2. Differentiation as best Linear Approximation

We can use linear algebra tools to study smooth functions, because smooth functions are almost linear in small regions. A feature of expressing vector valued linear functions in terms of matrices is that function composition is expressed as matrix multiplication (via the associative law). Below, we show how this feature propagates as the chain rule for derivatives of smooth functions.

The Taylor series shows exactly how to use the derivative to express the best linear approximation to the change in function value.

For a function of a single variable:

$$f(x)-f(x_0)=f'(x_0)\Delta x \ + \ \text{terms with higher derivatives}$$

For a scalar function of a vector:

$$f(x,y)-f(x_0,y_0)=\left(\frac{\partial f}{\partial x},\frac{\partial f}{\partial y}\right)\begin{pmatrix}\Delta x\\ \Delta y\end{pmatrix} \ + \ \text{terms with higher derivatives}$$

For a vector function of a vector, we use the Jacobian matrix:

$$\vec{f}(x,y)-\vec{f}(x_0,y_0)=\begin{bmatrix}\frac{\partial f_x}{\partial x} & \frac{\partial f_x}{\partial y}\\ \frac{\partial f_y}{\partial x} & \frac{\partial f_y}{\partial y}\end{bmatrix}\begin{pmatrix}\Delta x\\ \Delta y\end{pmatrix} \ + \ \text{terms with higher derivatives}$$

The chain rule results in matrix multiplication because the best linear approximation to composed functions is the composition of the best linear approximations of the individual functions.

In the following examples, all the functions involved are simple. They are identified by a constant plus its constant derivative. No higher derivatives.

Noting that derivatives of constants are zero, the single variable chain rule can be demonstrated as:

$$\left.\frac{d}{dt}f(u(t))\right|_{t_0}=\left.\frac{d}{dt}\left(f_0+\frac{df}{du}\Delta u\right)\right|_{t_0}=\left.\frac{d}{dt}\left(f_0+\frac{df}{du}\left(\frac{du}{dt}\ \Delta t\right)\right)\right|_{t_0}=\left.\frac{df}{du}\frac{du}{dt}\right|_{t_0}$$

A similar demonstration with multiple variables becomes:

$$\frac{d}{dt}\vec{f}(\vec{u}(t))\Big|_{t_0} = \frac{d}{dt}\left(\vec{f}_0 + \begin{bmatrix} \frac{\partial f_1}{\partial u_1} & \frac{\partial f_1}{\partial u_2} \\ \frac{\partial f_2}{\partial u_1} & \frac{\partial f_2}{\partial u_2} \end{bmatrix}\begin{pmatrix} \Delta u_1 \\ \Delta u_2 \end{pmatrix}\right)\Bigg|_{t_0} = \frac{d}{dt}\left(\vec{f}_0 + \begin{bmatrix} \frac{\partial f_1}{\partial u_1} & \frac{\partial f_1}{\partial u_2} \\ \frac{\partial f_2}{\partial u_1} & \frac{\partial f_2}{\partial u_2} \end{bmatrix}\begin{pmatrix} \frac{du_1}{dt}\Delta t \\ \frac{du_2}{dt}\Delta t \end{pmatrix}\right)\Bigg|_{t_0}$$

$$= \begin{bmatrix} \frac{\partial f_1}{\partial u_1} & \frac{\partial f_1}{\partial u_2} \\ \frac{\partial f_2}{\partial u_1} & \frac{\partial f_2}{\partial u_2} \end{bmatrix}\begin{pmatrix} \frac{du_1}{dt} \\ \frac{du_2}{dt} \end{pmatrix}\Bigg|_{t_0}$$

Similarly $\frac{d}{dt}\vec{f}(\vec{x}(\vec{u}(t))) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{bmatrix}\begin{bmatrix} \frac{\partial x_1}{\partial u_1} & \frac{\partial x_1}{\partial u_2} \\ \frac{\partial x_2}{\partial u_1} & \frac{\partial x_2}{\partial u_2} \end{bmatrix}\begin{pmatrix} \frac{du_1}{dt} \\ \frac{du_2}{dt} \end{pmatrix}$

Einstein Summation Notation

Linear algebra involves expressing sums of products. The Einstein summation convention is a compressed notation useful for expressing sums. It is characterized by the use of both superscripts and subscripts together with the rule that repeated indices are summed.

