Analysis 2 Notes

Khaled Allen
Mark Coffey
MATH 4001
University of Colorado
Text -

1 The directional derivative

We want to get a parallel notion of derivative for multivariable functions as that for single-variable functions. We will define a directional derivative, explore its relationship to the gradient vector and the partial derivative, and then examine why it is insufficient to give us the same power as the one-dimensional derivative.

Definition 1.1. The derivative of a vector-valued function $f : \mathbb{R}^n \to \mathbb{R}$ in the direction $\vec{u}$ is denoted $D_u f$ or $\dfrac{\partial f}{\partial \vec{u}}$ and is given by

$$D_u f = \lim_{t \to \infty} \frac{f(\vec{x} + t\vec{u}) - f(\vec{x})}{t}$$

where $\vec{u}$ is a unit vector, $\vec{x}, \vec{u} \in \mathbb{R}$.

Remark. This essentially means restricting the function along the vector $\vec{u}$ and then taking the derivative of the resulting single variable function. Gianquinta gives it as a completely new function, effectively expressing f as part of a single variable function.

In general, this is a common technique in analysis 2: express a multivariable function as a single variable and then perform a familiar operation from single-variable calculus.

1.1 Relationship to the Gradient and Partial Derivative

$D_u f(\vec{x}$ is related to the gradient via

$$D_u f(\vec{x} = \nabla f(\vec{x}) \cdot \vec{u})$$

We can therefore express $\frac{\partial f}{\partial x_i}$ as $D_{x_i} f(\vec{x}) = \nabla f(\vec{x}) \cdot e^i$.

1.2 Property: Cauchy Inequality

The norm of the directional derivative in the direction $\vec{u}$ is less than or equal to the norm of the gradient times the norm of the direction vector, which is always 1.

$$|D_u f| \leq |\nabla f(\vec{x})||\vec{u}| = |\nabla f(\vec{x})| \tag{1}$$

1

1.3 Property: Along the gradient itself

If you take the derivative in the direction of the gradient, the direction of greatest change, the derivative is actually just the norm of the gradient. That is, it is the maximum of 1 For $\vec{u} = \frac{\nabla f}{|\nabla f|}$, one has

$$D_u f = \frac{\nabla f \cdot \nabla f}{|\nabla f|} = \frac{|\nabla f|^2}{|\nabla f|} = |\nabla f|$$

.

1.4 In the opposite direction

Proposition 1.0.1. $D_{-u} f = -D_u f$

Proof.

$$D_{-\vec{u}} f = \nabla f \cdot -\vec{u} = \sum_i^n D_i f(-u_i)$$
$$= -\sum_i^n D_i f u_i$$
$$= -\nabla f \cdot \vec{u} = -D_u f$$

$\square$

1.5 Continuity

The existence of a (or even all) directional derivatives does not imply continuity.

1.5.1 Examples

1. $f : \mathbb{R}^2 \to \mathbb{R}$ that takes the value 0 on the axis but 1 everywhere else has all directional derivatives, but is discontinuous at $(0,0)$.

2. Let
$$f(x,y) = \begin{cases} \frac{x^2 y}{x^2 + y^2} & x \neq 0 \\ 0 & x = 0 \end{cases}.$$

Then f is continuous at the origin (since $\lim_{x,y \to (0,0)} f(x,y) = f(0,0)$) but has unequal partial derivatives.

$$D_1 f(0,0) = D_{(1,0)} f(0,0) = 0 = D_2 f(0,0) = D_{(0,1)} f(0,0)$$

However, $D_{(1,1)} f(0,0) = \frac{1}{2}$

QUESTION 1. I don't really understand why this shows that the function is discontinuous. The example in the book (Gian. p3) is confusing, and I'm not even sure how you would evaluate the partial derivatives in our notes, since it gives $\frac{0}{0}$. You could take the limit, but that's sort of different.

2 Differentiable Equations & Linear Tangent Map

Definition 2.1. (Diff. functions & tangent map) Let $A \subset \mathbb{R}^n, f : A \to \mathbb{R}, x_0 \in \text{interior}(A)$. We say f is differentiable at x_0 if there exists some linear function $L : \mathbb{R}^n \to \mathbb{R}$ such that

$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0) - L(h)}{|h|} = 0 \tag{2}$$

L is called the differential or tangent map (linear tangent map) of f at x_0 denoted df_{x_0} or $df(x_0)$.

