Calculus - Differentiation in Several Variables

Andrew Andreas

July 31, 2024

1 Introduction

Please report any errors to andrew.andreas0@gmail.com

2 Limits

Definition. Limit of a function Let $\mathbf{f} : X \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be a function then the limit of the function $\mathbf{f}$

$$\lim_{\mathbf{x} \longrightarrow \mathbf{a}} \mathbf{f}(\mathbf{x}) = \mathbf{L}$$

means that for any $\epsilon > 0$ you can find a $\delta > 0$ such that if $\mathbf{x} \in X$ and $0 < ||\mathbf{x} - \mathbf{a}|| < \delta$ then $||\mathbf{f}(\mathbf{x}) - \mathbf{L}|| < \epsilon$

Example 1. Consider the limit of the function below

$$\lim_{(x,y,z) \longrightarrow (1,-1,2)} (3x - 5y + 2z) = 12$$

According to the definition of the limit, then for any $\epsilon > 0$ there must exist $\delta > 0$ such that $0 < ||\mathbf{x} - \mathbf{a}|| < \delta$

$$\begin{aligned} 0 &< ||(x, y, z) - (1, -1, 2)|| < \delta \\ 0 &< ||(x - 1, y + 1, z - 2)|| < \delta \\ 0 &< \sqrt{(x-1)^2 + (y+1)^2 + (z-2)^2} < \delta \end{aligned}$$

and since all quantities are positive, we know that each component is less than δ

$$\begin{aligned} \sqrt{(x-1)^2} &= |x - 1| < \delta \\ \sqrt{(y+1)^2} &= |y + 1| < \delta \\ \sqrt{(z-2)^2} &= |z - 2| < \delta \end{aligned}$$

Then lets consider $|f(\mathbf{x}) - L| < \epsilon$

$$|(3x - 5y + 2z) - 12| < \epsilon$$

If we rewrite the ϵ inequality and use the triangle inequality $|a + b| \leq |a| + |b|$

$$\begin{aligned} |3x - 5y + 2z - 12| &= |3(x-1) - 5(y+1) + 2(z-2)| \\ &\leq |3(x-1)| + |5(y+1)| + |2(z-2)| \\ &\leq 3|(x-1)| + 5|(y+1)| + 2|(z-2)| \\ &< 3\delta + 5\delta + 2\delta \\ &= 10\delta \leq 10\frac{\epsilon}{10} = \epsilon \end{aligned}$$

So that $\delta \leq \frac{\epsilon}{10}$. We can see then that this then satisfies the proof.

2.1 Intuition of the Limit Definition

Intuitively, we can think about the definition of a limit as an experiment where $\mathbf{x} \in X$ represents measurement of some independent variable and $\mathbf{f}(\mathbf{x})$ represents the measurement of the outcome of the experiment. We want to know that for a chosen experimental error ϵ, what error in measurement δ will be tolerated?

Specifically, imagine we are running an experiment in a chemistry lab where we wish to measure the viscosity of some liquid $f(\mathbf{x})$ that is the result of a chemical reaction between equal volumes of A and B (assume we are given B in the correct volume, $x_B = v_B$).

$$A + B \longrightarrow f(\mathbf{x})$$

We know from reference material that the viscosity of the liquid is L. Then we wish to understand for a chosen experimental error ϵ, to what extent can we inaccurately measure the volume of A, v_A? We represent the measurement error as $|x_A - v_A|$.

Then for some experimental error $|f(\mathbf{x}) - L| < \epsilon$ there should exist some δ which satisfies the error in measuring out $A, 0 < |x_A - v_A| < \delta$

2.2 Sets

Definition. Closed ball A closed ball is defined as the set of points $\mathbf{x} \in \mathbb{R}^n$ that fall within the space enclosed by the ball with radius r and centred at $\mathbf{a}$

$$||\mathbf{x} - \mathbf{a}|| \leq r$$

Definition. Open ball An open ball is defined as the set of points $\mathbf{x} \in \mathbb{R}^n$ that fall within the space enclosed by the ball with radius *less than* r and centred at $\mathbf{a}$

$$||\mathbf{x} - \mathbf{a}|| < r$$

Definition. Open Set A set $X \subseteq \mathbb{R}^n$ is said to be **open** in $\mathbb{R}^n$ if for any point $\mathbf{x} \in X$ there is an **open ball** centred at $\mathbf{x}$ that lies entirely within X

Definition. Boundary Point A point is said to be in the boundary of the set $X \subseteq \mathbb{R}^n$ if every open ball centred at $\mathbf{x}$ contains both elements of X and $/\{X\}$

Definition. Closed Set A set $X \subseteq \mathbb{R}^n$ is said to be **closed** in $\mathbb{R}^n$ if it contains all its boundary points

Definition. Neighbourhood The neighbourhood of a point $\mathbf{x} \in X$ is an open set containing $\mathbf{x}$ and is contained within the set X

2.3 The geometric interpretation of a Limit

Given $\epsilon > 0$ then you can find a corresponding $\delta > 0$ such that if points $\mathbf{x} \in X$ are inside an open ball of radius δ, then $\mathbf{f}(\mathbf{x})$ will remain inside an open ball of radius ϵ centred at $\mathbf{L}$

2.4 Properties of Limits

Theorem. Uniqueness of Limits If a limit exists, it is unique. That is, let $\mathbf{f} : X \subseteq \mathbb{R}^n \to \mathbb{R}^m$. If $\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{f}(\mathbf{x}) = \mathbf{L}$ and $\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{f}(\mathbf{x}) = \mathbf{M}$, then $\mathbf{L} = \mathbf{M}$

Theorem. Algebraic Properties Let $\mathbf{F}, \mathbf{G} : X \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be vector-valued functions and $f, g : X \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$ be scalar-valued functions, and let $k \in \mathbb{R}$ be a scalar

1. If $\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{F}(\mathbf{x}) = \mathbf{L}$ and $\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{G}(\mathbf{x}) = \mathbf{M}$ then $\lim_{\mathbf{x}\to\mathbf{a}} (\mathbf{F} + \mathbf{G})(\mathbf{x}) = \mathbf{L} + \mathbf{M}$
2. If $\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{F}(\mathbf{x}) = \mathbf{L}$ then $\lim_{\mathbf{x}\to\mathbf{a}} k\mathbf{F}(\mathbf{x}) = k\mathbf{L}$
3. If $\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = L$ and $\lim_{\mathbf{x}\to\mathbf{a}} g(\mathbf{x}) = M$, then $\lim_{\mathbf{x}\to\mathbf{a}}(fg)(\mathbf{x}) = LM$
4. If $\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = L, g(\mathbf{x}) \neq 0$ for $\mathbf{x} \in X$, and $\lim_{\mathbf{x}\to\mathbf{a}} g(\mathbf{x}) = M \neq 0$, then $\lim_{\mathbf{x}\to\mathbf{a}}(f/g)(\mathbf{x}) = L/M$

Theorem. Suppose $\mathbf{f} : X \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is a vector-valued function. Then $\lim_{\mathbf{x}\longrightarrow\mathbf{a}} \mathbf{f}(\mathbf{x}) = \mathbf{L}$, where $\mathbf{L} = (L_1, L_2, ..., L_m)$, if and only if $\lim_{\mathbf{x}\longrightarrow\mathbf{a}} f_i(\mathbf{x}) = L_i$ for $i = 1, ..., m$

2.5 Continuous Functions

Definition. Continuous Function Let $\mathbf{f} : X \subseteq \mathbb{R}^n \to \mathbb{R}^m$ and let $\mathbf{a} \in X$. Then $\mathbf{f}$ is said to be **continuous at a** if

$$\lim_{\mathbf{x} \longrightarrow \mathbf{a}} \mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{a})$$

Simply this means that if a function is continuous, then the limit of a function as it approaches $\mathbf{a}$ is just the value of the function evaluated at $\mathbf{a}$, $f(\mathbf{a})$

If we extend the Properties of Limits in the context of continuity, then they tell us the following

1. The sum of two continuous functions $\mathbf{F} + \mathbf{G}$, where $\mathbf{F}, \mathbf{G} : X \subseteq \mathbb{R}^n \to \mathbb{R}^m$ are continuous at $\mathbf{a} \in X$, is also continuous at $\mathbf{a}$
2. For all $k \in \mathbb{R}$, the scalar multiple $k\mathbf{F}$ of a function $\mathbf{F} : X \subseteq \mathbb{R}^n \to \mathbb{R}^m$ that is continuous at $\mathbf{a} \in X$, is also continuous at $\mathbf{a}$
3. The product fg and quotient $f/g, (g \neq 0)$ of two scalar valued functions $f, g : X \subseteq \mathbb{R}^n \to \mathbb{R}$ that are continuous at $\mathbf{a} \in X$ is also continuous at $\mathbf{a}$
4. $\mathbf{F} : X \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is continuous at $\mathbf{a} \in X$ if and only if its component functions $F_i : X \subseteq \mathbb{R}^n \to \mathbb{R}, i = 1, ..., m$ are all continuous at $\mathbf{a}$

Theorem. If $\mathbf{f} : X \subseteq \mathbb{R}^n \to \mathbb{R}^m$ and $\mathbf{g} : Y \subseteq \mathbb{R}^m \to \mathbb{R}^p$ are continuous functions, such that the range of $\mathbf{f} \subseteq Y$, then the composite function $\mathbf{g} \circ \mathbf{f} : X \subseteq \mathbb{R}^n \to \mathbb{R}^p$ is also defined and continuous.