A column vector is represented with superscripts: $x^i \approx \begin{pmatrix} x^1 \\ x^2 \end{pmatrix}$

A row vector (functional, co-vector) is represented with subscripts
$a_i \approx (a_{1,}a_2)$

The summation convention applied to a functional acting on a vector is denoted as: $a_k x^k \approx \sum_{k=1}^{n} a_k x^k$ ($\approx$ so that you don't think this a recursive statement)

A dot product of vectors is denoted as $x^i y^i \approx \sum_{i=1}^{n} x^i y^i$

To express matrix multiplication:

$$C^i_j = A^i_k B^k_j \quad \text{means} \quad C^i_j \approx \sum_{k=1}^{n} A^i_k B^k_j \quad \text{for any i and j.}$$

This example shows free indices i and j and the dummy index k.

Unlike matrix multiplication, the order of multiplication can change without affecting the value. The positions of the dummy indices identifies the order of multiplication and addition operations for non-commutative matrices.

For example $(A\,B)^i_j = A^i_k B^k_j$ and $(B\,A)^i_k = B^i_k A^k_j = A^k_j B^i_k$

We will want to convert to other bases like $\vec{u}_1, \vec{u}_{2,} \ldots = u^i_1, u^i_2, \ldots = u^i_k$ and use the associated matrix multiplication. If we want to transform a vector, we write

$$x^i = u^i_k x'^k \quad \text{Instead of} \quad \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} = \begin{bmatrix} u^1_1 & u^1_2 \\ u^2_1 & u^2_2 \end{bmatrix} \begin{pmatrix} x'^1 \\ x'^2 \end{pmatrix}$$

Transforming a Dot Product

We have a systematic way of writing the dot product of two vectors in transformed coordinates so that the result is unchanged:

$$x^i y^i = \left(x'^m u^i_m\right)\left(y'^n u^i_n\right) = x'^m \left(u^i_m u^i_n\right) y'^n = x'^m g_{mn} y'^n$$

where

$$g_{11} = u^1_1 u^1_1 + u^2_1 u^2_1 = \vec{u}_1 \cdot \vec{u}_1 \qquad g_{12} = u^1_1 u^1_2 + u^2_1 u^2_2 = \vec{u}_1 \cdot \vec{u}_2$$
$$g_{21} = u^1_2 u^1_1 + u^2_2 u^2_1 = \vec{u}_2 \cdot \vec{u}_1 \qquad g_{22} = u^1_2 u^1_2 + u^2_2 u^2_2 = \vec{u}_2 \cdot \vec{u}_2$$

The matrix g_{ij} is the array of all pairwise dot products of the alternative basis elements (not necessarily orthonormal). We will see g_{ij} again when we describe a Riemannian metric.

Einstein Notation and Derivatives

Partial derivative arrays are easily expressed with Einstein notation. Expressing the chain rule is especially convenient:

$$\frac{d}{dt} \vec{f}\left(\vec{x}\left(\vec{u}(t)\right)\right) = \frac{\partial f^i}{\partial x^j} \frac{\partial x^j}{\partial u^k} \frac{d\,u^k}{d\,t}$$

Note that an upper index in the denominator is a lower index.

Here we compare matrix and Einstein notation for expressing the change in value of a scalar valued function using Taylor series:

$$f(x, y) - f(x_{0,} y_0)$$

$$= \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)\begin{pmatrix}\Delta x \\ \Delta y\end{pmatrix} + \frac{1}{2} (\Delta x, \Delta y)\begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}\begin{pmatrix}\Delta x \\ \Delta y\end{pmatrix} + \frac{1}{6} ? + \cdots$$

$$f(x^i) - f(x_0^i) = \frac{\partial f}{\partial x^i} \Delta x^i + \frac{1}{2} \Delta x^i \frac{\partial^2 f}{\partial x^i \partial x^j} \Delta x^j + \frac{1}{6} \frac{\partial^3 f}{\partial x^i \partial x^j \partial x^k} \Delta x^i \Delta x^j \Delta x^k + \cdots$$

Notice that the summation loop structure is clearer in the repeated index notation when compared to matrix notation.

The Road Ahead

We will use the simplicity of linear math to study the complications of surfaces having multiple variables. The linear math tools are valid because a smooth surface is locally linear. Whenever we speak of functions or vector fields, we will always assume differentiability and hence local linearity.

Most of our analysis of a surface will involve working with the tangent planes in a region. Each tangent plane is a linear approximation to the surface. We will examine methods for working with the tangent planes.