An alternate notation

$$f(x_0 + h) = f(x_0) + df_{x_0}(h) + o(|h|) \tag{3}$$

This latter notation states that a function evaluated h away from a known point x_0 is equivalent to the function evaluated at x_0, plus the "differential" at x_0 (the immediate rate of change), plus some error that is less than the offset.

Definition 2.2. (Alternate way to get differentiable) Let $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ with a point $c \in A$. We say that f is differentiable at c if there exists from linear function $L : \mathbb{R}^n \to \mathbb{R}^m$ such that for every $\epsilon > 0$ there exists some $\delta(\epsilon) > 0$ such that when $|x - c| < \delta, x \in \mathbb{R}^n$ then $|f(x) - f(c) - L(x - c)| \le \epsilon|x - c|$.

If $u = x - c$, then for all $\epsilon > 0$, there exists some $\delta = \delta(\epsilon) > 0$ such that for $u \in \mathbb{R}^m, |u| \le \delta(\epsilon)$. Then we have

$$|f(u + c) - f(c) - L(u)| \le \epsilon|u| \tag{4}$$

2.1 Example

Since L is linear, $f(c) + L(x - c) = f(c) - L(c) + L(x)$ and we have

$$\lim_{|u| \to 0} \frac{|f(u + c) - f(c) - L(u)|}{|u|} = 0$$

QUESTION 2.1. I don't understand why this is the case or what its significance is?

Remark. on the tangent map The tangent map sort of creates a new, linear space at the point of a function. It is like taking the behavior of the function at that point, and making a whole new (linear) vector space at that point.

Proposition 2.0.1. *A function f has at most one derivative at a point.*

Proof. Let $L(v)$ be a differential of f at x_0. Then by the properties of the limits of composite functions and continuity, we have

$$\lim_{t \to 0} \frac{f(x_0 + tv) - f(x_0)}{t} - L(v) = \lim_{t \to 0} \frac{f(x_0 + tv) - f(x_0) - L(tv)}{t} = 0$$

Then f has a directional derivative in the direction v and $\frac{\partial f}{\partial v}(x_0) = L(v)$, implying the uniqueness of L by the uniquness of the directional derivative. □

2.2 An Example

If f is linear, for $A \in \mathbb{R}^n, f : A \to \mathbb{R}^m, c \in A, x \in A$, then we have

$$f(x) - f(c) - f(x - c) = f(x) - f(c) - f(x) + f(c) = 0$$

, implying $f(x) - f(c) = f(x - c)$ and that f is differentiable (since we can put $L(u) = f(u)$ for $u = x - c$) and is itself.

2.3 An more sophisticated way to obtains the gradient vector

We need to build up some groundwork first, and familiarize ourselves with a theorem.

Definition 2.3. (Linear Functional). f is a **linear functional** from $\mathbb{R}^n$ to $\mathbb{R}$ if $f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$ for $\alpha, \beta \in \mathbb{R}$ and $x, y \in \mathbb{R}^n$.

More generally, f is a **linear functional** from a vector field V to its field F if the functional requirement is met.

We also define an **algebraic dual space** as the space of all linear functionals from V to its field F. Note that this dual space is also a vector space in its own right (meaning that linear functionals themselves can be summed and scaled).

The Riez representation theorem **characterizes** all the linear functionals, i.e. the dual space, when the associated vector space is a Hilbert Space.

Definition 2.4. (Hilbert space) A vector space is called **Hilbert space** if it has an inner product defined on it and it is complete.

Theorem 2.1. *(Riesz representation theorem in $\mathbb{R}^n$) For any linear functional $f : \mathbb{R}^n \to \mathbb{R}$ there exists a vector $x \in \mathbb{R}^n$ such that $f(y) = < x, y >$. Conversely, if we define $f(y) = < x, y >$ for some $x \in \mathbb{R}^n$ then f is a linear functional.*

Remark. This essentially means that for each functional in a dual space, there is some vector in the vector space that you could pair via inner product to the operand of the functional to achieve the same effect as the functional.

We see this in the connection between the gradiant and the derivative: $f'(y) = < \nabla f(x), y >$. In this case, the derivative is the functional, and its operand is being dotted with the gradient vector to achieve the same effect.

This also shows how ∇f has its meaning derived from its properties as a covector (linear functional).

Proposition 2.1.1. $L(\vec{h}) = \vec{h} \cdot \vec{x}_n, \forall \vec{h} \in \mathbb{R}^n$. *We will show that this $\vec{x}$ is actually the gradient vector, showing that the linear tangent map L is equivalent to the notion of a multivariable derivative.*

Proof. Since $L(\vec{h})$ is a linear functional, by the Riesz representation theorem for bounded linear functionals, there exists some vector call it $\vec{x_L} \in \mathbb{R}^n$, such that $L(\vec{h}) = \vec{h} \cdot \vec{x_L}$.

