The Jacobian, Tensors, and an Introduction to Nonlinear Transformations

Cody Vig

February 14, 2022

1 Review of the Change of Basis Matrix

Suppose we have bases $\beta := \{\beta_1, \dots, \beta_n\}$ and $\bar{\beta} := \{\bar{\beta}_1, \dots, \bar{\beta}_n\}$ for an n-dimensional vector space $\mathbf{V}$. If we have a vector $v \in \mathbf{V}$ and we know its decomposition in the β-basis i.e., we know $v = \sum_{i=1}^n x_i \beta_i$, then we can compute its expansion in the $\bar{\beta}$-basis if we can compute the change of basis matrix $Q = [\mathbf{I_V}]_\beta^{\bar{\beta}}$. In particular, we define Q so that

$$\bar{x}_i = \sum_{j=1}^n Q_{ij} x_j.$$

In matrix notation, $[v]_{\bar{\beta}} = Q[v]_\beta$. We usually compute this matrix by recognizing that $[v]_{\bar{\beta}} = [\mathbf{I_V}(v)]_{\bar{\beta}} = [\mathbf{I_V}]_\beta^{\bar{\beta}}[v]_\beta \equiv Q[v]_\beta$ implies that the jth column of Q is $[\mathbf{I_V}(\beta_j)]_{\bar{\beta}}$. That is, the n^2 numbers Q_{ij} satisfy

$$\beta_j = \sum_{i=1}^n Q_{ij} \bar{\beta}_i.$$

Note the asymmetry here! Q maps *the coordinates* of β into *the coordinates* of $\bar{\beta}$, but Q maps the *basis* $\bar{\beta}$ into the *basis* β:

$$\bar{x}_i = \sum_{j=1}^n Q_{ij} x_j \tag{1a}$$

$$\beta_j = \sum_{i=1}^n Q_{ij} \bar{\beta}_i. \tag{1b}$$

It is useful to have expressions for $\bar{x}_i$ in terms of the $\{x_i\}$ and for $\bar{\beta}_i$ in terms of the $\{\beta_i\}$. To accomplish this, we use a standard trick in differential geometry called *index contraction*. Effectively, we "multiply" both side of (1b) by Q_{jk}^{-1} and sum over j. We get:

$$\sum_{j=1}^n Q_{jk}^{-1} \beta_j = \sum_{j=1}^n Q_{jk}^{-1} \sum_{i=1}^n Q_{ij} \bar{\beta}_i = \sum_{i=1}^n \left\{ \sum_{j=1}^n Q_{ij} Q_{jk}^{-1} \right\} \bar{\beta}_i = \sum_{i=1}^n \delta_{ik} \bar{\beta}_i = \bar{\beta}_k.$$

In the penultimate step, I introduced the notation δ_{ij} to denote the $(i, j)^{\text{th}}$ element of the identity $\mathbf{I}_n$. This symbol is called the *Kronecker delta*, and it equals 1 if $i = j$ and zero otherwise. This gives us the following relations:

$$\bar{x}_i = \sum_{j=1}^n Q_{ij} x_j \tag{2a}$$

$$\bar{\beta}_j = \sum_{i=1}^{n} Q^{-1}_{ij}\beta_i. \tag{2b}$$

The next section is an aside, as it is not strictly relevant, but it gives an introduction to objects called *tensors* and the mathematical notation used in differential geometry.

1.1 Contravariant and Covariant Objects

The fact that coordinates and vectors transform oppositely under a change of basis is purely because the object they are describing should be invariant under that change of basis. For instance, if v is an arrow in the plane (i.e., a representation of a vector in $\mathbb{R}^2$) and we swap bases for the space, the vector itself did not change, our perspective did. In particular, it should certainly be the case that $\sum_{i=1}^n x_i\beta_i = v = \sum_{i=1}^n \bar{x}_i\bar{\beta}_i$, or that $0 = \sum_{i=1}^n x_i\beta_i - \sum_{i=1}^n \bar{x}_i\bar{\beta}_i$, and indeed,

$$\sum_{\nu=1}^{n} x_\nu\beta_\nu - \sum_{\mu=1}^{n} \bar{x}_\mu\bar{\beta}_\mu = \sum_{\nu=1}^{n} x_\nu \sum_{\mu=1}^{n} Q_{\mu\nu}\bar{\beta}_\mu - \sum_{\mu=1}^{n}\sum_{\nu=1}^{n} Q_{\mu\nu}x_\nu\bar{\beta}_\mu = \sum_{\mu=1}^{n}\sum_{\nu=1}^{n}\{Q_{\mu\nu}x_\nu - Q_{\mu\nu}x_\nu\}\bar{\beta}_\mu = 0,$$

where I used Equations (2) in the second equality. This is why coordinates and bases must change oppositely under a change of basis. This leads naturally to the notion of *covariant* and *contravariant* objects. This motivates the following (informal) definitions.

