UNSW Mathematics Society

MathSoc

(Higher) Several Variable Calculus

Differential Calculus

Presented by: Henry Lam

Term 1, 2020

Differential Calculus

Functions in $\mathbb{R}^n$

Up till now we have mainly dealt with single-valued functions, but we'll now generalise some key concepts from these to higher-dimension functions.

Definition 1

A **curve** in $\mathbb{R}^n$ is a set of vectors where the components are dependent on a single variable, i.e. from $\mathbb{R}$ to $\mathbb{R}^n$. They are generally described parametrically by:

$$\mathbf{c}(t) = (c_1(t), c_2(t), \ldots, c_n(t))^T$$

Each $c_i(t)$ is a function from $\mathbb{R}$ to $\mathbb{R}$.

The most common dimensions that you'll be working with are $\mathbb{R}^2$ and $\mathbb{R}^3$.

Surfaces

In $\mathbb{R}^3$, we deal with surfaces.

Definition 2

A **surface** is the image of $D \subset \mathbb{R}^2$ under a function $\mathbf{f}$ from D to $\mathbb{R}^3$.

Essentially, this is just an extension of curves, with two parameters to describe the domain instead.
The following three are the most common methods to describe surfaces:

- **Explicitly**: $z = f(x, y)$
- **Implicitly**: $g(x, y, z) = 0$
- **Parametrically**: $x = x(u, v), y = y(u, v)$ and $z = z(u, v)$.

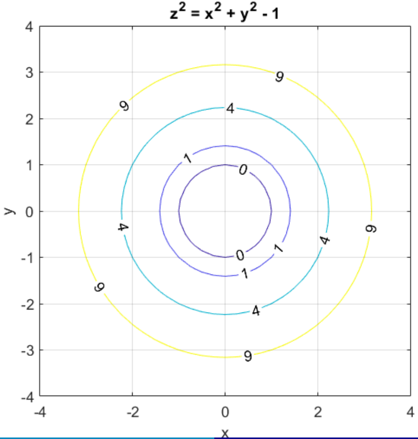
Contour Plots

As it is quite difficult to draw surfaces directly, one way we could visualise their shape is by plotting horizontal slices.

Contour Plots Example

Example 1

Plot the contour plots of $x^2 + y^2 - z = 1$ at the following values: $z = 0, 1, 2, 3$.



Partial Derivatives

From first year, we have the following definition for partial derivatives.

Definition 3

The $\mathbf{k^{th}}$ **first-order partial derivative** of $f : \mathbb{R}^n \to \mathbb{R}$ is denoted by $\frac{\partial f}{\partial x_k}, \partial_k f, D_k f, f_k$ or f_{x_k} is:

$$\frac{\partial f}{\partial x_k}(\mathbf{x}) = \lim_{h \to 0} \frac{f(\mathbf{x} + h\mathbf{e}_k) - f(\mathbf{x})}{h},$$

where $\mathbf{e}_k$ is the $\mathrm{k^{th}}$ standard basis for $\mathbb{R}^n$ (all zeroes except for the $\mathrm{k^{th}}$ value being 1).

We'll be using $\frac{\partial f}{\partial x_k}$ or f_{x_k} throughout this session.

Partial Derivatives

For higher-order partial derivatives, we must note the order of differentiation. For example:

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)$$

Here we differentiate with respect to y first, then x. Generally speaking, we look to the right-most variable and move left. Throughout this session, we'll be denoting this by f_{yx}.

Partial Derivatives

As stated in the previous slide, we need to worry about the order of differentiation, however this can quickly become very cumbersome. So we'll introduce a very useful theorem:

Theorem 1: Clariaut's Theorem

If $f, f_{x_i}, f_{x_j}, f_{x_i x_k}, f_{x_k x_i}$ all exist and are continuous on an open set around $\mathbf{a}$ then

$$f_{x_i x_k}(\mathbf{a}) = f_{x_k x_i}(\mathbf{a})$$

i.e. both commute to the same value.

As the majority of the functions that we deal with are infinitely differentiable (e.g. sine, cosine, polynomials, etc.) we can abuse this theorem.

