BITS Pilani
Pilani Campus

Mathematical Foundations for Data Science

MFDS Team

DSECL ZC416, MFDS

Lecture No. 9

Agenda

- Concept of limits, continuity and differentiability for 1-D
- Properties of continuous functions
- Maxima and minima for one variable
- Function of several variables
- Partial derivatives and Directional derivatives
- Extremum in several variables
  - Lagrange multipliers (theory and an example)
  - Method of steepest descent (theory and an example).

Concept of Limits

Limit of the function – Value that f(x) gets closer to as x approaches some number

$f(x) = \frac{4}{3}x - 4,$

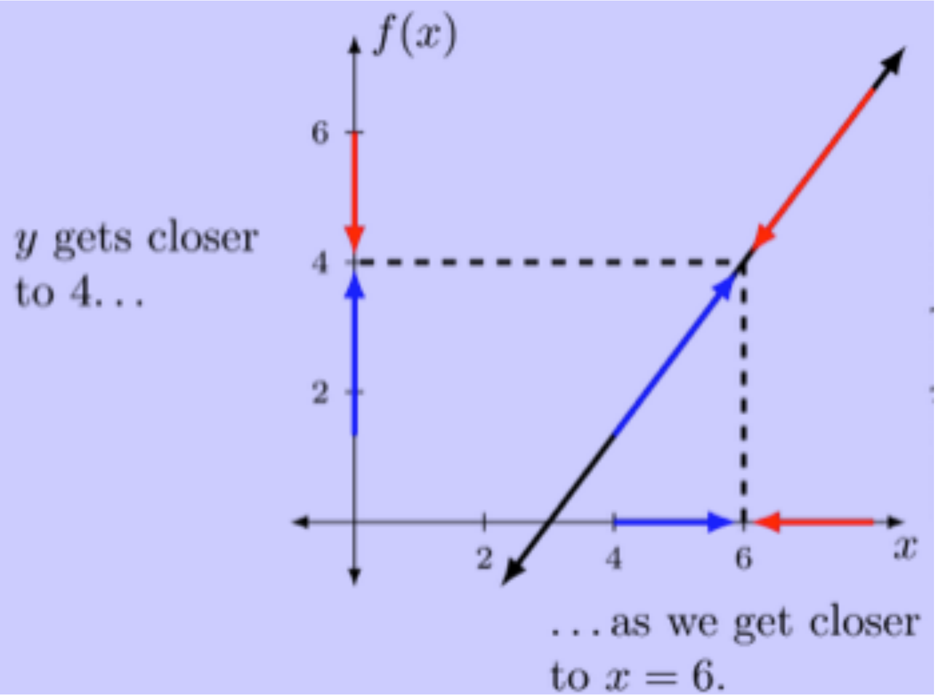
The "lim" tells us we're looking for a limit value, not a function value.

This tells us which function we're working with.

$$\lim_{x \to 6} f(x) = 4$$

This tells us what the variable is, and what it is approaching.

This is the value the function is approaching.



One sided Limits

Right Handed Limit $$\lim_{x \to a^+} f(x) = L$$

Can make f(x) as close to L as we want for all x sufficiently close to a with x > a without actually letting x be a

Left Handed Limit $$\lim_{x \to a^-} f(x) = L$$

Can make f(x) as close to L as we want for all x sufficiently close to a with x < a without actually letting x be a

NOTE: If two one sided limits **have different value**, **then normal limit will not exist**

$$\lim_{x \to a^+} f(x) \neq \lim_{x \to a^-} f(x)$$

Limits

Check the existence of the limit for function $f(x) = |x| - |x-1|$ at $x = 0$.

The limit of a function $f(x)$ at $x = a$ exists only when its left hand limit (LHL) and right hand limit (RHL) exist and are equal , $\lim_{x \to a^-} f(x) = \lim_{x \to a^+} f(x)$.

Given, $\lim_{x \to 0} f(x) = \lim_{x \to 0} (|x| - |x-1|)$

LHL : $\lim_{x \to 0^-} (|x| - |x-1|)$

x is negative when $x \to 0^-$. Therefore $|x| = -x$

$= \lim_{x \to 0^-} (-x - |x-1|)$

$x - 1$ is negative when $x \to 0^-$. Therefore $|x-1| = -x+1$

$= \lim_{x \to 0^-} (-x - (-x+1)) = \lim_{x \to 0^-} (-1)$

$= -1$

Limits

RHL : $\lim_{x \to 0^+} (|x| - |x-1|)$
x is positive when $x \to 0^+$. Therefore $|x| = x$
$= \lim_{x \to 0^+} (x - |x-1|)$
$x - 1$ is negative when $x \to 0^+$. Therefore $|x-1| = -x+1$
$= \lim_{x \to 0^+} (x - (-x+1)) = \lim_{x \to 0^+} (2x - 1)$
$= -1.$

Figure: $f(x) = |x| - |x-1|$

Continuity

Let f be a function. The f is **continuous** at $x = a$ if

$f(a)$ is defined This means a is in the domain of f.

$\lim_{x \to a} f(x)$ exists This means there exists a finite limit at $x = a$.

$\lim_{x \to a} f(x) = f(a)$ This means at $x = a$ the limit is equal to the functional value.

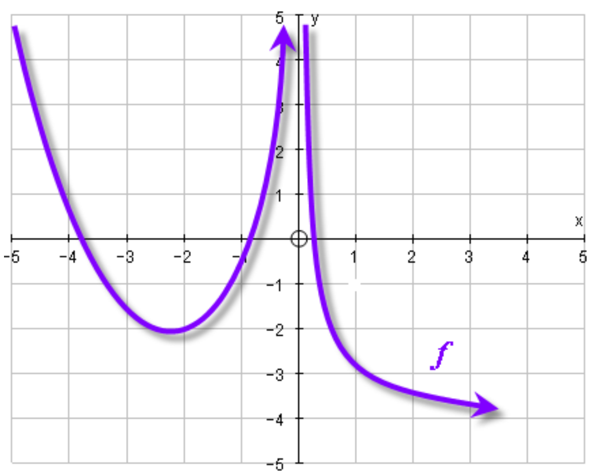
If a function is not continuous at a we say f is **discontinuous** at a.

A function is **continuous on its domain** if it is continuous at each point in the domain of f.

Continuity

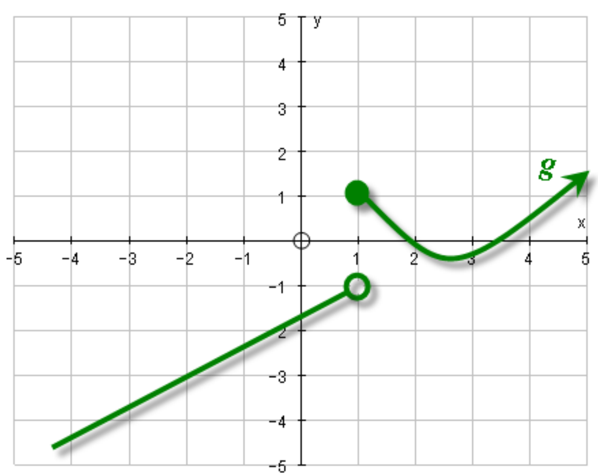
f(0) is not defined.

