Contents

1 Spaces of Higher Dimensions

To extend calculus to higher dimensions, we first introduce some preliminary knowledge about the various constructs and their behaviours in higher-dimensional spaces. Specifically, we consider the **Euclidean n-spaces**, or simply denoted by $\mathbb{R}^n$. Recall that $\mathbb{R}^n$ is just the set of all **ordered** n-tuple of real numbers, which represents the coordinates of a point in $\mathbb{R}^n$.

1.1 Curves in n-Dimensional Spaces

1.1.1 Curves

Intuitively, we view a curve as the locus of a moving point. However, from the perspective of functions, we can see a curve as a set of points (equivalently, a set of vectors) in $\mathbb{R}^n$ with every point (vector) being the image of some real number. Therefore, we can view a curve as the **image** of some interval $D \subseteq \mathbb{R}$ under some mapping $R: D \to \mathbb{R}^n$. For $t \in D$, we can write

$$R(t) = \begin{bmatrix} f_1(t) \\ f_2(t) \\ \vdots \\ f_n(t) \end{bmatrix}$$

as the position vector of a point in the curve parametrised by $R(t)$, where $f_1, f_2, \cdots, f_n$ are real-evaluated functions, known as the **component functions**.

> *Remark.* Note that the parametrisation for a curve is **not unique**.

Observe that in the above example, if $f_1, f_2, \cdots, f_n$ are all **linear** functions, i.e., $f_i(x) = ax+b$ for some real constants a and b for $i = 1, 2, \cdots, n$, then the curve becomes a straight line.

1.1.2 Lines

In an axiomatic formulation, a line is said to be such that any two distinct points in a space uniquely determines a line. Therefore, we can say that a line itself is an undefined structure which fulfills a set of axioms. However, to make things simple and concrete, we can define several ways that describe a line.

Note that every point in $\mathbb{R}^n$ can be uniquely equated to a vector known as its **position vector**. Therefore, every line can be uniquely determined by two **distinct vectors** in $\mathbb{R}^n$.

Definition 1.1.1 ▸ Line

Let $\boldsymbol{a}$, $\boldsymbol{b}$ be two distinct vectors in $\mathbb{R}^n$. The line L determined by $\boldsymbol{a}$ and $\boldsymbol{b}$ is defined to be the set

$$L = \{\boldsymbol{v} : \boldsymbol{v} = \boldsymbol{a} + k\boldsymbol{u},\ k \in \mathbb{R},\ \boldsymbol{u} = \boldsymbol{a} - \boldsymbol{b}\}.$$

In other words, a line is uniquely determined by a **point** and a **direction**. Fix a point with position vector $\boldsymbol{a}$ and a direction vector $\boldsymbol{u}$ for a line L, we can thus parametrise the position vector (or the coordinates) of an arbitary point in L as

$$\boldsymbol{r} = R(t) = \boldsymbol{a} + t\boldsymbol{u}$$

We can also define the relations between lines in $\mathbb{R}^n$. Note that in plane geometry, two lines are either intersecting or parallel. However, in $\mathbb{R}^n$ where $n > 2$, non-parallel lines may not intersect.

Theorem 1.1.2 ▸ Parallel lines

Two lines are parallel if and only if their direction vectors are parallel, i.e., if L_1 and L_2 are parametrised by $R_1(t) = \boldsymbol{a} + t\boldsymbol{u}_1$ and $R_2(s) = \boldsymbol{b} + s\boldsymbol{u}_2$ respectively, then $L_1 \parallel L_2$ if and only if $\boldsymbol{u}_1 = k\boldsymbol{u}_2$ for some $k \in \mathbb{R}$.

Theorem 1.1.3 ▸ Intersecting lines

Let L_1 and L_2 be lines parametrised by $R_1(t) = \boldsymbol{a}+t\boldsymbol{u}_1$ and $R_2(s) = \boldsymbol{b}+s\boldsymbol{u}_2$ respectively. Then L_1 and L_2 intersect, i.e., $L_1 \cap L_2 \neq \varnothing$, if and only if the linear system

$$R_1(t) - R_2(s) = \mathbf{0}$$

has solutions.

Two intersecting lines may not necessarily have a unique intersection. Specifically, if two lines have more than one intersection, they are known to be **coincident**, i.e., they completely overlap on one another.

If two lines are neither parallel nor intersecting, they are called to be **skew** lines.

> *Remark.* Multiple lines with the same intersection are known to be **concurrent**.

1.1.3 Tangent Vectors

Suppose we are given a curve C parametrised by

$$R(t) = \begin{bmatrix} f_1(t) \\ f_2(t) \\ \vdots \\ f_n(t) \end{bmatrix},$$

and we are interested in the **rate of change** of the coordinates of the points in C with respect to t. Naturally, we would fix position vectors $\boldsymbol{r}_1, \boldsymbol{r}_2 \in C$, and consider the vector

$$\frac{\boldsymbol{r}_2 - \boldsymbol{r}_1}{\Delta t} = \frac{R(t_2) - R(t_1)}{\Delta t} = \begin{bmatrix} \frac{f_1(t_2)-f_1(t_1)}{\Delta t} \\ \frac{f_2(t_2)-f_2(t_1)}{\Delta t} \\ \vdots \\ \frac{f_n(t_2)-f_n(t_1)}{\Delta t} \end{bmatrix}$$

for some change of t, Δt. We can write the above more concisely as

$$\frac{\boldsymbol{r}_2 - \boldsymbol{r}_1}{\Delta t} = \frac{R(t+\Delta t) - R(t)}{\Delta t} = \begin{bmatrix} \frac{f_1(t+\Delta t)-f_1(t)}{\Delta t} \\ \frac{f_2(t+\Delta t)-f_2(t)}{\Delta t} \\ \vdots \\ \frac{f_n(t+\Delta t)-f_n(t)}{\Delta t} \end{bmatrix}.$$

Note that $\lim_{\boldsymbol{r}_2 \to \boldsymbol{r}_1} \frac{\boldsymbol{r}_2 - \boldsymbol{r}_1}{\Delta t}$ is exactly the vector for the rate of change of coordinates in C, so we have the following definition:

Definition 1.1.4 ▸ Tangent vector

Let C be a curve in $\mathbb{R}^n$ parametrised by

$$R(t) = \begin{bmatrix} f_1(t) \\ f_2(t) \\ \vdots \\ f_n(t) \end{bmatrix},$$

then the **tangent vector** of C at t is defined to be the vector

$$\lim_{\Delta t \to 0} \begin{bmatrix} \frac{f_1(t+\Delta t)-f_1(t)}{\Delta t} \\ \frac{f_2(t+\Delta t)-f_2(t)}{\Delta t} \\ \vdots \\ \frac{f_n(t+\Delta t)-f_n(t)}{\Delta t} \end{bmatrix} = \begin{bmatrix} f_1'(t) \\ f_2'(t) \\ \vdots \\ f_n'(t) \end{bmatrix},$$

denoted by $R'(t)$.

