# MAT 257, Handout 5: October 10-14, 2011.

## I. Assignment.

**Over the next week.** Continue reading Chapter 2 of Spivak, reviewing as necessary and continuing through the end of page 39. Continue to work exercises from Chapter 2. Exercise 3-27 is a sort of warm-up for the Implicit Function Theorem, so you will get the most out of it if you look at it before we get to the Implicit function Theorem.

## II. About second derivatives.

**Preliminaries**. Let $\mathcal{L}(\mathbb{R}^n;\mathbb{R}^m)$ denote the space of linear maps from $\mathbb{R}^n$ to $\mathbb{R}^m$. This is clearly a vector space, and it is easy to see that it has dimension $mn$, since there is a one-to-one linear mapping between $\mathcal{L}(\mathbb{R}^n;\mathbb{R}^m)$ and the space of $m \times n$ matrices, which in turn can be naturally identified with $\mathbb{R}^{m\times n}$. We will write $\lambda$ to denote a generic element of $\mathcal{L}(\mathbb{R}^n;\mathbb{R}^m)$, and we will use the notation

$$\|\lambda\| := \sup\{|\lambda(x)| \ : \ x \in \mathbb{R}^n, |x| = 1\} \qquad \text{and} \tag{1}$$

$$|\lambda| := \Big( \sum_{i=1}^n |\lambda(e_i)|^2 \Big)^{1/2}. \tag{2}$$

It is straightforward to verify that these are both norms on $\mathcal{L}(\mathbb{R}^n;\mathbb{R}^m)$. Also, note that if $A$ is the $m \times n$ matrix that represents $\lambda$, with entries $(a_{ij})$, then $\lambda(e_i)$ is the $i$th column of $A$., so that $|\lambda| = (\sum_{i,j} a_{ij}^2)^{1/2}$. In other words, $|\lambda|$ is the euclidean norm of the point in $\mathbb{R}^{m\times n}$ whose components are the entries $(a_{ij})$ of $A$ listed in some fixed order. For this reason we will call $|\lambda|$ the Euclidean norm of $\lambda$, whereas $\|\lambda\|$ is called the operator norm of $\lambda$.

It follows from exercises (3) and (4) on handout 2 (which were discussed at length in the lecture) that there exist constants $k < K$ such that

$$k|\lambda| \le \|\lambda\| \le K|\lambda| \tag{3}$$

for all $\lambda \in \mathcal{L}(\mathbb{R}^n;\mathbb{R}^m)$. (In particular, feel free to use this fact in the homework problems.) In fact one can also check directly that $n^{-1/2}|\lambda| \le \|\lambda\| \le |\lambda|$. for all $\lambda \in \mathcal{L}(\mathbb{R}^n;\mathbb{R}^m)$.

**Definition of second derivative**: Suppose that $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at every point $x$ of an open set $U \subset \mathbb{R}^n$. Then for every $x \in U$, $Df(x)$ is an element of $\mathcal{L}(\mathbb{R}^n;\mathbb{R}^m)$, so we can view $Df$ as a function from $U$ to $\mathcal{L}(\mathbb{R}^n;\mathbb{R}^m)$. In this situation, we say that $Df$ is *continuous* at a point $a \in U$ if $\forall\varepsilon > 0 \exists\delta > 0$ such that $|Df(x) - Df(a)| < \varepsilon$ whenever $x \in U$ and $|x - a| < \delta$.

We emphasize the distinction between $Df(x)$, an element of $\mathcal{L}(\mathbb{R}^n;\mathbb{R}^m)$, and $Df$, a function $U \to \mathcal{L}(\mathbb{R}^n;\mathbb{R}^m)$.

We say that $f$ is *twice differentiable* at a point $a \in \mathbb{R}^n$ if $Df$ exists at every point in a neighbourhood $U$ of $a$, and if there exists a linear map $\Lambda : \mathbb{R}^n \to \mathcal{L}(\mathbb{R}^n;\mathbb{R}^m)$ such that

$$\lim_{h\to 0} \frac{|Df(a+h) - Df(a) - \Lambda(h)|}{|h|} = 0 \tag{4}$$

When this holds we write $D(Df)(a)$ to denote $\Lambda$, and we call it the second derivative of $f$.

We have used the Euclidean norm (2) on $\mathcal{L}(\mathbb{R}^n;\mathbb{R}^m)$ to define continuity and differentiability of $Df : U \to \mathcal{L}(\mathbb{R}^n;\mathbb{R}^m)$. But in view of (3), we could have used the operator norm, and the resulting definitions would have been completely equivalent to the definitions we have given above.

**Remark**. In Spivak, $f : \mathbb{R}^n \to \mathbb{R}^m$ is said to be continuously differentiable in an open set $U$ is all partial derivatives of all components of $f$ exist everywhere in $U$ and are continuous functions from $U$ to $\mathbb{R}$. One can check, using Theorems 2-7 and 2-8 in Spivak, that $f$ is continuously differentiable in $U$ if and only if $f$ is differentiable at every point of $U$, and in addition $Df$ is a continuous function from $U$ to $\mathcal{L}(\mathbb{R}^n;\mathbb{R}^m)$.