The function is continuous on its domain, but not continuous at every real number.



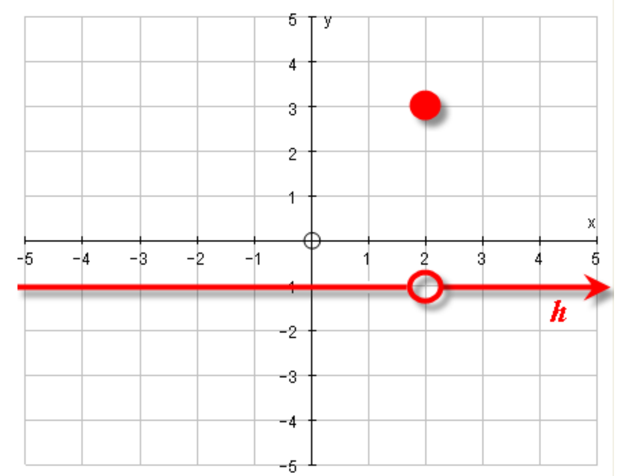
$\lim_{x\to 1} g(x)$ does not exist

The function does not have a limit at x = a.

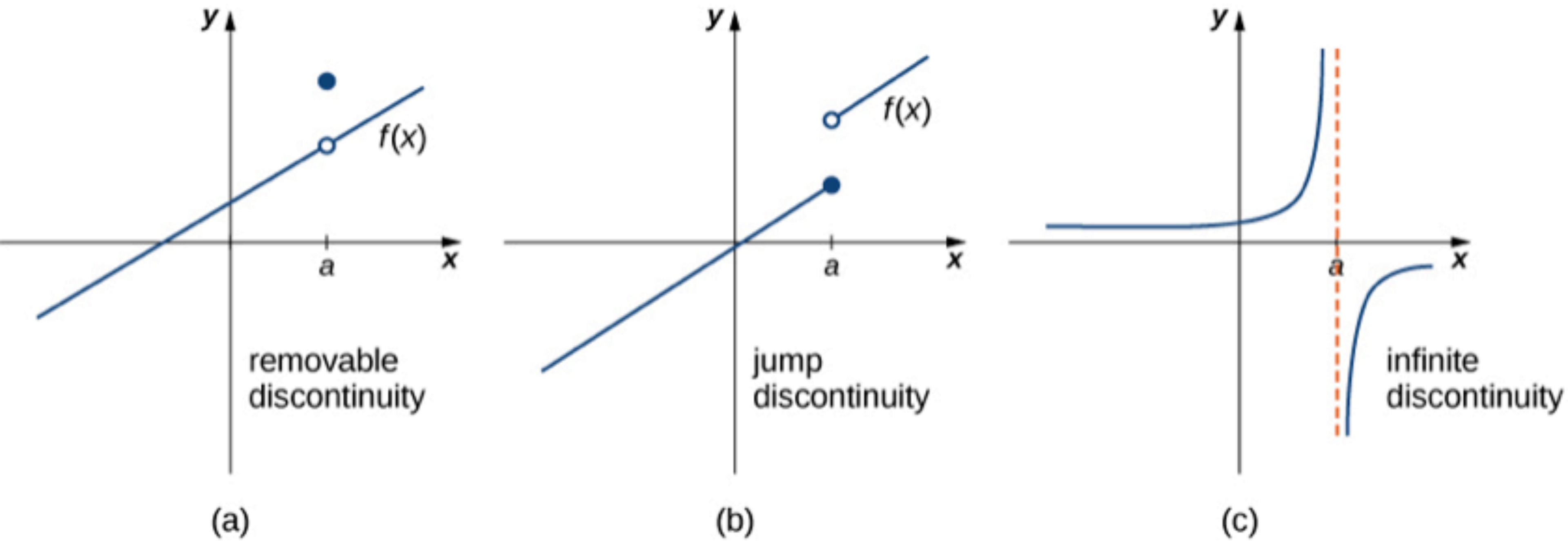


$\lim_{x\to 2} h(x) \neq h(2)$

The limit and the functional value are not equal x = a.



Types of Discontinuities



Hole in the graph

- Left hand Limit and Right Hand Limits are not same
- Non Infinite Discontinuity for which sections of function do no meet up

Discontinuity located at Vertical Asymptote

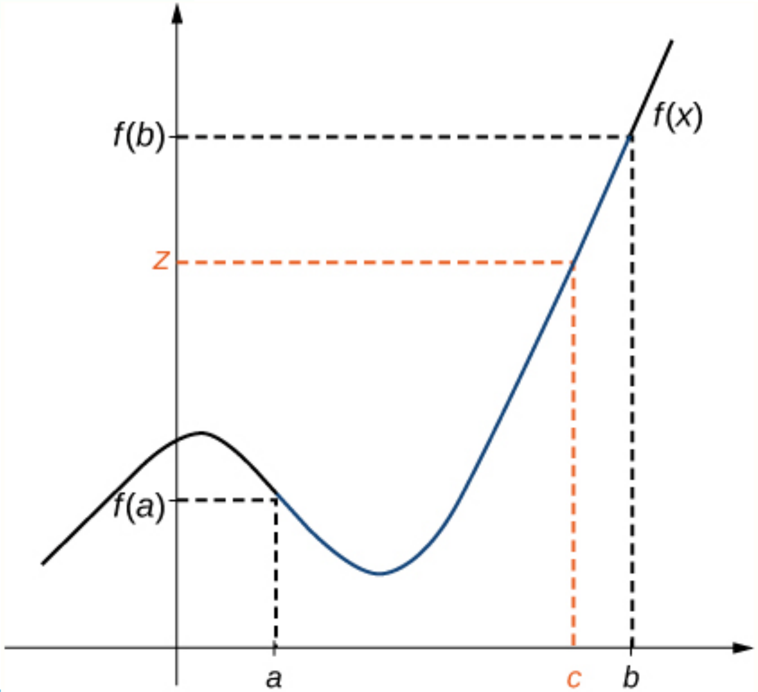
Properties of Continuous Function

Let f and g be continuous functions. Then

1. $f + g$ is a continuous function.
2. fg is a continuous function.
3. $\dfrac{f}{g}$ is a continuous function, whenever $g(x) \neq 0$.

INTERMEDIATE VALUE THEOREM

Let f be continuous over a closed, bounded interval $[a, b]$. If z is any real number between $f(a)$ and $f(b)$, then there is a number c in $[a, b]$ satisfying $f(c) = z$ in Figure.