By the definition of $L(\vec{h})$, $L(\vec{h}) = \lim_{h \to 0} \frac{f(y+h) - f(h)}{|h|} = D_h f(y) = \nabla f(y) \cdot h = \vec{h} \cdot x_L$.

Thus, $x_L = \nabla f(y)$ and the linear tangent map at h is equivalent to the directional derivative. $\square$

QUESTION 2. According to our notes, this only applies if $L(h)$ is bounded. How can we show that it is bounded?

QUESTION 3. According to the internet, a Hilbert space is a complete vector space with an inner product. In our notes, we call that a Banach Space. What is the distiction?

2.4 Examples of Inner Products

Definition 2.5. (Inner product) Some (thing?) map/function/operation is an inner product if it is:

- It is linear

- It is symmetric

- If one element is 0, the output is 0

1.
$$\cdot : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$

defined as $x \cdot y = \sum_{i=1}^n x_i y_i = x^T y$

QUESTION 4. How does this translate to the linear functional? What is the functional and what is the x_L? I thought that maybe the vector was x^T and the dot operator was itself the functional, but maybe it's the other way around?

2. Generalizing, let G be a symmetric square matrix ($G = G^T$) that is singular (i.e. $\det G \neq 0$, it has non-zero eigenvalues, and it is non-invertible). Then, for $x, y \in \mathbb{R}^n$ put $(x \cdot y)_G = x^T G y$ (row vector × square matrix × column vector goes to $\mathbb{R}$).

Proposition 2.1.2. $x \cdot y = x^T G y$ *is an inner product.*

Remark. In 1 $G = \text{Id}$. In $\mathbb{R}^n \times \mathbb{R}^n$, a square matrix G allows us to generate all possible inner products.

QUESTION 5. Please explain how this is an inner product and what this means?

3. Let V be the space of all continuous functions on $(-\pi, \pi]$. Then for $f, g \in V, < f, g >= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) g(x) dx$. Then $f \cdot g =< f, g >= V \times V \to \mathbb{R}$ is an inner product.

2.5 Example of a Bounded Linear Functional

Let X be a Banach Space of functions continuous in $[a, b]$ with norm as the sup norm. For each $x \in X$, let $G(x)(t) = \int_a^t x(s)ds$. We seek to show that G is bounded and to find its norm.

To show that it is bounded, we need to find M such that $\forall x \in X, |G(x)|_\infty \leq M|x|_\infty$

$$\left| \int_a^t x(s)ds \right| = \int_a^t |x(s)ds|$$
$$\leq |t - a||x|_\infty$$
$$\leq (b - a)|x|_\infty$$
$$\implies |G(x)|_\infty \leq (b - a)|x|_\infty$$

The norm, being the sup norm, comes from the lowest possible M:

$$|G(x)|_\infty = b - a$$

QUESTION 6. Why do we have $|G(x)|_\infty$? What's the ∞ mean/for?

6

3 Big-O and Little-o Notation

These are tools we can use to compare the "size" of functions. More accurately, they compare the relative growth rates of functions.

Definition 3.1. We write $f(x) = O[h(x)], x > x_1, x_1 \geq x_0$ if there is some constant $c > 0$ and some $x_1 \in \mathbb{R}$ such that $|f(x)| \leq ch(x)$ when $x > x_1$. f is said to be **Big-O** of h.

3.1 Examples

1. $\sin(x) = O(1)$

2. $x^2 \sin(x) = O(x^2)$

3. $\frac{\sin x}{x} = O(\frac{1}{x})$

3.2 Properties of Big-O Function

1. If f, g are both $O[h(x)]$ for $x > x_1$ then so is the sum $f(x) + g(x)$.

2. If $f = O(g)$ and $g = O(h)$ then $f(x) = O(h)$.

3. If $f(x) = O[h(x)]$ and $g(x) > 0$ then $f(x)g(x) = O[h(x)g(x)]$.

Definition 3.2. We write $f(x) = o[h(x)]$ for all $x > x_1$ if $\lim_{x \to x_1} \frac{f(x)}{h(x)} = 0$. Note that x_1 can go to infinity, which implies that f is dominated by h.