Example For functions which are not defined at certain points, these functions can still be continuous if we define the function so that at these points, they take on the limiting value. For example, consider the function

$$f(x, y) = \frac{x^3 + y^3}{x^2 + y^2}$$

We know that at the origin this function is not defined. However, if we define the function so that at the origin the function has value 0, then this function is indeed continuous.

$$f(x, y) = \begin{cases} \frac{x^3+y^3}{x^2+y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0) \end{cases}$$

3 The Derivative

Definition. Partial Function Suppose $\mathbf{f} : X \subseteq \mathbb{R}^n \to \mathbb{R}$ is a scalar-valued function, where X is open in $\mathbb{R}^n$, and let $\mathbf{x} = (x_1, x_2, ..., x_n)$ denote a point in $\mathbb{R}^n$. A partial function F with respect to the variable x_i is a single-variable function which is obtained by holding all other variables $x_{j \neq i}, j = 1, ..., n$ as constant

Definition. Partial Derivative The partial derivative of f with respect to x_i is the derivative of the partial function with respect to x_i

$$\begin{aligned} \frac{\partial f}{\partial x_i} &= \lim_{h \to 0} \frac{f(x_1, ..., x_i + h, ..., x_n) - f(x_1, ..., x_n)}{h} \\ &= \lim_{h \to 0} \frac{f(\mathbf{x} + h\mathbf{e_i}) - f(\mathbf{x})}{||h\mathbf{e_i}||} \end{aligned}$$

which is defined as the instantaneous rate of change of the function f when all other variables are held constant, except the specified variable.

The requirement of the domain X to be **Open** is a necessary condition for several reasons:

- **Local Behaviour and Neighbourhood** A neighbourhood around a given point $\mathbf{a} \in X$ must exist. If a set is closed, it will contain its boundaries and if $\mathbf{a}$ is at the boundary then there does not exist a neighbourhood around $\mathbf{a}$ that is **entirely** within X

- **Differentiability** The definition of differentiability relies on the limit of the difference quotient as **h** approaches zero. A closed set might obstruct the evaluation of the limit from all directions

- **Continuity of Partial Derivatives** In an open domain, a differentiable function is also continuous. If it were closed then the neighbourhood would exist outside the domain

Example 2. Consider $f : X \subseteq \mathbb{R}^2 \to \mathbb{R} = x^2 + y^2$, where X is open in $\mathbb{R}^2$, then its partial derivative at $(1, 1, 2)$ with respect to x and y can be visualised as per the below

- The blue tangent line represents the partial derivative with respect to x, holding y constant

- The orange tangent line represents the partial derivative with respect to y, holding x constant

The tangent lines visualise how the function changes along the lines that pass through $x = 1$ and $y = 1$ respectively.

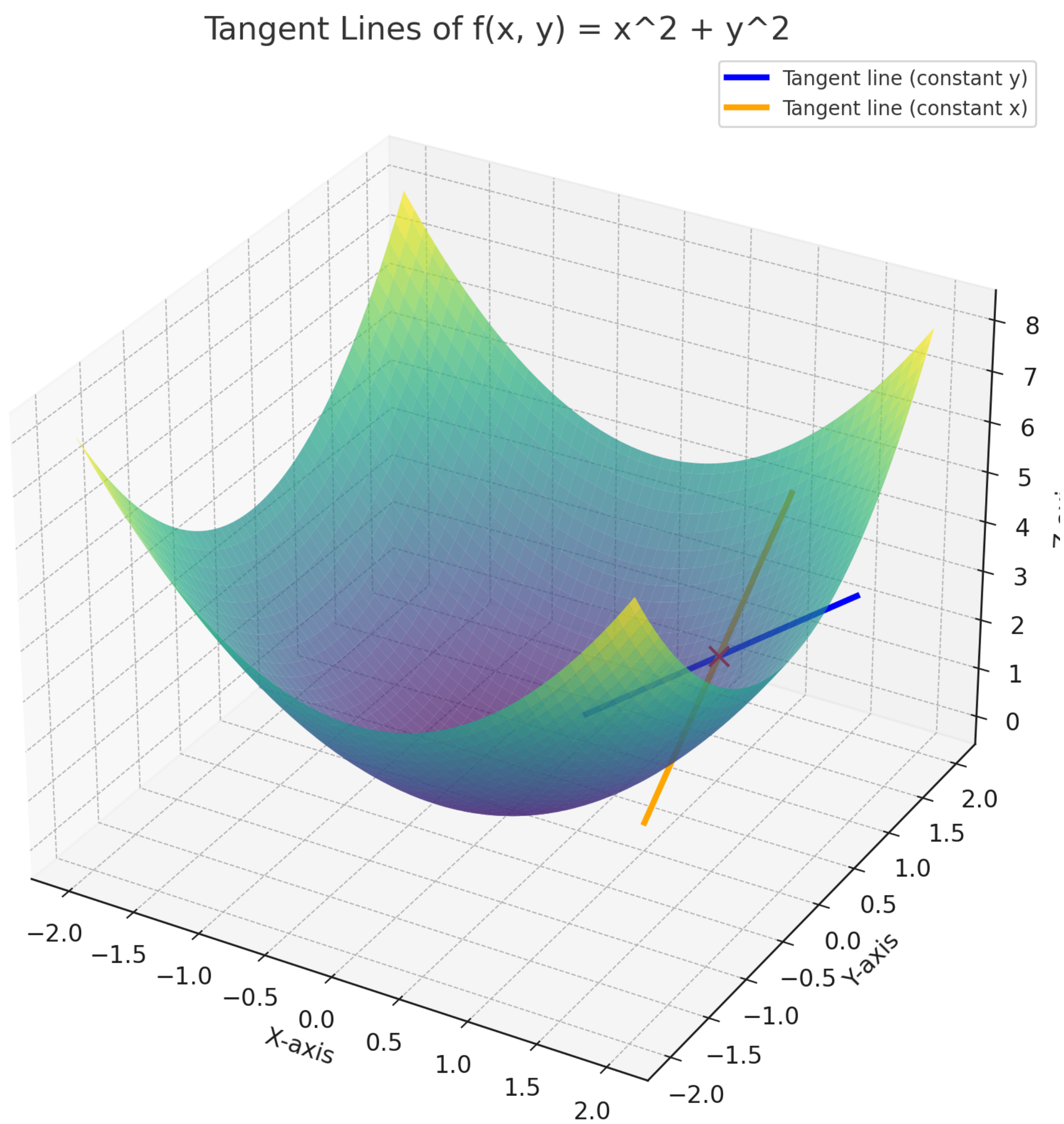

Figure 1: $f : X \subseteq \mathbb{R}^2 \to \mathbb{R} = x^2 + y^2$ with its partial derivatives at $(1, 1, 2)$

3.1 The Tangent Plane

We briefly mentioned tangent lines in Figure 1. For a scalar-valued function of a single variable, $F : X \subseteq \mathbb{R} \to \mathbb{R}$ then to have F be differentiable at a we know this means there must exist a tangent line at $\big(a, F(a)\big)$. The tangent line is given by the equation

$$y = F(a) + F'(a)(x - a)$$

This can be derived from the point slope form of a line, requiring a point which the line passes through and the slope of the line

$$m = \frac{y - y_1}{x - x_1}$$

where $m = F'(a)$, $y_1 = F(a)$ and $x_1 = a$

The **tangent hyperplane** is a generalization of the tangent line to higher dimensions.

Consider $f : X \subseteq \mathbb{R}^2 \to \mathbb{R}$ which is defined by $z = f(x, y)$ and forms a surface in 3-D. Then at the point $\big(a, b, f(a, b)\big)$, we know that the partial derivative with respect to x is the slope of the line tangent to the surface at the point $\big(a, b, f(a, b)\big)$ defined by the intersection of the plane $y = b$ with the surface $z = f(x, y)$.