Note: If a function is continuous over a closed interval, then function takes on every value between the values at its end points

Intermediate Value Theorem

Show that there is a root of equation $4x^3 - 6x^2 + 3x - 2 = 0$ between 1 and 2

We are looking for a solution of given equation that is number c between 1 and 2 such that $f(c) = 0$

Given : $a = 1$, $b = 2$

$f(1) = 4 - 6 + 3 - 2 = -1 < 0$

$f(2) = 32 - 24 + 6 - 2 = 12 > 0$

$f(1) < 0 < f(2) \Rightarrow d = 0$ is number between $f(1)$ and $f(2)$

f is a continuous since it is a polynomial. Intermediate value theorem says that there is a number c between 1 and 2 such that $f(c) = 0$

Equation has atleast one root c in the interval (1,2)

NOTE : We can precisely locate root using Intermediate Value Theorem

Intermediate Value Theorem fails for Discontinuous Functions

Derivative

The derivative of f at x is $f'(x) \equiv \frac{df}{dx}(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}.$

- The derivative is the slope of the tangent
- When the limit exists the function is called DIFFERENTIABLE at x.
- The term differentiable function is used to denote functions differentiable at every point in the domain
- If a function is differentiable at a point, it is continuous at that point

Functions which are not differentiable
Cube Root Function
1/x in domain [0, ∞]
Floor and Ceiling Functions at integer values

Differentiability

$$\lim_{h\to 0_+}\frac{f(h)-f(0)}{h} = \lim_{h\to 0_+}\frac{h-0}{h} \qquad \text{cancellation of } h \text{ okay, since } h \neq 0 \text{ for limit at } 0$$
$$= \lim_{h\to 0_+} 1$$
$$= 1$$

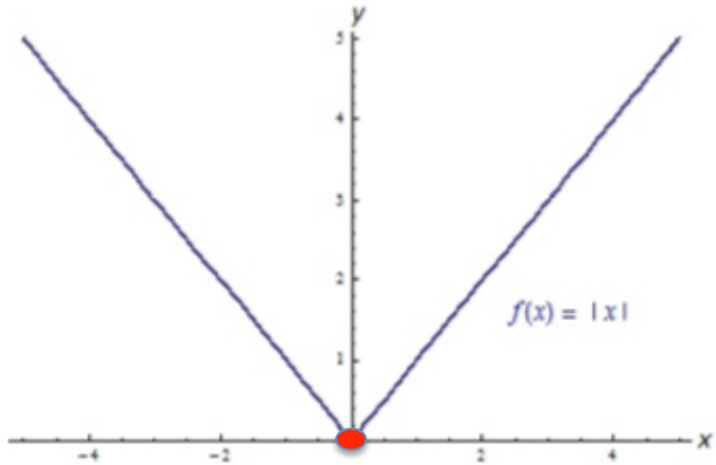
But

$$\lim_{h\to 0_-}\frac{f(h)-f(0)}{h} = \lim_{h\to 0_-}\frac{-h-0}{h} \qquad \text{cancellation of } h \text{ okay, since } h \neq 0 \text{ for limit at } 0$$
$$= \lim_{h\to 0_-} -1$$
$$= -1$$

Since the left- and right-hand limits do not agree,

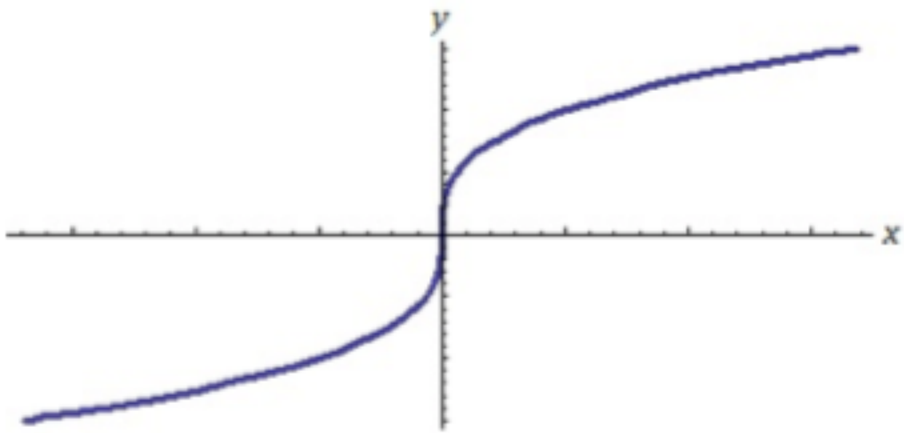
$$\lim_{h\to 0}\frac{f(h)-f(0)}{h}$$

Does not exits, and so $|\,x\,|$ is not differentiable at $x=0$.



$$f(x) = |\,x\,| = \begin{cases} x & x \geq 0 \\ -x & x < 0 \end{cases}$$

Differentiability



$$g(x) = x^{1/3}$$

The graph is smooth at $x = 0$, but does appear to have a vertical tangent.

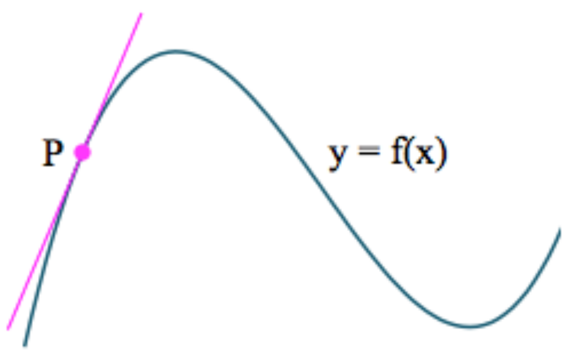
$$\lim_{h \to 0} \frac{(0+h)^{1/3} - 0^{1/3}}{h} = \lim_{h \to 0} \frac{(h)^{1/3}}{h} = \lim_{h \to 0} \frac{1}{h^{2/3}}$$

As $h \to 0$, the denominator becomes small, so the fraction grows without bound. Hence g is not differentiable at $x = 0$.

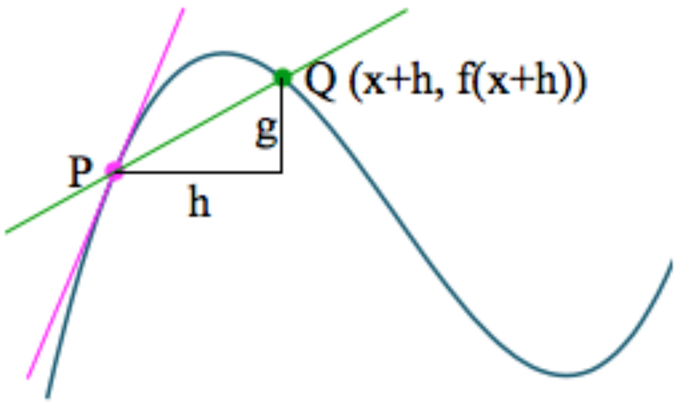
Differentiation from First Principle

$\frac{dy}{dx}$ or $f'(x)$ or y'.

$$\frac{dy}{dx} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$



Slope of the tangent at P.