Definition 1. A vector is called *covariant* if its components transform in accordance to (2b). We write these components with subscript, e.g., x_i.

Definition 2. A vector is called *contravariant* if its components transform in accordance to (2a). We write these components with superscript, e.g., x^i.[1]

All of the vectors we have seen so far have been contravariant. The reason for this is that "covariant vectors" are actually not vectors in $\mathbf{V}$, but dual vectors in $\mathbf{V}^*$, the dual space to $\mathbf{V}$. We will see examples of covariant vectors when we study dual spaces, so I might append this document later. If we assign[2] a single covariant and a single contravariant index to a matrix so that the elements of Q are $Q^i{}_j$, then (2) becomes:

$$\bar{x}^i = \sum_{j=1}^{n} Q^i{}_j x^j \qquad \equiv Q^i{}_j x^j \tag{3a}$$

$$\bar{\beta}_j = \sum_{i=1}^{n} (Q^{-1})^i{}_j\beta_i \equiv (Q^{-1})^i{}_j\beta_i \tag{3b}$$

In the last step of each equality, I have stopped writing the sum. This is a convention in differential geometry called the *Einstein summation convention.* It turns out that any time a sum is performed in linear algebra, it's always over a repeated covariant and contravariant index, e.g., $\bar{x}^i = Q^i{}_j x^j$. Since the sum will always be over the set $i \in \{1, \dots, n\}$, there is no loss of clarity in just *not* writing the sum. This convention may seem lazy now, but it will prove to be immensely helpful in the following definition.

[1] This is *not* an exponent.

[2] This is *not* a definition. This is because matrices are constructed from a tensor product between (contravariant) vectors and (covariant) dual vectors, i.e., the set $\{e^i \otimes e_j\}$ for $i = 1, \dots, m$, $j = 1, \dots, n$, is the standard basis for $M_{m\times n}(\mathbb{F})$. We will discuss this more when we talk about dual spaces.

Definition 3. An object is said to be an *(r,s)-tensor*, or a *type (r, s) tensor*, if it has r covariant indices and s contravariant indices, each of which transform in accordance to (2). We write such an object like $T^{i_1,\ldots,i_s}_{j_1,\ldots,j_r}$.

It helps to explicitly write out the transformation law. If each of the m covariant indices transform with (2b) and each of the n contravariant indices transform with (2a), the tensor transforms like

$$\bar{T}^{i'_1,\ldots,i'_s}_{j'_1,\ldots,j'_r} = \sum_{i_1=1}^{n} Q^{i'_1}{}_{i_1} \cdots \sum_{i_n=1}^{n} Q^{i'_s}{}_{i_s} \sum_{j_1=1}^{n} (Q^{-1})^{j_1}{}_{j'_1} \cdots \sum_{j_n=1}^{n} (Q^{-1})^{j_r}{}_{j'_r} T^{i_1,\ldots,i_r}_{j_1,\ldots,j_s}$$
$$= \prod_{p=1}^{s} \sum_{i_p=1}^{n} Q^{i'_p}{}_{i_p} \prod_{q=1}^{r} \sum_{j_q=1}^{n} (Q^{-1})^{j_q}{}_{j'_q} T^{i_1,\ldots,i_s}_{j_1,\ldots,j_r}.$$

Using the Einstein notation defined in (3), the transformation law becomes:

$$\bar{T}^{i'_1,\ldots,i'_s}_{j'_1,\ldots,j'_r} = Q^{i'_1}{}_{i_1} \cdots Q^{i'_s}{}_{i_s} (Q^{-1})^{j_1}{}_{j'_1} \cdots (Q^{-1})^{j_r}{}_{j'_r} T^{i_1,\ldots,i_s}_{j_1,\ldots,j_r}. = \prod_{p=1}^{s} Q^{i'_p}{}_{i_p} \prod_{q=1}^{r} (Q^{-1})^{j_q}{}_{j'_q} T^{i_1,\ldots,i_s}_{j_1,\ldots,j_r}$$

