Jacobians, Directional Derivatives, and the Chain Rule.

Suppose $f_1, f_2, ... f_q$ are functions of p variables $x_1, ..., x_p$; thus for $1 \leq i \leq q$, $f_i(x_1, x_2, ..., x_p)$ is some real number. For a given point $\mathbf{x} = (x_1, x_2, ..., x_p) \in \mathbf{R}^p$, we can assemble the numbers $f_1(\mathbf{x}), f_2(\mathbf{x}), ..., f_q(\mathbf{x})$ into a q-component column vector $\mathbf{f(x)}$. (We will also write $\mathbf{x}$ as a column vector below.) Thus we obtain a map

$$\mathbf{f} : \mathbf{R}^p \to \mathbf{R}^q.$$

(If one or more of the f_i's is not defined at every point of $\mathbf{R}^p$, we actually get a function whose domain is just a subset of $\mathbf{R}^p$, not all of $\mathbf{R}^p$.) This is an important and sophisticated perspective on which much of advanced calculus is based. For what we do below, it is best to write points in $\mathbf{R}^p$ and $\mathbf{R}^q$ as column vectors, rather than row vectors.

We call $\mathbf{f} : \mathbf{R}^p \to \mathbf{R}^q$ *differentiable* at a point $\mathbf{a} \in \mathbf{R}^p$ if all the partial derivatives $\partial f_i / \partial x_j$ $(1 \leq i \leq q, 1 \leq j \leq p)$ exist at $\mathbf{x}$=$\mathbf{a}$ and are continuous there. (Technically this definition is not quite right, but it will suffice for us.) At each such $\mathbf{a}$, we define a $q \times p$ matrix $J_{\mathbf{f}}(\mathbf{a})$, called the *Jacobian* of $\mathbf{f}$ at $\mathbf{a}$; it is defined by

$$(J_{\mathbf{f}}(\mathbf{a}))_{ij} = \frac{\partial f_i}{\partial x_j}(\mathbf{a}).$$

The phrase "$J_{\mathbf{f}}$ at $\mathbf{a}$" means $J_{\mathbf{f}}(\mathbf{a})$. If $A = J_{\mathbf{f}}(\mathbf{a})$, the associated linear map $L_A : \mathbf{R}^p \to \mathbf{R}^q$ is called the *derivative* (or sometimes the *differential*) of $\mathbf{f}$ at $\mathbf{a}$, and we will write it as

$$D_{\mathbf{a}}\mathbf{f} : \mathbf{R}^p \to \mathbf{R}^q.$$

<u>Exercises.</u>

1. Let $\mathbf{f} : \mathbf{R}^3 \to \mathbf{R}^2$ be defined by $f_1\begin{pmatrix} x \\ y \\ z \end{pmatrix} = x + y + z, \quad f_2\begin{pmatrix} x \\ y \\ z \end{pmatrix} = x^2 + y^3 - z^4.$
(i) Compute $J_{\mathbf{f}}$ at a general point $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbf{R}^3$. (ii) Compute $J_{\mathbf{f}}$ at the point $\begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}$.

2. Let $\mathbf{g} : \mathbf{R} \to \mathbf{R}^3$ be defined by $g_1(t) = 3t^2, \quad g_2(t) = 2t, \quad g_3(t) = \cos(t-1)$. (i) Compute $J_{\mathbf{g}}(t)$ for general t. (ii) Compute $J_{\mathbf{g}}(t)$ at $t = 1$.

3. Let A be a $q \times p$ matrix and let $h = L_A : \mathbf{R}^p \to \mathbf{R}^q$ be the associated linear map. Compute $J_{\mathbf{h}}(\mathbf{a})$ for all $\mathbf{a} \in \mathbf{R}^p$.