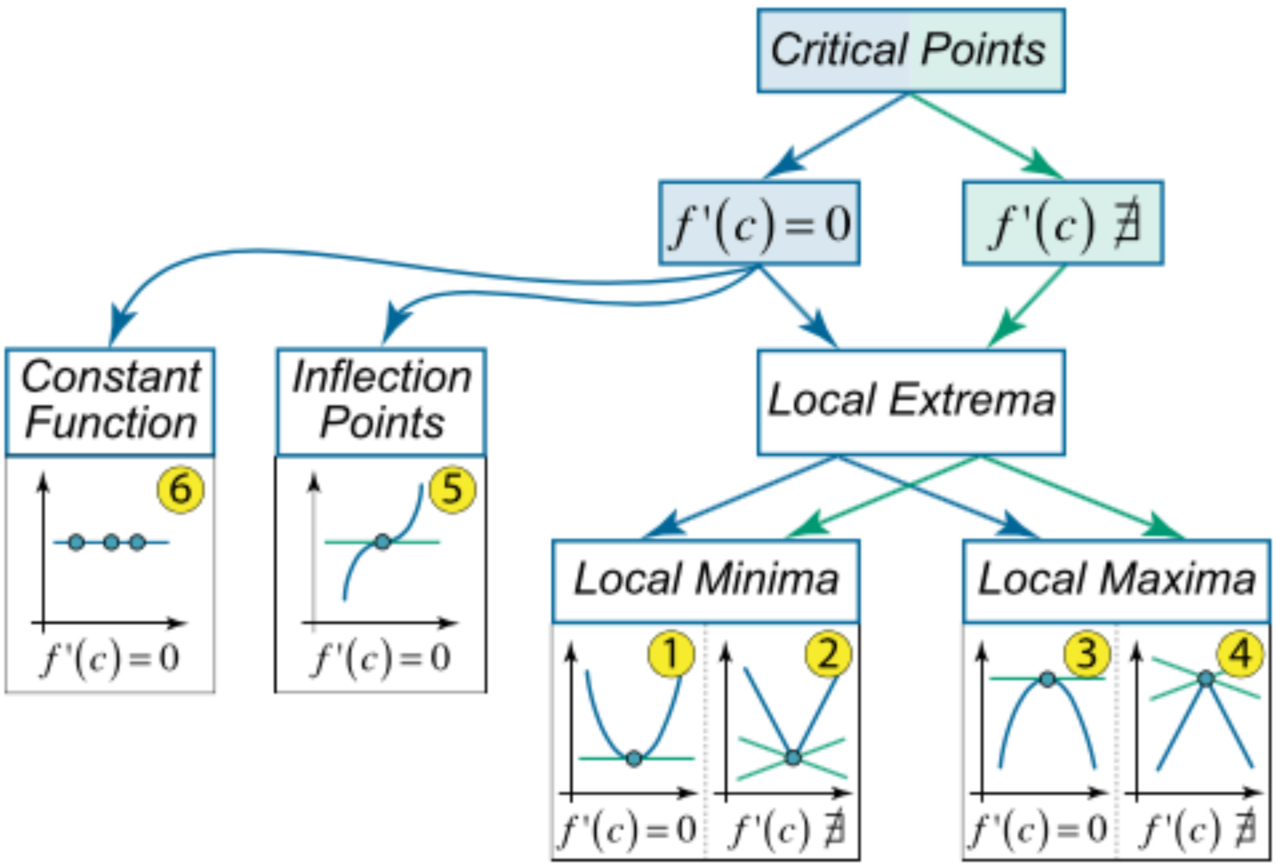


Slope of the line PQ.

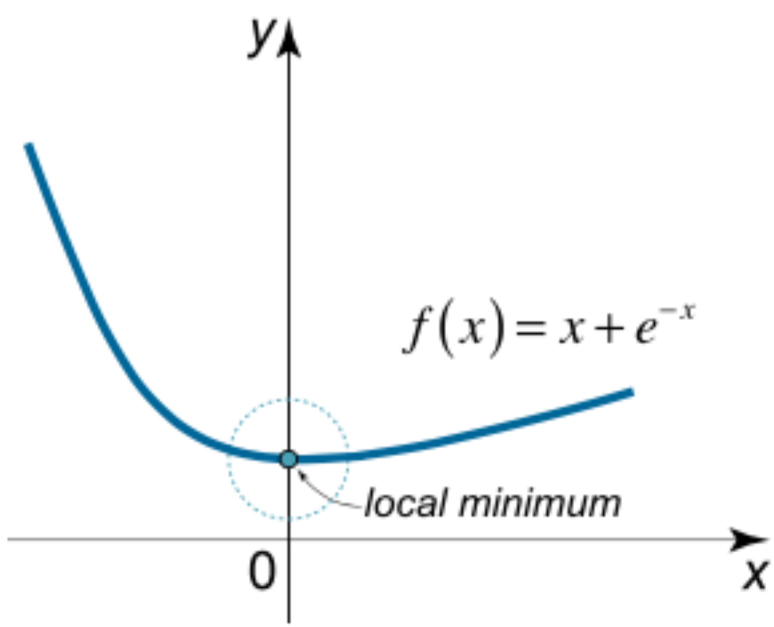
The value $\frac{g}{h}$ is an approximation to the slope of the tangent which we require.

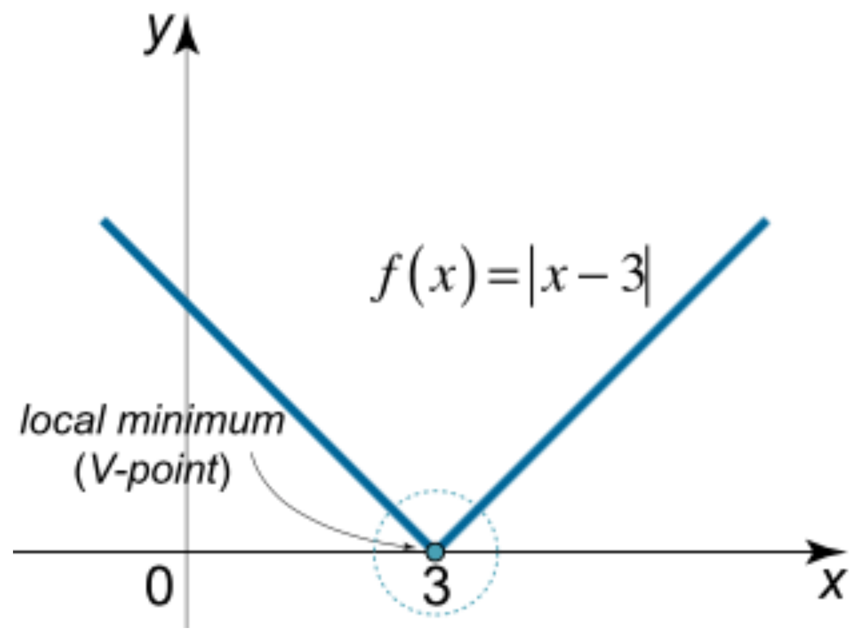
Critical Points

Let $f(x)$ be a function and let c be a point in the domain of the function. The point c is called a critical point of f if either $f'(c) = 0$ or $f'(c)$ does not exist.