If you've ever talked to a physicist, they'll usually write something like

$$\bar{T}^{\rho}_{\mu\nu} = \frac{\partial \bar{x}^{\rho}}{\partial x^k} \frac{\partial x^i}{\partial \bar{x}^{\mu}} \frac{\partial x^j}{\partial \bar{x}^{\mu}} T^k_{ij}.$$

In the next section, we will see why $Q^i{}_j = \frac{\partial \bar{x}^i}{\partial x^j}$, and as a result, why $(Q^{-1})^i{}_j = \frac{\partial x^i}{\partial \bar{x}^j}$. This is the notion of *the Jacobian*. Since this section is optional, I will stop using Einstein notation and covariant and contravariant indices in what follows.

2 The Jacobian

In linear algebra, we usually construct Q by writing the elements of one basis in terms of the other. It is occasionally useful, however, to write Q in terms of the coordinates themselves. If we differentiate (2a) with respect to x_k, we get

$$\frac{\partial \bar{x}_i}{\partial x_k} = \sum_{j=1}^{n} \frac{\partial}{\partial x_k} \{Q_{ij} x_j\} = \sum_{j=1}^{n} \left[\frac{\partial Q_{ij}}{\partial x_k} x_j + Q_{ij} \frac{\partial x_j}{\partial x_k} \right].$$

If the transformation is linear, then Q_{ij} does not depend on the x_k's and so the derivative vanishes. Also, $\partial x_j / \partial x_k = \delta_{jk}$, the Kronecker delta defined above. Then:

$$\frac{\partial \bar{x}_i}{\partial x_k} = \sum_{j=1}^{n} \left[\frac{\partial Q_{ij}}{\partial x_k} x_j + Q_{ij} \frac{\partial x_j}{\partial x_k} \right] = \sum_{j=1}^{n} Q_{ij} \delta_{jk} = Q_{ik}.$$

This gives us the identity

$$Q_{ij} = \frac{\partial \bar{x}_i}{\partial x_j}; \tag{4a}$$

that is, the j^{th} column of Q is the derivative of the old coordinate $\{\bar{x}_i\}$ with respect to the new coordinate x_j. It follows that

$$Q^{-1}_{ij} = \frac{\partial x_i}{\partial \bar{x}_j}. \tag{4b}$$

One way to verify (4a) is to perform an index contraction on (2a). We "multiply" both sides of (2a) by Q^{-1}_{ki} and sum over i to get

$$\sum_{i=1}^{n} Q^{-1}_{ki}\bar{x}_i = \sum_{i=1}^{n} Q^{-1}_{ki}\sum_{j=1}^{n} Q_{ij}x_j = \sum_{j=1}^{n}\left\{\sum_{i=1}^{n} Q^{-1}_{ki}Q_{ij}\right\}x_j = \sum_{j=1}^{n}\delta_{kj}x_j = x_k.$$

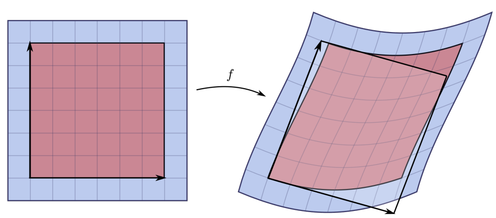
Hence $x_i = \sum_{j=1}^{n} Q^{-1}_{ij}\bar{x}_j$, and so $\frac{\partial x_i}{\partial \bar{x}_k} = \sum_{j=1}^{n} Q^{-1}_{ij}\frac{\partial \bar{x}_j}{\partial \bar{x}_k} = \sum_{j=1}^{n} Q^{-1}_{ij}\delta_{jk} = Q^{-1}_{ij}$, as expected. For the sake of completeness, this means that the transformation matrix Q from (2) is given by

$$Q = [\mathbf{I_V}]_{\bar{\beta}}^{\beta} = \begin{pmatrix} \frac{\partial \bar{x}_1}{\partial x_1} & \cdots & \frac{\partial \bar{x}_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial \bar{x}_n}{\partial x_1} & \cdots & \frac{\partial \bar{x}_n}{\partial x_n} \end{pmatrix}$$

In this notation, this matrix is called *the Jacobian* of the transformation. The determinant of this matrix is called *The Jacobain determinant* and is usually denoted by $\frac{\partial(\bar{x}_1,\ldots,\bar{x}_n)}{\partial(x_1,\ldots,x_n)}$.