Similarly, the partial derivative with respect to y is the slope of the line tangent to the surface at $(a, b, f(a, b))$ that is defined by the intersection of the plane $x = a$ with the surface $z = f(x, y)$.

We have two tangent lines and since both these lines are tangent to the surface of $z = f(x, y)$ at the point $\big(a, b, f(a, b)\big)$, then these lines must intersect and form the tangent plane in $\mathbb{R}^2$.

Vector Parametric Equation of a Line Given a line that goes through the point $P_0(\mathbf{b})$, whose position vector is $\overrightarrow{OP_0} = \mathbf{b}$ and which is parallel to $\mathbf{a}$, then the vector parametric form is

$$\mathbf{r}(t) = \mathbf{b} + t\mathbf{a}$$

which is parameterized by the variable t. The tangent lines can be written as

$$\begin{aligned}\mathbf{l}_1(t) &= \big(a, b, f(a, b)\big) + t\big(1, 0, f_x(a, b)\big)\\ \mathbf{l}_2(t) &= \big(a, b, f(a, b)\big) + t\big(0, 1, f_y(a, b)\big)\end{aligned}$$

We see directly that this means that for incremental movements in $\mathbf{i}$ and $\mathbf{j}$, this results in a change in the function value of $f_x(a, b)$ and $f_y(a, b)$, respectively in the z direction. See Figure 1 for a visualisation.

Two vectors that are parallel to these lines are

$$\begin{aligned}\mathbf{u} &= \mathbf{i} + f_x(a, b)\mathbf{k}\\ \mathbf{v} &= \mathbf{j} + f_y(a, b)\mathbf{k}\end{aligned}$$

To derive the formula for the tangent plane, we will use the fact that the normal vector, $\mathbf{n}$, at the point $\big(a, b, f(a, b)\big)$ must be perpendicular to the tangent plane $\overrightarrow{P_0P}$

$$\begin{aligned}&\mathbf{n} \cdot \overrightarrow{P_0P} = 0\\ &\mathbf{n} \cdot \big((x - a)\mathbf{i} + (y - b)\mathbf{j} + (z - c)\mathbf{k}\big) = 0\end{aligned}$$

where $\overrightarrow{P_0P}$ is the general form of an equation of a plane, not necessarily with respect to the origin

The normal vector $\mathbf{n}$ can be found via the cross product of the vectors $\mathbf{u}$ and $\mathbf{v}$

$$\mathbf{n} = \mathbf{u} \times \mathbf{v} = -f_x(a, b)\mathbf{i} - f_y(a, b)\mathbf{j} + \mathbf{k}$$

Then substituting this into the above expression for the plane

$$\begin{aligned}&(-f_x(a, b), -f_y(a, b), 1) \cdot \overrightarrow{P_0P} = 0\\ &(-f_x(a, b), -f_y(a, b), 1) \cdot (x - a, y - b, z - f(a, b)) = 0\\ &-f_x(a, b)(x - a) - f_y(a, b)(y - b) + (z - f(a, b)) = 0\end{aligned}$$

Rearranging for z we see that the general equation for the tangent plane in $\mathbb{R}^2$ is

$$z = f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b)$$

Where the general vector form of the equation for the tangent plane is

$$\mathbf{z} = f(\mathbf{a}) + \nabla f(\mathbf{a})(\mathbf{x} - \mathbf{a})$$

Consider $f : X \subseteq \mathbb{R}^2 \to \mathbb{R} = x^2 + y^2$, as depicted in example 1, then the tangent plane at $(1,1,2)$ can be visualised below.

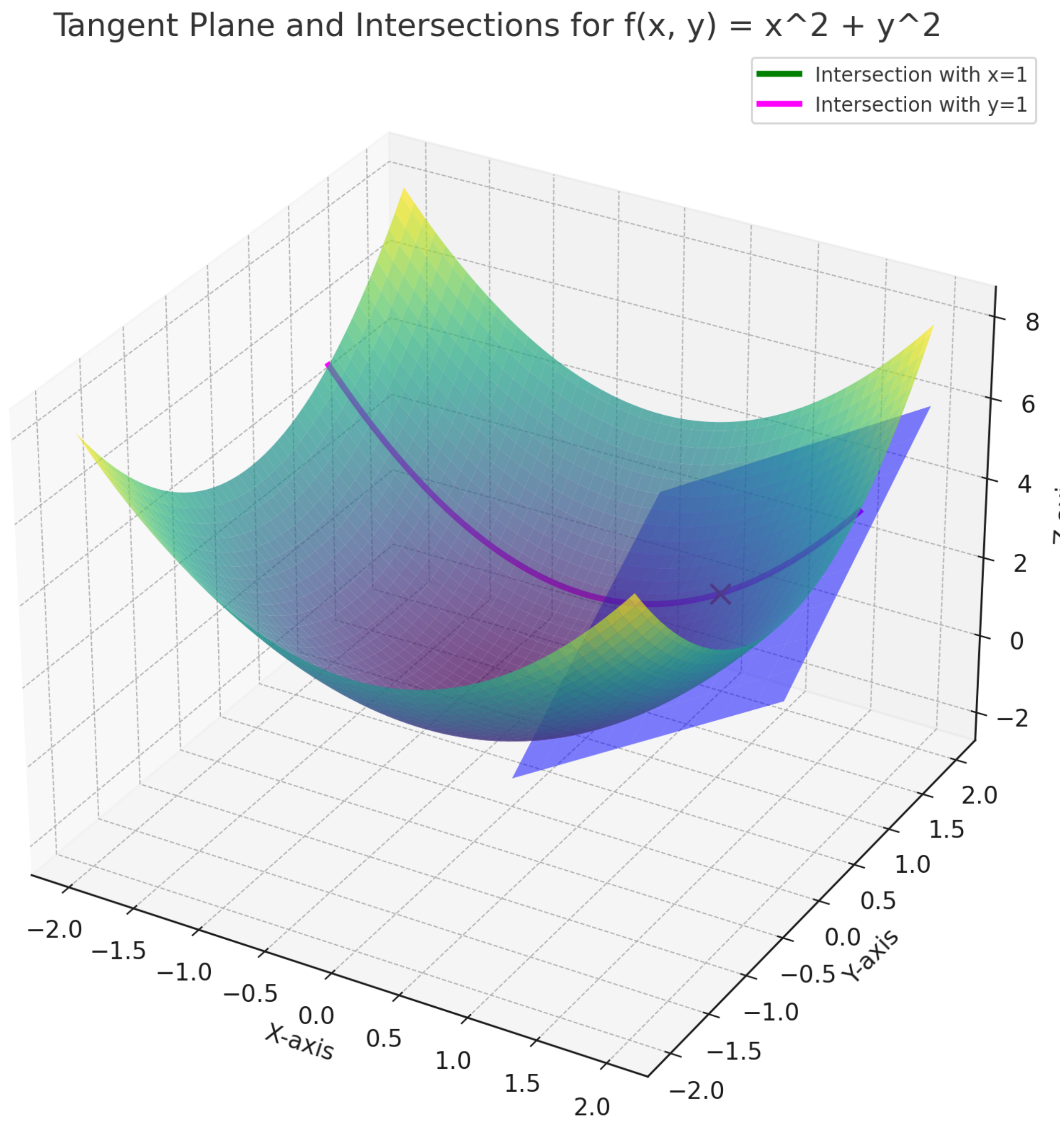

Figure 2: The tangent plane of $f : X \subseteq \mathbb{R}^2 \to \mathbb{R} = x^2 + y^2$ with its partial derivatives at $(1,1,2)$

We revisit tangent planes later to deal with functions of the form $f(\mathbf{x}) = c$ to provide a form for their solution

3.2 Differentiability of Scalar-Valued Functions of a Single Variable

If we consider the definition of the derivative in a single variable

$$F'(a) = \lim_{h \to 0} \frac{F(a+h) - F(a)}{h}$$

and let $x = a + h$, then the limit condition $h \to 0$ is equivalent to $x \to a$

$$\begin{aligned}
F'(a) &= \lim_{x \to a} \frac{F(x) - F(a)}{x - a} \\
0 &= \lim_{x \to a} \left[\frac{F(x) - F(a)}{x - a}\right] - F'(a) \\
0 &= \lim_{x \to a} \left[\frac{F(x) - F(a) - F'(a)(x-a)}{x - a}\right] \\
0 &= \lim_{x \to a} \left[\frac{F(x) - H(x)}{x - a}\right]
\end{aligned}$$

Where $H(x) = F(a) + F'(a)(x - a)$ is the 1-dimensional equivalent of the equation of the tangent plane - the tangent line, as we saw previously.