Jacobian Matrix

Definition 4

If all the partial derivatives of $\mathbf{f} : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$ exist at $\mathbf{a} \in \Omega$ then the **Jacobian** matrix at $\mathbf{a}$ is:

$$J_{\mathbf{a}}\mathbf{f} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{a}) & \frac{\partial f_1}{\partial x_2}(\mathbf{a}) & \dots & \frac{\partial f_1}{\partial x_n}(\mathbf{a}) \\ \frac{\partial f_2}{\partial x_1}(\mathbf{a}) & \frac{\partial f_2}{\partial x_2}(\mathbf{a}) & \dots & \frac{\partial f_2}{\partial x_n}(\mathbf{a}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{a}) & \frac{\partial f_m}{\partial x_2}(\mathbf{a}) & \dots & \frac{\partial f_m}{\partial x_n}(\mathbf{a}) \end{pmatrix}$$

Note

Each column of the Jacobian corresponds to the derivative of each x_k.

Differentiability of $f : \mathbb{R}^n \to \mathbb{R}^m$

Analogy: From the one-dimensional case, we know that the definition of a derivative at a point, say a is:

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

$$0 = \lim_{x \to a} \frac{f(x) - f(a) - f'(a)(x - a)}{x - a}$$

This can be interpreted as a comparison between the euclidean distance between the actual function and its first-order approximation, against the x-axis difference.

Differentiability of $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$

We can extend this idea further to apply to higher-dimensions, through the following:

Definition 5

A function $\mathbf{f} : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ is differentiable at $\mathbf{a} \in \Omega$ if there is a linear map $\mathbf{L} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ such that

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

The matrix of $\mathbf{L}$ is called the **derivative** of $\mathbf{f}$ at $\mathbf{a}$ and is denoted as $D_{\mathbf{a}}\mathbf{f}$.

For the majority of the cases in this course, the **derivative** is just the **Jacobian matrix**, $\mathbf{J_a f}$.

Differentiation Example

Example 2

Show that the function $\mathbf{f}(\mathbf{x}) = (x^2 + y^2, xy^2)$ is differentiable at $\mathbf{a} = (1, 0)$, with derivative of $\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$.

Solution Q2

Subbing everything into the limit yields:

$$\lim_{\mathbf{x}\to(1,0)} \frac{\|(x^2+y^2-1-2(x-1), xy^2)\|}{\|(x-1,y)\|}$$

$$= \lim_{\mathbf{x}\to(1,0)} \frac{\sqrt{[(x-1)^2+y^2]^2+x^2y^4}}{\sqrt{(x-1)^2+y^2}}$$

Using the triangle inequality, i.e. $|\mathbf{a}+\mathbf{b}| \le |\mathbf{a}|+|\mathbf{b}|$,

$$\le \lim_{\mathbf{x}\to(1,0)} \frac{\sqrt{[(x-1)^2+y^2]^2}}{\sqrt{(x-1)^2+y^2}} + \lim_{\mathbf{x}\to(1,0)} \frac{|x|y^2}{\sqrt{(x-1)^2+y^2}}$$

$$= \lim_{\mathbf{x}\to(1,0)} \frac{|x|y^2}{\sqrt{(x-1)^2+y^2}}$$

Solution Q2 Cont.

As $(x-1)^2 + y^2 \geq y^2 \geq 0 \iff [(x-1)^2 + y^2]^{-1} \leq y^{-2}$,

$$
\begin{aligned}
&\leq \lim_{\mathbf{x}\to(1,0)} \frac{|x|y^2}{|y|} \\
&= 0
\end{aligned}
$$

As this limit is strictly non-negative, then by the pinching theorem, its limit is 0.

General Tips for Differentiation

- Whenever you decompose the function using inequalities, make sure that the relevant component goes to zero!
- Also make sure that you keep the $\leq$ order, as we want to show that it is bounded above by 0.
- The three main inequalities are:
  - Triangle Inequality: $\|\mathbf{a} - \mathbf{b}\| \leq \|\mathbf{a}\| + \|\mathbf{b}\|$
  - Cauchy-Schwartz: $|\mathbf{a} \cdot \mathbf{b}| \leq \|\mathbf{a}\|\|\mathbf{b}\|$
  - Manipulating the numerator/denominator.