We can then define the notion of a tangent line:

Definition 1.1.5 ▸ Tangent line

The **tangent line** to a curve parametrised by $R(t)$ at t_0 is the line passing through the point $R(t_0)$ in the direction of the tangent vector to C at t_0, i.e., it is the Line

$$R(t_0) + kR'(t_0) \quad k \in \mathbb{R}.$$

Remark. There are two things to take note based on the above definitions:

1. The equation of the tangent line is **independent** of the parametrisation of C, but the tangent vector is **dependent** on the parametrisation which determines its magnitude.
2. A line is the tangent line to itself.

From the above, we can easily see that $R'(t)$ exists if and only if each of the $f_1, f_2, \cdots, f_n$ are differentiable. With that, we introduce a simple method to determine the continuity and differentiability of a curve given its parametrisation:

Theorem 1.1.6 ▸ Continuity and differentiability of a curve

A curve C is **continuous** (and respectively, **differentiable**) if its parametrisation is **continuous** (and respectively, **differentiable**).

1.1.4 Arc Length

Recall that if curve in $\mathbb{R}^2$ is parametrised by

$$\begin{cases} x = f(t) \\ y = g(t) \end{cases}$$

which is integrable, then the arc length from $t = a$ to $t = b$ is

$$\int_a^b \sqrt{\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2 + \left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2}\,\mathrm{d}t.$$

Analogously, we derive the formula for arc length in $\mathbb{R}^n$ as follows:

Theorem 1.1.7 ▸ Arc length

Let C be a curve in $\mathbb{R}^n$ parametrised by

$$R(t) = \begin{bmatrix} f_1(t) \\ f_2(t) \\ \vdots \\ f_n(t) \end{bmatrix},$$

then the **arc length** of C between $R(a)$ and $R(b)$ is given by

$$\int_a^b \sqrt{\sum_{i=1}^n f_i'(t)^2}\,\mathrm{d}t.$$

Proof. Let n be a positive integer, and $\Delta t \coloneqq \frac{b-a}{n}$. Let $t_j = a + j\Delta t$, then

$$a = t_0 < t_1 < t_2 < \cdots < t_n = b.$$

Let s_j be the distance between $R(t_{j-1})$ and $R(t)$, then

$$s_j = \sqrt{\sum_{i=1}^n \big(f_i(t_j) - f_i(t_{j-1})\big)^2} = \sqrt{\sum_{i=1}^n \big(f_i'(t_j)\Delta t\big)^2}.$$

Therefore, the arc length between $R(a)$ and $R(b)$ is given by

$$\begin{aligned}\lim_{\Delta t \to 0} \sum_{j=0}^n s_j &= \lim_{\Delta t \to 0} \sum_{j=0}^n \sqrt{\sum_{i=1}^n \big(f_i'(t_j)\Delta t\big)^2} \\ &= \lim_{\Delta t \to 0} \sum_{j=0}^n \sqrt{\sum_{i=1}^n \big(f_i'(t_j)\big)^2}\,\Delta t \\ &= \int_a^b \sqrt{\sum_{i=1}^n f_i'(t)^2}\,\mathrm{d}t.\end{aligned}$$

□

1.2 Surfaces in n-Dimensional Spaces

Intuitively, we view the notion of a **surface** as a structure "swept" out by one or more curves. We introduce two ways to describe a surface.

1.2.1 Surfaces as Graphs of Functions

Just like how we can describe a curve using a mapping, a surface can also be viewed as the graph of a certain mapping (i.e., the set of all vectors in the image of a domain under a mapping).

Definition 1.2.1 ▸ Graph of functions

Let $f : D \to \mathbb{R}^n$ be a mapping where $D \subseteq \mathbb{R}^m$, the set

$$\{ f(x) : x \in D \}$$

is the surface known as the **graph** of f.

Remark. Note that $g : \mathbb{R} \to \mathbb{R}^n$ and $h : \mathbb{R}^m \to \{\boldsymbol{v}\}$ both fulfill the above definition, which means that **curves** and **points** are also technically "surfaces". They are known as **degenerate** surfaces.

In particular, let $f(x_1, x_2, \cdots, x_{n-1})$ be a function in $n-1$ variables, then the surface which is the graph of f is given by the set

$$\{ (x_1, x_2, \cdots, x_{n-1}, f(x_1, x_2, \cdots, x_{n-1})) : x_1, x_2, \cdots, x_{n-1} \in D \}.$$

1.2.2 Surfaces as Level Sets of Functions

We introduce the concept of level sets of functions:

Definition 1.2.2 ▸ Level set

Let $f(x_1, x_2, \cdots, x_{n-1})$ be a function in n variables, then the **k-level set** of f is defined as the set

$$\{ (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n : f(x_1, x_2, \cdots, x_n) = k \}.$$

We can view the k-level set as the "projection" of the graph of f at $f(x_1, x_2, \cdots, x_n) = k$ from $\mathbb{R}^{n+1}$ to $\mathbb{R}^n$. As such, a surface in $\mathbb{R}^n$ can be described as a level set for some function whose graph is in $\mathbb{R}^{n+1}$.

1.2.3 Planes

In coordinate plane geometry, we conventionally define a plane to be a Euclidean plane, i.e., a 2-dimensional Euclidean space. We can now abstract the notion of plane as follows:

Definition 1.2.3 ▸ Plane

A plane is a space (or flat surface) of dimension 2.

It is easy to see that a plane is a special case for a 2-dimensional surface. Note that for any plane, we can always find a vector which is orthogonal to the plane, so we can describe a plane using this orthogonal vector.