We note that for the limit to equal 0 then we require that the numerator to equal 0 not the denominator, otherwise the limit will not be defined. This means that the difference between the function at its limiting value and the tangent line must approach zero **faster** than $x - a$ approaches zero. This is what is meant by "$H(x)$ is a good linear approximator to F near $\mathbf{a}$".

3.3 Differentiability of Scalar-Valued Functions of n Variables

Definition. Differentiable To formalize this to higher dimensions, let X be open in $\mathbb{R}^n$, $f : X \to \mathbb{R}$ be a scalar-valued function and let $\mathbf{a} = (a_1, a_2, ..., a_n) \in X$. f is said to be differentiable at $\mathbf{a}$ if

- All partial derivatives $f_{x_i}(\mathbf{a}), i = 1, ..., n$ exist
- If the function $h(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ is a good linear approximation of f near to $\mathbf{a}$

In the multivariate case, $h(\mathbf{x})$ can be defined as

$$h(\mathbf{x}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \circ (\mathbf{x} - \mathbf{a})$$

Where $\nabla f(\mathbf{a})$ is a column vector, hence $\nabla f(\mathbf{a})^T$ is a row vector multiplied by the column vector $\mathbf{x} - \mathbf{a}$. The extension of the idea of $h(\mathbf{x})$ being a good linear approximator of f near a is then

$$\lim_{\mathbf{x} \to \mathbf{a}} \frac{f(\mathbf{x}) - h(\mathbf{x})}{||\mathbf{x} - \mathbf{a}||} = 0$$

The geometric meaning here is the same as in our understanding in 1 dimension. The graph of the function $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ is the **hypersurface** in $\mathbb{R}^{n+1}$ and $h(\mathbf{x}) = f(\mathbf{a}) + \nabla f(\mathbf{a})^T(\mathbf{x} - \mathbf{a})$ is the **tangent hyperplane** at $(\mathbf{a}, f(\mathbf{a}))$

3.4 Differentiability of Vector-Valued Function of n Variables

Definition. General Case of Differentiability For the most general setting whereby $X \subseteq \mathbb{R}^n$ is open in $\mathbb{R}^n$, $\mathbf{f} : X \to \mathbb{R}^m$ be a vector-valued function and let $\mathbf{a} = (a_1, a_2, ..., a_n) \in X$. $\mathbf{f}$ is said to be differentiable at $\mathbf{a}$ if

- All partial derivatives $D\mathbf{f}(\mathbf{a})$ exists, where $D\mathbf{f}(\mathbf{a})$ is the $m \times n$ matrix of the collection of partial derivatives, where each row of the matrix is given by $Df_i(\mathbf{a}) = \nabla f_i(\mathbf{a})$
- If the function $h(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}^m$ is a good linear approximation of $\mathbf{f}$ near to $\mathbf{a}$

$$\lim_{\mathbf{x} \to \mathbf{a}} \frac{||\mathbf{f}(\mathbf{x}) - \mathbf{h}(\mathbf{x})||}{||\mathbf{x} - \mathbf{a}||} = \lim_{\mathbf{x} \to \mathbf{a}} \frac{||\mathbf{f}(\mathbf{x}) - \left[\mathbf{f}(\mathbf{a}) + D\mathbf{f}(\mathbf{a})(\mathbf{x} - \mathbf{a})\right]||}{||\mathbf{x} - \mathbf{a}||} = 0$$

The vector norm in the numerator is used to ensure that there is a notion of "closeness" in vector-valued functions and is independent of the coordinate system.

Theorem 3.9 If $\mathbf{f} : X \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\mathbf{a}$, then it is continuous at $\mathbf{a}$

Theorem 3.10 If $\mathbf{f} : X \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is such that for $i = 1, ..., m$ and $j = 1, ..., n$, all $\frac{\partial f_i}{\partial x_j}$ exist and are continuous in a neighbourhood of $\mathbf{a}$ of X, then $\mathbf{f}$ is differentiable at $\mathbf{a}$

Note that the converse is **not true**. If the partial derivatives are not continuous then this does not imply the function is $\mathbf{f}$ is not differentiable at $\mathbf{a}$

Theorem 3.11 A function $\mathbf{f} : X \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\mathbf{a} \in X$ if and only if each of its component functions $f_i : X \subseteq \mathbb{R}^n \to \mathbb{R}, i = 1, ..., m$, is differentiable at $\mathbf{a}$

The proof of these theorems can be found in the appendix.

3.5 The Derivative

The derivative is related closely to the idea of differentiability insofar that the in the univariate case for a scalar-valued function, then the derivative which is a real-valued scalar is the slope of the tangent line.

In the multivariate case of a vector-valued function, then the derivative is the matrix of partial derivatives $D\mathbf{f}$ as it relates to $\mathbf{h}(\mathbf{x})$ is a good linear approximation of $\mathbf{f}(\mathbf{x})$ near $\mathbf{a}$

4 Properties of Higher Order Partial Derivatives

Theorem. Linearity of Partial Derivatives For two functions $\mathbf{f}, \mathbf{g} : X \subseteq \mathbb{R}^n \to \mathbb{R}^m$, both differentiable at a point $\mathbf{a}$ and $c \in \mathbb{R}$ be any scalar. Then

- The function $\mathbf{h} = \mathbf{f} + \mathbf{g}$ is also differentiable at $\mathbf{a}$ and is given by

$$\mathrm{D}\mathbf{h}(\mathbf{a}) = \mathrm{D}\mathbf{f}(\mathbf{a}) + \mathrm{D}\mathbf{g}(\mathbf{a})$$

- The function $\mathbf{k} = c\mathbf{f}$ is differentiable at $\mathbf{a}$ and is given by

$$\mathrm{D}\mathbf{k}(\mathbf{a}) = \mathrm{D}(c\mathbf{f})(\mathbf{a}) = c\mathrm{D}\mathbf{f}(\mathbf{a})$$

Proof of the first claim is provided below. We must prove that firstly that all partial derivatives $\mathrm{D}\mathbf{h}(\mathbf{a})$ exist and secondly, that its tangent plane is a good linear approximator of $\mathbf{h}$ near $\mathbf{a}$. See the **General Case of Differentiability**

Part 1 We show that the matrix of partial derivatives of $\mathbf{h}$ is the sum of those of $\mathbf{f}$ and $\mathbf{g}$. $\mathbf{h}(\mathbf{x}) = \big(h_1(\mathbf{x}), h_2(\mathbf{x}), ..., h_m(\mathbf{x})\big)$ and so the ij^{th} entry of $\mathrm{D}\mathbf{h}(\mathbf{x})$ is $\frac{\partial h_i}{\partial x_j}, j = 1, ..., n$. By definition we have that $h_1(\mathbf{x}) = f_1(\mathbf{x}) + g_1(\mathbf{x})$ and since f_i and g_i are scalars

$$\frac{\partial h_i(\mathbf{a})}{\partial x_j} = \frac{\partial}{\partial x_j}(f_i(\mathbf{a}) + g_i(\mathbf{a})) = \frac{\partial f_i(\mathbf{a})}{\partial x_j} + \frac{\partial g_i(\mathbf{a})}{\partial x_j}$$

Which arises due to the linearity of differentiation in one dimension, since each value is a scalar. Furthermore, since $\mathbf{f}, \mathbf{g}$ were both given to be differentiable at $\mathbf{a}$ we can conclude that $\mathrm{D}\mathbf{f}(\mathbf{a})$ and $\mathrm{D}\mathbf{g}(\mathbf{a})$ both exist and therefore $D\mathbf{h}(\mathbf{a})$ does as well.

$$D\mathbf{h}(\mathbf{a}) = \mathrm{D}\mathbf{f}(\mathbf{a}) + \mathrm{D}\mathbf{g}(\mathbf{a})$$

Part 2 We know the matrix of partial derivatives exists, now we must show the the function is differentiable.

$$\lim_{\mathbf{x} \to \mathbf{a}} \frac{||\mathbf{h}(\mathbf{x}) - \big[\mathbf{h}(\mathbf{a}) + D\mathbf{h}(\mathbf{a})(\mathbf{x} - \mathbf{a})\big]||}{||\mathbf{x} - \mathbf{a}||} = 0$$

