NOTES ON HIGHER DIMENSIONAL DERIVATIVES

JAKE MARCINEK
HARVARD UNIVERSITY
CAMBRIDGE, MA
MARCINEK@MATH.HARVARD.EDU

1. Single-variable recap

Let $f(x)$ be a differentiable function of one variable.

Definition 1.1. The *derivative of f at a* is

$$f'(a) = f_x(a) = \frac{\partial f}{\partial x}(a) = \lim_{x\to a} \frac{f(x)-f(a)}{x-a} = \lim_{t\to 0} \frac{f(a+t)-f(a)}{t}. \tag{1}$$

Remark 1.2. The derivative of f at a is the instantaneous rate of change of f when increasing x at unit speed near a. It can also be thought of as the linear approximation (i.e. first order approximation) to f.

Definition 1.3. The *Taylor series approximation* $T = T_{f,a}$ for f at a is

$$T(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x-a)^k. \tag{2}$$

The *n^{th}-partial Taylor series approximation* $T_n = T_{n,f,a}$ for f at a is

$$T_n(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k. \tag{3}$$

Here $f^{(k)}$ is the k^{th} derivative of f and $f^{(0)} = f$ by convention.

Remark 1.4. The idea here is that T_n and f have the same k^{th} derivative at a for all $0 \le k \le n$.

Definition 1.5. The *linearization of f at a* is $L = T_1$. That is,

$$L(x) = f(a) + f'(a)(x-a). \tag{4}$$

Remark 1.6. The linearization of f and f itself share the same order 0 and order 1 information at a. That is, $L(a) = f(a)$ and $L'(a) = f'(a)$.

2. More variables

We want a notion of derivative in higher dimensions which generalizes the single variable derivative, so let's begin by looking at higher variable functions in only one direction at a time. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function of multiple variables, $P \in \mathbb{R}^n$ any point in the domain of f, and $v \in \mathbb{R}^n$ any vector (thought of as emanating from P).

Date: October 22, 2016.

Definition 2.1. The *directional derivative in the direction v of f at P* is

$$D_v f(a,b) = \frac{\partial f}{\partial x}(a) = \lim_{h \to 0} \frac{f(P+tv) - f(P)}{t}. \tag{5}$$

Remark 2.2. The directional derivative in the direction v of f at P is the instantaneous rate of change of f when moving (in the domain) through P with velocity v.

Remark 2.3. Consider the path through the domain $\mathbb{R}^n$ given by $\gamma(t) = P + tv$. Then $D_v f(P) = \frac{\partial}{\partial t}[f \circ \gamma](0)$; that is, the directional derivative in the direction v of f at P can be thought of as the single variable derivative of f restricted to the line determined by P and v.

Theorem 2.4. *As a function of v, the directional derivative of f at P*

$$\begin{aligned} \mathbb{R}^n &\to \mathbb{R} \\ v &\mapsto D_v f(P) \end{aligned}$$

is linear. This means $D_{sv+tw} f(P) = sD_v f(P) + tD_w f(P)$ for any v, w vectors and any s, t scalars. In particular, there is a unique vector $\nabla f(P)$ with the property that $D_v f(P) = \nabla f(P) \cdot v$ for every vector v.

Proof. For the first part, plug $sv + tw$ into the definition of derivative. The "in particular" part is linear algebra magic. □

Definition 2.5. The vector $\nabla f(P)$ is the *gradient* of f at P. The gradient of f is a *vector field*; i.e. a function $\mathbb{R}^n \to \mathbb{R}^n$ where the input is thought of as a point and the output is thought of as a vector originating at the input point.

