# MAT 257, Handout 6: October 17-21, 2011.

## I. Assignment.

Finish reading Chapter 2 of Spivak, rereading earlier sections as necessary. Also, read this handout and fill in some missing details!

## II. Higher derivatives.

**a. (multi)-linear algebra preliminaries**. Let $\mathcal{L}(\mathbb{R}^n;\mathbb{R}^m)$ denote the space of linear maps from $\mathbb{R}^n$ to $\mathbb{R}^m$, and more generally, let $\mathcal{L}^k(\mathbb{R}^n;\mathbb{R}^m)$ denote the space of $k$-linear maps from $(\mathbb{R}^n)^k \to \mathbb{R}^m$. Thus, an element of $\mathcal{L}^k(\mathbb{R}^n;\mathbb{R}^m)$ is a function $\Lambda(\mathbb{R}^n)^k \to \mathbb{R}^m$ such that

$$\begin{aligned}\Lambda(v_1,\ldots,v_{i-1},av_i+bw_i,\ldots,v_{i+1},\ldots,v_k) &= a\Lambda(v_1,\ldots,v_{i-1},v_i,\ldots,v_{i+1},\ldots,v_k)\\ &\quad + b\Lambda(v_1,\ldots,v_{i-1},w_i,\ldots,v_{i+1},\ldots,v_k)\end{aligned}$$

for all vectors $v_1,\ldots,v_k,w_i \in \mathbb{R}^n$, scalars $a,b\in\mathbb{R}$, and $i\in\{1,\ldots,k\}$.

We have seen that there is a natural isomorphism between the space of linear maps $\mathbb{R}^n \to \mathcal{L}(\mathbb{R}^n;\mathbb{R}^m)$ and the space $\mathcal{L}^2(\mathbb{R}^n;\mathbb{R}^m)$ of bilinear maps $\mathbb{R}^n\times\mathbb{R}^n \to \mathbb{R}^m$. This is an instance of a more general fact:

**Lemma 1.** *For $k \geq 2$, is a natural isomorphism between the space of linear maps $\mathbb{R}^n \to \mathcal{L}^{k-1}(\mathbb{R}^n;\mathbb{R}^m)$ and $\mathcal{L}^k(\mathbb{R}^n;\mathbb{R}^m)$.*

*Proof.* Given a linear map $\mathbf{\Lambda} : \mathbb{R}^n \to \mathcal{L}^{k-1}(\mathbb{R}^n;\mathbb{R}^m)$, define $\Lambda \in \mathcal{L}^k(\mathbb{R}^n;\mathbb{R}^m)$ by

$$\Lambda(v_1,\ldots,v_k) = \mathbf{\Lambda}(v_1)(v_2,\ldots,,v_k). \tag{1}$$

The right-hand side denotes the element of $\mathbb{R}^m$ obtained as before in the case $k=2$, that is: $\mathbf{\Lambda}(v_1)$ is by definition an element of $\mathcal{L}^{k-1}(\mathbb{R}^n;\mathbb{R}^m)$ and so we can let it act on the $k-1$ vectors $v_2,\ldots,v_k$ to get an element of $\mathbb{R}^m$. One can then verify that the function $\Lambda : (\mathbb{R}^n)^k \to \mathbb{R}^m$ defined in this way is a $k$-linear map $\mathbb{R}^n\to\mathbb{R}^m$.

Conversely, given $\Lambda \in \mathcal{L}^k(\mathbb{R}^n;\mathbb{R}^m)$, we can use (1) to define $\mathbf{\Lambda} : \mathbb{R}^n \to \mathcal{L}^{k-1}(\mathbb{R}^n;\mathbb{R}^m)$ (where now the left-hand side is given and we are defining the right-hand side). It is then routine to verify that $\mathbf{\Lambda}$ is indeed a linear map $\mathbb{R}^n \to \mathcal{L}^{k-1}(\mathbb{R}^n;\mathbb{R}^m)$. □