**Reformulation of the second derivative.** By definition, if $f$ is twice differentiable at $a$, then $D(Df)(a)$ is a linear map $\mathbb{R}^n \to \mathcal{L}(\mathbb{R}^n;\mathbb{R}^m)$.

Thus, for any $v \in \mathbb{R}^n$, $D(Df)(a)(v)$ is a linear map $\mathbb{R}^n \to \mathbb{R}^m$. So for any $w \in \mathbb{R}^n$, $D(Df)(a)(v)(w)$ is an element of $\mathbb{R}^m$.

So we can identify the second derivative of $f$ at $a$ with the function, which we will write $D^2 f(a) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$, defined by

$$D^2 f(a)(v,w) = D(Df)(a)(v)(w).$$

It is straightforward to check from the definitions that $D^2 f(a)$ is bilinear. The following lemma clarifies the "meaning" of $D^2 f(a)$.

**Lemma 1.** *Suppose that $f : \mathbb{R}^n \to \mathbb{R}^m$ is twice differentiable at $a \in \mathbb{R}^n$. Then for any $v$ and $w$ in $\mathbb{R}^n$,*

$$D^2 f(a)(v,w) = \lim_{h\to 0} \frac{f(a+hv+hw) - f(a+hv) - f(a+hw) + f(a)}{h^2}. \tag{5}$$

*(In particular, the limit on the right-hand side exists.)*

The proof is an exercise (with hints) that is part of the third homework assignment, see below.

Lemma 1 should be compared with Exercise 2-29(c) in Spivak, which establishes a parallel relationship between the first derivative and directional derivatives.

Note in particular that it follows from (5) and the bilinearity of $D^2 f$ that

$$D^2 f(v,v) = -D^2 f(v,-v) = \lim_{h\to 0} \frac{f(a+hv) - 2f(a) + f(a-hv)}{h^2}$$

The expression on the right-hand side in naturally interpreted as the "second directional derivative of $f$ in the $v$ direction", in view of its similarity to the familiar formula for the second derivative from single-variable calculus.

Another consequence of (5) is

**Corollary 1.** *If $f : \mathbb{R}^n \to \mathbb{R}^m$ is twice differentiable at a point $a \in \mathbb{R}^n$, then $D^2 f(a)(v,w) = D^2 f(a)(w,v)$.*

*Proof.* Since the right-hand side of (5) is symmetric with respect to $v$ and $w$, the left-hand side must be as well. ☐

Note that problems 1 and 2 below together establish that $D_i(D_j f)(x) = D_j(D_i f)(x)$ at points $x$ where $f$ is twice differentiable. (The same conclusion is stated, under somewhat different hypotheses, in Spivak's Theorem 2-5, although the proof in Spivak is deferred to the exercises in Chapter 3.)

## III. Homework, due October 21.

**1**. Spivak, exercise 2-14.

Although it is not part of the assignment, please also take a look at Spivak, exercise 2-15 (a), which is an easy consequence of exercise 2-14.

**2**. Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable at $a \in \mathbb{R}^n$. Prove that $D(Df)(a)(e_i)(e_j) = D_i(D_j f)(a)$

**Hint**: For every pair of vectors $v$ and $w$,

$$\frac{Df(a+hv) - Df(a) - D(Df)(a)(hv)}{h}(w) \to 0$$

as $h \to 0$. Give the very short proof that this follows from the definitions, and use it to prove the desired identity.

**3**. Prove Lemma 1 above by filling in the steps below. Part of the exercise is deciphering the notation. (The same is true for problem 1 above.)

**a**. Fix $f, a, v, w$ as in the statement of the lemma, and introduce the notation

$$\Delta(h) = \frac{f(a+hv+hw) - f(a+hv) - f(a+hw) + f(a)}{h^2}.$$

By applying the 1-dimensional mean-value theorem to the function $g(s) = \frac{f(a+hv+shw)-f(a+shw)}{h^2}$ and using Spivak's exercise 2-29(c) (which you can take to be known), or otherwise, show that there exists some $\theta \in (0,1)$ such that

$$\Delta(h) = \frac{Df(a+hv+\theta hw) - Df(a+\theta hw)}{h}(w)$$

whenever $h$ is small enough.

**b**. Let $\Lambda = D^2f(a)$. By adding and subtracting and using the definition (4) of $\Lambda$, rewrite $\frac{Df(a+hv+\theta hw)-Df(a+\theta hw)}{h}$ as expressions involving $\Lambda$ and terms that tend to zero as $h \to 0$, and use the resulting identity to complete the proof of the Lemma.

**4**. Spivak, exercise 2-38.

**marking**: 10 marks each. In problems with two parts, the marks will be split equally between the two halves, except in problem 3, where part (a) is worth 3 marks and part (b) is worth 7 marks.

In this class, marks are not in general strongly correlated with the difficulty of a problem.