Differentiation

The previous example illustrates how you would go about proving differentiability by definition. While this is nice, it can quickly become very tedious and/or extremely difficult to do. Luckily for us, we can use the following theorem:

Theorem 2: Differentiability from Derivatives

Consider an open set $\Omega \subset \mathbb{R}^n$ and $\mathbf{f} : \Omega \to \mathbb{R}^m$. If $\frac{\partial f_i}{\partial x_j}$ exists and is continuous on Ω for all $i = 1, 2, ..., n$ and $j = 1, 2, ..., m$, then $\mathbf{f}$ is differentiable on Ω.

As we mostly deal with infinitely differentiable functions, this theorem will apply in most cases.

Multivariate Chain Rule

Definition 6

Suppose that **f** is a function in terms of $u_1, u_2, ..., u_m$, with each u_i being functions in terms of $x_1, x_2, ..., x_n$. Then, the following holds:

$$\frac{\partial f}{\partial x_j} = \sum_{i=1}^{m} \frac{\partial f}{\partial u_i} \times \frac{\partial u_i}{\partial x_j}$$

Multivariate Chain Rule Example

Example 3

Suppose $f(u,v) = uv^3 + 2v$ with $u(x,y) = 2x^3 + y$ and $v(x,y) = 9e^{xy}$. Find $\frac{\partial f}{\partial x}$.

Q3 Solution

We want to find f_x and so we focus on u_x, v_x.

$$f_u = v^3 \qquad f_v = 3uv^2 + 2$$
$$u_x = 6x^2 \qquad v_x = 9ye^{xy}$$

Subbing it all into the chain rule yields:

$$\begin{aligned} f_x &= v^3 \times 6x^2 + (3uv^2 + 2) \times 9ye^{xy} \\ &= (9e^{xy})^3 \times 6x^2 + (3(2x^3 + y)(9e^{xy})^2 + 2) \times 9ye^{xy} \end{aligned}$$

Multivariate Chain Rule

As we generalised the derivative of multi-variable functions, we'll also do the same for the chain rule.

Theorem 3: Multivariate Chain Rule

Consider $\mathbf{f} : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$ and $\mathbf{g} : \Omega' \subset \mathbb{R}^m \to \mathbb{R}^p$. If both $\mathbf{g}$ and $\mathbf{f}$ are differentiable, then so is $\mathbf{g} \circ \mathbf{f} : \Omega \to \mathbb{R}^p$. In particular,

$$D_{\mathbf{a}}(\mathbf{g} \circ \mathbf{f}) = D_{\mathbf{f}(\mathbf{a})}(\mathbf{g}) D_{\mathbf{a}}(\mathbf{f})$$

where $\mathbf{a} \in \Omega$ and $\mathbf{f}(\mathbf{a}) \in \Omega'$.

Note

This is just the matrix equivalent of the chain rule that we are familiar with! In fact, it closely resembles the univariate case, $(g \circ f)'(x) = g'(f(x))f'(x)$.

Multivariate Chain Rule Example

Example 4

Find $D_{\mathbf{a}}(\mathbf{g} \circ \mathbf{f})$, where $\mathbf{f}(x, y) = \begin{pmatrix} x^2 + y \\ x - 2y^2 \end{pmatrix}$, $\mathbf{g}(u, v) = \begin{pmatrix} 2u + v \\ \sin u \\ u + 2v^2 \end{pmatrix}$

and $\mathbf{a} = (1, 1)$.

Q4 Solution

Using the multivariate chain rule, we'll find the Jacobians of each function first.

$$D\mathbf{g} = \begin{pmatrix} 2 & 1 \\ \cos u & 0 \\ 1 & 4v \end{pmatrix} \qquad D\mathbf{f} = \begin{pmatrix} 2x & 1 \\ 1 & -4y \end{pmatrix}$$

Noting that $\mathbf{f}(1,1) = (2,-1)$ leads to:

$$D_{(2,-1)}\mathbf{g} = \begin{pmatrix} 2 & 1 \\ \cos 2 & 0 \\ 1 & -4 \end{pmatrix} \qquad D_{(1,1)}\mathbf{f} = \begin{pmatrix} 2 & 1 \\ 1 & -4 \end{pmatrix}$$

$$\implies D_{(1,1)}(\mathbf{g} \circ \mathbf{f}) = \begin{pmatrix} 5 & -2 \\ 2\cos 2 & \cos 2 \\ -2 & 17 \end{pmatrix}.$$

Directional Derivatives

Earlier we briefly went over the idea of partial derivatives. We'll be extending this idea further, by considering any direction, rather than just the standard basis vectors.

