i, since it is an arbitrary summation index, we have the transformation law of a *scalar*:

$$\tilde{A}^i \tilde{B}_i = A^i B_i.$$

Additionally, we can form tensors of the second rank by multiplying components of vectors. For example, using two contravariant vectors, we can form the quantities $A^i B^j$. These transform as

$$\tilde{A}^i \tilde{B}^j = \left(\frac{\partial \tilde{x}^i}{\partial x^k} A^k \right) \left(\frac{\partial \tilde{x}^j}{\partial x^l} B^l \right) = \frac{\partial \tilde{x}^i}{\partial x^k} \frac{\partial \tilde{x}^j}{\partial x^l} A^k B^l.$$

This is the transformation law for a contravariant tensor of the second rank. Likewise, the components $A_i B_j$ form a covariant tensor transforming according to

$$\tilde{A}_i \tilde{B}_j = \left(\frac{\partial x^k}{\partial \tilde{x}^i} A_k \right) \left(\frac{\partial x^l}{\partial \tilde{x}^j} B_l \right) = \frac{\partial x^k}{\partial \tilde{x}^i} \frac{\partial x^l}{\partial \tilde{x}^j} A_k B_l.$$

We can also form a mixed tensor, with one index contravariant and the other covariant:

$$\tilde{A}^i \tilde{B}_j = \left(\frac{\partial \tilde{x}^i}{\partial x^k} A^k \right) \left(\frac{\partial x^l}{\partial \tilde{x}^j} B_l \right) = \frac{\partial \tilde{x}^i}{\partial x^k} \frac{\partial x^l}{\partial \tilde{x}^j} A^k B_l.$$

This is also called a type $(1,1)$ tensor. In general, a tensor with r contravariant indices and s covariant indices is called type (r,s). Note that a second rank tensor need not be a product of vectors, but a product of vectors is a second rank tensor.

We can now, finally, state the general transformation law of a type (r,s) tensor of dimension n (and rank $r+s$). Following by analogy to the specific cases we've examined, we have

$$\tilde{T}^{\alpha_1 \alpha_2 \cdots \alpha_r}_{\beta_1 \beta_2 \cdots \beta_s} = \frac{\partial \tilde{x}^{\alpha_1}}{\partial x^{\gamma_1}} \cdots \frac{\partial \tilde{x}^{\alpha_r}}{\partial x^{\gamma_r}} \frac{\partial x^{\delta_1}}{\partial \tilde{x}^{\beta_1}} \cdots \frac{\partial x^{\delta_s}}{\partial \tilde{x}^{\beta_s}} T^{\gamma_1 \gamma_2 \cdots \gamma_r}_{\delta_1 \delta_2 \cdots \delta_s}.$$

> **Example 3.1.** If B^h_{ij} is an arbitrary type $(1,2)$ tensor and $A^{ijk} B^h_{ij}$ is a type $(2,0)$ tensor, show that the components A^{ijk} form a type $(3,0)$ tensor.
>
> **Solution:** We are given that
>
> $$\tilde{A}^{ijk} \tilde{B}^h_{ij} = \frac{\partial \tilde{x}^k}{\partial x^\mu} \frac{\partial \tilde{x}^h}{\partial x^\nu} A^{ij\mu} B^\nu_{ij}$$
>
> and
>
> $$\tilde{B}^h_{ij} = \frac{\partial x^\alpha}{\partial \tilde{x}^i} \frac{\partial x^\beta}{\partial \tilde{x}^j} \frac{\partial \tilde{x}^h}{\partial x^\gamma} B^\gamma_{\alpha\beta}.$$

Substituting, we have

$$\frac{\partial x^\alpha}{\partial \tilde{x}^i}\frac{\partial x^\beta}{\partial \tilde{x}^j}\frac{\partial \tilde{x}^h}{\partial x^\gamma}\tilde{A}^{ijk}B^\gamma_{\alpha\beta} = \frac{\partial \tilde{x}^k}{\partial x^\mu}\frac{\partial \tilde{x}^h}{\partial x^\nu}A^{ij\mu}B^\nu_{ij}.$$