Critical Point - Local Minimum

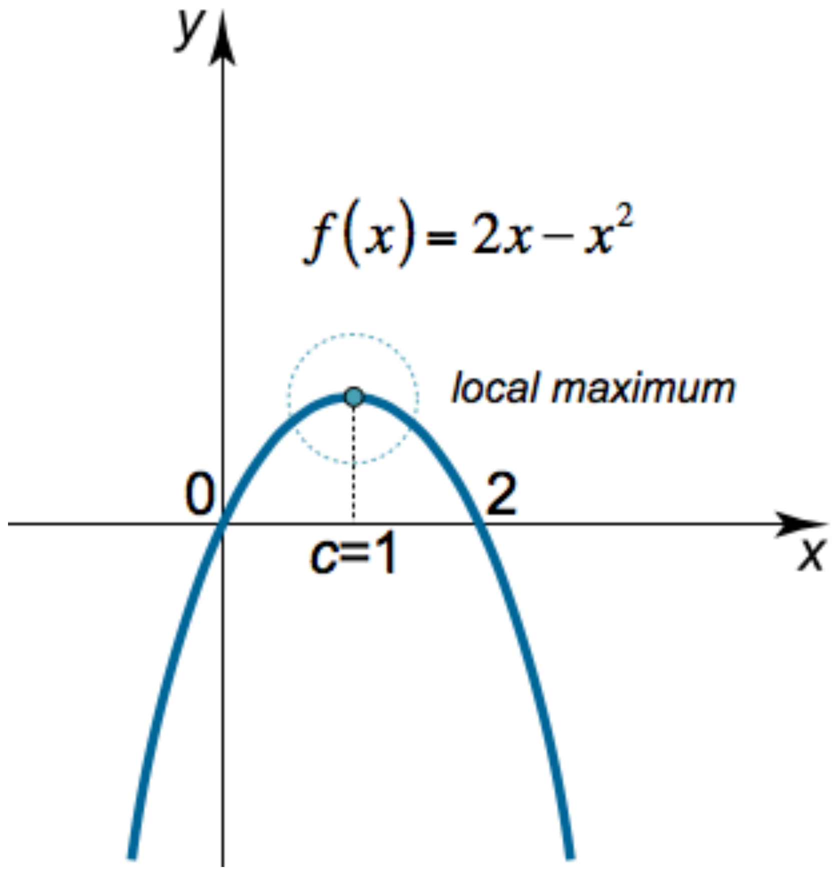


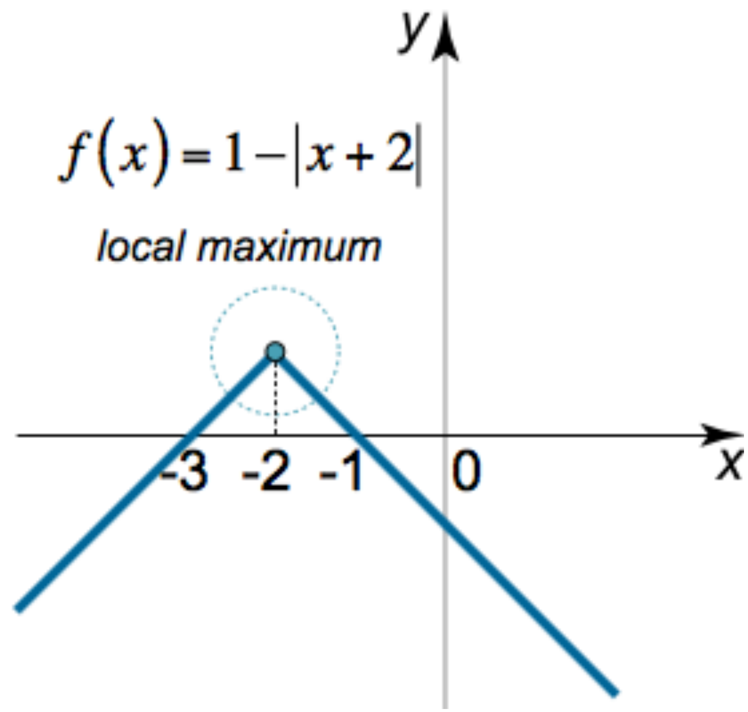


x = c is local minimum if function changes from decreasing to increasing at that point

Derivative does not exist at this point

Critical Point – Local Maximum

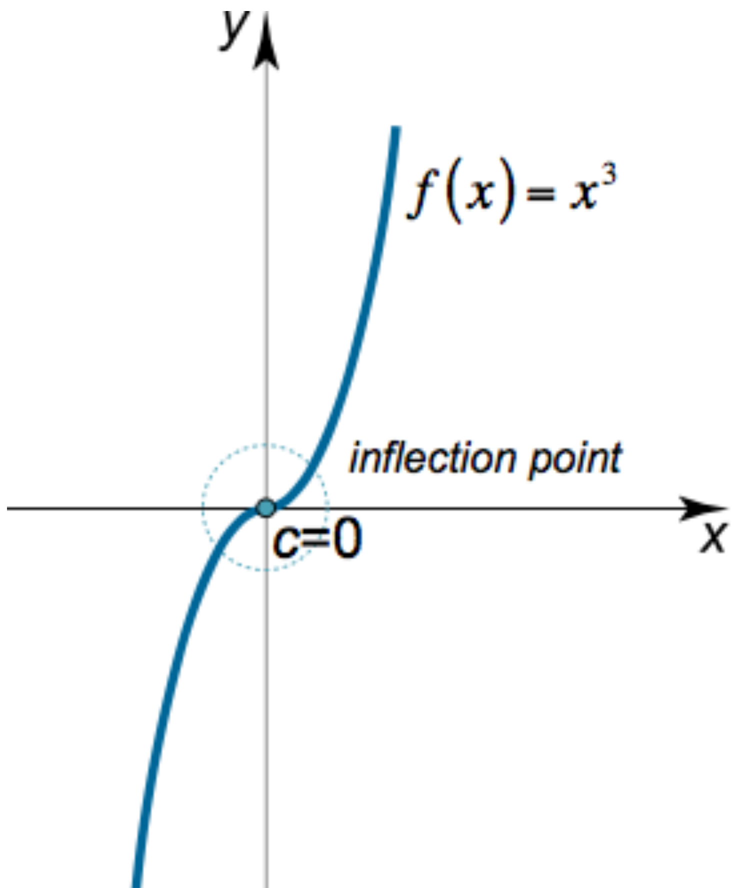




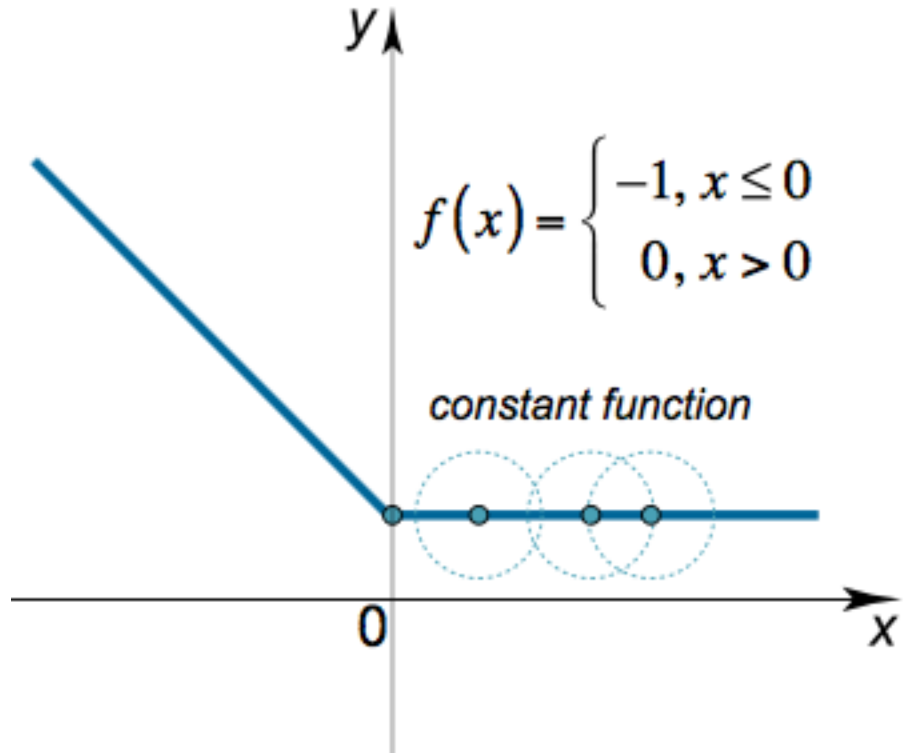
Critical Point x = c is local maximum if the function changes from increasing to decreasing at that point