3.3 Examples

$\log^2 x = o(x)$ as $x \to \infty$ because $\lim_{x \to \infty} \frac{\log^2 x}{x} = 0$ via LHR.

3.4 Relavance to the linear map

If $f : A \subset \mathbb{R}^2 \to \mathbb{R}$ is differentiable at $x_0 \in A$ then

$$f(u_0 + h, v_0 + k) = f(u_0, v_0) + \frac{\partial f}{\partial u}(u_0, v_0)h + \frac{\partial f}{\partial v}(u_0, v_0)k + o(\sqrt{h^2 + k^2})$$

.

4 Jacobian

Definition 4.1. For a scalar function $f : A \subset \mathbb{R}^n \to \mathbb{R}$, the **Jacobian** is

$$Df(x_0) = \left(\frac{\partial f}{\partial x^1}, \frac{\partial f}{\partial x^2} + \cdots + \frac{\partial f}{\partial x^n} \right)\bigg|_{x_0} = \nabla f(x_0)$$

So we can say that $df_{x_0}(\vec{h}) = \vec{h} \cdot \nabla f = Df(\vec{x_0})\vec{h}$.

Remark. This means that the derivative in the direction of $\vec{h}$ of f at x_0 is equivalent to $\vec{h}$ dotted widh the Jacobian of f (the gradient when f maps to a single dimensional space).

4.1 Properties of ∇ operator

- $\nabla(uv) = u\nabla v + v\nabla u$. This is basically the product rule.

- $\nabla u^n = nu^{n-1}\nabla u$. Chain rule.

Definition 4.2. For a vector-valued function $f : \mathbb{R}^n \to \mathbb{R}^m (n, m \in \mathbb{Z}^{1+})$ the **Jacobian** $Df(x_0)$ is

$$Df(x_0) = \frac{\partial f^i}{\partial x^i}(x_0) = \begin{bmatrix} \frac{\partial f^1}{\partial x^1} & \frac{\partial f^1}{\partial x^2} & \cdots & \frac{\partial f^1}{\partial x^n} \\ \frac{\partial f^2}{\partial x^1} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{\partial f^m}{\partial x^1} & \cdots & \cdots & \frac{\partial f^m}{\partial x^n} \end{bmatrix}\Bigg|_{x_0}$$

8

Then the tangent matrix is given by

$$L(h) = Df(x_0)h = \begin{bmatrix} \nabla f^1 \cdot h \\ \nabla f^2 \cdot h \\ \vdots \\ \nabla f^m \cdot h \end{bmatrix}\Bigg|_{x_0}$$

Remark. In the case that $n = m$, the Jacobian will be square and $\det Df(x_0) = \frac{\partial(f^1, f^2, \cdots, f^m)}{\partial(x^1, x^2, \cdots x^n)}$.

QUESTION 7. Can you explain more the intuition of the linear map? Also, what is the importance (or even the meaning) of the statement about the determinant?

7

5 Hessian Matrices

Definition 5.1. If A is an open set in $\mathbb{R}^n$, and $f : A \to \mathbb{R}$ is differentiable at $x_0 \in A$, and each first derivative has partial derivatives at x_0, we say f has second derivatives at x_0. If the coordinates of $\mathbb{R}^n$ are given as $(x^1, x^2, \cdots x^n)$, then the partials are

$$\frac{\partial^2 f}{\partial x^i \partial x^j} \quad D_i D_j f(x_0) \quad D_{ij} f(x_0)$$

. Then the matrix of second partial derivatives is called the **Hessian** and is given by

$$\mathbf{H}f(x_0) = [D_i D_j f(x_0)]$$

.

QUESTION 8. What happens if f is from $\mathbb{R}^n \to \mathbb{R}^m, n, m > 1$? Even in the case of our matricised version of the second derivative, an $n \times n$ Hessian is still only for a function that outputs to a sigle dimension.

5.1 Examples

1. For $f : \mathbb{R}^2 \to \mathbb{R}$

$$\mathbf{H}f(x_0) = \begin{bmatrix} \frac{\partial^2 f}{\partial_{xx}} & \frac{\partial^2 f}{\partial_{xy}} \\ \frac{\partial^2 f}{\partial_{yx}} & \frac{\partial^2 f}{\partial_{yy}} \end{bmatrix}$$

5.2 Critical Points

Definition 5.2. If A is on an open set in $\mathbb{R}^n$, $f : A \to \mathbb{R}$ is differentiable in A then the points $x \in A$ such that $df_x = 0$ are **critical points** of f in A.