We can now multiply each side by $\frac{\partial x^\eta}{\partial \tilde{x}^h}$, obtaining

$$\frac{\partial x^\alpha}{\partial \tilde{x}^i}\frac{\partial x^\beta}{\partial \tilde{x}^j}\left(\frac{\partial x^\eta}{\partial \tilde{x}^h}\frac{\partial \tilde{x}^h}{\partial x^\gamma}\right)\tilde{A}^{ijk}B^\gamma_{\alpha\beta} = \frac{\partial \tilde{x}^k}{\partial x^\mu}\left(\frac{\partial x^\eta}{\partial \tilde{x}^h}\frac{\partial \tilde{x}^h}{\partial x^\nu}\right)A^{ij\mu}B^\nu_{ij}.$$

On the left in parentheses, we have a mixed second rank tensor with components 1 if $\eta = \gamma$ and 0 if $\eta \neq \gamma$. This is known as the Kronecker tensor δ^η_γ. The situation is similar on the right hand side. Thus, we have

$$\frac{\partial x^\alpha}{\partial \tilde{x}^i}\frac{\partial x^\beta}{\partial \tilde{x}^j}\delta^\eta_\gamma\tilde{A}^{ijk}B^\gamma_{\alpha\beta} = \frac{\partial \tilde{x}^k}{\partial x^\mu}\delta^\eta_\nu A^{ij\mu}B^\nu_{ij}.$$

We can easily carry out the sums in γ and ν based on the values of the Kronecker tensor, which gives

$$\frac{\partial x^\alpha}{\partial \tilde{x}^i}\frac{\partial x^\beta}{\partial \tilde{x}^j}\tilde{A}^{ijk}B^\eta_{\alpha\beta} = \frac{\partial \tilde{x}^k}{\partial x^\mu}A^{ij\mu}B^\eta_{ij}.$$

The next step illusrates a very important technique. Since B is arbitrary, we can choose $B^1_{11} = 1$ and all other components 0. Our relation then implies that

$$\frac{\partial x^1}{\partial \tilde{x}^i}\frac{\partial x^1}{\partial \tilde{x}^j}\tilde{A}^{ijk} = \frac{\partial \tilde{x}^k}{\partial x^\mu}A^{11\mu}.$$

This can be repeated for all choices of the indices, so we have in general

$$\frac{\partial x^\alpha}{\partial \tilde{x}^i}\frac{\partial x^\beta}{\partial \tilde{x}^j}\tilde{A}^{ijk} = \frac{\partial \tilde{x}^k}{\partial x^\mu}A^{\alpha\beta\mu}.$$

Finally, we multiply each side by $\frac{\partial x^\gamma}{\partial \tilde{x}^k}$, obtaining

$$\frac{\partial x^\alpha}{\partial \tilde{x}^i}\frac{\partial x^\beta}{\partial \tilde{x}^j}\frac{\partial x^\gamma}{\partial \tilde{x}^k}\tilde{A}^{ijk} = A^{\alpha\beta\gamma}.$$

We see that this is the transformation law for a contravariant third-rank tensor.

4 The Metric Tensor

In Euclidean space, we know that the square of a length of a vector is given by the sum of the squares of its components. In index notation, we write this as

$$|\boldsymbol{X}|^2 = X^i X^i.$$

It is useful to know this length even when the components of $\boldsymbol{X}$ are expressed in another coordinate system. We can accomplish this using the contravariant transformation law:

$$|\boldsymbol{X}|^2 = X^i X^i = \left(\frac{\partial x^i}{\partial \tilde{x}^j}\tilde{X}^j\right)\left(\frac{\partial x^i}{\partial \tilde{x}^k}\tilde{X}^k\right) = \frac{\partial x^i}{\partial \tilde{x}^j}\frac{\partial x^i}{\partial \tilde{x}^k}\tilde{X}^j\tilde{X}^k.$$

The quantities $\frac{\partial x^i}{\partial \tilde{x}^j}\frac{\partial x^i}{\partial \tilde{x}^k}$ form a covariant tensor, as can be easily verified. We call this the *metric tensor*:

$$g_{jk} = \frac{\partial x^i}{\partial \tilde{x}^j}\frac{\partial x^i}{\partial \tilde{x}^k}.$$

Example 4.1. Determine the metric tensor in polar, cylindrical, and spherical coordinates.

