MAT 257, Handout 8: October 31-November 4, 2011.

Assignment.

Start to read chapter 3 of Spivak for later in the week.

Some topics from the lectures

Level sets, gradients, regular and critical values. For continuously differentiable $f : \mathbb{R}^n \to \mathbb{R}$ we defined

- a *level set* of f is a set of the form $f^{-1}(\{p\})$, for some $p \in \mathbb{R}$. We generally abuse notation and write $f^{-1}(p)$.
- if f is differentiable at a, the *gradient* of f at a is the vector defined by the requirement that
$$\langle \nabla f(a), v\rangle = Df(a)(v) \quad \text{ for all } v \in \mathbb{R}^n$$
Note that $Df(a), f'(a)$ and $\nabla f(a)$ are all different objects: $Df(a)$ is a linear map $\mathbb{R}^n \to \mathbb{R}$; $f'(a)$ is a $1 \times n$ matrix representing that linear map, and $\nabla f(a)$ is a vector. In particular, note that to define ∇f we need to know the inner product on $\mathbb{R}^n$, whereas the other two objects are defined without knowing the inner product. However, they all look very similar when written down. For example, $f'(a) = (D_1 f(a), \dots, D_n f(a))$, whereas $\nabla f(a)$ is the vector whose ith component is exactly $D_i f(a)$.
- A *regular point* of f is a point $a \in \mathbb{R}^n$ such that $\nabla f(a) \neq 0$. A *critical point* is a point $a \in \mathbb{R}^n$ such that $\nabla f(a) = 0$.
- A *regular value* of f is a number $p \in \mathbb{R}$ such that every point $x \in f^{-1}(p)$ is regular. A *critical value* is a number $p \in \mathbb{R}$ that is not a regular value.

More generally, for a differentiable function $f : \mathbb{R}^n \to \mathbb{R}^m$ with $m \leq n$, we sometimes say a *regular value* of f is a point $p \in \mathbb{R}^m$ such that $Df(x)$ has rank m at every $x \in f^{-1}(p)$.

Informally,

- $\nabla f(a)$ is orthogonal to the level set of f that passes through a. (Note that the inner product is needed to talk about "orthogonality".)
- $\nabla f(a)$ points in the "direction of fastest increase" of f at a, and $|\nabla f(a)|$ is the slope of f in the direction of fastest increase. (Note that the inner product is needed to talk about "direction of fastest increase." Why is this true?)

Making these statements precise requires some definitions, including the following (revised somewhat from the lecture). Let us temporarily write $bla(f, a)$ to denote the "best linear approximation of f at a" so that $bla(f, a)(x) = f(a) + Df(a)(x - a) = f(a) + \langle \nabla f(a), x - a\rangle$.

- If a is a regular point of f and $f(a) = p$, then the *tangent plane* of $f^{-1}(p)$ at a is $bla(f, a)^{-1}(p)$, i.e. the level set of $bla(f, a)$ that passes through a.
- the *tangent space* to $f^{-1}(p)$ at a is $\{v \in \mathbb{R}^n : Df(a)(v) = 0\} = \{v \in \mathbb{R}^n : \langle \nabla f(a), v\rangle = 0\}$. Note that this is the set of vectors that are parallel to the tangent plane of $f^{-1}(p)$ at a.

(I did not carefully distinguish between the tangent plane and the tangent space in the lecture.) Then in the above informal statements, "orthogonal to the level set of f that passes through a" should be understood to mean "orthogonal to every vector v in the tangent space of $f^{-1}(p)$ at a." The other statements about the "direction of fastest increase" and "slope n the direction of fastest increase" can similarly be formulated precisely in terms of $bla(f, a)$.

We said vaguely that there is a connection between critical points of f and "changes in topology of the level sets of f". This connection is developed in depth in Morse Theory, a branch of topology.

Taylor's Theorem in $\mathbb{R}^n$. If $f : \mathbb{R}^n \to \mathbb{R}$ is k times differentiable at a, then

$$\boxed{f(x+v) = \sum_{j=0}^{k} \frac{1}{j!} D^j f(a)(v\dots,v) \ + R_k(v) \qquad \text{where } \lim_{v\to 0} \frac{|R_k(v)|}{|v|^k} = 0.} \tag{1}$$

The $k = 1$ case is just the *definition* of Df, since $R_1(v) = F(a+v) - f(a) - Df(a)(v)$.

A weaker version of the theorem states that $\lim_{t\to 0} R_k(tv)/t^k = 0$ for every $v \in \mathbb{R}^n$. (Why is this weaker?) This weaker assertion follows immediately from combining any standard 1-dimensional version of Taylor's Theorem with Lemma 3 from the Week 6 handout.

Below are a couple of exercises that related to Taylor's Theorem, one of which sketches the proof of the stronger assertion (1) above.

