Math 225 class notes

Elizabeth Denne
Last modified 2/26/09

1 Derivatives

Definition 1. A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $x = a$ if there is a linear map $\lambda : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{h \to 0} \frac{||f(a+h) - f(a) - \lambda(h)||}{||h||} = 0.$$

The linear map λ is often denoted $f'(a)$ and the matrix which represents the linear map is often denoted $Df(a)$.

So $f'(a) : \mathbb{R}^n \to \mathbb{R}^m$ is $f'(a)(h) = f'(a)(h_1, h_2, \dots, h_n)$, or $Df(a)$ is an $m \times n$ matrix.

Some derivatives we know

1. The **constant map** $f : \mathbb{R}^n \to \mathbb{R}^m$ defined by $f(x) = c$ a constant. Then derivative is the zero map, $f'(a)(h) = 0(h) = 0$ or $Df(a)(h) = 0$, the $m \times n$ zero matrix.

2. If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a **linear map**, then $f'(a) = T$, or $DF(a)$ **is** the matrix of T. (Remember a linear map is one where $T(x+y) = T(x) + T(y)$ and $T(kx) = kT(x)$ for a real number k.) Here are some exmples:

   (a) The **identity map** $I : \mathbb{R}^n \to \mathbb{R}^n$ is $I(x) = x$. So $I'(a) = I$ or $I'(a)(h) = I(h) = h$ and so $DI(a) = I_n$ the $n \times n$ identity matrix, or

   $$DI(a)(h) = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix} = \begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix}.$$

   (b) The i**th projection map** $\pi^i : \mathbb{R}^n \to \mathbb{R}$ defined by $\pi^1(x_1, \dots, x_n) = x_i$. Then $(\pi^i)'(a) = \pi^i$ or $(\pi^i)'(a)(h) = \pi^i(h) = h^i$ and (putting a 1 in the ith column of the $1 \times n$ matrix $D\pi^i(a)$)

   $$D\pi^i(a)(h) = \begin{bmatrix} 0 & \dots & 1 & \dots & 0 \end{bmatrix} \begin{bmatrix} h^1 \\ \vdots \\ h^n \end{bmatrix} = h^i.$$

(c) The **sum map** $s : \mathbb{R}^2 \to \mathbb{R}$ is $s(x, y) = x+y$. Then $s'(a, b) = s$ or $s'(a, b)(h, k) = s(h, k) = h + k$ and $Ds(a, b) \begin{bmatrix} h \\ k \end{bmatrix} = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} h \\ k \end{bmatrix} = h + k$.

3. The **product map** $p : \mathbb{R}^2 \to \mathbb{R}$ defined by $p(x, y) = xy$. Then $p'(a, b)(h, k) = bh+ak$ or $Dp(a, b)(h, k) = \begin{bmatrix} b & a \end{bmatrix} \begin{bmatrix} h \\ k \end{bmatrix} = bh + ak$.

2 More Derivatives

1. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be written in terms of its component functions, namely

$$f(x_1, \ldots x_n) = (f^1(x_1, \ldots x_n), f^2(x_1 \ldots, x_n), \ldots, f^m(x_1 \ldots, x_n)).$$

We proved in class that we can find $Df(a)$ in terms of $Df^i(a)$. How do we do this? (Don't forget we need either f to be differentiable at a or each f^i $(i = 1, 2 \ldots, m)$ to be differentiable at a.)

$Df(a) : \mathbb{R}^n \to \mathbb{R}^m$ is defined by $Df(a) = (Df^1(a), Df^2(a), \ldots, Df^m(a))$, where $Df^i(a) : \mathbb{R}^n \to \mathbb{R}$. Another way to write this is
$Df(a)(h) = (Df^1(a)(h), Df^2(a)(h), \ldots, Df^m(a)(h))$, yet another way to see this is $Df(a)(h^1, \ldots, h^n) = (y_1, \ldots, y_m)$ where each $y_i = Df^i(a)(h_1, \ldots, h_n)$ $i = 1, \ldots m$.