Solution: First, we note that in an orthogonal coordinate system, the metric tensor is diagonal. This follows directly from orthogonality: for example, since $\hat{\boldsymbol{r}} = \frac{\partial x}{\partial r}\hat{\boldsymbol{x}} + \frac{\partial y}{\partial r}\hat{\boldsymbol{y}}$ and $\hat{\boldsymbol{\theta}} = \frac{\partial x}{\partial \theta}\hat{\boldsymbol{x}} + \frac{\partial y}{\partial \theta}\hat{\boldsymbol{y}}$ are orthogonal, the components g_{12} and g_{21} of the metric tensor in polar coordinates vanish. It remains to compute the diagonal components. In polar coordinates with $\tilde{x}^1 = r$ and $\tilde{x}^2 = \theta$, we have

$$\begin{aligned} g_{11} &= \left(\frac{\partial x}{\partial r}\right)^2 + \left(\frac{\partial y}{\partial r}\right)^2 = \cos^2\theta + \sin^2\theta = 1, \\ g_{22} &= \left(\frac{\partial x}{\partial \theta}\right)^2 + \left(\frac{\partial y}{\partial \theta}\right)^2 = r^2\sin^2\theta + r^2\cos^2\theta = r^2. \end{aligned}$$

In cylindrical coordinates with $\tilde{x}^1 = r$, $\tilde{x}^2 = \theta$, and $\tilde{x}^3 = z$, similar computations give

$$\begin{aligned} g_{11} &= \left(\frac{\partial x}{\partial r}\right)^2 + \left(\frac{\partial y}{\partial r}\right)^2 + \left(\frac{\partial z}{\partial r}\right)^2 = \cos^2\theta + \sin^2\theta = 1, \\ g_{22} &= \left(\frac{\partial x}{\partial \theta}\right)^2 + \left(\frac{\partial y}{\partial \theta}\right)^2 + \left(\frac{\partial z}{\partial \theta}\right)^2 = r^2\sin^2\theta + r^2\cos^2\theta = r^2, \\ g_{33} &= \left(\frac{\partial x}{\partial z}\right)^2 + \left(\frac{\partial y}{\partial z}\right)^2 + \left(\frac{\partial z}{\partial z}\right)^2 = 1. \end{aligned}$$

Finally, in spherical coordinates with $\tilde{x}^1 = r$, $\tilde{x}^2 = \theta$, and $\tilde{x}^3 = \phi$, we have

$$\begin{aligned} g_{11} &= \left(\frac{\partial x}{\partial r}\right)^2 + \left(\frac{\partial y}{\partial r}\right)^2 + \left(\frac{\partial z}{\partial r}\right)^2 = \sin^2\theta\cos^2\phi + \sin^2\theta\sin^2\phi + \cos^2\theta = 1, \\ g_{22} &= \left(\frac{\partial x}{\partial \theta}\right)^2 + \left(\frac{\partial y}{\partial \theta}\right)^2 + \left(\frac{\partial z}{\partial \theta}\right)^2 = r^2\cos^2\theta\cos^2\phi + r^2\cos^2\theta\sin^2\phi + r^2\sin^2\theta = r^2, \\ g_{33} &= \left(\frac{\partial x}{\partial \phi}\right)^2 + \left(\frac{\partial y}{\partial \phi}\right)^2 + \left(\frac{\partial z}{\partial \phi}\right)^2 = r^2\sin^2\theta\sin^2\phi + r^2\sin^2\theta\cos^2\phi = r^2\sin^2\theta. \end{aligned}$$

Note that in each of these cases, the determinant of the metric tensor is the square of the familiar Jacobian determinant. This is not a coincidence: you can verify that $g = J^{\mathrm{T}}J$.

The metric tensor is of paramount importance for understanding a curvilinear coordinate system. As a practical example, we will use it to determine the gradient, divergence, and curl in orthogonal curvilinear coordinates.