For $x_0 \in A$ a critical point

$$f(x_0 + h) = f(x_0) + \frac{1}{2}(\mathbf{H}f(x_0)h) \cdot h + o(|h|^2)$$

QUESTION 9. Why is the above statement true? Where does it come from?

Proposition 5.0.1. *If x_0 is a local minimizer (local minimum), then $\boldsymbol{H}f(x_0)\gamma \cdot \gamma \geq 0, \forall \gamma \in \mathbb{R}^n$.*

Proof. If x_0 is a local minimizer of f we have

$$\begin{aligned} 0 &\leq f(\vec{x_0} + \lambda\vec{\gamma}) - f(\vec{x_0}) \\ &= \frac{1}{2}\mathbf{H}f(\vec{x_0})(\lambda\vec{\gamma}) \cdot \lambda\vec{\gamma} + o(\lambda^2) \\ &= \frac{1}{2}\mathbf{H}f(\vec{x_0})(\vec{\gamma}) \cdot \vec{\gamma} + o(1) \end{aligned}$$

$\square$

Remark. If x_0 is a local minimizer, then the eigenvalue of $\mathbf{H}f(x_0)$ are non-negative. If the eigenvalues are positive, then x_0 is an isolated minimizer. That motivates the following discussion on spectral decomposition.

QUESTION 10. Explain the meaning of this in terms of $\mathbb{R}^3$ and what happens if x_0 is a local maximizer?

6 Spectral Decomposition

Note that a real-valued, symmetric matrix A has real eigenvalues. Also, not that eigenvectors corresponding to distinct eigenvalues are $\perp$. That is, $u_i \cdot u_j = 0, \lambda_i \neq \lambda_j$. In this case, A may be decomposed into a characteristic form such that

$$A = \sum_i \lambda_i \mathcal{P}_i$$

where λ_i is the i-th eigenvalue.

Here, $\mathcal{P}_i$ is called the projector and has the properties $\mathcal{P}_i^2 = \mathcal{P}_i$ and $\mathcal{P}_i = u_i u_i^T$ where u_i is the i-th eigenvector. Also, $\sum_i \mathcal{P}_i = \text{Id}$.

This leads to

$$A\vec{h} = \sum_i \lambda_i \mathcal{P}_i \vec{h} = \sum_i \lambda_i \vec{u}_i \vec{u}_i^T \vec{h}$$

6.1 Example spectral decomposition

Let $X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Note that X is symmetric and real-valued, and thus has i) real-valued eigenvalue and ii) is an orthogonal matrix (i.e. its inverse is its transpose).

We have eigenvalues $\lambda_1 = 1, \lambda_{-1} = -1$. The eigenvectors are $u_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, u_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

From the eigenvectors we can get the projectors

$$\mathcal{P}_1 = \frac{1}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},$$

$$\mathcal{P}_{-1} = \frac{1}{2} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

$$\mathcal{P}_1 + \mathcal{P}_{-1} = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \text{Id}$$

Furthermore,

$$X = \sum_i^{n=2} \lambda_i \mathcal{P}_i = (1) \left(\frac{1}{2} \right) \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + (-1) \left(\frac{1}{2} \right) \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

Remark. 1. The diagonalization of X is given by $Q\Lambda Q^T$ where Q is the column matrix formed of the eigenvectors and Λ is the matrix formed by taking the eigenvalues on the main diagonal.

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

2. If $A = \sum_i u_i \mathcal{P}_i$, suppose $f(t) = \sum_m a_m t^m, m > 0$ and $f(A) = \sum_m f(\lambda_i)\mathcal{P}_i, f = \sqrt{A}$. Then

$$X^{1/2} = 1 \cdot \mathcal{P}_1 + i\mathcal{P}_{-1} = \frac{1}{2} \begin{bmatrix} 1+i & 1-i \\ 1-i & 1+i \end{bmatrix}$$

11

such that $\left(X^{1/2}\right)^2 = X$

QUESTION 11. What do the various f's refer to? Are they the same function or are they nested? When is the product of projectors a projector? What are the eigenvalues of projectors? Are projectors generally invertible?