Theorem 1.2.4 ▸ Equation of planes

Let P be a plane with a basis, and let $\boldsymbol{n} \perp P$. If $\boldsymbol{p} \in P$, then for any $\boldsymbol{r} \in P$, we have

$$\boldsymbol{r} \cdot \boldsymbol{n} = \boldsymbol{p} \cdot \boldsymbol{n},$$

where $\boldsymbol{n}$ is known as the **normal vector** to P.

With the notion of the normal vector, we are able to describe several relations between planes.

Theorem 1.2.5 ▸ Parallel planes

Two planes are parallel if and only if their normal vectors are parallel.

Theorem 1.2.6 ▸ Orthogonal planes

Two planes are orthogonal if and only if their normal vectors are orthogonal.

Theorem 1.2.7 ▸ Angle between planes

Let P_1, P_2 be two planes with normal vectors $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$ respectively, and let θ be the angle between P_1 and P_2, then

$$\cos\theta = \frac{|\boldsymbol{n}_1 \cdot \boldsymbol{n}_2|}{\|\boldsymbol{n}_1\| \, \|\boldsymbol{n}_2\|}.$$

2 Multivariable Functions

2.1 Limits of Multivariable Functions

Recall that for a 1-variable function $f(x)$ over $\mathbb{R}$, we view the limit of $f(x)$ at $x = a$ to be the value which $f(x)$ approaches as x gets arbitrarily close to a. We can generalise limits for n-variable functions.

Note that the domain of an n-variable function f is some set $D \subseteq \mathbb{R}^n$. Let $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ and $\boldsymbol{y} = (y_1, y_2, \cdots, y_n)$ be vectors in $\mathbb{R}^n$, we define the "closeness" between $\boldsymbol{x}$ and $\boldsymbol{y}$ by considering their distance

$$d(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}.$$

Therefore, we can write the following definition:

Definition 2.1.1 ▸ Limit of n-variable functions

Let f be an n-variable function whose domain $D \subseteq \mathbb{R}^n$ contains some neighbourhood of $\boldsymbol{a} = (a_1, a_2, \cdots, a_n)$. For $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ We say that

$$\lim_{\boldsymbol{x} \to \boldsymbol{a}} f(\boldsymbol{x}) = L$$

if for all $\epsilon > 0$, there is some $\delta > 0$ such that $|f(\boldsymbol{x}) - L| < \epsilon$ whenever $d(\boldsymbol{x}, \boldsymbol{a}) < \delta$.

Note that for 1-variable functions, we can easily determine the existence of their limits at some value, and thus compute the limits, by checking the equality of their left- and right-limits. However, in $\mathbb{R}^n$, a vector $\boldsymbol{x}$ may approach $\boldsymbol{a}$ in **infinitely many** distinct paths, so we have to check that for all mappings $p, q : \mathbb{R} \to \mathbb{R}^n$ with $p(0) = q(0) = \boldsymbol{a}$, $\lim_{t\to 0} f(p(t))$ and $\lim_{t\to 0} f(q(t))$ exist and are equal in order to prove the existence of $\lim_{\boldsymbol{x}\to\boldsymbol{a}} f(\boldsymbol{x})$.

Notice that the above reasoning provides a convenient way to **disprove** the existence of $\lim_{\boldsymbol{x}\to\boldsymbol{a}} f(\boldsymbol{x})$.

Theorem 2.1.2 ▸ Disprove the existence of limits n-variable functions

Let f be an n-variable function whose domain $D \subseteq \mathbb{R}^n$ contains some neighbourhood of $\boldsymbol{a} = (a_1, a_2, \cdots, a_n)$. Then $\lim_{\boldsymbol{x}\to\boldsymbol{a}} f(\boldsymbol{x})$ **does not exist** if and only if there are mappings $p, q : \mathbb{R} \to \mathbb{R}^n$ with $p(0) = q(0) = \boldsymbol{a}$ such that $\lim_{t\to 0} f(p(t)) \neq \lim_{t\to 0} f(q(t))$

Note that we can perform basic arithmetic operations on limits for 1-variable functions. Similarly, we can prove the following theorem for multivariable functions:

Theorem 2.1.3 ▸ Limit laws for multivariable functions

Let f and g both be functions in n variables. If $\lim_{\boldsymbol{x}\to\boldsymbol{a}} f(\boldsymbol{x})$ and $\lim_{\boldsymbol{x}\to\boldsymbol{a}} g(\boldsymbol{x})$ both exist, then

1. $\lim_{\boldsymbol{x}\to\boldsymbol{a}} (f(\boldsymbol{x}) + g(\boldsymbol{x})) = \lim_{\boldsymbol{x}\to\boldsymbol{a}} f(\boldsymbol{x}) + \lim_{\boldsymbol{x}\to\boldsymbol{a}} g(\boldsymbol{x})$;
2. $\lim_{\boldsymbol{x}\to\boldsymbol{a}} f(\boldsymbol{x})g(\boldsymbol{x}) = (\lim_{\boldsymbol{x}\to\boldsymbol{a}} f(\boldsymbol{x}))(\lim_{\boldsymbol{x}\to\boldsymbol{a}} g(\boldsymbol{x}))$;
3. $\lim_{\boldsymbol{x}\to\boldsymbol{a}} \frac{f(\boldsymbol{x})}{g(\boldsymbol{x})} = \frac{\lim_{\boldsymbol{x}\to\boldsymbol{a}} f(\boldsymbol{x})}{\lim_{\boldsymbol{x}\to\boldsymbol{a}} g(\boldsymbol{x})}$, provided that $\lim_{\boldsymbol{x}\to\boldsymbol{a}} g(\boldsymbol{x}) \neq 0$.