We begin by considering

$$\begin{aligned}
&\lim_{\mathbf{x}\to\mathbf{a}} \frac{||\mathbf{h}(\mathbf{x}) - \big[\mathbf{h}(\mathbf{a}) + D\mathbf{h}(\mathbf{a})(\mathbf{x}-\mathbf{a})\big]||}{||\mathbf{x}-\mathbf{a}||} \\
&= \lim_{\mathbf{x}\to\mathbf{a}} \frac{||\big[\mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})\big] - \big[\mathbf{f}(\mathbf{a}) + \mathbf{g}(\mathbf{a}) + D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a}) + D\mathbf{g}(\mathbf{a})(\mathbf{x}-\mathbf{a})\big]||}{||\mathbf{x}-\mathbf{a}||} \\
&\leq \lim_{\mathbf{x}\to\mathbf{a}} \frac{||\mathbf{f}(\mathbf{x}) - \big[\mathbf{f}(\mathbf{a}) + D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a})\big]||}{||\mathbf{x}-\mathbf{a}||} + \lim_{\mathbf{x}\to\mathbf{a}} \frac{||\mathbf{g}(\mathbf{x}) - \big[\mathbf{g}(\mathbf{a}) + D\mathbf{g}(\mathbf{a})(\mathbf{x}-\mathbf{a})\big]||}{||\mathbf{x}-\mathbf{a}||}
\end{aligned}$$

where we achieve the inequality on the last line via the triangle inequality

Given the differentiability of $\mathbf{f}$, $\mathbf{g}$ at $\mathbf{a}$ then we know that the limit exists, meaning that we can find δ_1, δ_2 and ϵ_1, ϵ_2 that satisfies

$$0 < ||\mathbf{x}-\mathbf{a}|| < \delta_1, \quad \frac{||\mathbf{f}(\mathbf{a}) - \big[\mathbf{f}(\mathbf{a}) + D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a})\big]||}{||\mathbf{x}-\mathbf{a}||} < \epsilon_1$$

and

$$0 < ||\mathbf{x}-\mathbf{a}|| < \delta_2, \quad \frac{||\mathbf{g}(\mathbf{a}) - \big[\mathbf{g}(\mathbf{a}) + D\mathbf{g}(\mathbf{a})(\mathbf{x}-\mathbf{a})\big]||}{||\mathbf{x}-\mathbf{a}||} < \epsilon_2$$

So that if we were to choose $\epsilon_1 + \epsilon_2 = \epsilon/2$ and δ to be the smaller of δ_1, δ_2 such that for $0 < ||\mathbf{x}-\mathbf{a}|| < \delta$ then both limits hold. Therefore,

$$\begin{aligned}
&\lim_{\mathbf{x}\to\mathbf{a}} \frac{||\mathbf{h}(\mathbf{x}) - \big[\mathbf{h}(\mathbf{a}) + D\mathbf{h}(\mathbf{a})(\mathbf{x}-\mathbf{a})\big]||}{||\mathbf{x}-\mathbf{a}||} \\
&\leq \lim_{\mathbf{x}\to\mathbf{a}} \frac{||\mathbf{f}(\mathbf{x}) - \big[\mathbf{f}(\mathbf{a}) + D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a})\big]||}{||\mathbf{x}-\mathbf{a}||} + \lim_{\mathbf{x}\to\mathbf{a}} \frac{||\mathbf{g}(\mathbf{x}) - \big[\mathbf{g}(\mathbf{a}) + D\mathbf{g}(\mathbf{a})(\mathbf{x}-\mathbf{a})\big]||}{||\mathbf{x}-\mathbf{a}||} \\
&< \epsilon_1 + \epsilon_2 \\
&= \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon
\end{aligned}$$

Which then provides that $\mathbf{h}$ is differentiable at $\mathbf{a}$

Theorem. Product and Quotients In higher dimensions the product and chain rule do not have simple functional forms but for scalar-valued functions we can define

- Let $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ and $\mathbf{g} : X \subseteq \mathbb{R}^n \to \mathbb{R}^m$

$$Df\mathbf{g}(\mathbf{a}) = f(\mathbf{a})D\mathbf{g}(\mathbf{a}) + \mathbf{g}(\mathbf{a})Df(\mathbf{a})$$

- Let $f, g : X \subseteq \mathbb{R}^n \to \mathbb{R}$. Then if $g(\mathbf{a}) \neq 0$, then the quotient function f/g is differentiable at $\mathbf{a}$, then

$$\frac{g(\mathbf{a})Df(\mathbf{a}) - f(\mathbf{a})Dg(\mathbf{a})}{g(\mathbf{a})^2}$$

5 The Chain Rule

Chain Rule for Scalar-Valued Functions Let $f : X \subseteq \mathbb{R}^2 \to \mathbb{R}$ be differentiable at $\mathbf{x}_0 = \mathbf{x}(t_0) = (x_0, y_0) \in X$ and let $\mathbf{x} : T \subseteq \mathbb{R} \to \mathbb{R}^2$ be differentiable at $t_0 \in T$, where $f, \mathbf{x}$ are open in $\mathbb{R}$ and $\mathbb{R}^2$, respectively, and the range of $\mathbf{x}$ is contained within X. If f is of class C^1, then $f \circ \mathbf{x} : T \to \mathbb{R}$ is differentiable at t_0 and

$$\frac{df}{dt}(t_0) = \frac{\partial f}{\partial x}(\mathbf{x}_0)\frac{dx}{dt}(t_0) + \frac{\partial f}{\partial y}(\mathbf{x}_0)\frac{dy}{dt}(t_0)$$

Note that this represents an abuse of notation since f actually represents a composite function $f\big(\mathbf{x}(t)\big)$ but is common practice since it avoids the use of too many variables in such functions.

The proof of the chain rule is provided below. Let $z = f \circ \mathbf{x}$ then we are after dz/dt at the point t_0. z is a scalar-valued function so that its derivative is defined as

$$\frac{dz}{dt}(t_0) = \lim_{t \to t_0} \frac{z(t) - z(t_0)}{t - t_0}$$
$$\frac{dz}{dt}(t_0) = \lim_{t \to t_0} \frac{f(x,y) - f(x_0,y_0)}{t - t_0}$$

Then if we add 0 into the numerator in the form of $-f(x_0,y) + f(x_0,y)$ we can rewrite this as

$$\frac{dz}{dt}(t_0) = \lim_{t \to t_0} \frac{f(x,y) - f(x_0,y) + f(x_0,y) - f(x_0,y_0)}{t - t_0}$$
$$\frac{dz}{dt}(t_0) = \lim_{t \to t_0} \frac{f(x,y) - f(x_0,y)}{t - t_0} + \lim_{t \to t_0} \frac{f(x_0,y) - f(x_0,y_0)}{t - t_0}$$

If we then consider the result of the **Mean Value Theorem**, that

$$f_x(c,y)(a-b) = f(a,y) - f(b,y)$$

then we can rewrite the numerator as $f(x,y) - f(x_0,y) = f_x(c,y)(x - x_0)$ and $f(x_0,y) - f(x_0,y_0) = f_y(x_0,d)(y - y_0)$, which is then

$$\frac{dz}{dt}(t_0) = \lim_{t \to t_0} \frac{f_x(c,y)(x - x_0)}{t - t_0} + \lim_{t \to t_0} \frac{f_y(x_0,d)(y - y_0)}{t - t_0}$$

As a reminder, $x = x(t)$ and $x_0 = x(t_0)$ so rewriting this for clarity

$$\begin{aligned}\frac{dz}{dt}(t_0) &= \lim_{t \to t_0} f_x(c,y)\frac{x(t) - x(t_0)}{t - t_0} + \lim_{t \to t_0} f_y(x_0,d)\frac{y(t) - y(t_0)}{t - t_0} \\ &= f_x(x_0,y_0)\lim_{t \to t_0} \frac{x(t) - x(t_0)}{t - t_0} + f_y(x_0,y_0)\lim_{t \to t_0} \frac{y(t) - y(t_0)}{t - t_0} \\ &= f_x(x_0,y_0)\frac{dx}{dt}(t_0) + f_y(x_0,y_0)\frac{dy}{dt}(t_0)\end{aligned}$$

As $t \to t_0$, $f_x(c,y) \to f_x(x_0,y_0)$ and $f_y(x_0,d) \to f_y(x_0,y_0)$ because c is a point on the line between x and x_0, then $\lim_{t \to t_0} c = x_0$ with this also being true for $d \in [y, y_0]$.