Definition 7

The directional derivative of $\mathbf{f} : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ in the direction of the unit vector $\hat{\mathbf{u}}$ at $\mathbf{a} \in \Omega$ is

$$D_{\hat{\mathbf{u}}} f(\mathbf{a}) = f'_{\hat{\mathbf{u}}}(\mathbf{a}) = \lim_{t \to 0} \frac{f(\mathbf{a} + t\hat{\mathbf{u}}) - f(\mathbf{a})}{t}.$$

Note

$D_{\hat{\mathbf{u}}} f(\mathbf{a})$ may be confusing, as it could be interpreted as the derivative of f, evaluated at $\hat{\mathbf{u}}$. Because of this, I recommend using the second notation, $f'_{\hat{\mathbf{u}}}(\mathbf{a})$.

Directional Derivatives (MATH2111)

In most cases, we can simplify this limit into the following.

Theorem 4

If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\mathbf{a}$ and $\hat{\mathbf{u}}$ is a unit vector, then:

$$f'_{\hat{\mathbf{u}}}(\mathbf{a}) = \nabla f(\mathbf{a}) \cdot \hat{\mathbf{u}}.$$

Directional Derivative Example

Example 5

A bushwalker is climbing a mountain, of which the equation is $h(x,y) = 400 - \frac{x^2+4y^2}{10000}$. Here x, y and h are measured in metres, the x-axis points East and the y-axis points North. The bushwalker is at a point P, 1600 metres West and 400 metres South of the peak.

(i) What is the slope of the mountain of P in the direction of the peak?

(ii) In which direction at P is the slope the greatest?

Q5 Solution

(i) Direction of the peak is towards the origin, i.e. $\mathbf{u} = (4, 1)$. Finding the gradient at $P = (-1600, -400)$ and the unit vector $\hat{\mathbf{u}}$:

$$\nabla h(P) = \frac{1}{25}(8, 8) \qquad \hat{\mathbf{u}} = \frac{1}{\sqrt{17}}(4, 1).$$

Thus, $h'_{\hat{\mathbf{u}}}(P) = \hat{\mathbf{u}} \cdot \nabla h(P) = \dfrac{8}{5\sqrt{17}}$.

Q5 Solution

(i) Direction of the peak is towards the origin, i.e. $\mathbf{u} = (4, 1)$. Finding the gradient at $P = (-1600, -400)$ and the unit vector $\hat{\mathbf{u}}$:

$$\nabla h(P) = \frac{1}{25}(8, 8) \qquad \hat{\mathbf{u}} = \frac{1}{\sqrt{17}}(4, 1).$$

Thus, $h'_{\hat{\mathbf{u}}}(P) = \hat{\mathbf{u}} \cdot \nabla h(P) = \dfrac{8}{5\sqrt{17}}$.

(ii) This is in the direction of the gradient, i.e. North-East.

Gradient

The gradient is just the Jacobian of a function $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$. One of the special properties of the gradient is that it points in the direction that the function experiences maximal change at any point.

Tangent Planes

Just like in the 1-Dimensional case, sometimes we are interested in finding the tangent line to the curve, however here we are interested in a tangent plane to the surface.

General Procedure

Suppose that you have a function that **implicitly** describes the surface, i.e. $g(x, y, z) = c$, then you can find the tangent plane to the surface at point $\mathbf{a}$, through the following:

$$\nabla g(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) = 0.$$

Gradient being tangential/normal

From the above note, the gradient is actually normal to the surface, rather than being tangential. The key difference is in the way **we define the surface/curve**, here we defined it **implicitly**.

Tangent Planes Example

Example 6

Find the tangent plane to the surface $z = xy^3 - y$ at $\mathbf{a} = (1, 1, 0)$.