Proof. Let $\lim_{\boldsymbol{x}\to\boldsymbol{a}} f(\boldsymbol{x}) = L_f$ and $\lim_{\boldsymbol{x}\to\boldsymbol{a}} g(\boldsymbol{x}) = L_g$.
For all $\epsilon > 0$, there are $\delta_f, \delta_g > 0$ such that $|f(\boldsymbol{x}) - L_f| < \frac{\epsilon}{2}$ whenever $d(\boldsymbol{x}, \boldsymbol{a}) < \delta_f$ and $|g(\boldsymbol{x}) - L_g| < \frac{\epsilon}{2}$ whenever $d(\boldsymbol{x}, \boldsymbol{a}) < \delta_g$
For all $\epsilon > 0$, take $\delta = \min\{\delta_f, \delta_g\}$. Whenever $d(\boldsymbol{x}, \boldsymbol{a}) < \delta$, we have:

$$\begin{aligned}
|f(\boldsymbol{x}) + g(\boldsymbol{x}) - (L_f + L_g)| &= |f(\boldsymbol{x}) - L_f + g(\boldsymbol{x}) - L_g| \\
&\leq |f(\boldsymbol{x}) - L_f| + |g(\boldsymbol{x}) - L_g| \\
&< \frac{\epsilon}{2} + \frac{\epsilon}{2} \\
&= \epsilon.
\end{aligned}$$

Therefore, $\lim_{\boldsymbol{x}\to\boldsymbol{a}} (f(\boldsymbol{x}) + g(\boldsymbol{x})) = \lim_{\boldsymbol{x}\to\boldsymbol{a}} f(\boldsymbol{x}) + \lim_{\boldsymbol{x}\to\boldsymbol{a}} g(\boldsymbol{x})$.
For all $\epsilon > 0$, there are also $\delta_f', \delta_g' > 0$ such that $|f(\boldsymbol{x}) - L_f| < \sqrt{\epsilon}$ whenever $d(\boldsymbol{x}, \boldsymbol{a}) < \delta_f'$ and $|g(\boldsymbol{x}) - L_g| < \sqrt{\epsilon}$ whenever $d(\boldsymbol{x}, \boldsymbol{a}) < \delta_g'$.
For all $\epsilon > 0$, take $\delta' = \min\{\delta_f', \delta_g'\}$. Whenever $d(\boldsymbol{x}, \boldsymbol{a}) < \delta'$, we have:

$$|(f(\boldsymbol{x}) - L_f)(g(\boldsymbol{x}) - L_g) - 0| = |f(\boldsymbol{x}) - L_f|\,|g(\boldsymbol{x}) - L_g| < \sqrt{\epsilon} \cdot \sqrt{\epsilon} = \epsilon.$$

Therefore, $\lim_{\boldsymbol{x}\to\boldsymbol{a}} (f(\boldsymbol{x}) - L_f)(g(\boldsymbol{x}) - L_g) = 0$. Note that

$$f(\boldsymbol{x})g(\boldsymbol{x}) = (f(\boldsymbol{x}) - L_f)(g(\boldsymbol{x}) - L_g) + L_g f(\boldsymbol{x}) + L_f g(\boldsymbol{x}) - L_f L_g,$$

so we have:

$$\begin{aligned}
\lim_{\boldsymbol{x}\to\boldsymbol{a}} f(\boldsymbol{x})g(\boldsymbol{x}) &= \lim_{\boldsymbol{x}\to\boldsymbol{a}} \left(\left(f(\boldsymbol{x})-L_f\right)\left(g(\boldsymbol{x})-L_g\right)+L_g f(\boldsymbol{x})+L_f g(\boldsymbol{x})-L_f L_g\right)\\
&= \lim_{\boldsymbol{x}\to\boldsymbol{a}} \left(f(\boldsymbol{x})-L_f\right)\left(g(\boldsymbol{x})-L_g\right)+L_g \lim_{\boldsymbol{x}\to\boldsymbol{a}} f(\boldsymbol{x})+L_f \lim_{\boldsymbol{x}\to\boldsymbol{a}} g(\boldsymbol{x})+L_f L_g\\
&= 0+L_g L_f+L_f L_g-L_f L_g\\
&= L_f L_g\\
&= \lim_{\boldsymbol{x}\to\boldsymbol{a}} f(\boldsymbol{x})g(\boldsymbol{x}) = \left(\lim_{\boldsymbol{x}\to\boldsymbol{a}} f(\boldsymbol{x})\right)\left(\lim_{\boldsymbol{x}\to\boldsymbol{a}} g(\boldsymbol{x})\right).
\end{aligned}$$

□

Finally, we can extend the squeeze theorem to n-variable functions:

Theorem 2.1.4 ▶ Squeeze theorem in n variables

Let f, g, h be functions in n variables. If $g(\boldsymbol{x}) \leq f(\boldsymbol{x}) \leq h(\boldsymbol{x})$ whenever $d(\boldsymbol{x},\boldsymbol{a}) < c$ for some real constant c, and $\lim_{\boldsymbol{x}\to\boldsymbol{a}} g(\boldsymbol{x}) = \lim_{\boldsymbol{x}\to\boldsymbol{a}} h(\boldsymbol{x}) = L$, then $\lim_{\boldsymbol{x}\to\boldsymbol{a}} f(\boldsymbol{x}) = L$.

2.2 Continuity of Multivariable Functions

We define continuity for multivariable functions similarly to the case of 1-variable functions.

Definition 2.2.1 ▶ Continuity of n-variable functions

A function $f : \mathbb{R}^n \to \mathbb{R}$ is **continuous** at $\boldsymbol{a}$ if

$$\lim_{\boldsymbol{x}\to\boldsymbol{a}} f(\boldsymbol{x}) = f(\boldsymbol{a}).$$

If f is not continuous at $\boldsymbol{a}$, we say that $\boldsymbol{a}$ is a **discontinuity** of f.
In particular, f is said to be **continuous on** $D \subseteq \mathbb{R}^n$ if it is continuous at every point in D.

Just like 1-variable functions, continuity is preserved under simple arithmetic operations for multivariable functions.

Theorem 2.2.2 ▶ Continuous n-variable functions under arithmetic operations

If f and g are functions in n variables which are continous at $\boldsymbol{a}$, then $f \pm g$ and $f \cdot g$ are both continuous at $\boldsymbol{a}$. In particular, if $g(\boldsymbol{a}) \neq 0$, then $\frac{f(x)}{g(x)}$ is continuous at $\boldsymbol{a}$ as well.

Continuity for n-variable functions is also preserved under function composition similarly to 1-variable functions.

Theorem 2.2.3 ▸ Continuity of n-variable functions under composition

If f is an n-variable function which is continous at $\boldsymbol{a}$, and g is a 1-variable function which is continuous at $f(\boldsymbol{a})$, then the function

$$h(\boldsymbol{x}) = (g \circ f)(\boldsymbol{x}) = g(f(\boldsymbol{x}))$$

is continuous at $\boldsymbol{a}$.