For a function of n variables $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ and $\mathbf{x} : T \subseteq \mathbb{R} \to \mathbb{R}^n$, then the chain rule follows a simple extension and can be written as

$$\begin{aligned}\frac{df}{dt}(t_0) &= \begin{bmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}_0) & \frac{\partial f}{\partial x_2}(\mathbf{x}_0) & \cdots & \frac{\partial f}{\partial x_n}(\mathbf{x}_0) \end{bmatrix} \begin{bmatrix} \frac{dx_1}{dt}(t_0) \\ \frac{dx_2}{dt}(t_0) \\ \vdots \\ \frac{dx_n}{dt}(t_0) \end{bmatrix} \\ &= Df(\mathbf{x}_0)D\mathbf{x}(t_0)\end{aligned}$$

Which can also be written as $\nabla f(\mathbf{x}_0) \cdot \mathbf{x}'(t_0)$

5.1 General Case in Higher Dimensions

If we extend this case to higher dimensions, and say let $\mathbf{f} : X \subseteq \mathbb{R}^m \to \mathbb{R}^p$ be differentiable at $\mathbf{x}_0 = \mathbf{x}(\mathbf{t_0}) \in X$ and $\mathbf{x} : T \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be differentiable at $\mathbf{t_0} \in T$, where X and T are open in $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively and the range of $\mathbf{x} \in X$. Then $\mathbf{f} \circ \mathbf{x} : T \to \mathbb{R}^p$ is differentiable at $\mathbf{t_0}$, so the chain rule is

$$D(\mathbf{f} \circ \mathbf{x}) = D(\mathbf{f})(\mathbf{x_0})D(\mathbf{x})(\mathbf{t_0})$$

It is necessary to appreciate that each of the component derivatives are matrices. For example, we know that the function $\mathbf{f}$ is a mapping from m dimensional space to a p dimensional space and that this mapping is realized through a linear map, which can be represented as a $p \times m$ matrix.

However, the output $\mathbf{f}(\mathbf{x_0})$ is a p dimensional vector where each of the p entries in the vector (the component functions) is a scalar value $f_i(\mathbf{x_0}) : X \to \mathbb{R}$

$$\mathbf{f}(\mathbf{x_0}) = \begin{bmatrix} f_1(\mathbf{x_0}) \\ f_2(\mathbf{x_0}) \\ \vdots \\ f_p(\mathbf{x_0}) \end{bmatrix}$$

Now if we consider the derivative of $\mathbf{f}$ evaluated at $\mathbf{x_0}$, $D\mathbf{f}(\mathbf{x_0})$, with respect to its inputs $\mathbf{x_0}$, then what we are doing is going to consider how each of the component functions changes with respect to each of the m inputs in $\mathbf{x_0}$, such that

$$D\mathbf{f}(\mathbf{x_0}) = \begin{bmatrix} \frac{\partial f_1(\mathbf{x_0})}{\partial x_1} & \frac{\partial f_1(\mathbf{x_0})}{\partial x_2} & \dots & \frac{\partial f_1(\mathbf{x_0})}{\partial x_m} \\ \frac{\partial f_2(\mathbf{x_0})}{\partial x_1} & \frac{\partial f_2(\mathbf{x_0})}{\partial x_2} & \dots & \frac{\partial f_2(\mathbf{x_0})}{\partial x_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_p(\mathbf{x_0})}{\partial x_1} & \frac{\partial f_p(\mathbf{x_0})}{\partial x_2} & \dots & \frac{\partial f_p(\mathbf{x_0})}{\partial x_m} \end{bmatrix} \in \mathbb{R}^{p \times m}$$

Similar reasoning yields the fact that $D(\mathbf{x})(\mathbf{t_0})$ is a $m \times n$ matrix, so we can see that the derivative of the composite function is

$$D(\mathbf{f} \circ \mathbf{x}) = \begin{bmatrix} \frac{\partial f_1(\mathbf{x_0})}{\partial x_1} & \frac{\partial f_1(\mathbf{x_0})}{\partial x_2} & \dots & \frac{\partial f_1(\mathbf{x_0})}{\partial x_m} \\ \frac{\partial f_2(\mathbf{x_0})}{\partial x_1} & \frac{\partial f_2(\mathbf{x_0})}{\partial x_2} & \dots & \frac{\partial f_2(\mathbf{x_0})}{\partial x_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_p(\mathbf{x_0})}{\partial x_1} & \frac{\partial f_p(\mathbf{x_0})}{\partial x_2} & \dots & \frac{\partial f_p(\mathbf{x_0})}{\partial x_m} \end{bmatrix} \begin{bmatrix} \frac{\partial x_1(\mathbf{t_0})}{\partial t_1} & \frac{\partial x_1(\mathbf{t_0})}{\partial t_2} & \dots & \frac{\partial x_1(\mathbf{t_0})}{\partial t_n} \\ \frac{\partial x_2(\mathbf{t_0})}{\partial t_1} & \frac{\partial x_2(\mathbf{t_0})}{\partial t_2} & \dots & \frac{\partial x_2(\mathbf{t_0})}{\partial t_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_m(\mathbf{t_0})}{\partial t_1} & \frac{\partial x_m(\mathbf{t_0})}{\partial t_2} & \dots & \frac{\partial x_m(\mathbf{t_0})}{\partial t_n} \end{bmatrix} \in \mathbb{R}^{p \times n}$$

6 Directional Derivatives and the Gradient

6.1 The Directional Derivative

Let X be open in $\mathbb{R}^n$, $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ is a scalar-valued function and the point $\mathbf{a} \in X$. Let $\mathbf{v} \in \mathbb{R}^n$ be a unit vector, then the directional derivative of f at $\mathbf{a}$ in the direction of $\mathbf{v}$, denoted $D_v f(\mathbf{a})$ is

$$D_v f(\mathbf{a}) = \lim_{t \to 0} \frac{f(\mathbf{a} + t\mathbf{v}) - f(\mathbf{a})}{t}$$

We can define another form for the directional derivative if we define $F(t) = f(\mathbf{a}+t\mathbf{v})$ and consider the definition

$$D_v f(\mathbf{a}) = \lim_{t \to 0} \frac{f(\mathbf{a} + t\mathbf{v}) - f(\mathbf{a})}{t} = \lim_{t \to 0} \frac{F(t) - F(0)}{t} = F'(0)$$

Which we can rewrite as

$$F'(0) = \frac{dF}{dt}\bigg|_{t=0} = \frac{d}{dt} f(\mathbf{a} + t\mathbf{v})\big|_{t=0}$$

If we parameterize $\mathbf{x}(t) = \mathbf{a} + t\mathbf{v}$, then we can use the chain rule

$$\begin{aligned} \frac{d}{dt} f(\mathbf{a} + t\mathbf{v}) =& Df(\mathbf{x})D\mathbf{x}(t) \\ =& Df(\mathbf{x})\mathbf{v} \end{aligned}$$

Evaluating this at $t = 0$ yields

$$Df(\mathbf{a})\mathbf{v} = \nabla f(\mathbf{a}) \circ \mathbf{v}$$

Which has the geometric interpretation that the directional derivative in the direction of the unit vector $\mathbf{v}$ is just the dot product of the gradient vector and the vector itself. This highlights the dependence of the directional derivative on the partial derivatives since we are projecting the gradient vector onto the unit vector $\mathbf{v}$.

6.2 The Gradient

From the directional derivative we can calculate the direction of steepest ascent if we consider the definition of the dot product

$$\begin{aligned}\nabla f(\mathbf{x}) \circ \mathbf{v} =& ||\nabla f(\mathbf{x})||\ ||\mathbf{v}||\cos\theta \\ =& ||\nabla f(\mathbf{x})||\cos\theta\end{aligned}$$

Since $-1 \leq \cos\theta \leq 1$, then we know that

$$-||\nabla f(\mathbf{x})|| \leq \nabla f(\mathbf{x}) \circ \mathbf{v} \leq ||\nabla f(\mathbf{x})||$$

This function will be maximised and minimized when the vectors are parallel and anti-parallel, respectively.

6.3 Tangent Planes Revisited

Generally, given a surface described by an equation of the form $f(x, y, z) = c$ where c is a constant, it may be impractical to solve for z even if done so as several functions of x and y. There is a way to find the tangent plane which will be shown below.