Derivative does not exist at this point

Critical Point



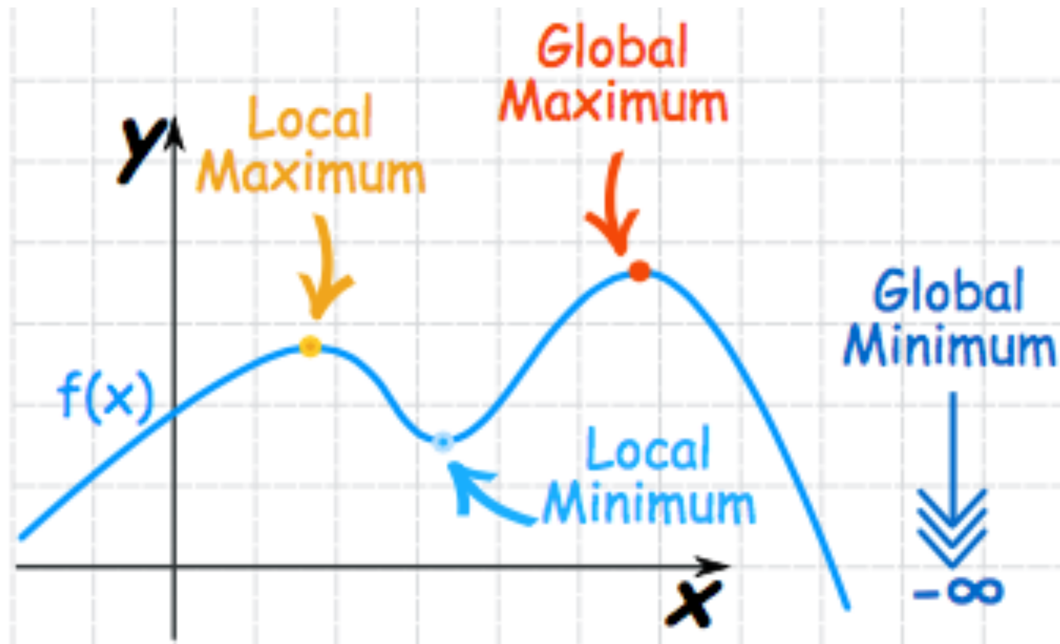
A critical point x = c is an inflection point if function changes concavity at that point



Each point of a constant function is critical

Global Maximum and Minimum

- The maximum or minimum over the entire function is called Global maximum or minimum
- There is only one global maximum and global minimum but there can be more than one local maximum or minimum



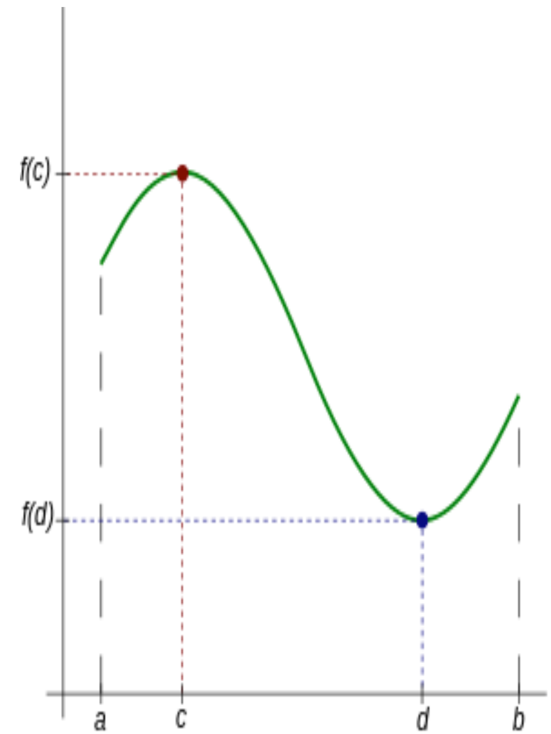
Extreme Value Theorem

Extreme Value Theorem guarantees that every function must have absolute(global) maximum and minimum

Assumptions:

1. f is continuous on an interval
2. Interval is a closed interval [a,b]

If either assumption fails, we are not allowed to draw conclusion that f hits minimum and maximum value on that closed interval



A continuous function $f(x)$ on the closed interval $[a,b]$ showing the absolute max (red) and the absolute min (blue).

Second Derivative Test

First Derivative Test is used in conjunction with extreme value theorem to find the absolute maximum and minimum of a real valued function defined on a closed, bounded interval

After establishing the critical points of a function, second derivative test uses the value of second derivative at those points to determine whether such points are a local maximum or local minimum.
If a function f is twice differentiable at a critical point x (f´ (x) = 0 then

- If $f''(x) < 0$ then f has a local maximum at x.
- If $f''(x) > 0$ then f has a local minimum at x.
- If $f''(x) = 0$, the test is inconclusive.

Second Derivative Test

Example : $f(x) = 2x^3 - 3x^2 - 36x + 2$

$f'(x) = 0 \rightarrow 6x^2 - 6x - 36 = 0$

$x = 3, \ -2$

Compute $f''(x) = 12x - 6$

At $x = 3 \rightarrow f''(3) = 12(3) - 6 = 30 > 0$

At $x = -2 \rightarrow f''(-2) = 12(-2) - 6 = -30 < 0$

X = -2 Point is Maxima

X = 3 Point is Minima

- If $f''(x) < 0$ then f has a local maximum at x.
- If $f''(x) > 0$ then f has a local minimum at x.
- If $f''(x) = 0$, the test is inconclusive.

Partial Derivative

When input of function is made up of multiple variables, we want to see how function changes as we let just one of those variable change while holding all the others constant

$$\frac{\partial f}{\partial x}$$

f is a multivariable function

∂f Tiny change in functions output

∂x Tiny change in x

$$\frac{\partial f}{\partial x} = \underbrace{\frac{\partial}{\partial x} x^2 y}_{\text{Treat } y \text{ as constant; take derivative.}} = 2xy$$

$$\frac{\partial f}{\partial y} = \underbrace{\frac{\partial}{\partial y} x^2 y}_{\text{Treat } x \text{ as constant; take derivative.}} = x^2 \cdot 1$$

Hessian Matrix

Square matrix which is a way of organizing all second order partial derivative information of a multivariable function
Hessian matrix is always symmetric matrix →entries of the matrix are symmetric across its main diagonal

$$Hf(x,y) = \begin{bmatrix} f_{xx}(x,y) & \underline{f_{xy}}(x,y) \\ \underline{f_{yx}}(x,y) & f_{yy}(x,y) \end{bmatrix}$$

$$H\,f\,(x,\,y) \equiv \begin{vmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x\,\partial y} \\ \frac{\partial^2 f}{\partial y\,\partial x} & \frac{\partial^2 f}{\partial y^2} \end{vmatrix}.$$

Define D(x,y) to be the determinant

Note : We expect the eigenvalues of the Hessian to be positive at local minimum and negative at local maximum
If the Hessian has both positive and negative eigen values the corresponding point must be saddle point

Hessian Matrix

The function $f(x, y) = x^3 + 2(x - y)^2 - 3x$ has a critical point at $(1, 1)$. Classify this critical point as a local maximum, a local minimum, or a saddle point.