As a consequence of the above theorems, the following functions are continuous over their entire domains:

- Multivariable polynomials;
- Multivariable trigonometric functions;
- Multivariable exponential functions;
- Multivariable rational functions.

2.3 Differentiability of Multivariable Functions

A natural next step from continuity is differentiability for n-variable functions, which is a bit more complicated than 1-variable functions, as we can differentiate with respect to each of the variables for a function with more than one single independent variable.

2.3.1 Partial Derivatives

The notion of **partial derivatives** can be interpreted as follows: suppose we have a function f in n variables $x_1, x_2, \cdots, x_n$, we wish to find the rate of change of f with respect to some x_i only while keeping the other $n-1$ variables constant.

Formally, we have the following definition:

Definition 2.3.1 ▸ Partial derivative

Suppose f is an n-variable function, we define the **partial derivative** of f **with respect to x_i** as the function

$$f_{x_i}(\boldsymbol{x}) = \frac{\partial f}{\partial x_i} := \lim_{\Delta x_i \to 0} \frac{f(x_1, x_2, \cdots, x_{i-1}, x_i + \Delta x_i, x_{i+1}, \cdots, x_n) - f(x_1, x_2, \cdots, x_n)}{\Delta x_i}.$$

In other words, suppose we define a function $g(x_i) = f(x_1, x_2, \cdots, x_i, \cdots, x_n)$, then the partial derivative of f with respect to x_i is just the derivative of g with respect to x_i, i.e., $f_{x_i}(\boldsymbol{x}) = g'(x_i)$.

Note that if f is an n-variable function, then the partial derivatives of f are also n-variable functions, which we can still differentiate with respect to each of the n-variables. Performing partial differentiation of f yields the **n-th order partial derivatives** of f.

Conventionally, we denote an n-th order partial derivative of f by writing in the subscript of f the variables we differentiate it with respect to **in the same order** of these differentiation. For example, f_{xy} means the second order partial derivative of f obtained by differentiating f first with respect to x and then with respect to y.

We have the following theorem for n-th order derivatives:

Theorem 2.3.2 ▸ Clairaut's theorem

Let f be an n-variable function defined on D and let $\boldsymbol{a} \in D$. If the functions f_{xy} and f_{yx} are continuous on D, then

$$f_{xy}(\boldsymbol{a}) = f_{yx}(\boldsymbol{a}).$$

2.3.2 Differentiability

To define differentiability of n-variable functions rigorously, we first introduce the following preliminary definition:

Definition 2.3.3 ▸ Interior point

Let $P \in D \subseteq \mathbb{R}^n$. P is known as an **interior point** of D if there exists some $\epsilon > 0$ such that the set

$$B_\epsilon(P) := \{\, Q \in \mathbb{R}^n \,:\, d(P, Q) < \epsilon \,\}$$

is a subset of D. The set of all interior points of D is known as the **interior** of D. In particular, if $P \in D$ is not an interior point of D, then P is a **boundary point** of D. The set of all boundary points of D is known as the **boundary** of D.

Remark. If every point in D is an interior point of D, i.e., D equals its interior, then D is said to be **open**.

And so we define differentiability as follows:

Definition 2.3.4 ▸ Differentiability of n-variable functions

Let $D \subseteq \mathbb{R}^n$ and let P with position vector $\boldsymbol{p}$ be an interior point of D. A function $f : D \to \mathbb{R}$ is **differentiable** at $\boldsymbol{p}$ if there exists a **linear mapping** $L : \mathbb{R}^n \to \mathbb{R}$ such that

$$\lim_{\Delta \boldsymbol{p} \to \boldsymbol{0}} \frac{f(\boldsymbol{p} + \Delta \boldsymbol{p}) - f(\boldsymbol{p}) - L(\Delta \boldsymbol{p})}{\|\Delta \boldsymbol{p}\|} = 0.$$

The linear mapping L is known as the **(total) derivative** of f at $\boldsymbol{p}$, which is denoted as $\mathrm{D}f_{\boldsymbol{p}}$.
f is said to be **differentiable on D** if it is **continuous on D** and **differentiable at every interior point** in D.

Recall that the graph of an n-variable function f is a surface in $\mathbb{R}^{n+1}$, which is analogous to a curve in $\mathbb{R}^2$ which is the graph of a 1-variable function. Therefore, we can define the notion of a **tangent plane** analogously to that of a tangent line.

Let $T_{\boldsymbol{a}}\mathbb{R}^n$ denote the set of all vectors in $\mathbb{R}^n$ with initial point whose position vector is $\boldsymbol{a}$. Then the position vector in $\mathbb{R}^n$ of any vector $\boldsymbol{b}$ in $T_{\boldsymbol{a}}\mathbb{R}^n$ is $\boldsymbol{a} + \boldsymbol{b}$.

Thus, we can think $\mathrm{D}f_{\boldsymbol{a}}$ as a linear mapping $\boldsymbol{b} \mapsto \mathrm{D}f_{\boldsymbol{a}} \in \mathbb{R}$ for all vectors $\boldsymbol{b} \in T_{\boldsymbol{a}}\mathbb{R}^n$. Geometrically, this is the change in "height" between the initial and terminal points of $\boldsymbol{b}$.

Definition 2.3.5 ▸ Tangent plane

Let f be an n-variable function defined on $D \subseteq \mathbb{R}^n$. Let $f(\boldsymbol{x}) = f(x_1, x_2, \cdots, x_n) = x_{n+1}$, then the **tangent plane** to f at $(x_1, x_2, \cdots, x_{n+1})$ is defined to be the graph of the mapping $\boldsymbol{y} \mapsto f(\boldsymbol{x}) + \mathrm{D}f_{\boldsymbol{x}}(\boldsymbol{y} - \boldsymbol{x})$.

Next, we shall introduce a way to systematically find this linear mapping $\mathrm{D}f$.