Q6 Solution

Rearrange this to implicitly define this function, i.e. $g(x, y, z) = xy^3 - y - z = 0$. The gradient is then:

$$\nabla g(\mathbf{x}) = (y^3, 3xy^2 - 1, -1).$$

Now we just sub everything in:

$$\begin{aligned} \nabla g(1, 1, 0) \cdot (\mathbf{x} - \mathbf{a}) &= 0 \\ (1, 3 - 1, -1) \cdot (x - 1, y - 1, z) &= 0 \\ x + 2y - z &= 3 \end{aligned}$$

Normal Lines Example

Example 7

Find the normal line to the curve $y^2 = x^2 - 3$ at the point $a = (2, 1)$.

Q7 Solution

We'll arrange this to define the function implicitly, as this gradient is **normal** to the curve.

$$f(x,y) = x^2 - y^2 = 3 \implies \nabla f(x,y) = (2x, -2y).$$

Then the normal line to $f(x,y) = 3$ at $(2,1)$ is given by:

$$\mathbf{x} = (2,1) + t(4,-2) \qquad \text{for all real } t.$$

Affine Approximations

One application of tangent planes is that sometimes we want to approximate a function, f, close to a particular value, say $\mathbf{a}$. To do this, we just find the tangent plane to f at $\mathbf{a}$ and then sub in the point we want in.

Example 8

Approximate the value of $f(x, y) = xy^3 - y$ at $(1.1, 0.92)$ using an affine approximation.

Q8 Solution

From Q6, we know the tangent plane to $(1,1)$ which is quite close to $(1.1, 0.92)$:

$$x + 2y - z = 3 \iff z = x + 2y - 3$$

So we can just sub the value in,

$$f(1.1, 0.92) \approx 1.1 + 2(0.92) - 3 = -0.06.$$

Taylor Series

Just like about everything we have covered so far, Taylor series can also be generalised to multi-variable cases as well. We'll firstly focus on the 2-variable case, and from there it is quite straightforward to generalise to higher variable cases.

Taylor Series

Definition 8

Consider a function $f : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$ which is C^k on Ω. Then the k$^{\text{th}}$-order Taylor series at (a, b) can be expressed as:

$$
\begin{aligned}
P_k(x,y) \quad = \quad & f(a,b) + \frac{1}{1!}[f_x(a,b)(x-a) + f_y(a,b)(y-b)] \\
& + \frac{1}{2!}[f_{xx}(a,b)(x-a)^2 + 2f_{xy}(a,b)(x-a)(y-b) \\
& + f_{yy}(a,b)(y-b)^2] + \cdots \\
& + \frac{1}{k!}[\text{terms involving partial derivatives of order k}]
\end{aligned}
$$

Taylor Series

Pattern of coefficients

From the previous slide, you may have noticed that the coefficients of the kth-order derivatives were from a binomial expansion. This is generally true as most of the functions that you deal with has its partial derivatives commute, e.g. $f_{xy} = f_{yx}$. As such, you can think of the coefficient as the number of unique patterns that we get from arranging the differentiated terms.

Taylor Series Example

Example 9

Find the 5^{th} degree Taylor polynomial of $\sin(x+y)$ at $(0,0)$.

Q9 Solution

Finding all partial derivatives up to the fifth-degree at $(0, 0)$:

$$f = 0$$
$$f_x = f_y = 1$$
$$f_{xx} = f_{xy} = f_{yy} = 0$$
$$f_{xxx} = \cdots = f_{yyy} = -1$$
$$f_{xxxx} = \cdots = f_{yyyy} = 0$$
$$f_{xxxxx} = \cdots = f_{yyyyy} = 1$$

Thus, the fifth-order Taylor polynomial is:

$$P_5(x, y) = \frac{1}{1!}[x + y] + \frac{1}{3!}\left[-x^3 - 3x^2y - 3xy^2 - y^3\right]$$
$$\frac{1}{5!}\left[x^5 + 5x^4y + 10x^3y^2 + 10x^2y^3 + 5xy^4 + y^5\right]$$

Remainder Term

The remainder term in a Taylor series approximation is simply the next-order term (so from the previous example, it would be order 6), but rather than being evaluated at $\mathbf{a}$, its evaluated on some point lying on the line segment between $\mathbf{a}$ and $\mathbf{x}$.

Example 10

Find the corresponding remainder term to a second-order Taylor series approximation of $\sin(x + y)$ at $(0, 0)$.