SOLUTION The Hessian of f is

$$Hf(x, y) = \begin{bmatrix} 6x + 4 & -4 \\ -4 & 4 \end{bmatrix}$$

and in particular

$$Hf(1, 1) = \begin{bmatrix} 10 & -4 \\ -4 & 4 \end{bmatrix}$$

The eigenvalues of this matrix are 2 and 12, so $(1, 1)$ is a local minimum.

The function $f(x, y) = 6\cos x + 4x \sin y$ has a critical point at $(0, 0)$. Classify this critical point as a local maximum, a local minimum, or a saddle point.

SOLUTION The Hessian of f is

$$Hf(x, y) = \begin{bmatrix} -6\cos x & 4\cos y \\ 4\cos y & -4x \sin y \end{bmatrix}$$

and in particular

$$Hf(0, 0) = \begin{bmatrix} -6 & 4 \\ 4 & 0 \end{bmatrix}$$

The eigenvalues of this matrix are -8 and 2, so $(0, 0)$ is a saddle point.

Gradient

The gradient of a scalar-valued multivariable function $f(x, y, \dots)$, denoted ∇f, packages all its partial derivative information into a vector:

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \vdots \end{bmatrix}$$

In particular, this means ∇f is a vector-valued function.

If you imagine standing at a point $(x_0, y_0, \dots)$ in the input space of f, the vector $\nabla f(x_0, y_0, \dots)$ tells you which direction you should travel to increase the value of f most rapidly.

These gradient vectors—$\nabla f(x_0, y_0, \dots)$—are also perpendicular to the contour lines of f.

Lagrange Multiplier

1. Constrained Optimization problem
2. Lagrange Multiplier is developed to figure out the maxima/ minima of an objective function f under a constraint function g

Core Idea : Look for points where the contour lines of f and g are tangent to each other

Suppose you want to maximize this function:

$$f(x, y) = 2x + y$$

Function that needs to be optimized

But let's also say you limited yourself to inputs (x, y) which satisfy the following equation:

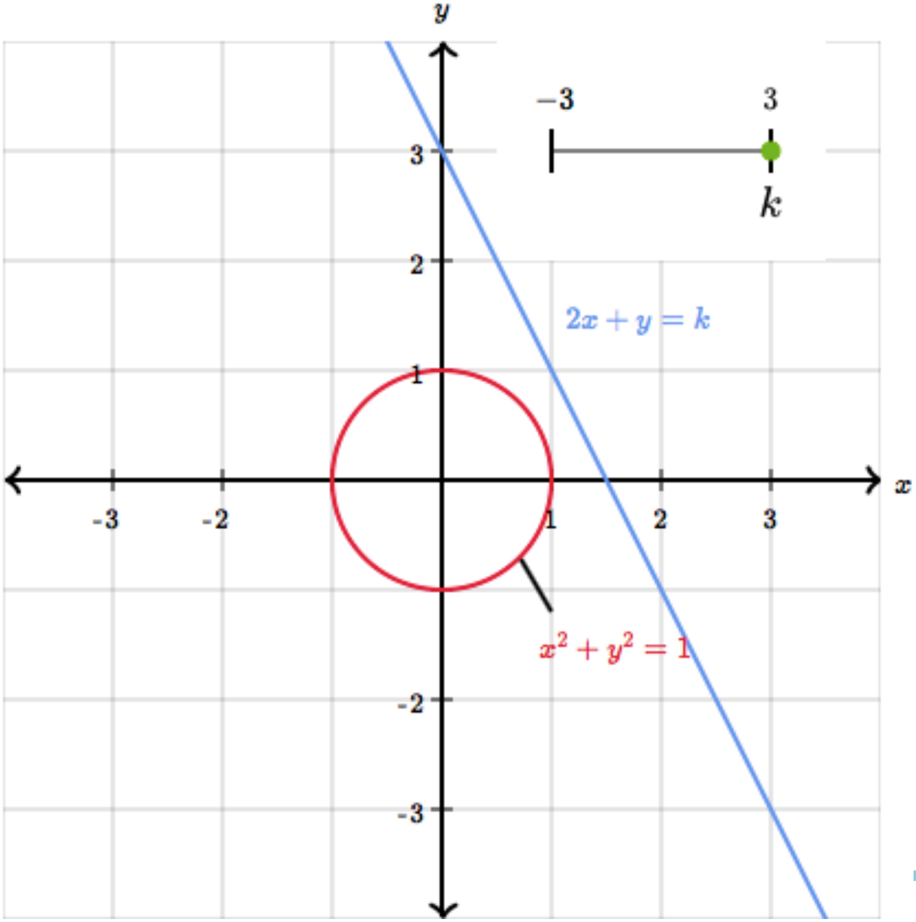
$$x^2 + y^2 = 1$$

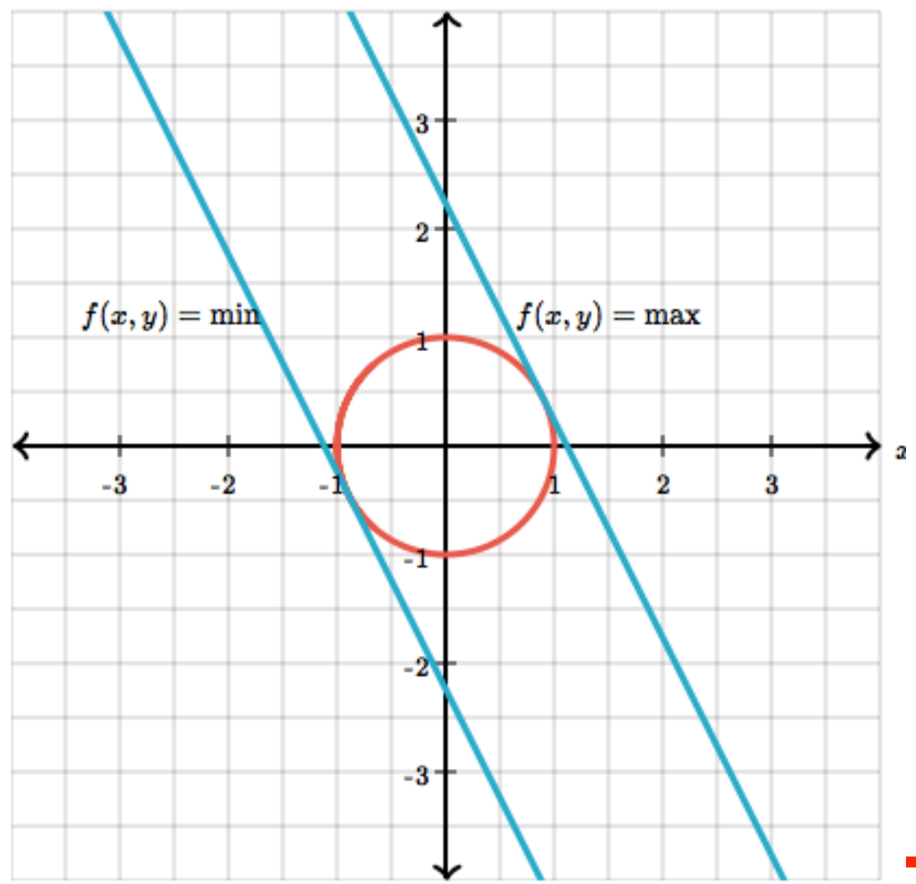
Constraint

For which point (x,y) on the unit circle is the value 2x +y biggest?