Theorem 2.3.6 ▸ Formula for total derivative

If f is an n-variable function which is differentiable at $\boldsymbol{a} = (a_1, a_2, \cdots, a_n)$, then

$$\mathrm{D}f_{\boldsymbol{a}}(\boldsymbol{x}) = \mathrm{D}f_{\boldsymbol{a}}(x_1, x_2, \cdots, x_n) = \sum_{i=1}^{n} f_{x_i}(\boldsymbol{a}) x_i.$$

Proof. Note that $\mathrm{D}f_{\boldsymbol{a}}$ is a linear transformation, so it suffices to prove that $\mathrm{D}f_{\boldsymbol{a}}(\boldsymbol{e}_i) = f_{x_i}(\boldsymbol{x})$ for $i = 1, 2, \cdots, n$ where $\boldsymbol{e}_i$ is the i-th vector in the standard basis for $\mathbb{R}^n$.
Let $\Delta \boldsymbol{p} = h\boldsymbol{e}_i$, then by Definition 2.3.4, we have:

$$\lim_{h \to 0} \frac{f(\boldsymbol{a} + h\boldsymbol{e}_i) - f(\boldsymbol{a}) - h\mathrm{D}f_{\boldsymbol{a}}(\boldsymbol{e}_i)}{h} = 0.$$

Re-arranging the above equation, we have:

$$\mathrm{D}f_{\boldsymbol{a}}(\boldsymbol{e}_i) = \lim_{h\to 0} \frac{f(\boldsymbol{a}+h\boldsymbol{e}_i) - f(\boldsymbol{a})}{h} = f_{x_i}(\boldsymbol{a}).$$

□

Remark. Note that even though we compute the total derivative using the partial derivatives, the existence of partial derivatives **does not imply** differentiability.

Analogously to 1-variable calculus, we can similarly prove the following laws for total derivatives:

Theorem 2.3.7 ▸ Arithmetic operations on total derivatives

Let f and g be n-variable functions which are differentiable at $\boldsymbol{a}$, then

1. $f \pm g$ is differentiable at $\boldsymbol{a}$ and $\mathrm{D}(f \pm g)_{\boldsymbol{a}}(\boldsymbol{x}) = \mathrm{D}f_{\boldsymbol{a}}(\boldsymbol{x}) \pm \mathrm{D}g_{\boldsymbol{a}}(\boldsymbol{x})$;
2. fg is differentiable at $\boldsymbol{a}$ and $\mathrm{D}(fg)_{\boldsymbol{a}}(\boldsymbol{x}) = g(\boldsymbol{x})\mathrm{D}f_{\boldsymbol{a}}(\boldsymbol{x}) + f(\boldsymbol{x})\mathrm{D}g_{\boldsymbol{a}}(\boldsymbol{x})$;
3. cf is differentiable at $\boldsymbol{a}$ for all $c \in \mathbb{R}$, and $\mathrm{D}(cf)_{\boldsymbol{a}}(\boldsymbol{x}) = c\mathrm{D}f_{\boldsymbol{a}}(\boldsymbol{x})$;
4. $\frac{f}{g}$ is differentiable at $\boldsymbol{a}$ if $g(\boldsymbol{a}) \neq 0$ and $\mathrm{D}\left(\frac{f}{g}\right)_{\boldsymbol{a}}(\boldsymbol{x}) = \frac{1}{g(\boldsymbol{a})^2}(g(\boldsymbol{x})\mathrm{D}f_{\boldsymbol{a}}(\boldsymbol{x}) - f(\boldsymbol{x})\mathrm{D}g_{\boldsymbol{a}}(\boldsymbol{x}))$.

Theorem 2.3.8 ▸ Chain rule for multivariable functions

Let u be a differentiable function in n variables $x_1, x_2, \cdots, x_n$, and let each of the x_i's be differentiable functions in m variables $t_1, t_2, \cdots, t_m$, then

$$\frac{\partial u}{\partial t_j} = \sum_{i=1}^{n} \frac{\partial u}{\partial x_i}\frac{\partial x_i}{\partial t_j}$$

for $j = 1, 2, \cdots, m$.

Lastly, here is a more straight-forward way to check differentiability:

Theorem 2.3.9 ▸ Differentiability theorem

Let f be an n-variable function defined on $D \subseteq \mathbb{R}^n$ and let $\boldsymbol{a} \in D$. If all first order partial derivatives of f are defined on D and continuous at $\boldsymbol{a}$, then f is differentiable at $\boldsymbol{a}$.

Remark. The converse of the above theorem is **false**, i.e., a differentiable function might have discontinuous partial derivatives!

2.3.3 Gradient Vectors

Note that in an n-dimensional space, we can describe a direction with the **unit vector** in that direction. With this, we are able to compute the rate of change of a function f at some point with position vector $\boldsymbol{a}$ in the direction of some unit vector $\boldsymbol{u}$, i.e., the change in $f(\boldsymbol{x})$ per unit length from $\boldsymbol{a}$ in the direction of $\boldsymbol{u}$. More formally, we have the following definition:

Definition 2.3.10 ▸ Directional derivative

Let f be an n-variable function and $\boldsymbol{u}$ be a unit vector in $\mathbb{R}^n$. The **direction derivative** of f at $\boldsymbol{a}$ in the direction of $\boldsymbol{u}$ is defined as

$$\mathrm{D}f_{\boldsymbol{a}}(\boldsymbol{u}) = \lim_{h \to 0} \frac{f(\boldsymbol{a} + h\boldsymbol{u}) - f(\boldsymbol{a})}{h}$$

provided that the limit exists.

Remark. The partial derivatives of f is just special cases of directional derivatives in the directions of the vectors in the standard basis of $\mathbb{R}^n$.

Recall that the existence of partial derivatives does not imply differentiability, so a function can still be not differentiable even if all the directional derivatives are defined at a point. However, conversely, differentiability does imply the existence of all directional derivatives.

Theorem 2.3.11 ▸ Directional derivatives of differentiable functions

Let f be a function in n variables $x_1, x_2, \cdots, x_n$ which is differentiable at $\boldsymbol{a}$, then all of the directional derivatives of f at $\boldsymbol{a}$ exist, and for all unit vectors $\boldsymbol{u} \in \mathbb{R}^n$,

$$\mathrm{D}f_{\boldsymbol{a}}(\boldsymbol{u}) = \sum_{i=1}^{n} f_{x_i}(\boldsymbol{a}) u_i,$$

where

$$\boldsymbol{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

Note that the formula in 2.3.11 resembles the dot product between two vectors, which gives us motivation to define the following:

Definition 2.3.12 ▸ Gradient vector

et f be a function in n variables $x_1, x_2, \cdots, x_n$, then the **gradient vector** of f is defined as

$$\nabla f(x_1, x_2, \cdots, x_n) = \begin{bmatrix} f_{x_1}(x_1, x_2, \cdots, x_n) \\ f_{x_2}(x_1, x_2, \cdots, x_n) \\ \vdots \\ f_{x_n}(x_1, x_2, \cdots, x_n) \end{bmatrix}.$$

With the notion of the gradient vector, we are able to re-write the formula for directional derivative as $\mathrm{D} f_{\boldsymbol{a}}(\boldsymbol{u}) = \nabla f(\boldsymbol{a}) \cdot \boldsymbol{u}$.