In what follows, we will abuse notation somewhat by identifying $\Lambda \in \mathcal{L}^k(\mathbb{R}^n;\mathbb{R}^m)$ with the space of linear maps $\mathbb{R}^n \to \mathcal{L}^{k-1}(\mathbb{R}^n;\mathbb{R}^m)$. So for example, if $\Lambda \in \mathcal{L}^k(\mathbb{R}^n;\mathbb{R}^{n-1})$ and $v_1$ is a vector, we will write $\Lambda(v_1)$ (instead of $\mathbf{\Lambda}(v_1)$, as above, which would be more correct) to denote the element of $\mathcal{L}^{k-1}(\mathbb{R}^n;\mathbb{R}^m)$ defined by (1).

Some norms on $\mathcal{L}^k(\mathbb{R}^n;\mathbb{R}^m)$ include the Euclidean norm

$$|\Lambda| = \left(\sum_{i_1,\ldots,i_k=1}^{n} |\Lambda(e_{i_1},\ldots,e_{i_k})|^2\right)$$

and the operator norm

$$\|\Lambda\| = \sup\{\Lambda(v_1,\ldots,v_k) \ : |v_i| \leq 1 \text{ for all } i \in \{1,\ldots,k\}\}.$$

(You can check that these are norms. *exercise!*) General considerations tell us that there exist constants $0 < c \leq C$ (depending on $k$) such that

$$c|\Lambda| \leq \|\Lambda\| \leq C|\Lambda| \tag{2}$$

for all $\Lambda \in \mathcal{L}^k(\mathbb{R}^n;\mathbb{R}^m)$.

**b. definition of the $k$th derivative.** We now define the $k$th derivative inductively.

**Definition 1.** *A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is said to be $k$ times differentiable at a point $a \in \mathbb{R}^n$ if there exists an open neighbourhood $U$ of $a$ such that the $(k-1)$st derivative $D^{k-1}f(x) \in \mathcal{L}^{k-1}(\mathbb{R}^n;\mathbb{R}^m)$ exists at every $x \in U$, and if there exists $\Lambda \in \mathcal{L}^k(\mathbb{R}^n;\mathbb{R}^m)$ such that*

$$\lim_{h\to 0} \frac{|D^{k-1}f(a+h) - D^{k-1}f(a) - \Lambda(v)|}{|h|} = 0.$$

*When this holds, $\Lambda$ is said to be the $k$th derivative of $f$ at $a$, and is written $D^k f(a)$.*

We have used the Euclidean norm in the definition of the derivative, but as in the case $k = 2$, in view of (2) we could just as well have used the operator norm. (Indeed any other norm on $\mathcal{L}^k(\mathbb{R}^n;\mathbb{R}^m)$ would be fine as well and would lead to a completely equivalent definition.)

**c. facts about higher derivatives.**

The proofs of some of these facts are very smiler to the $k = 2$ cases, which we have already seen. Proofs of others appear at the end of this handout.

An informal summary is:

**1**. when a function $f$ is $k$ times differentiable at a point $a$, all $k$th order partial derivatives exist, and we can essentially identify $D^k f$ with the collection of all $k$th order partial derivatives (made into a $k$-linear operator in a particular way, see formula (6) below.)

**2**. It can however happen that all $k$th-order partial derivatives of a function $f$ exist at a point $a$, but that the function is not $k$ times differentiable at that point. In this situation, all bets are off. However, we will almost never encounter this situation. And if all $k$th order partial derivatives exist and are continuous in an open set $U$, then $f$ is $k$ times differentiable everywhere in $U$.