Q10 Solution

We need the third-order term of the Taylor series, which is:

$$\frac{1}{3!}\left[f_{xxx}x^3 + 3f_{xxy}x^2y + 3f_{xyy}xy^2 + f_{yyy}y^3\right]$$

where each of the derivatives are evaluated at some $\mathbf{z}$ in between the segment from $(0,0)$ and (x,y), i.e. $\mathbf{z} = \mathbf{0} + t\mathbf{x}$ where $t \in [0,1]$.

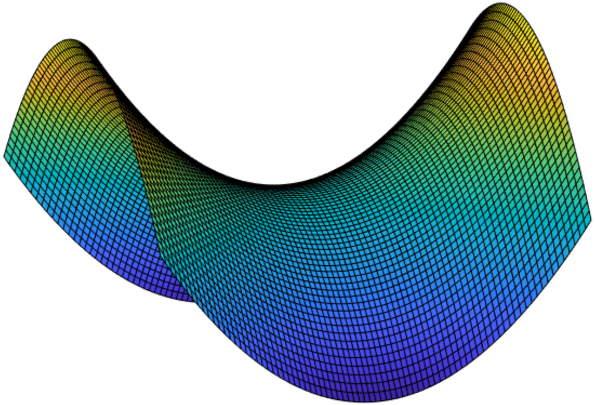
Classification of Extrema

Definition 9

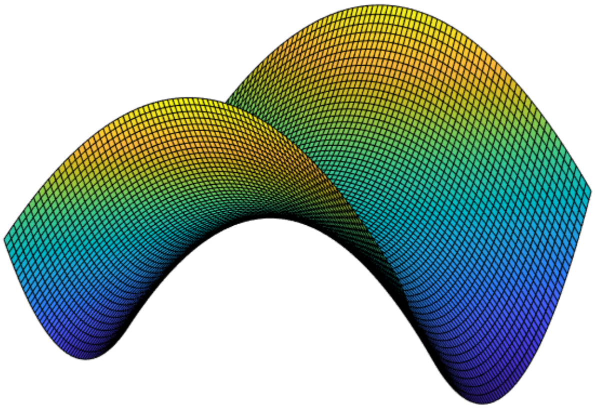
Suppose $f : \mathbb{R}^n \to \mathbb{R}$. Then $\mathbf{a}$ is a:

- critical point if $\nabla f(\mathbf{a}) = \mathbf{0}$ or it doesn't exist;
- global maximum if $f(\mathbf{x}) \leq f(\mathbf{a})$ for all $\mathbf{x} \in \mathbb{R}^n$;
- global minimum if $f(\mathbf{x}) \geq f(\mathbf{a})$ for all $\mathbf{x} \in \mathbb{R}^n$;
- local maximum if $f(\mathbf{x}) \leq f(\mathbf{a})$ for all $\mathbf{x}$ in some neighbourhood around $\mathbf{a}$;
- local minimum if $f(\mathbf{x}) \geq f(\mathbf{a})$ for all $\mathbf{x}$ in some neighbourhood around $\mathbf{a}$;
- stationary point if f is differentiable at $\mathbf{a}$ and $\nabla f(\mathbf{a}) = \mathbf{0}$;
- saddle point if it is a stationary point but is neither a local maximum or local minimum.

Saddle Point



Saddle Point



Extrema on bounded domains

When dealing with functions with bounded domains, extrema points may also occur along the boundary, so we add them to the criteria of critical points.

Theorem 5: Criticial Points

Suppose $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$, where Ω is a bounded set. Then, local extrema points can only be found at the critical points of Ω.

So whenever we are interested in finding extrema points, we only need to consider the critical points over the relevant domain.

Classifying Extrema Points (MATH2011)

As 2011 and 2111 handle this in different ways, we're going to cover each approach individually.

Theorem 6: Classifying Stationary Points

Suppose $f : \mathbb{R}^2 \to \mathbb{R}, \nabla f(\mathbf{a}) = 0$, and both f_x and f_y are continuous around some disk around $\mathbf{a}$. Let

$$D = (f_{xy}(\mathbf{a}))^2 - f_{xx}(\mathbf{a})f_{yy}(\mathbf{a}).$$

Then if:

- $D < 0$ and $f_{xx}(\mathbf{a}) > 0, f(\mathbf{a})$ is a local minimum;
- $D < 0$ and $f_{xx}(\mathbf{a}) < 0, f(\mathbf{a})$ is a local maximum;
- $D > 0, f(\mathbf{a})$ is a saddle point.