The gradient is the unique vector satisfying $df = (\boldsymbol{\nabla} f)\cdot d\boldsymbol{r}$. We know that

$$d\boldsymbol{r} = \left|\frac{\partial \boldsymbol{r}}{\partial \tilde{x}^1}\right| d\tilde{x}^1\hat{\boldsymbol{e}}_1 + \left|\frac{\partial \boldsymbol{r}}{\partial \tilde{x}^2}\right| d\tilde{x}^2\hat{\boldsymbol{e}}_2 + \left|\frac{\partial \boldsymbol{r}}{\partial \tilde{x}^3}\right| d\tilde{x}^3\hat{\boldsymbol{e}}_3 = \sqrt{g_{11}}d\tilde{x}^1\hat{\boldsymbol{e}}_1 + \sqrt{g_{22}}d\tilde{x}^2\hat{\boldsymbol{e}}_2 + \sqrt{g_{33}}d\tilde{x}^3\hat{\boldsymbol{e}}_3.$$

Likewise,

$$df = \frac{\partial f}{\partial \tilde{x}^1}d\tilde{x}^1 + \frac{\partial f}{\partial \tilde{x}^2}d\tilde{x}^2 + \frac{\partial f}{\partial \tilde{x}^3}d\tilde{x}^3.$$

Therefore, we must have

$$\boldsymbol{\nabla} f = \frac{1}{\sqrt{g_{11}}}\frac{\partial f}{\partial \tilde{x}^1}\hat{\boldsymbol{e}}_1 + \frac{1}{\sqrt{g_{22}}}\frac{\partial f}{\partial \tilde{x}^2}\hat{\boldsymbol{e}}_2 + \frac{1}{\sqrt{g_{33}}}\frac{\partial f}{\partial \tilde{x}^3}\hat{\boldsymbol{e}}_3.$$

For example, in spherical coordinates, the gradient is

$$\boldsymbol{\nabla} f = \frac{\partial f}{\partial r}\hat{\boldsymbol{r}} + \frac{1}{r}\frac{\partial f}{\partial \theta}\hat{\boldsymbol{\theta}} + \frac{1}{r\sin\theta}\frac{\partial f}{\partial \phi}\hat{\boldsymbol{\phi}}.$$

To determine the divergence, we can use the divergence theorem on an infinitesimal "cube" between vertices $(\tilde{x}^1 - d\tilde{x}^1, \tilde{x}^2 - d\tilde{x}^2, \tilde{x}^3 - d\tilde{x}^3)$ and $(\tilde{x}^1 + d\tilde{x}^1, \tilde{x}^2 + d\tilde{x}^2, \tilde{x}^3 + d\tilde{x}^3)$. The divergence theorem tells us that

$$\oiint \boldsymbol{v}\cdot d\boldsymbol{A} = \iiint \boldsymbol{\nabla}\cdot\boldsymbol{v}\, dV,$$

and since our cube is infinitesimal, we can replace the Riemann sum on the right with the single term $\boldsymbol{\nabla}\cdot\boldsymbol{v}(\tilde{x}^1, \tilde{x}^2, \tilde{x}^3)\, dV$. On the left, we have three terms, one for each coordinate direction. The surface integral on the faces of constant $\tilde{x}^2$ and $\tilde{x}^3$ is given by

$$4\, v^1\sqrt{g_{22}g_{33}}d\tilde{x}^2 d\tilde{x}^3\big|_{(\tilde{x}^1+d\tilde{x}^1,\tilde{x}^2,\tilde{x}^3)} - 4\, v^1\sqrt{g_{22}g_{33}}d\tilde{x}^2 d\tilde{x}^3\big|_{(\tilde{x}^1-d\tilde{x}^1,\tilde{x}^2,\tilde{x}^3)} = 8\frac{\partial}{\partial\tilde{x}^1}\left(v^1\sqrt{g_{22}g_{33}}\right)d\tilde{x}^1 d\tilde{x}^2 d\tilde{x}^3.$$