7 Mean Value Theorems

Recall the one-dimensional Mean Value Theorem...but first, Rolle's Theorem:

Lemma 7.1. *(Rolle's Theorem) Suppose that f is continuous on closed interval $[a, b]$ and f' exists on (a, b) and $f(a) = f(b) = 0$. Then there exists $c \in (a, b)$ such that $f'(c) = 0$.*

Theorem 7.2. *(Mean value theorem) Suppose f is continuous on $[a, b]$, f' exists on (a, b). Then there exists $c \in (a, b)$ such that $f(b) - f(a) = f'(c)(b - a)$.*

(Generalized mean value theorem) Let f, g be continuous on $[a, b]$ with f', g' existing on (a, b). Then there exists a $c \in (a, b)$ such that $f'(c)[g(b) - g(a)] = g'(c)[f(b) - f(a)]$

Proof. Put $h(x) = [f(b) - f(a)]g(x) - [g(b) - g(a)]f(x)$ Then $h(a) = f(b)g(a) - g(b)f(a) = h(b)$. We need to show that $h'(x) = 0$ for some $x \in (a, b)$ since $h'(x) = f'(c)[g(b) - g(a)] = g'(c)[f(b) - f(a)]$.

If $h(x)$ is constant, $h'(x) = 0, \forall x \in (a, b)$. If there exists some $x_1 > a$ such that $h(x_0) > h(a)$, take x' as the value where h attains its maximum. Then by Rolle's Theorem (and use of another function $\psi(x) = h(x) - \max\{h(a), h(b)\}$) $h(x') = 0$. Else, take x' as the value where h attains its minimum and do the same. $\qquad\square$

Remark. If f is differentiable, do we generally have that $f(b) - f(a) = Df(c)(b - a)$? No we don't.

7.1 Example

For $f(x) = (x - x^2, x - x^3), a = 0, b = 1$, $f(a) = 0 = f(b)$. However, $Df = \nabla f \cdot \vec{u} = (1 - 2x, 1 - 3x^2)$. So we would need

$$\langle 0, 0 \rangle = \langle 1 - 2x, 1 - 3x^2 \rangle \cdot \vec{u}$$

QUESTION 7.1. Wouldn't $u = \langle 0, 0 \rangle$ work for this?

7.2 Mean value theorem for multivariable functions

We begin by "flattening" the situation to a single dimension, along a line.

Proposition 7.2.1. *Let $\Omega \subset \mathbb{R}^n$ be open and $f : \Omega \to \mathbb{R}$. Suppose that Ω contains a, b and line S between them. Suppose that f is differentiable at every point of this segment.*
Then there exists some $c \in S$ such that $f(b) - f(a) = Df(c)(b - a)$.

Proof. Let $\psi : \mathbb{R} \to \mathbb{R}^n$, $\psi(t) = (1 - t)\vec{a} + t\vec{b} = a + t(b - a)$. Then $\psi'(t) = b - a$. Then $\psi(0) = a, \psi(1) = b$.
 Let $F(t) = f(\psi(t)) = f((1 - t)a + tb)$. By the chain rule

$$F'(t) = f'(\psi(t))\psi'(t)$$
$$= f'(\psi(t))(b - a)$$

Now, we apply the 1-Dimenstional MVT to F to obtain $F(1) - F(0) = F'(t_0)$ for some $t_0 \in (0, 1)$.
 Then let $c = \psi(t_0)$.

$$f(b) - f(a) = F(1) - F(0)$$
$$= F'(t_0)$$
$$= f'(c)(b - a)$$

$\square$

7.3 Extension of the Mean Value Theorem

Let $\Omega \subset \mathbb{R}^n$ be open, $f : \Omega \to \mathbb{R}^m$. Suppost that Ω contains a, b and a line segment S between them. Suppost that f is differentiable on every point on S.
 Then there is a point c on S such that $|f(b) - f(a)| \leq |\mathbf{D}f(c)(b - a)|$

QUESTION 7.2. How is this different from the last proposition? Also, how can you guarantee that $\|y\| = 1$ in the next line? Is this theorem just a corrected version of the last one?

Proof. Let $y = \frac{f(b)-f(a)}{|f(b)-f(a)|}$. $\|y\| = 1$ Put $h(x) = f(x) \cdot \vec{y} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix} \begin{bmatrix} y_1 & y_2 & \cdots & y_m \end{bmatrix}$ Then

$h(b) - h(a) = (f(b) - f(a)) \cdot \vec{y} = |f(b) - f(a)|$ Then $\mathbf{D}h(x) = \mathbf{D}f(x) \cdot \vec{y}$. From the 1-Dimensional MVT, you proceed

$$h(b) - h(a) = \mathbf{D}h(c)(b-a)$$
$$|f(b) - f(a) = \mathbf{D}f(c)(b-a) \cdot y$$
$$= |\mathbf{D}f(x)(b-a)|\|y\|$$
$$\leq |\mathbf{D}f(x)(b-a)| \qquad \text{(Since } x \cdot y \leq |x|\|y|)$$

$\square$

8 Derivatives of Composite Functions & Chain Rule

Definition 8.1. For functions F and G such that the range of F is in the domain of G, the composition $G \circ F$ may be defined as $(G \circ F)(x) = (G(F(x)))$.