D is called the **discriminant** of the critical point.
The case of $D = 0$ isn't covered in MATH2011.

Classifying Extrema Points (MATH2011)

Example 11

Classify the stationary points of $f(x, y) = x^3 - y^3 - 2xy + 4$.

Q11 Solution

Firstly, we need to find the stationary points, i.e.

$$\nabla f(x,y) = (3x^2 - 2y, -3y^2 - 2x) = (0,0).$$

Solving these equations simultaneously leads to the following points: $(0,0)$, $\left(-\frac{2}{3}, \frac{2}{3}\right)$.
Next, we need to find the second-order partial derivatives, i.e.

$$f_{xx}(\mathbf{x}) = 6x \quad f_{yy}(\mathbf{x}) = -6y \quad f_{xy}(\mathbf{x}) = -2$$

For $(0,0)$ we have:

$$D = (-2)^2 - 0 \times 0 = 4 > 0$$

So $(0,0)$ is a saddle point.

Q11 Solution Cont.

For $\left(-\frac{2}{3}, \frac{2}{3}\right)$:

$$D = (-2)^2 - 6 \times \left(-\frac{2}{3}\right) \times -6 \times \frac{2}{3} = -12 < 0$$
$$f_{xx}\left(-\frac{2}{3}, \frac{2}{3}\right) = -4 < 0.$$

Thus $\left(-\frac{2}{3}, \frac{2}{3}\right)$ is a local maximum.

Classifying Extrema Points (MATH2011)

Example 12

Consider the set S which describes the triangle with vertices at $(0,1),(1,-1)$ and $(-1,-1)$. Find the maximum of $f(x,y) = x^2 - xy + y^2$ over this set.

Q12 Solution

This surface is differentiable everywhere, and so we only focus on stationary points and the boundary.
The only stationary point is $(0, 0)$, so $f(0, 0) = 0$.

1. For $\underline{y = 1 - 2x, x \in [0, 1]}$: $f(x, 1 - 2x) = 7x^2 - 5x + 1$ which has a maximum of 3 at $(1, -1)$
2. For $\underline{y = 2x + 1, x \in [-1, 0]}$: $f(x, 2x + 1) = 3x^2 + 3x + 1$ which has a maximum of 1 at $(-1, -1)$
3. For $\underline{y = -1, x \in [-1, 1]}$: $f(x, -1) = x^2 + x + 1$ which has a maximum of 3 at $(1, -1)$

Hence, the maximum of f on S is 3 at $(1, -1)$.

Classifying Extrema Points (MATH2111)

For MATH2111, we'll be using the Hessian

Definition 10

For $f : \mathbb{R}^n \to \mathbb{R}$ the Hessian of f at $\mathbf{a}$ is the $n \times n$ matrix,

$$H(f,\mathbf{a}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{a}) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{a}) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_2^2}(\mathbf{a}) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2}(\mathbf{a}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\mathbf{a}) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\mathbf{a}) \end{pmatrix}$$

Order of Hessian

Just like the Jacobian each column of the Hessian is differentiated with respect to the same variable **last**.

Classifying Extrema Points (MATH2111)

The nature of the stationary points are dependent on the eigenvalues of the Hessian matrix at **a**. If all eigenvalues are positive, negative, or some mix in between, then its a minimum, maximum or a saddle point, respectively.
As eigenvalues can become quite tedious to find, we'll utilise an alternative test for classifying stationary points.

Theorem 7: Sylvester's Criterion

Denote H_k as the upper left $k \times k$ sub-matrix of H and $\Delta_k = \det H_k$, then H is:

Positive definite	$\Leftrightarrow$	$\Delta_k > 0$ for all k
Positive semidefinite	$\Rightarrow$	$\Delta_k \geq 0$ for all k
Negative definite	$\Leftrightarrow$	$\Delta_k < 0$ for all odd k and
		$\Delta_k > 0$ for all even k
Negative semidefinite	$\Rightarrow$	$\Delta_k \leq 0$ for all odd k and
		$\Delta_k \leq 0$ for all even k

Classifying Extrema Points (MATH2111)