**Theorem 1.** *Assume that $f; \mathbb{R}^n \to \mathbb{R}^m$ is $k$ times differentiable at $a$. Then for every $i_1, \ldots, i_k \in \{1, \ldots, n\}$*

$$D^k f(a)(e_{i_1}, \ldots, e_{i_k}) = D_{i_1}(D_{i_2} \cdots (D_{i_k} f) \cdots)) = D_{i_k,\ldots,i_1} f(a). \tag{3}$$

*In particular, the derivative on the right-hand side (called a $k$th order partial derivative) exists. More generally,*

$$D^k f(v_1, \ldots, v_k) = D_{v_1}(D_{v_2}(\cdots (D_{v_k} f) \cdots)), \tag{4}$$

*where the right-hand side denotes the result of iterated direction differentiation of $f$ in the directions $v_k, \ldots, v_1$. Moreover, partial derivatives are independent of the order of differentiation, so that if $\sigma$ is any permutation of $\{1, \ldots, k\}$, then*

$$D_{i_k,\cdots,i_1} f(a) = D_{i_{\sigma(k)},\cdots,i_{\sigma(1)}} f(a). \tag{5}$$

*Finally, for any vectors $v_1, \ldots, v_k$, where $v_j = (v_i^1, \ldots, v_j^n)$*

$$D^k f(a)(v_1, \ldots, v_k) = \sum_{i_1,\ldots,i_k=1}^{n} v_1^{i_1} \cdots v_k^{i_k} D_{i_k,\ldots,i_1} f(a). \tag{6}$$

Note that both sides of equations (3) - (6) are vectors with $m$ components.

The second equality in (3) is true by our notation, so the point is the first equality. Equation (5) is a special case of the more general fact that $D^k f(a)(v_1, \ldots, v_k) = D^k f(a)(v_{\pi(1)}, \ldots, v_{\pi(k)})$ for all $v_1, \ldots, v_k$. In other words, $D^k f$ is a *symmetric* $k$-linear map $(\mathbb{R}^n)^k \to \mathbb{R}^m$,

If $k$th order partial and/or directional derivatives exist but $f$ is not $k$ times differentiable at $a$, then various pathologies can occur, e.g. partial derivatives depending on the order of differentation. However, we will rarely encounter such situations in this class. And none of these unpleasant situations occur if the $k$th order partial derivatives are continuous, as follows from the next theorem.

**Theorem 2.** *Assume that $f : \mathbb{R}^n \to \mathbb{R}^m$ and that every partial derivative of order $k$ of every component of $f$ exists and is continuous in an open set $U \subset \mathbb{R}^n$. Then $f$ is $k$-times differentiable at every point of $U$.*

It is worth pointing out the following result, which we will need later. (The proof follows easily from (4).)

**Lemma 2.** *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is $k-1$ times differentiable at every point in an open set $U \subset \mathbb{R}^n$, and continuously differentiable at a point $a \in U$. For any $v \in \mathbb{R}^n$ if we define $g(t) = f(a+tv)$, then the $k$th derivative of $g$ exists at $t = 0$, and*

$$(\frac{d}{dt})^k g(0) = D^k f(a)(v, \ldots, v) \tag{7}$$

The left-hand side of (7) denotes the $k$th derivative of $g$, evaluated at $t = 0$. (I am not sure what notation you are used to. Below I will also use the notation $(\frac{d}{dt})^k g(t)|_{t=0}$ for the same thing.)

The same result holds for functions $\mathbb{R}^n \to \mathbb{R}^m$, where (7) is then understood to hold for each component.

Finally, we also state the following result, which expresses $D^k f(v_1, \ldots, v_k)$ as a $t \to 0$ limit of a certain linear combination of values of $f$ at the vertices of a parallelepiped generated by the vectors $tv_1, \ldots, tv_k$. This is a generalization of problem 3 on homework 3. Before reading the statement, think about what you expect it to be, based on the $k = 2$ case. (And after reading the statement, which is a little opaque, think about what it says.....)

**Lemma 3.** *If $f : \mathbb{R}^n \to \mathbb{R}^m$ is $k$ times differentiable at $a$, for every $v_1, \ldots, v_k \in \mathbb{R}^n$*

$$D^k f(a)(v_1, \ldots, v_k) = \lim_{t\to 0} \frac{1}{t^k} \sum_{\sigma \in \{0,1\}^k} (-1)^{k - \sum_{k=1}^k \sigma^i} f(a + t\sum_{i=1}^k \sigma^i v_i) \tag{8}$$

**d. Some proofs**.

*Proof of (parts of) Theorem 1. Step 1.* First, we prove (4). We will argue by induction on $k$. The case $k = 1$ is Exercise 2-29(c) in Spivak. We may assume that $v_1, \ldots, v_k$ are all nonzero, as otherwise both sides of (4) vanish and it obviously holds.