Theorem Let $X \subseteq \mathbb{R}^n$ be open and $f : X \to \mathbb{R}$ be a function of class C^1. If $\mathbf{x}_0$ is a point on the level set $S = \{\mathbf{x} \in X : f(\mathbf{x}) = c\}$, then a vector $\nabla f(\mathbf{x}_0)$ is perpendicular to S

We need to establish that for any curve on the level set, the gradient $\nabla f(\mathbf{x}_0)$ will be perpendicular to the tangent to the curve at any point $\mathbf{x}_0 \in S$

If we let a curve C be given parametrically by the path $\mathbf{x} : I \to \mathbb{R}^n$ where I is the interval (a, b) and $t \in I$, then $\mathbf{x}(t) = \big(x_1(t), x_2(t), ..., x_n(t)\big)$ where $\mathbf{x}(t_0) = \big(x_1(t_0), x_2(t_0), ..., x_n(t_0)\big)$. The function is then actually a function of t, so that for any $t \in I \Rightarrow f(\mathbf{x}(t)) = f\big(x_1(t), x_2(t), ..., x_n(t)\big) = c$. Then the derivative of of the function along any curve $C \in S$

$$\frac{d}{dt} f(\mathbf{x}(t)) = 0$$

but by the chain rule we can write this as

$$\frac{d}{dt} f(\mathbf{x}(t)) = \nabla f(\mathbf{x}(t)) \circ \mathbf{x}'(t)$$

Which when evaluated at the point t_0

$$\nabla f(\mathbf{x}(t_0)) \circ \mathbf{x}'(t_0) = 0$$

shows that the gradient vector is perpendicular to any point on the surface.

In general, for a surface $S \in \mathbb{R}^n$ defined by the equation

$$f(\mathbf{x}) = c$$

and if $\mathbf{x}_0 \in X \subseteq \mathbb{R}^n$ and the gradient vector $\nabla f(\mathbf{x}_0)$ exists, then $\nabla f(\mathbf{x}_0)$ is perpendicular to the tangent plane at $\mathbf{x}_0$. Since we have identified the normal vector $\mathbf{n} = \nabla f(\mathbf{x}_0)$, then we can use the definition of a plane to identify the tangent plane

$$\nabla f(\mathbf{x}_0) \circ (\mathbf{x} - \mathbf{x}_0) = 0$$

To relate the original equation we have for the tangent plane and the more general definition that we found here. If $S \in \mathbb{R}^3$ given by the equation $z = f(x, y)$ where f is differentiable, then our original formula for the tangent plane would be

$$z = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b)$$

To view this in the more general form, S may be written as

$$f(x, y) - z = 0$$

such that if we define a new function $F(x, y, z) = f(x, y) - z$ we see that this is a level set of F with height 0. So we can use the general equation to find the tangent plane at the point $(a, b, f(a, b))$

$$F_x(a, b, f(a, b))(x - a) + F_y(a, b, f(a, b))(y - b) + F_z(a, b, f(a, b))(z - f(a, b)) = 0$$

noting that $\frac{dF}{dx} = \frac{\partial f}{\partial x}$ and so on.

$$\begin{aligned} &f_x(a, b)(x - a) + f_y(a, b)(y - b) - 1(z - f(a, b)) = 0 \\ &z = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b) \end{aligned}$$

So we see that the more general form for the level set is equivalent to the original equation

6.4 The Normal Line

Given a surface S defined by the equation $F(x, y, z) = c$, then then normal line at a point $\mathbf{a} = (x_0, y_0, z_0)$ is a line that is perpendicular to the surface at $\mathbf{a}$ and therefore also perpendicular to the tangent hyperplane.

Since the gradient vector, if it exists, is perpendicular to the level set $F(x, y, z) = c$, then this is also the normal vector. Thus we can easily define the equation for the normal line parametrically

$$\begin{aligned} x &= x_0 + t f_x(x, y, x) \\ y &= y_0 + t f_y(x, y, x) \\ z &= z_0 + t f_z(x, y, x) \end{aligned}$$

7 The Newton-Raphson Method

The Newton-Raphson method is an iterative algorithm used to find the roots of the equation of the form $\mathbf{f}(\mathbf{x}) = \mathbf{0}$ through the use of the derivative. For a map $\mathbf{f} : X \subseteq \mathbb{R}^n \to \mathbb{R}^n$ we can approximate the map

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{x}_0) + D\mathbf{f}(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)$$

Then if we consider where the function approximation is zero

$$\begin{aligned} &\mathbf{f}(\mathbf{x}_0) + D\mathbf{f}(\mathbf{x}_0)(\mathbf{x}_1 - \mathbf{x}_0) = 0 \\ &- \mathbf{f}(\mathbf{x}_0) = D\mathbf{f}(\mathbf{x}_0)(\mathbf{x}_1 - \mathbf{x}_0) \\ &- \left(D\mathbf{f}(\mathbf{x}_0)\right)^{-1}\mathbf{f}(\mathbf{x}_0) = (\mathbf{x}_1 - \mathbf{x}_0) \\ &\mathbf{x}_1 = \mathbf{x}_0 - \left(D\mathbf{f}(\mathbf{x}_0)\right)^{-1}\mathbf{f}(\mathbf{x}_0) \end{aligned}$$

provided that the inverse matrix exists.

This method does not always converge but when it does it always converges to $\mathbf{f}(\mathbf{x}) = 0$. To see this, let $L = \lim_{k\to\infty} \mathbf{x}_k$ which also implied that $\lim_{k\to\infty} \mathbf{x}_{k-1}$. Taking limits to equation (96)

$$\begin{aligned} L =& L - \left(D\mathbf{f}(L)\right)^{-1}\mathbf{f}(L) \\ L =& L - \left(D\mathbf{f}(L)\right)^{-1} \times \mathbf{0} \\ L =& L \end{aligned}$$

Hence L is a root of the equation.

8 Appendix

8.1 Mean Value Theorem

$$f_x(c, y)(a - b) = f(a, y) - f(b, y)$$

The mean value theorem tells us that over a straight line, then the average difference between two points is equivalent to the partial derivative over the axis (variable) which the points lie, evaluated at some point c between the two points. Formally, if f is continuous on $[a, b]$ and differentiable on the open interval (a, b) then there exists at least one point c, such that $f'(c) = \frac{f(a)-f(b)}{(q-b)}$ where the RHS is the average gradient over the interval

8.2 Theorem 3.9

This proof utilizes the Cauchy–Schwarz inequality, which is provided as a further addendum below. It is utilized in the proof of a claim that for now we will assume to be true. Let $\mathbf{x} \in \mathbb{R}^n$ and B a $m \times n$ matrix. Then if $\mathbf{y} = B\mathbf{x}$, then

$$||\mathbf{y}|| \leq K||\mathbf{x}||$$

where $K = \left(\sum_{i,j}(b_{i,j})^2\right)^{1/2}$

Given that $\mathbf{f}$ is differentiable at $\mathbf{a}$, we must prove that it is also continuous at $\mathbf{a}$, which means that

$$\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{a})$$
$$\lim_{\mathbf{x}\to\mathbf{a}} ||\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a})|| = 0$$

We can rewrite this and use the triangle inequality to show that

$$\begin{aligned} &||\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a}) + D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a}) - D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a})|| \\ &\qquad \leq ||\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a}) - D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a})|| + ||D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a})|| \end{aligned} \quad (1)$$

If we now note that the second term on the RHS is a vector, then we can apply the our claim that

$$||D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a})|| \leq K||\mathbf{x}-\mathbf{a}||$$

Such that

$$\begin{aligned} &||\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a}) - D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a})|| + ||D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a})|| \\ &\qquad \leq ||\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a}) - D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a})|| + K||\mathbf{x}-\mathbf{a}|| \end{aligned} \quad (2)$$

If we recall the definition of differentiability, we require that the numerator of the limit must approach zero **faster** than $||\mathbf{x}-\mathbf{a}||$, then we can say that

$$||\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a}) - D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a})|| \leq ||\mathbf{x}-\mathbf{a}||$$

So that we can rewrite equation (83) as

$$\begin{aligned} ||\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a}) - D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a})|| + K||\mathbf{x}-\mathbf{a}|| \leq& ||\mathbf{x}-\mathbf{a}|| + K||\mathbf{x}-\mathbf{a}|| \\ =& (1+K)||\mathbf{x}-\mathbf{a}|| \end{aligned}$$

Then in the limit as $\mathbf{x} \to \mathbf{a}$, $(1+K)||\mathbf{x}-\mathbf{a}|| \to 0$ so that

$$\begin{aligned} ||\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a})|| \leq& (1+K)||\mathbf{x}-\mathbf{a}|| \\ ||\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a})|| =& 0 \end{aligned}$$

8.3 Theorem 3.10

Proof of **theorem 3.10** will be provided for when $\mathbf{f}$ is a scalar-valued function and in the case where it is a vector-valued function. We know from the **general case of differentiability** that it suffices to show that all partial derivatives exist and that $\mathbf{h}(\mathbf{x})$ is a good linear approximator of $\mathbf{f}$ near $\mathbf{a}$. In the theorem, we are given the existence of all partial derivatives and that they are continuous. With this in mind, we must prove the limit exists.