We now follow up by discussing some useful properties of the gradient vector.

Theorem 2.3.13 ▸ Orthogonality between the gradient vector and level sets

Let f be a differentiable function in n variables and let $\boldsymbol{a} \in \mathbb{R}^n$. Let S be the level set of f containing $\boldsymbol{a}$. If $\nabla f(\boldsymbol{a}) \neq \mathbf{0}$, then $\nabla f(\boldsymbol{a}) \perp S$.

Proof. Let S be the k-level set of f and parametrised by

$$R(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix},$$

then $f(R(t)) = k$. Differentiating both sides with respect to t yields

$$\frac{\mathrm{d}}{\mathrm{d}t} f(R(t)) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \frac{\partial x_i}{\partial t} = \nabla f(R(t)) \cdot R'(t) = 0.$$

Therefore, $\nabla f(R(t)) \perp R'(t)$ for all t, i.e., for all $\boldsymbol{a} \in S$, $\nabla f(\boldsymbol{a}) \perp S$ at $\boldsymbol{a}$. □

Note that Theorem 2.3.13 offers another way to find the tangent plane to a function f at $\boldsymbol{a}$. Let this tangent plane be T and let $\boldsymbol{r} \in T$ be an arbitrary vector, then $(\boldsymbol{r} - \boldsymbol{a}) \parallel T$ and so

$$\nabla f(\boldsymbol{a}) \cdot (\boldsymbol{r} - \boldsymbol{a}) = 0.$$

Furthermore, we can also prove the following theorem:

Theorem 2.3.14 ▸ Computing directional derivatives with the gradient vector

Let f be a differentiable function in n variables and let P be a point with position vector $\boldsymbol{p}$ such that $\nabla f(\boldsymbol{p}) \neq \boldsymbol{0}$. If $\boldsymbol{u}$ is a unit vector with initial point P and θ is the angel between $\boldsymbol{u}$ and $\nabla f(\boldsymbol{p})$, then

$$\mathrm{D}f_{\boldsymbol{p}}(\boldsymbol{u}) = \|\nabla f(\boldsymbol{p})\| \cos\theta.$$

Proof.

$$\begin{aligned}\mathrm{D}f_{\boldsymbol{p}}(\boldsymbol{u}) &= \nabla f(\boldsymbol{p}) \cdot \boldsymbol{u} \\ &= \|\nabla f(\boldsymbol{p})\| \, \|\boldsymbol{u}\| \cos\theta \\ &= \|\nabla f(\boldsymbol{p})\| \cos\theta.\end{aligned}$$

□

The above theorem implies that

$$-\|\nabla f(\boldsymbol{p})\| \leq \mathrm{D}f_{\boldsymbol{p}}(\boldsymbol{u}) \leq \|\nabla f(\boldsymbol{p})\|.$$

Note that $\mathrm{D}f_{\boldsymbol{p}}(\boldsymbol{u})$ attains maximum and minimum at $\theta = 0$ and $\theta = \pi$ respectively, so $\pm\nabla f(\boldsymbol{p})$ points to the directions of fastest and slowest changes of f respectively.

2.3.4 Implicit Differentiation in n-Variables

Given variables $x_1, x_2, \cdots, x_n$, sometimes it may not be easy or even possible to define a function relating one of the n variables to the rest $n-1$ variables. Therefore, to analyse the derivatives between these variables, we need to perform differentiation implicitly.

Let F be a function in n variables. Let $x_1, x_2, \cdots, x_n$ be such that $F(x_1, x_2, \cdots, x_n) = k$, then the set of all points $(x_1, x_2, \cdots, x_n)$ is exactly the k-level set of F.

Note that this relationship helps us **implicitly define** each of the x_i's as a function in the other $n-1$ variables. It is thus reasonable to differentiate each of the x_i's with respect to some x_j for $i \neq j$.

Theorem 2.3.15 ▸ Implicit differentiation in n variables

Let F be a differentiable function in n variables $x_1, x_2, \cdots, x_n$ and let k be a real constant. If $F(x_1, x_2, \cdots, x_{i-1}, x_i, x_{i+1}, \cdots x_n) = k$ defines x_i implicitly as a function of

$x_1, x_2, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n$ and $F_{x_i}(\boldsymbol{x}) \neq 0$, then

$$\frac{\partial x_i}{\partial x_j}(\boldsymbol{x}) = -\frac{F_{x_j}(\boldsymbol{x})}{F_{x_i}(\boldsymbol{x})}.$$

Proof. Differentiating both sides of $F(x_1, x_2, \cdots, x_i, \cdots x_n) = k$ with respect to x_j, by Theorem 2.3.8, we have:

$$\sum_{k=1}^{n} F_{x_k}(\boldsymbol{x}) \frac{\partial x_k}{\partial x_j} = 0,$$

which simplifies to

$$\begin{aligned}\sum_{\substack{1 \leq k \leq n \\ k \neq i,j}} F_{x_k}(\boldsymbol{x}) \frac{\partial x_k}{\partial x_j} + F_{x_j}(\boldsymbol{x}) \frac{\partial x_j}{\partial x_j} + F_{x_i}(\boldsymbol{x}) \frac{\partial x_i}{\partial x_j} &= F_{x_j}(\boldsymbol{x}) + F_{x_i}(\boldsymbol{x}) \frac{\partial x_i}{\partial x_j} \\ &= 0.\end{aligned}$$

Since $F_{x_i}(\boldsymbol{x}) \neq 0$, we have:

$$\frac{\partial x_i}{\partial x_j}(\boldsymbol{x}) = -\frac{F_{x_j}(\boldsymbol{x})}{F_{x_i}(\boldsymbol{x})}.$$

□

Applying implicit differentiation, we can conveniently compute the tangent plane to the graph of a function at some point in the 3-dimensional Euclidean space.