2. The **Chain Rule.** Given $f : \mathbb{R}^n \to \mathbb{R}^m$ differentiable at a and $g : \mathbb{R}^m \to \mathbb{R}^p$ differentiable at $f(a)$, then $g \circ f : \mathbb{R}^n \to \mathbb{R}^p$ is differentiable at a and the derivative $(g \circ f)'(a) : \mathbb{R}^n \to \mathbb{R}^p$ is $(g \circ f)'(a)(h) = [g'(f(a))) \circ f'(a)](h) = g'(f(a))(f'(a)(h))$. In other words, we take $h \in \mathbb{R}^n$, then compute $f'(a)(h)$. We'll get something in $\mathbb{R}^m$, let's call this $(b_1, \ldots, b_m)$, then we compute $g'(f(a))(b_1, \ldots, b_m)$. So the linear map $(f \circ g)'(a)$ is a composition of the linear map $f'(a)$ with the linear map $g'(f(a))$.

The linear map $f'(a)$ is represented by the $m \times n$ matrix $Df(a)$ and the linear map $g'(f(a)$ is represented by the $p \times m$ matrix $Dg(f(a))$. So the linear map $(g \circ f)'(a)$ is represented by a $p \times n$ matrix, which can be computed by multiplying the matrices $Dg(f(a))Df(a)$.

3. The **Sum Rule** (or the derivative of the sum is the sum of the derivative). Given $f, g : \mathbb{R}^n \to \mathbb{R}$ both differentiable at a, then $f + g : \mathbb{R}^n \to \mathbb{R}$ is also differentiable at a and $(f + g)'(a) = f'(a) + g'(a)$ or $D(f + g)(a) = Df(a) + Dg(a)$.

We can prove this by using chain rule and the sum map from above in a sneaky way. First recall $s : \mathbb{R}^2 \to \mathbb{R}$ is $s(y, z) = y + z$ and $Ds(a, b) = s$. Second, define $F : \mathbb{R}^n \to \mathbb{R}^2$, by $F(x) = (f(x), g(x))$, or $F = (f, g)$. Observe F has been expressed in terms of component functions as in (1) of this section, and we deduce that $DF(a) = (Df(a), Dg(a))$. Third, note that $f + g = s \circ F$. [Observe $s \circ F(x) = s(F(x)) = s(f(x), g(x)) = f(x) + g(x) = (f + g)(x)$.] Finally, put all these ideas together to get

$$D(f+g)(a) = D(s \circ F)(a) = Ds(F(a))DF(a) = Ds(f(a), g(a))DF(a)$$

$$Ds(f(a), g(a))DF(a) = s \circ DF(a) = s(Df(a), Dg(a)) = Df(a) + Dg(a)$$

4. The **Product Rule**.

   Given $f, g : \mathbb{R}^n \to \mathbb{R}$ both differentiable at a, then $fg : \mathbb{R}^n \to \mathbb{R}$ is also differentiable at a and $(fg)'(a) = f'(a)g(a) + f(a)g'(a)$ or $D(fg)(a) = f(a)Dg(a) + g(a)Df(a)$. The proof is very similar to the above. We write$fg = p \circ F$, where $F : \mathbb{R}^n \to \mathbb{R}^2$ is defined by $F(x) = (f(x), g(x))$, or $F = (f, g)$ and $p : \mathbb{R}^2 \to \mathbb{R}$ is given by $p(y, z) = yz$. Using the chain rule we see

$$D(fg)(a) = D(p \circ F)(a) = Dp(F(a))DF(a) = Dp(f(a), g(a))DF(a)$$

$$Dp(f(a), g(a))DF(a) = Dp(f(a), g(a))(Df(a), Dg(a)) = g(a)Df(a) + f(a)Dg(a).$$

5. The **Jacobian matrix**.

3 Proof of Chain Rule