9 The derivative as a linear transformation

9.1 Redefining the derivative

Matrices appear in many different contexts in mathematics, not just when we need to solve a system of linear equations. An important instance is linear approximation. Recall from your calculus course that a differentiable function f can be expanded about any point a in its domain using Taylor's theorem. We can write

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(c)}{2!}(x-a)^2,$$

where c is some point between x and a. The **remainder** term $\frac{f''(c)}{2!}(x-a)^2$ is the "error" made by using the linear approximation to f at $x = a$,

$$f(x) \approx f(a) + f'(a)(x-a).$$

That is, $f(x)$ minus the approximation is exactly equal to the error (remainder) term. In fact, we can write Taylor's theorem in the more suggestive form

$$f(x) = f(a) + f'(a)(x-a) + \epsilon(x,a),$$

where the remainder term has now been renamed the **error term** $\epsilon(x,a)$ and has the important property

$$\lim_{x\to a} \frac{\epsilon(x,a)}{x-a} = 0.$$

(The existence of this limit is another way of saying that the error term "looks like" $(x-a)^2$.)

This observation gives us an alternative (and in fact, much better) definition of the derivative:

Definition: The real-valued function f is said to be **differentiable** at $x = a$ if there exists a number A and a function $\epsilon(x,a)$ such that

$$f(x) = f(a) + A(x-a) + \epsilon(x,a),$$

where

$$\lim_{x\to a} \frac{\epsilon(x,a)}{x-a} = 0.$$

Remark: the error term $\epsilon = \frac{f''(c)}{2}(x-a)^2$ clearly depends on a, and it depends on x as well since the number c varies with x.

Theorem: This is equivalent to the usual calculus definition.

Proof: If the new definition holds, then if we compute $f'(x)$ by the usual definition, we find

$$\lim_{x\to a} \frac{f(x)-f(a)}{x-a} = A + \lim_{x\to a} \frac{\epsilon(x,a)}{x-a} = A + 0 = A,$$

and $A = f'(a)$ according to the standard definition. Conversely, if the standard definition of differentiability holds, then we can *define* $\epsilon(x,a)$ to be the error made in the linear approximation:

$$\epsilon(x,a) = f(x) - f(a) - f'(a)(x-a).$$

Then

$$\lim_{x\to a} \frac{\epsilon(x,a)}{x-a} = \lim_{x\to a} \frac{f(x)-f(a)}{x-a} - f'(a) = f'(a) - f'(a) = 0,$$

so f can be written in the new form, with $A = f'(a)$.

Example: Let $f(x) = 4 + 2x - x^2$, and let $a = 2$. So $f(a) = f(2) = 4,$ and $f'(a) = f'(2) = 2 - 2a = -2.$. Now subtract $f(2) + f'(2)(x-2)$ from $f(x)$ to get

$$4 + 2x - x^2 - (4 - 2(x-2)) = -4 + 4x - x^2 = -(x-2)^2.$$

This is the error term, which is quadratic in $x-2$, as advertised. So $8-2x (= f(2)+f'(2)(x-2))$ is the correct linear approximation to f at $x = 2$.

Suppose we try some other linear approximation - for example, we could try $f(2)-4(x-2) = 12-4x$. Subtracting this from $f(x)$ gives $-8+6x-x^2 = -2(x-2)-(x-2)^2$, which is our new error term. But this won't work, since

$$\lim_{x\to 2} \frac{-2(x-2)-(x-2)^2}{(x-2)} = -2,$$

which is clearly not 0. The only "linear approximation" that leaves a quadratic remainder as the error term is the one formed in the usual way, using the derivative.

Exercise: Interpret this geometrically in terms of the slope of various lines passing through the point $(2, f(2))$.

9.2 Generalization to higher dimensions

Our new definition of derivative is the one which generalizes to higher dimensions. We start with an

Example: Consider a function from R^2 to R^2, say

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} u(x,y) \\ v(x,y) \end{pmatrix} = \begin{pmatrix} 2 + x + 4y + 4x^2 + 5xy - y^2 \\ 1 - x + 2y - 2x^2 + 3xy + y^2 \end{pmatrix}$$

By inspection, as it were, we can separate the right hand side into three parts. We have

$$\mathbf{f}(\mathbf{0}) = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

and the linear part of $\mathbf{f}$ is the vector

$$\begin{pmatrix} x + 4y \\ -x + 2y \end{pmatrix},$$

which can be written in matrix form as

$$A\mathbf{x} = \begin{pmatrix} 1 & 4 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

By analogy with the one-dimensional case, we might guess that

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{0}) + A\mathbf{x} + \text{ an error term of order 2 in } x,\ y.$$

where A is the matrix

$$A = \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix}(0,0).$$

And this suggests the following

Definition: A function $f : R^n \to R^m$ is said to be **differentiable** at the point $\mathbf{x} = \mathbf{a} \in R^n$ if there exists an $m \times n$ matrix A and a function $\epsilon(\mathbf{x}, \mathbf{a})$ such that

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{a}) + A(\mathbf{x} - \mathbf{a}) + \epsilon(\mathbf{x}, \mathbf{a}),$$

where

$$\lim_{\mathbf{x} \to \mathbf{a}} \frac{\epsilon(\mathbf{x}, \mathbf{a})}{||\mathbf{x} - \mathbf{a}||} = \mathbf{0}.$$

The matrix A is called the **derivative of f at $\mathbf{x} = \mathbf{a}$**, and is denoted by $D\mathbf{f}(\mathbf{a})$.

Generalizing the one-dimensional case, it can be shown that if

$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} u_1(\mathbf{x}) \\ \vdots \\ u_m(\mathbf{x}) \end{pmatrix},$$

is differentiable at $\mathbf{x} = \mathbf{a}$, then the derivative of $\mathbf{f}$ is given by the $m \times n$ matrix of partial derivatives

$$D\mathbf{f}(\mathbf{a}) = \begin{pmatrix} \dfrac{\partial u_1}{\partial x_1} & \cdots & \dfrac{\partial u_1}{\partial x_n} \\ \vdots & \vdots & \vdots \\ \dfrac{\partial u_m}{\partial x_1} & \cdots & \dfrac{\partial u_m}{\partial x_n} \end{pmatrix}_{m \times n} (\mathbf{a}).$$

Conversely, if all the indicated partial derivatives exist and are continuous at $\mathbf{x} = \mathbf{a}$, then the approximation

$$\mathbf{f}(x) \approx \mathbf{f}(\mathbf{a}) + D\mathbf{f}(\mathbf{a})(\mathbf{x} - \mathbf{a})$$

is accurate to the second order in $\mathbf{x} - \mathbf{a}$.

Exercise: Find the derivative of the function $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^3$ at $\mathbf{a} = (1, 2)^t$, where

$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} (x+y)^3 \\ x^2 y^3 \\ y/x \end{pmatrix}$$