Generally $G \circ F \neq F \circ G$.

However, if $H \circ (G \circ F)$ and $(H \circ G) \circ F$ can be determined, then $H \circ (G \circ F) = (H \circ G) \circ F$.

Definition 8.2. (Recall linear maps) For all $\epsilon > 0 \exists \delta(\epsilon, f)$ such that for $\vec{x} \in \mathbb{R}^n, |\vec{x} - \vec{c}| \leq \delta(\epsilon, f) \implies |f(x) - f(c) - L_f(\vec{x} - \vec{c})| \leq \epsilon|x - c|$ for linear map L_f.

Theorem 8.1. *(Chain Rule) Let f have domain $A \Subset \mathbb{R}^n$ and range in $\mathbb{R}^m$. Let g have domain $B \Subset \mathbb{R}^m$ and range in $\mathbb{R}^n$. Suppose that f is differentiable at c and that g is differentiable at $b = f(c)$.*

Then the composition $h = g \circ f$ is differentiable at c and $\mathbf{D}(h(c)) = \mathbf{D}g(b) \circ \mathbf{D}f(c)$.

Proof. Let $\epsilon > 0$. By continuity of f, $|f(x) - f(c)| \leq k(x - c), k > 0$ when $|x - c| < \gamma, \gamma > 0$.

Let $L_f = Df(c), L_g = Dg(b)$. Since these are linear maps, there exists some M such that, for $u \in \mathbb{R}^m$

$$|L_g(u)| \leq M|u|, \qquad (*)$$

Now, require $|x - c| \leq \min(\gamma, \frac{1}{k}\delta(\epsilon, g))$. Then $|f(x) - f(c)| \leq \delta(\epsilon, g)$ from the continuity. Obtain then $|g(f(x)) - g(f(c)) - L_g[f(x) - f(c)]| \leq \epsilon|f(x) - f(c)|$.

Require further $|x - c| \leq \delta(\epsilon, f)$. Then

$$|L_g[f(x) - f(c) - L_f(x - c)] \leq \epsilon M|x - c| \qquad \text{(from } *)$$

Now put $\delta_1 = \min(\gamma, \frac{1}{k}\delta(\epsilon, g), \delta(\epsilon, f))$.

Then

$$|g(f(x)) - g(f(c)) - L_g(L_f(x - c)) \leq \epsilon(k + M)|x - c|$$
$$|g \circ f(x) - g \circ f(c) - L_g L_f(x - c)| \leq \epsilon(k + M)|x - c|$$
$$\implies \mathbf{D}h(c) = L_g \circ L_f$$

$\square$

9 Taylor's Theorem

9.1 Preliminaries

Proposition 9.0.1. *(Leibnitz formula for the nth derivative of a product) Let $h = fg$. Then*

$$h^{(n)}(x) = \sum_{k=0}^{n} \binom{n}{k} f^{(n)}(x) g^{(n-k)}(x)$$

where the binomial coefficient $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

Proof. By induction. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Theorem 9.1. *(Taylor) Let f have $f^{(n)}$ everywhere in (a,b) that $f^{(n-1)}$ continuous on $[a,b]$, and for every $x \in [a,b]$ there is an $x_1(c,x)$. Then*

$$f(x) = f(c) + \underbrace{\sum_{k=1}^{n-1} \frac{f^{(k)}(c)}{k!}(x-c)^k}_{(n-1)\,degree\ approx.\ polynomial} + \underbrace{\frac{f^{(n)}(x_1)}{n!}(x-c)^n}_{error\ term} \qquad (5)$$

Proof. Using another theorem

Theorem 9.2. *Let f, g have n derivatives in (a,b) and $n-1$ derivatives in $[a,b]$. Assume $c \in [a,b]$. Then for $x \in [a,b]$ there is a point x_1 between x and c such that*

$$\left[f(x) - \sum_{k=0}^{n-1} \frac{f^{(k)}(c)}{k!}(x-c)^k \right] g^{(n)}(x_1) = f^{(n)}(x_1) \left[g(x) - \sum_{k=0}^{n-1} \frac{g^{(k)}(c)}{k!}(x-c)^k \right]$$

For Taylor's Theorem, we consider the special case $g(x) = (x-c)^n$ so that $g^{(j)} = (n)(n-1)\cdots(n-j+1)(x-c)^{n-j}$. Then $g^{(i)}(x)$ evaluated at $x = c$ is always 0, and $g^{(n)}(x) = n!$ for all x.