Some exercises

1. In this exercise we will prove (1).

Because we only assume that f is k times differentiable at a, our proof can only refer to $D^k f(a)$, as well as lower derivatives of f at points near a; we have not assumed that $D^k f(x)$ exists for any $x \neq a$. So our first task is to develop an expression for $R_k(v)$ that involves only these quantities. For present purposes, this rules out standard formulas for $R_k(v)$ as found for example on Wikipdia.

a. Prove that if $g : \mathbb{R} \to \mathbb{R}$ is $k-1$ times differentiable in an interval $(-\delta, \delta)$ (with $\delta > 0$) and k times differentiable at 0, then for $t \in (-\delta, \delta)$,

$$\begin{aligned} g(t) - \sum_{j=0}^{k} \frac{t^j}{t!} g^{(j)}(0) &= \int_0^t \int_0^{s_1} \cdots \int_0^{s_{k-1}} \left(g^{(k)}(s_k) - g^{(k)}(0) \right) ds_k \cdots ds_2 \ ds_1 \\ &= \int_0^t \int_0^{s_1} \cdots \int_0^{s_{k-2}} \left(g^{(k-1)}(s_{k-1}) - g^{(k-1)}(0) - s_{k-1} g^{(k)}(0) \right) ds_{k-1} \cdots ds_2 \ ds_1. \end{aligned}$$

where $g^{(j)}$ denotes the jth derivative of g.

b. Using part **(a)** and Lemma 3 on Handout 6, prove that if $f : \mathbb{R}^n \to \mathbb{R}$ is k times differentiable at $a \in \mathbb{R}^n$, then if v is sufficiently small[1] then

$$R_k(v) = \int_0^1 \int_0^{s_1} \cdots \int_0^{s_{k-2}} \left(D^{(k-1)} f(a + s_{k-1} v) - D^{(k-1)} f(a) - D^{(k)} f(a)(s_{k-1} v) \right) (v, \dots, v) ds_{k-1} \cdots ds_2 \ ds_1.$$

where $R_k(v)$ is *defined* to be

$$R_k(v) := f(a+v) - \sum_{j=0}^{k} \frac{1}{j!} D^j f(a)(v, \dots, v).$$

(As usual, part of what you must do is decipher the notation.)

c. Use part **(b)** and the definition of $D^k f(a)$ to prove that $\lim_{v\to 0} R_k(v)/|v|^k = 0$.

2. In this exercise, we introduce notation that allows for a rather elegant explicit formula (in terms of partial derivatives of f) for the kth order Taylor polynomial of f.

When f is a function whose domain is $\mathbb{R}^n$ (or an open subset of $\mathbb{R}^n$), we will use the notation

$$D^\alpha f(a) = D_1^{\alpha_1} D_2^{\alpha_2} \cdots D_n^{\alpha_n} f(a)$$

where α is an n-tuple of the form $(\alpha_1, \dots, \alpha_n)$, with α_i a nonnegative integer for every i. Thus, the ith derivative is iterated α_i times.

[1] that is, if $a + v \in R$ for some open rectangle R containing a in which in which f is $k-1$ times continuously differentiable

For example if $n = 2$, and $\alpha = (2, 1)$, then $D^\alpha f = D_1 D_1 D_2 f(a)$.

Note, if we wanted to, we could write "$D_{(2,1)} f$" to mean "$D^\alpha f$, for the multi-index $\alpha = (2, 1)$." But we will generally not do so, because $D_{(2,1)} f$ looks enough like $D_{2,1} f = D_1 D_2 f$ that it is easy to get confused. So we will generally only use the notation $D^\alpha f$ for "generic" α.

If $f : \mathbb{R}^n \to \mathbb{R}^m$, then $D^\alpha f(x)$ denotes the element of $\mathbb{R}^m$ with components $D^\alpha f^i$, $i = 1, \ldots, m$.

We call α a *multi-index*, and we use the notation

$$|\alpha| = \sum_{i=1}^n \alpha_i, \qquad \alpha! := \alpha_1! \cdots \alpha_n!.$$

so that in particular $|\alpha|$ is the "order" of the partial derivative $D^\alpha f$. (For this reason we sometimes call $|\alpha|$ the order of the multi-index α.) Also, if $v = (v^1, \ldots, v^n)$ is a vector in $\mathbb{R}^n$, then we will use the notation

$$v^\alpha = (v^1)^{\alpha_1} \cdots (v^n)^{\alpha_n}$$

Exercise **2** is to prove that if f is k times differentiable, then

$$\boxed{\sum_{j=0}^k \frac{1}{j!} D^j f(a)(v, \ldots, v) \;=\; \sum_{|\alpha| \le k} \frac{v^\alpha}{\alpha!} D^\alpha f(a).} \tag{2}$$

The sum on the right-hand side means "the sum over all multi-indices α of order at most k".

remark: The other expression that we have for a kth order Taylor series, following directly from results in handout 6, is

$$\sum_{j=0}^k \left(\sum_{i_1, \ldots, i_j = 1}^n v^{i_1} \cdots v^{i_j} D_{i_j, \cdots, i_1} f(a) \right) \tag{3}$$

(We could also write (2) as $\sum_{j=0}^k \left(\sum_{|\alpha| = j|} \frac{v^\alpha}{\alpha!} D^\alpha f(a) \right)$, making it look a bit more like (3).)

In general I find it clearer to write Taylor series of order k in the form (3) if $k \le 2$, and using multi-indices if $k \ge 3$, and especially for a generic positive intiger k.