The assumption that $f$ is $k$-times differentiable, and the equivalence of the euclidean norm and the operator norm, implies that for any vector $v_1$,

$$\lim_{t\to 0} \frac{\|D^{k-1}f(a+tv_1) - D^{k-1}f(a) - D^k f(a)(tv_1)\|}{t} = 0.$$

Note also that for any $T \in \mathcal{L}^{k-1}(\mathbb{R}^n, \mathbb{R})$, and any nonzero vectors $v_2, \ldots, v_k$,

$$|T(v_2, \ldots, v_k)| \; = \; |v_2| \cdots |v_k| \; T(\frac{v_2}{|v_2|}, \ldots, \frac{v_k}{|v_k|}) \; \le \; |v_2| \cdots |v_k| \; \|T\|$$

using the definition of the operator norm $\|T\|$. Applying this to $T = D^{k-1}f(a+tv_1) - D^{k-1}f(a) - D^k f(a)(v_1)$, which is an element of $\mathcal{L}^{k-1}(\mathbb{R}^n; \mathbb{R}^m)$, we find that

$$\begin{aligned} 0 &= |v_1| \cdots |v_k| \lim_{t\to 0} \frac{\|D^{k-1}f(a+tv_1) - D^{k-1}f(a) - D^k f(a)(tv_1)\|}{t|v_1|} \\ &\ge \lim_{t\to 0} \frac{\left|D^{k-1}f(a+tv_1)(v_2, \ldots, v_k) - D^{k-1}f(a)(v_2, \ldots v_k) - D^k f(a)(tv_1, v_2, \ldots, v_k)\right|}{t} = 0, \end{aligned}$$

We rewrite the last term, using the linearity of $D^k f(a)$, to obtain

$$\lim_{t\to 0} \frac{D^{k-1}f(a+tv_1)(v_2,\dots,v_k) - D^{k-1}f(a)(v_2,\dots v_k)}{t} - D^k f(a)(v_1,v_2,\dots,v_k) = 0.$$

This says that the directional derivative $D_{v_1}h(a)$ exists, for $h(x) = D^{k-1}f(x)(v_2,\dots,v_k)$. Identity (4) now follows by invoking the induction hypothesis.

*Step 2.* Conclusion (3) is a special case of (4), as noted above.

*Step 3.* Next, (5) can be proved by using the $k=2$ case to switch the order of differentiation, a pair of (adjacent) derivatives at a time. We omit the details. As an exercise, think about how to write up this argument correctly.

*Step 4.* Finally, to prove (6), we can write each $v_j$ as $v_j = \sum_{i_k=1}^n v_j^{i_i} e_{i_j}$, where $e_{i_j}$ denotes the standard basis vector in the $i_j$ direction. Then multi-linearity of $D^k f$,

$$D^k f(a)(v_1,\dots,v_k) = D^k f(a)(\sum_{i_1=1}^n v_i^{i_1} e_{i_i},\dots,\sum_{i_k=1}^n v_k^{i_k} e_{i_k}) = \sum_{i_1,\dots,i_k=1}^n v_1^{i_1}\cdots v_k^{i_k} D^k f(a)(e_{i_1},\dots,e_{i_k}).$$

Thus the conclusion follows from (3). □

Next we give the

*Proof of Theorem 2.* We will give the proof for $m=1$, As in the case of first derivatives, this implies the general case.