As a result,

$$\left[f(x) - \sum_{k=0}^{n-1} \frac{f^{(k)}(c)}{k!}(x-c)^k \right] n! = f^{(n)}(x_1)(x-c)^n$$

$$\implies f(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(c)}{k!}(x-c)^k + \frac{f^{(n)}(x_1)}{n!}(x-c)^n$$

Proceeding with the main proof:

Put

$$F(t) = f(t) + \sum_{k=1}^{n-1} \frac{f^{(k)}(t)}{k!}(x-t)^k$$

$$G(t) = g(t) + \sum_{k=1}^{n-1} \frac{g^{(k)}(t)}{k!}(x-t)^k$$

15

Since F, G have finite derivatives and are continuous on the interval, theorem 7.2 (generalized mean value theorem) applies and we get

$$F'(t)[G(x) - G(c)] = G'(t)[F(x) - F(c)]$$

Consider $F'(t)$.

$$F'(t) = f'(t) + \sum \frac{1}{k!}\left(f^{(k+1)}(t)(x-t)^k + f^{(k)}(t)(-k)(x-t)^{k-1}\right)$$

(By chain and product rule)

$$= f'(t) + \sum \frac{1}{k!}f^{(k+1)}(t)(x-t)^k - \sum \frac{1}{k!}f^{(k)}(t)(k)(x-t)^{k-1}$$

$$= f'(t) + \sum_{}^{n-1} \frac{1}{k!}f^{(k+1)}(t)(x-t)^k - \sum_{}^{n-1} \frac{1}{k-1!}f^{(k)}(t)(x-t)^{k-1}$$

$$= f'(t) + \sum_{k=1}^{n-1} \frac{1}{k!}f^{(k+1)}(t)(x-t)^k - \sum_{k=0}^{n-2} \frac{1}{k!}f^{(k+1)}(t)(x-t)^{k-1}$$

$$= \frac{f^{(n)}(t)}{(n-1)!}(x-t)^{n-1}$$

Similarly

$$G'(t) = \frac{f^{(n)}(t)}{(n-1)!}(x-t)^{n-1}$$

Plugging these back into the original equation, and using $t = x$, we get

$$F'(t)[G(x) - G(c)] = G'(t)[F(x) - F(c)]$$

$$\frac{f^{(n)}(t)}{(n-1)!}(x-t)^{n-1}[G(x) - G(c)] = \frac{f^{(n)}(t)}{(n-1)!}(x-t)^{n-1}[F(x) - F(c)]$$

$\square$

QUESTION 7.5. *In the notes, we finish this proof by settings $t = x$. However, that will negate these derivatives.*

9.2   Taylor's Theorem for Functions from $\mathbb{R}^n$ to $\mathbb{R}$

We want to get a formula that is somewhat similar to (5).

9.2.1 Multiindices

We start by defining a new notation calling **multiindices**.

Definition 9.1. Let $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_n), \alpha_j \geq 0$. We call α and **multiindex** and define the following operations on α:

- The norm: $|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_n$

- The factorial: $\alpha! = \alpha_1! \alpha_2! \cdots \alpha_n!$

- The monomial term: Let $x = (x_1, x_2, \cdots, x_n)$. Then $x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$

- Coefficients: Let C_α be a coefficient that depetsn on $\{\alpha_i\}$.

- Polynomials: $P(x) = \sum_{|\alpha|=m} C_\alpha x^\alpha$. So for $m = 2$, there are three possibilities for the terms: $\alpha_1, \alpha_2 = 1, 1; 2, 0; 0, 2$.

Letting $x \in \mathbb{R}^n$, and supposing that the needed partials exist at x and that $t = (t_1, t_2, \cdots, t_n) \in \mathbb{R}^n$.

Then we write

$$f'(\vec{x}, \vec{t}) = \sum_{i=1}^{n} D_i f(x) t_i$$

$$f''(\vec{x}, \vec{t}) = \sum_{i=1}^{n} \sum_{j=1}^{n} D_i D_j f(x) t_i$$

$$f'''(\vec{x}, \vec{t}) = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} D_i D_j D_k f(x) t_i$$