Theorem 2.6. *The gradient is orthogonal to level sets of f.*

Proof. For any direction v tangent to a level set of f at P,

$$0 = D_v f(P) = \nabla f(P) \cdot v \tag{6}$$

since f is constant along level sets. □

Corollary 2.7. *The direction of greatest increase (respectively, steepest descent) of f at P is given by $\nabla f(P)$ (resp. $-\nabla f(P)$).*

Proof. For any direction $|v|^2 = 1$, $D_v f(P) = \nabla f(P) \cdot v = |\nabla f(P)||v| \cos\theta$ is maximized (resp. minimized) when v and $\nabla f(P)$ are (resp. anti-)parallel. □

3. Linearization

Again, let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function of multiple variables and $P \in \mathbb{R}^n$ a point in the domain of f. Just as in the single variable case, the linearization of f and f itself should share the same order 0 and 1 information at P.

Definition 3.1. The linearization L of f at P is the unique linear function on $\mathbb{R}^n$ such that $L(P) = f(P)$ and $\nabla L(P) = \nabla f(P)$.

Remark 3.2. To construct such a function, let $L(Q) = f(P) + \nabla f(P) \cdot (Q - P)$. Then

$$L(P) = f(P) + \nabla f(P) \cdot 0 = f(P) \text{ and} \tag{7}$$

$$D_v L(P) = \lim_{t \to 0} \frac{\nabla f(P) \cdot tv}{t} = \nabla f(P) \cdot v \tag{8}$$

so $\nabla L(P) = \nabla f(P)$ by the defining property of the gradient.

4. In coordinates

So far we have done everything without using coordinates (which is hopefully a little satisfying). This is similar to how I defined dot/cross products in the first week using only algebraic properties with no coordinates. However, for the sake of homework and exams, where you are asked to do computations in a certain coordinate system, let's cover the coordinate dependent versions of these notions.

In this section, let's assume we are in the $x : y : z$ coordinates of $\mathbb{R}^3$. This coordinate system has a "standard basis" comprised of the three standard unit vectors: $\hat{\imath}, \hat{\jmath}, \hat{k}$.

Definition 4.1. The *partial derivatives* of f at $P = (a, b, c)$ are the directional derivatives in the standard directions:

$$\frac{\partial f}{\partial x}(a,b,c) = f_x(a,b,c) = D_{\hat{\imath}} f(P)$$
$$\frac{\partial f}{\partial y}(a,b,c) = f_y(a,b,c) = D_{\hat{\jmath}} f(P)$$
$$\frac{\partial f}{\partial z}(a,b,c) = f_z(a,b,c) = D_{\hat{k}} f(P).$$

Remark 4.2. We can write any vector v in coordinates by

$$v = (v \cdot \hat{\imath})\hat{\imath} + (v \cdot \hat{\jmath})\hat{\jmath} + (v \cdot \hat{k})\hat{k} = \langle v \cdot \hat{\imath}, v \cdot \hat{\jmath}, v \cdot \hat{k} \rangle. \tag{9}$$

The gradient is no different. However, because of its defining property, the gradient takes on a nice coordinate dependent form.

$$\nabla f = \langle \nabla f \cdot \hat{\imath}, \nabla f \cdot \hat{\jmath}, \nabla f \cdot \hat{k} \rangle = \langle D_{\hat{\imath}} f, D_{\hat{\jmath}} f, D_{\hat{k}} f \rangle = \langle f_x, f_y, f_z \rangle. \tag{10}$$

Lastly, let's look at the linearization L of f at $P = (a, b, c)$ in coordinates. Plugging $Q = (x, y, z)$ and $P = (a, b, c)$ into Remark 3.2, we get

$$\begin{aligned} L(x,y,z) &= f(a,b,c) + \nabla f(a,b,c) \cdot (\langle x,y,z \rangle - \langle a,b,c \rangle \\ &= f(a,b,c) + \langle f_x(a,b,c), f_y(a,b,c), f_z(a,b,c) \rangle \cdot \langle x-a, y-b, z-c \rangle \\ &= f(a,b,c) + f_x(a,b,c)(x-a) + f_y(a,b,c)(y-b) + f_z(a,b,c)(z-c). \end{aligned}$$