Theorem 8

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is C^2 and $\nabla f(\mathbf{a}) = \mathbf{0}$. Then

- $H(f, \mathbf{a})$ is positive definite $\Rightarrow$ f has a local minimum at $\mathbf{a}$
- $H(f, \mathbf{a})$ is negative definite $\Rightarrow$ f has a local maximum at $\mathbf{a}$
- f has a local minimum at $\mathbf{a} \Rightarrow H(f, \mathbf{a})$ is positive semidefinite
- f has a local maximum at $\mathbf{a} \Rightarrow H(f, \mathbf{a})$ is negative semidefinite

We can combine this theorem with Sylvester's Criterion to help classify stationary points. We'll be mostly working with the 2-dimensional case.

Classifying Extrema Points (MATH2111)

Example 13

Classify the stationary point $\mathbf{a} = (1, 1, 1)$ of $f(x, y, z) = x^4 + y^4 + z^4 - 4xyz$ using Sylvester's Criterion.

Q13 Solution

Want to classify a given stationary point, and so we need the Hessian.

$$H(f,(1,1,1)) = \begin{pmatrix} 12 & -4 & -4 \\ -4 & 12 & -4 \\ -4 & -4 & 12 \end{pmatrix}$$

From this, we can calculate that:
$\Delta_1 = 12 > 0, \Delta_2 = 128 > 0, \Delta_3 = 1024 > 0$. And so, by Sylvester's Criterion, $(1,1,1)$ is a local minimum of f.

Sylvester's Criterion Pattern

Method to remember classifications

You can remember the patterns by thinking of each sub-matrix determinant as the product of eigenvalues. So if $\Delta_2 > 0$, we could have 2 negative or 2 positive eigenvalues, or if $\Delta_1 < 0$ we have a negative eigenvalue. This **isn't** what these values actually mean, but could help with remembering the patterns.

Lagrange Multipliers

So far we have dealt with how to classify stationary points and finding extrema points on simple boundaries, e.g. lines, circles. Now, we are going to consider a larger variety of boundaries, through the method of lagrange multipliers.

Theorem 9: Lagrange Multipliers

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ are differentiable and

$$S = \{\mathbf{x} \in \mathbb{R}^n : g(\mathbf{x}) = c\}$$

Then to find any maximum and/or minimum points of f constrained to S, we need to find $\mathbf{a} \in S$ s.t. $\nabla f(\mathbf{a}) = \lambda \nabla g(\mathbf{a})$, for some real λ.

Lagrange Multiplier Examples

Example 14

Find the maximum and minimum values of $f(x, y) = 8x^2 - 2y$ subject to the constraint $x^2 + y^2 = 1$.

Q14 Solution

Define the constraint as $g(x, y) = x^2 + y^2 = 1$, and using Lagrange's Multiplier Theorem:

$$\nabla f(x, y) = \lambda \nabla g(x, y) \tag{1}$$

$$16x = 2\lambda x \tag{2}$$

$$-2 = 2\lambda y \tag{3}$$

From (2) we have $x = 0$ or $\lambda = 8$. Subbing these both in lead to the following: $(0, \pm 1)$, $\left(\frac{\pm 3\sqrt{7}}{8}, -\frac{1}{8}\right)$

Now we have to compare the values of each of these points, which are: $\mp 2, 8\frac{1}{8}$.

Hence, the maximum and minimum values of f over the unit circle are $8\frac{1}{8}$ and -2, respectively.

Lagrange Multiplier Examples

Example 15

Find the closest point(s) to the origin on the surface $xyz = 27$.

Q15 Solution

Function of interest is $d(x, y, z) = x^2 + y^2 + z^2$ with constraint of $g(x, y, z) = xyz = 27$.

$$2x = \lambda yz$$
$$2y = \lambda xz$$
$$2z = \lambda xy$$

Noting the pattern in the three equations and the given constraint, we can combine all of the above to yield:

$$2^3 xyz = \lambda^3 (xyz)^2 \Leftrightarrow \lambda = \frac{2}{3}.$$

Subbing this into the three equations and solving all of them simultaneously leads to:
$(3, 3, 3), (3, -3, -3), (-3, 3, -3), (-3, -3, 3)$ as none of them can be 0 (as $xyz = 27$).