The surface integrals of the other four faces of the cube are similar. The total integral is

$$\oiint \boldsymbol{v}\cdot d\boldsymbol{A} = 8d\tilde{x}^1 d\tilde{x}^2 d\tilde{x}^3\left(\frac{\partial}{\partial\tilde{x}^1}\left(v^1\sqrt{g_{22}g_{33}}\right) + \frac{\partial}{\partial\tilde{x}^2}\left(v^2\sqrt{g_{11}g_{33}}\right) + \frac{\partial}{\partial\tilde{x}^3}\left(v^3\sqrt{g_{11}g_{22}}\right)\right).$$

Equating this with $\boldsymbol{\nabla}\cdot\boldsymbol{v}\, dV = 8(\boldsymbol{\nabla}\cdot\boldsymbol{v})\sqrt{g_{11}g_{22}g_{33}}d\tilde{x}^1 d\tilde{x}^2 d\tilde{x}^3$, we obtain the divergence:

$$\boldsymbol{\nabla}\cdot\boldsymbol{v} = \frac{1}{\sqrt{g_{11}g_{22}g_{33}}}\left(\frac{\partial}{\partial\tilde{x}^1}\left(v^1\sqrt{g_{22}g_{33}}\right) + \frac{\partial}{\partial\tilde{x}^2}\left(v^2\sqrt{g_{11}g_{33}}\right) + \frac{\partial}{\partial\tilde{x}^3}\left(v^3\sqrt{g_{11}g_{22}}\right)\right).$$

Finally, for the curl, we use Stokes' theorem, which tells us that

$$\iint(\boldsymbol{\nabla}\times\boldsymbol{v})\cdot d\boldsymbol{S} = \oint \boldsymbol{v}\cdot d\boldsymbol{r}.$$

If we allow the surface to be an infinitesimal "square" in the $\tilde{x}^1\tilde{x}^2$ plane, then this becomes

$$4(\boldsymbol{\nabla}\times\boldsymbol{v})^3\sqrt{g_{11}g_{22}}d\tilde{x}^1 d\tilde{x}^2 = \oint \boldsymbol{v}\cdot d\boldsymbol{r}.$$

The line integral is

$$\begin{aligned}\oint \boldsymbol{v}\cdot d\boldsymbol{r} &= \left(2\, v^1\sqrt{g_{11}}\big|_{(\tilde{x}^1,\tilde{x}^2-d\tilde{x}^2,\tilde{x}^3)} - 2\, v^1\sqrt{g_{11}}\big|_{(\tilde{x}^1,\tilde{x}^2+d\tilde{x}^2,\tilde{x}^3)}\right)d\tilde{x}^1 \\ &+ \left(2\, v^2\sqrt{g_{22}}\big|_{(\tilde{x}^1+d\tilde{x}^1,\tilde{x}^2,\tilde{x}^3)} - 2\, v^2\sqrt{g_{22}}\big|_{(\tilde{x}^1-d\tilde{x}^1,\tilde{x}^2,\tilde{x}^3)}\right)d\tilde{x}^2 \\ &= 4\left(\frac{\partial(v^2\sqrt{g_{22}})}{\partial\tilde{x}^1} - \frac{\partial(v^1\sqrt{g_{11}})}{\partial\tilde{x}^1}\right)d\tilde{x}^1 d\tilde{x}^2.\end{aligned}$$

Therefore,

$$(\boldsymbol{\nabla}\times\boldsymbol{v})^3 = \frac{1}{\sqrt{g_{11}g_{22}}}\left(\frac{\partial(v^2\sqrt{g_{22}})}{\partial\tilde{x}^1} - \frac{\partial(v^1\sqrt{g_{11}})}{\partial\tilde{x}^1}\right).$$

We can determine the other components in a similar way, obtaining

$$\boldsymbol{\nabla}\times\boldsymbol{v} = \begin{vmatrix}\sqrt{g_{11}}\hat{\boldsymbol{e}}_1 & \sqrt{g_{22}}\hat{\boldsymbol{e}}_2 & \sqrt{g_{33}}\hat{\boldsymbol{e}}_3 \\ \frac{\partial}{\partial\tilde{x}^1} & \frac{\partial}{\partial\tilde{x}^2} & \frac{\partial}{\partial\tilde{x}^3} \\ v^1\sqrt{g_{11}} & v^2\sqrt{g_{22}} & v^3\sqrt{g_{33}}\end{vmatrix}.$$