3 Nonlinear Transformations and the Jacobian Determinant

Up to this point, we've exclusively studied *linear* transformations, but this is a rather limited class of functions. It turns out, however, that the methods developed in the study of linear maps are still useful in the study of nonlinear maps. One such exemplifying example of this idea is the interpretation of the Jacobian determinant. Suppose we have a (continuous) nonlinear map $f : \mathbb{R}^2 \to \mathbb{R}^2$ which sends $(u, v) \mapsto (x(u, v), y(u, v))$. We are interested in studying how a small area $\Delta u\Delta v$ changes under this transformation.



In the figure, the left represents the domain. Call the bottom-left corner $A : (u, v)$, the bottom right $B : (u + \Delta u, v)$, and the top left $C : (u, v + \Delta v)$, so that the area in red is $A_{uv} = \Delta u\Delta v$. The image of these points (on the right) are

$$\begin{aligned} &f(A) : (x(u, v), y(u, v)),\\ &f(B) : (x(u + \Delta u, v), y(u + \Delta u, v)),\\ &f(C) : (x(u, v + \Delta v), y(u, v + \Delta v)). \end{aligned}$$

We would like to calculate the red area A_{xy} on the right, but since the map is nonlinear, it is not obvious how to do so. Instead, we can approximate the area A_{xy} using the area of the parallelogram generated by the vectors $\overrightarrow{f(A)f(B)}$ and $\overrightarrow{f(A)f(C)}$. We compute these vectors as follows:

$$\begin{aligned} \overrightarrow{f(A)f(B)} &= \overrightarrow{f(B)} - \overrightarrow{f(A)} = \begin{pmatrix} x(u+\Delta u, v) \\ y(u+\Delta u, v)\end{pmatrix} - \begin{pmatrix} x(u, v) \\ y(u, v)\end{pmatrix} \approx \begin{pmatrix} \frac{\partial x}{\partial u} \\ \frac{\partial y}{\partial u}\end{pmatrix}\Delta u,\\ \overrightarrow{f(A)f(C)} &= \overrightarrow{f(C)} - \overrightarrow{f(A)} = \begin{pmatrix} x(u, v+\Delta v) \\ y(u, v+\Delta v)\end{pmatrix} - \begin{pmatrix} x(u, v) \\ y(u, v)\end{pmatrix} \approx \begin{pmatrix} \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial v}\end{pmatrix}\Delta v. \end{aligned}$$

The area of this parallelogram is the magnitude of the cross product of these vectors. Indeed,

$$A_{xy} \approx \left\| \overrightarrow{f(A)f(B)} \times \overrightarrow{f(A)f(C)} \right\| \approx \left\| \begin{pmatrix} \frac{\partial x}{\partial u} \\ \frac{\partial y}{\partial u} \end{pmatrix} \Delta u \times \begin{pmatrix} \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial v} \end{pmatrix} \Delta v \right\| = \left| \det \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} \end{pmatrix} \right| A_{uv}.$$

The matrix whose determinant is being calculated is the Jacobian of the transformation $\frac{\partial(x,y)}{\partial(u,v)}$. The discrepancy between the area we calculated and the true area A_{xy} vanishes as Δu and Δv shrink, which gives us the differential relation

$$\mathrm{d}A_{xy} = \left| \frac{\partial(x,y)}{\partial(u,v)} \right| \mathrm{d}A_{uv}.$$

This result holds in higher dimensions, too. We see that the Jacobian determinant gives the factor by which areas change under a transformation, which was the same property the determinant of the matrix representation of a linear transformation had. Indeed, the Jacobian is a matrix representation of a special type of linear transformation. It is the best linear approximation to f, in the same way that $L(x) := g(x_0) + g'(x_0)(x - x_0)$ is the best linear approximation to $g(x)$ at x_0. We can see this in the figure above, too. The columns of the Jacobian are vectors which are tangent to the images of $\overrightarrow{AB}$ and $\overrightarrow{AC}$ at the point $f(A)$, respectively.