***

Suppose $\mathbf{f} : \mathbf{R}^p \to \mathbf{R}^q$ is differentiable at a point $\mathbf{a} \in \mathbf{R}^p$, and that $\mathbf{u} \in \mathbf{R}^p$ is some vector. (In this handout, "points in $\mathbf{R}^p$" are no different from "vectors in $\mathbf{R}^p$"; we're calling $\mathbf{a}$ a point and $\mathbf{u}$ a vector only because of the very different roles that $\mathbf{a}$ and $\mathbf{u}$ play in the next sentence.) The *directional derivative of* $\mathbf{f}$ *at* $\mathbf{a}$ *in the direction* $\mathbf{u}$ is the vector in $\mathbf{R}^q$ given by the matirx product of the $q \times p$ matrix of $J_{\mathbf{f}}(\mathbf{a})$ and the $p \times 1$ matrix (i.e. column vector) $\mathbf{u}$:

$$(D_{\mathbf{a}}\mathbf{f})(\mathbf{u}) = J_{\mathbf{f}}(\mathbf{a})\ \mathbf{u}.$$

Note that *derivative* and *directional derivative*, as defined in this handout, are related but not identical concepts. (Further warning: the notation for directional derivative varies from book to book. Many authors write the above as $(D_{\mathbf{u}}\mathbf{f})(\mathbf{a})$ or $D_{\mathbf{u}}\mathbf{f}(\mathbf{a})$. Some use a lower case d instead of a D. Similarly, many authors write $D\mathbf{f}(\mathbf{a})$ or $d\mathbf{f}(\mathbf{a})$ for the derivative of $\mathbf{f}$ at $\mathbf{a}$.)

Exercises.

4. Check that when $q = 1$ and $p = 2$ or 3 the definition above gives the directional derivative you learned in calculus III. (You may have been told in calc. III that $\mathbf{u}$ should be a unit vector, but in more advanced calculus this restriction is dropped, so that the nature of the derivative as a linear transformation can show through.)

5. Let $\mathbf{f}$ be as in problem 1. Compute the directional derivative of $\mathbf{f}$ at $\begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}$ in the direction $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$.

***

Suppose $\mathbf{g} : \mathbf{R}^s \to \mathbf{R}^p$ and $\mathbf{f} : \mathbf{R}^p \to \mathbf{R}^q$ are functions. Then we can form the composition $\mathbf{f} \circ \mathbf{g} : \mathbf{R}^s \to \mathbf{R}^q$:

$$(\mathbf{f} \circ \mathbf{g})(\mathbf{x}) = \mathbf{f}(\mathbf{g}(\mathbf{x})).$$

The *chain rule* for functions of several variables is a theorem about derivatives of compositions of differentiable functions. In calc. I, you learned the chain rule for the case $p = q = s = 1$. In calc. III, you learned it for values of p, q, s from 1 to 3. These are baby versions of the true chain rule, which works for all p, q, s. The chain rule is most easily expressed in terms of matrix multiplication of Jacobians. Specifically, it says this:

Chain Rule Theorem: *Let* $\mathbf{g} : \mathbf{R}^s \to \mathbf{R}^p$ *be differentiable at* $\mathbf{a} \in \mathbf{R}^s$*, and let* $\mathbf{f} : \mathbf{R}^p \to \mathbf{R}^q$ *be differentiable at* $\mathbf{g}(\mathbf{a}) \in \mathbf{R}^p$*. Then* $\mathbf{f} \circ \mathbf{g} : \mathbf{R}^s \to \mathbf{R}^q$ *is differentiable at* $\mathbf{a}$*, and its Jacobian is given by the matrix product*

$$J_{\mathbf{f}\circ\mathbf{g}}(\mathbf{a}) = J_{\mathbf{f}}(\mathbf{g}(\mathbf{a}))\ J_{\mathbf{g}}(\mathbf{a}). \tag{1}$$

(We will not prove this theorem.) Again note that the matrices involved are of the right size to make the equation above sensible.

Exercises.

6. In the Chain Rule Theorem, suppose $f = L_A$ and $g = L_B$ are the linear maps associated to matrices A, B of the appropriate size. In view of your answer to exercise 3, to what familiar formula does equation (1) reduce in this case?

7. (i) Let **f**,**g** be as in exercises 1 and 2, and let **h**=**f**∘**g**. Without computing the function **h** explicitly, compute its Jacobian at $t = 1$. (ii) Let $h_i(t), 1 \leq i \leq 2$, be the component functions of **h**. From your answer to (i), evaluate dh_1/dt and dh_2/dt at $t = 1$.