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{||\mathbf{f}(\mathbf{x}) - \left[\mathbf{f}(\mathbf{a}) + D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a})\right]||}{||\mathbf{x}-\mathbf{a}||} = 0$$

For a scalar-valued function of two variables, the numerator can be rewritten as

$$\begin{aligned} &\mathbf{f}(\mathbf{x}) - \left[\mathbf{f}(\mathbf{a}) + D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a})\right] = \\ &\qquad f(x_1,x_2) - f(a_1,a_2) - f_{x_1}(a_1,a_2)(x_1-a_1) - f_{x_2}(a_1,a_2)(x_2-a_2) \end{aligned} \quad (3)$$

We can rewrite the first two terms and use the mean value theorem to simplify the term

$$\begin{aligned} f(x_1,x_2) - f(a_1,a_2) &= f(x_1,x_2) - f(a_1,x_2) + f(a_1,x_2) - f(a_1,a_2) \\ &= f_{x_1}(c_1,x_2)(x_1-a_1) + f_{x_2}(a_1,c_2)(x_2-a_2) \end{aligned} \quad (4)$$

Substituting this into the original equation for the numerator

$$\begin{aligned}
&f(x_1,x_2)-f(a_1,a_2)-f_{x_1}(a_1,a_2)(x_1-a_1)-f_{x_2}(a_1,a_2)(x_2-a_2)=\\
&\qquad f_{x_1}(c_1,x_2)(x_1-a_1)+f_{x_2}(a_1,c_2)(x_2-a_2)-f_{x_1}(a_1,a_2)(x_1-a_1)-f_{x_2}(a_1,a_2)(x_2-a_2)
\end{aligned}\tag{5}$$

Then its absolute value, which is the scalar equivalent of a norm, is given by

$$\begin{aligned}
&|f_{x_1}(c_1,x_2)(x_1-a_1)-f_{x_1}(a_1,a_2)(x_1-a_1)+f_{x_2}(a_1,c_2)(x_2-a_2)-f_{x_2}(a_1,a_2)(x_2-a_2)|\\
\leq\ &|f_{x_1}(c_1,x_2)(x_1-a_1)-f_{x_1}(a_1,a_2)(x_1-a_1)|+|f_{x_2}(a_1,c_2)(x_2-a_2)-f_{x_2}(a_1,a_2)(x_2-a_2)|
\end{aligned}\tag{6}$$

as a result of the triangle inequality. Creating a further inequality

$$\begin{aligned}
&=\big|f_{x_1}(c_1,x_2)-f_{x_1}(a_1,a_2)\big|\ |(x_1-a_1)|+\big|f_{x_2}(a_1,c_2)-f_{x_2}(a_1,a_2)\big|\ |(x_2-a_2)|\\
&\leq\Big[\big|f_{x_1}(c_1,x_2)-f_{x_1}(a_1,a_2)\big|+\big|f_{x_2}(a_1,c_2)-f_{x_2}(a_1,a_2)\big|\Big]||\mathbf{x}-\mathbf{a}||
\end{aligned}$$

Which is derived from the idea that for any $i=1,2$,

$$|x_i-a_i|\leq||\mathbf{x}-\mathbf{a}||=\big((x_1-a_1)^2+(x_2-a_2)^2\big)^{1/2}$$

Rewriting our equation

$$\begin{aligned}
&f(x_1,x_2)-f(a_1,a_2)-f_{x_1}(a_1,a_2)(x_1-a_1)-f_{x_2}(a_1,a_2)(x_2-a_2)=\\
&\qquad\leq\Big[\big|f_{x_1}(c_1,x_2)-f_{x_1}(a_1,a_2)\big|+\big|f_{x_2}(a_1,c_2)-f_{x_2}(a_1,a_2)\big|\Big]||\mathbf{x}-\mathbf{a}||
\end{aligned}\tag{7}$$

Then if we divide by $||\mathbf{x}-\mathbf{a}||$ and consider the limit as $\mathbf{x}\to\mathbf{a}$

$$\begin{aligned}
&\lim_{\mathbf{x}\to\mathbf{a}}\frac{f(x_1,x_2)-f(a_1,a_2)-f_{x_1}(a_1,a_2)(x_1-a_1)-f_{x_2}(a_1,a_2)(x_2-a_2)}{||\mathbf{x}-\mathbf{a}||}\\
&\qquad\leq\lim_{\mathbf{x}\to\mathbf{a}}\Big[\big|f_{x_1}(c_1,x_2)-f_{x_1}(a_1,a_2)\big|+\big|f_{x_2}(a_1,c_2)-f_{x_2}(a_1,a_2)\big|\Big]
\end{aligned}\tag{8}$$

Then as $\mathbf{x}\to\mathbf{a}$ then $c_i\to a_i$ because c_i lies on the line between a_i and x_i. Therefore the RHS of the equation approach zero. As a result of the squeeze theorem, which states that since the limit of the RHS approaches zero (which is the upper bound of the LHS), then the limit of the LHS must also equal zero. As a result, we are left with

$$\lim_{\mathbf{x}\to\mathbf{a}}\frac{f(x_1,x_2)-f(a_1,a_2)-f_{x_1}(a_1,a_2)(x_1-a_1)-f_{x_2}(a_1,a_2)(x_2-a_2)}{||\mathbf{x}-\mathbf{a}||}=0$$

which provides the scalar-valued proof.

For the vector-valued case where $\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^p$ we refer to the general case of differentiability again.

$$\lim_{\mathbf{x}\to\mathbf{a}}\frac{||\mathbf{f}(\mathbf{x})-\big[\mathbf{f}(\mathbf{a})+D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a})\big]||}{||\mathbf{x}-\mathbf{a}||}=0$$

$\mathbf{f}$ has p component functions $f_i, i=1,..,p$ which are scalar valued and we know from the scalar-valued proof above that the they are differentiable. Lets denote the component function of the numerator can be expressed as

$$G_i=f_i(\mathbf{x})-f_i(\mathbf{a})-Df_i(\mathbf{a})(\mathbf{x}-\mathbf{a})$$

So that we can express the limit as

$$\lim_{\mathbf{x}\to\mathbf{a}}\frac{||\big(G_1,G_2,...,G_p\big)||}{||\mathbf{x}-\mathbf{a}||}=0$$

which using the definition of the norm

$$\begin{aligned}
\frac{||\big(G_1,G_2,...,G_p\big)||}{||\mathbf{x}-\mathbf{a}||}&=\frac{\big(|G_1|^2+|G_2|^2+...+|G_p|^2\big)^{1/2}}{||\mathbf{x}-\mathbf{a}||}\\
&\leq\frac{|G_1|+|G_2|+...+|G_p|}{||\mathbf{x}-\mathbf{a}||}\\
&=\frac{|G_1|}{||\mathbf{x}-\mathbf{a}||}+\frac{|G_2|}{||\mathbf{x}-\mathbf{a}||}+...+\frac{|G_p|}{||\mathbf{x}-\mathbf{a}||}
\end{aligned}$$

We know that as $\mathbf{x}\to\mathbf{a}$ then $|G_i|/||\mathbf{x}-\mathbf{a}||\to 0$, by the differentiability of the component functions f_i and the definition of G_i, which proves the result.

8.4 Theorem 3.11

This proof follows directly from **Theorem 3.10** above. We showed that

$$\frac{||\mathbf{f}(\mathbf{x}) - \big[\mathbf{f}(\mathbf{a}) + D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a})\big]||}{||\mathbf{x}-\mathbf{a}||} \leq \frac{|G_1|}{||\mathbf{x}-\mathbf{a}||} + \frac{|G_2|}{||\mathbf{x}-\mathbf{a}||} + ... + \frac{|G_p|}{||\mathbf{x}-\mathbf{a}||}$$

and it follows that if the component functions are differentiable at $\mathbf{a}$, which is the upper bound of the LHS, then it follows that $\mathbf{f}$ is differentiable at $\mathbf{a}$. Conversely, for each of the component functions $i = 1, ..., p$, it follows that if $\mathbf{f}$ is differentiable at $\mathbf{a}$, then so is each of the component functions f_i

$$\frac{||\mathbf{f}(\mathbf{x}) - \big[\mathbf{f}(\mathbf{a}) + D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a})\big]||}{||\mathbf{x}-\mathbf{a}||} = \frac{||\big(G_1, G_2, ..., G_p\big)||}{||\mathbf{x}-\mathbf{a}||} \geq \frac{|G_i|}{||\mathbf{x}-\mathbf{a}||}$$