The idea is as follows: We have in (6) a formula for $D^k f(a)$ in terms of partial derivatives of $f$, a formula that must hold if, as we hope to prove, $f$ is $k$ times differentiable. So we will use this formula and the existence of partial derivatives to define a candidate for $D^k f$, and then verify (using the $k=1$ case of the result we are trying to prove, i.e. Theorem 2-8 in Spivak, which we already know) that our candidate for $D^k f$ in fact has the properties it is supposed to have.

The main difficulty is notation.

Fix a point $a \in U$. By assumption, the following definition makes sense, since all $k$th order partial derivatives exist everywhere in $U$:

$$\Lambda(v_1,\dots,v_k) = \sum_{i_1,\dots,i_k=1}^n v_1^{i_1}\cdots v_k^{i_k} D_{i_k,\dots,i_1} f(a).$$

(Compare (6).) We will show that $\Lambda = D^k f(a)$. In view of the definition of the Euclidean norm on $\mathcal{L}^k(\mathbb{R}^n;\mathbb{R})$, it suffices to prove that for every $i_1,\dots,i_k \in \{1,\dots,n\}$,

$$(9) \qquad \lim_{h\to 0}\frac{1}{|h|}\left|Df(a+h)(e_{i_1},\dots,,e_{i_{k-1}}) - Df(a)(e_{i_1},\dots,,e_{i_{k-1}}) - \Lambda(h,e_{i_1},\dots,,e_{i_{k-1}})\right| = 0.$$

(Proving this in detail is an exercise.) So we fix some such $i_1,\dots,i_{k-1}$ for the duration of the proof. By Theorem 1, we know that $D^{k-1}f(e_{i_1},\dots,e_{i_{k-1}}) = D_{i_{k-1},\dots,i_1}f(a)$. Let us simplify notation by writing $D_\alpha f(a)$ to denote $D_{i_{k-1},\dots,i_1}f$ for most of the rest of this proof. Now since $f$ is assumed to be $k-1$-times differentiable in a neighbourhood of $a$, we can use Lemma 1 to rewrite

$$\text{left-hand side of (9)} = \lim_{h\to 0}\frac{1}{|h|}\left|D_\alpha f(a+h) - D_\alpha f(a) - \Lambda(h,e_1,\dots,e_{k-1})\right|$$

Also, our hypotheses implies that $D_\alpha f$ is continuously differentiable in $U$, hence Theorem 2-8 in Spivak implies that $D_\alpha f$ is a differentiable function, and that the Jacobian "matrix" $(D_\alpha f)'(a)$

(here a row vector) is given by $(D_\alpha f)'(a) = (D_1 D_\alpha f(a), \ldots, D_n D_\alpha f(a))$. In particular,

$$
\begin{aligned}
D(D_\alpha f)(a)(h) = (D_\alpha f)'(a) &= \sum_{j=1}^{n} D_j D_\alpha f(a) h^j \\
&= \sum_{j=1}^{n} h^j D_{i_1 \cdots i_{k-1} j} f(a).
\end{aligned}
$$

If we scrutinize the definition of $\Lambda$, we see that the right-hand side is exactly $\Lambda(h, e_{i_1}, \ldots, e_{i_{k-1}})$. Thus

$$
\text{left-hand side of (9)} = \lim_{h \to 0} \frac{1}{|h|} \left| D_\alpha f(a+h) - D_\alpha f(a) - D(D_\alpha f)(a)(h) \right|
$$

And the above limit clearly equals 0, by definition of what it means for $D_\alpha f$ to be differentiable. Thus we have proved (9), and this completes the proof of the theorem (up to details left as exercises for the reader.) $\square$

*Proof of Lemma 2.* According to (4), if $f$ is $k$ times differentiable at $a$, then

$$
D^k f(v, \ldots, v) = D_v(D_v(\cdots D_v f)\cdots),
$$

and an easy induction argument shows that the right-hand side can be rewritten as $(\frac{d}{dt})^k g(0)$ for $g(t) = f(a + tv)$. $\square$

We omit the proof of Lemma 3. The idea is to use induction on $k$, and to argue in a way similar to the proof of Problem 3 in Homework 3. The notation is more complicated, but otherwise the proof is rather similar.