Let F be a function of 3 variables and let S be the k-level set of F for some real constant k, i.e., $S = \{(x, y, z) : F(x, y, z) = k\}$.

Suppose that $F(x, y, z) = k$ defines one of x, y, z implicitly as a function of the other two variables. Let $\boldsymbol{v}$ be the position vector of some point $(a, b, c) \in S$, then we can differentiate $\boldsymbol{v}$ with respect to x and y respectively to obtain two tangent vectors to S at (a, b, c) in the x- and y-directions respectively, given by

$$\begin{aligned}\frac{\partial \boldsymbol{v}}{\partial x} &= \begin{bmatrix} 1 \\ 0 \\ \frac{\partial z}{\partial x}(a, b, c) \end{bmatrix}; \\ \frac{\partial \boldsymbol{v}}{\partial y} &= \begin{bmatrix} 0 \\ 1 \\ \frac{\partial z}{\partial y}(a, b, c) \end{bmatrix}.\end{aligned}$$

Therefore, we compute a normal vector to S at (a,b,c) given by

$$\begin{bmatrix} 1 \\ 0 \\ \frac{\partial z}{\partial x}(a,b,c) \end{bmatrix} \times \begin{bmatrix} 0 \\ 1 \\ \frac{\partial z}{\partial y}(a,b,c) \end{bmatrix} = \begin{bmatrix} \frac{\partial z}{\partial x}(a,b,c) \\ \frac{\partial z}{\partial y}(a,b,c) \\ -1 \end{bmatrix}.$$

By 1.2.4, the equation for the tangent plane to S at (a,b,c) is

$$\frac{\partial z}{\partial x}(a,b,c)x + \frac{\partial z}{\partial y}(a,b,c)y - z = \frac{\partial z}{\partial x}(a,b,c)a + \frac{\partial z}{\partial y}(a,b,c)b - c.$$

2.4 Optimisation Problems in Multivariable Calculus

In this section, we discuss an important application of multivariable calculus in optimisation problems.

2.4.1 Extrema of Multivariable Functions

We first give the definition of extrema in multivariable functions.

Definition 2.4.1 ▸ Local extrema of n-variable functions

Let $f : D \to \mathbb{R}$ be a function in n variables, where $D \subseteq \mathbb{R}^n$. Let B be some disk centred at $C \in D$, then for all points $P \in B \cap D$:

- C is a **local maximum** of f if $f(P) \leq f(C)$;
- C is a **local minimum** of f if $f(P) \geq f(C)$.

A local minimum or local maximum is known as a **local extremum** of f.

Definition 2.4.2 ▸ Global extrema of n-variable functions

Let $f : D \to \mathbb{R}$ be a function in n variables, where $D \subseteq \mathbb{R}^n$. For all points $Q \in D$:

- C is a **global maximum** of f if $f(Q) \leq f(C)$;
- C is a **global minimum** of f if $f(Q) \geq f(C)$.

A global minimum or global maximum is known as a **global extremum** of f.

Remark. Note that all global extrema of a function are necessarily its local extrema, but the coverse is not true.

Observe that if P is a local extremum of some function f, then all directional derivatives of f at P must evaluate to 0. This is equivalent to having all partial derivatives of f evaluate to 0 at P, which motivates the following definition:

Definition 2.4.3 ▸ Critical point

Let $f : D \to \mathbb{R}$ be a function in n variables $x_1, x_2, \cdots, x_n$ which is differentiable at some point P in the interior of D. If $f_{x_i}(P) = 0$ for all i, then P is said to be a **critical point** of f.

Combining Definitions 2.4.1 and 2.4.3, we have:

Theorem 2.4.4 ▸ Relationship between local extrema and critical points

If a function f is differentiable at some point P and achieves a local extremum at P, then P is a critical point of f.

Note that the converse to the above theorem is false. We shall illustrate this with a counter example.

Consider the function $f(x, y) = y^2 - x^2$. Note that $f_x(x, y) = -2x$ and $f_y(x, y) = 2y$. Let $f_x(x, y) = f_y(x, y) = 0$, we have $x = y = 0$, so $(0, 0)$ is the only critical point of f.

Note that $f(0, 0) = 0$. However, for all $t \neq 0$, we have $f(t, 0) = -t^2 < 0$ and $f(0, t) = t^2 > 0$, which means that $f(0, 0)$ is neither a local minimum nor a local maximum.

Therefore, a function may not attain any local extremum at its critical points.

Definition 2.4.5 ▸ Saddle point

Let $f : D \to \mathbb{R}$ be a function and let P be a critical point of f. If for all disks B centred at P, there is some $Q_1 \in B$ such that $f(Q_1) > f(P)$ and there is some $Q_2 \in B$ such that $f(Q_2) < f(P)$, then P is called a **saddle point** of f.

Next, we shall discuss the notion of global extrema. Note that a function might be unbounded, so it is necessary to restrict the function to a certain subset of its domain to ensure the existence of global extrema.

Definition 2.4.6 ▸ Openness of a set

A set D is called **open** if for all $X \in D$ there is some disk B centred at X such that $B \subseteq D$.

Definition 2.4.7 ▸ Closed and bounded set

A set is called **closed** if its complement is open. A set D is called **bounded** if there is some disk B such that $D \subseteq B$.

Theorem 2.4.8 ▸ Extreme value theorem

If $f : D \to \mathbb{R}$ is continuous on D which is a closed and bounded set, then f has at least one global maximum and at least one global minimum.

We thus give the following algorithm in computing the global extrema of a function $f : D \to \mathbb{R}$ where D is closed and bounded:

An algorithm to compute global extrema

1. Find the critical points of f.
2. Evaluate f at each of the critical points.
3. Find the extreme values of f on the boundary of D.
4. Among all values computed in the previous two steps, the largest and the smallest are the global maximum and global minimum of f respectively.

2.4.2 Lagrange Multiplier

We now consider a special type of optimisation problems:

Let $f : D \to \mathbb{R}$ be an n-variable function and $C \subseteq D$ be a curve in D. Consider f restricted to C, i.e., the function $f\restriction_C$. Can we optimise this restriction of f, i.e., can we find the global extrema of f subject to the contraint C?

It turns out that solving the above problem is possible if C is given as some level set of an n-variable function.