Lagrange Multiplier

1. First draw the graph of f(x,y) which looks like slanted plane since f is linear.
2. Then project the circle $x^2 + y^2 = 1$ from xy plane vertically onto the graph of f
3. Maximum we are seeking corresponds with highest point of this projected circle on the graph





Lagrange Multiplier

$$\nabla f(x_0, y_0) = \lambda_0 \nabla g(x_0, y_0)$$

Here, λ_0 represents some constant. Some authors use a negative constant, $-\lambda_0$, but I personally prefer a positive constant, as it gives a cleaner interpretation of λ_0 down the road.

Let's see what this looks like in our example where $f(x, y) = 2x + y$ and $g(x, y) = x^2 + y^2$. The gradient of f is

$$\nabla f(x, y) = \begin{bmatrix} \frac{\partial}{\partial x}(2x + y) \\ \\ \frac{\partial}{\partial y}(2x + y) \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

and the gradient of g is

$$\nabla g(x, y) = \begin{bmatrix} \frac{\partial}{\partial x}(x^2 + y^2 - 1) \\ \\ \frac{\partial}{\partial y}(x^2 + y^2 - 1) \end{bmatrix} = \begin{bmatrix} 2x \\ 2y \end{bmatrix}$$

Lagrange Multiplier

Therefore, the tangency condition ends up looking like this:

$$\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \lambda_0 \begin{bmatrix} 2x_0 \\ 2y_0 \end{bmatrix}$$

$$x_0^2 + y_0^2 = 1$$

$$2 = 2\lambda_0 x_0$$

$$1 = 2\lambda_0 y_0$$

Three Equations and three unknowns

$$x_0 = \frac{1}{\lambda_0}$$

$$y_0 = \frac{1}{2\lambda_0}$$

$$x_0{}^2 + y_0{}^2 = 1$$

$$\left(\frac{1}{\lambda_0}\right)^2 + \left(\frac{1}{2\lambda_0}\right)^2 = 1$$

$$\frac{1}{\lambda_0^2} + \frac{1}{4\lambda_0^2} = 1$$

$$\pm\sqrt{\frac{5}{4}} = \lambda_0$$

$$\frac{\pm\sqrt{5}}{2} = \lambda_0$$

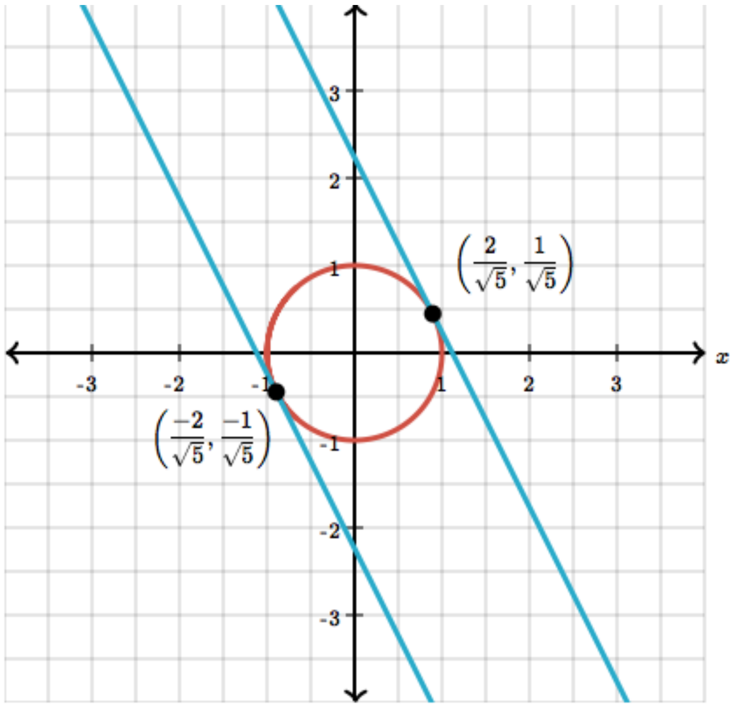
Lagrange Multiplier

$$(x_0, y_0) = \left(\frac{1}{\lambda_0}, \frac{1}{2\lambda_0}\right)$$
$$= \left(\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right) \quad \text{or} \quad \left(\frac{-2}{\sqrt{5}}, \frac{-1}{\sqrt{5}}\right)$$

We can see which of these is a maximum point and which is a minimum point by plugging these solutions into $f(x, y)$ and seeing which is bigger.

$$f\left(\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right) = 2\frac{2}{\sqrt{5}} + \frac{1}{\sqrt{5}}$$
$$= \frac{5}{\sqrt{5}}$$
$$= \sqrt{5} \quad \leftarrow \text{Maximum}$$

$$f\left(-\frac{2}{\sqrt{5}}, -\frac{1}{\sqrt{5}}\right) = 2\frac{-2}{\sqrt{5}} + \frac{-1}{\sqrt{5}}$$
$$= \frac{-5}{\sqrt{5}}$$
$$= -\sqrt{5} \quad \leftarrow \text{Minimum}$$



Method of Steepest Descent

- Steepest Descent is an iterative optimization method used to minimize an objective function by moving in the direction of steepest descent

Step 1. Start with an arbitrary initial point X_1

Step 2. The direction of steepest descent is $-\nabla f(X_i)$

Step 3. Determine the optimal step length τ(tau)

$X_{i+1} = X_i - \tau \nabla f_i$

Step 4. Test the new point X_{i+1}, for optimality. If X_{i+1} is optimum, stop the process. Otherwise set i =i+1, go to step 2

Method of Steepest Descent

Minimize Z = $x^2 + 3y^2$ starting from (2,2)

In this case, $\tau = \dfrac{x^2 + 9y^2}{2x^2 + 54y^2}$

```
j    x(j)       y(j)       f(x(j),y(j))
1    2.000000   2.000000   16.000000
2    1.285714   -0.142857   1.714286
3    0.214286   0.214286   0.183673
4    0.137755   -0.015306   0.019679
5    0.022959   0.022959   0.002108
6    0.014759   -0.001640   0.000226
7    0.002460   0.002460   0.000024
8    0.001581   -0.000176   0.000003
The value of j is 8 and x(8) and y(8) are 0.001581 and -0.000176
```