minima

or maxima

UnConstrained Optimization

For Multivariate Function

Recap : Calculus of one Variable

First Derivative Test

Suppose that a is a critical point of the function f, i.e., $f'(a) = 0$.

(a) If f' changes sign from positive to negative at a, then f has a local maximum at a.

(b) If f' changes sign from negative to positive at a, then f has a local minimum at a.

(c) If f' does not change sign at a, then f has no local extremum at a.

MSQ Question:

Suppose $f'(x) = x(x-2)^2(x+3)$, where $f'(x)$ is the derivative of $f(x)$.

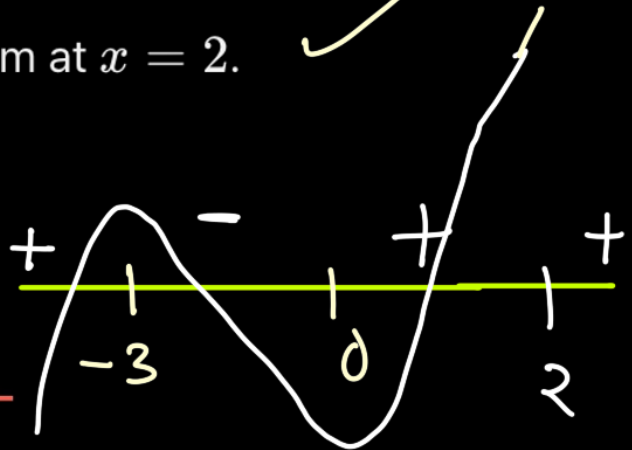
Which of the following is (are) true?

A. f has a local maximum at $x = -3$.

B. f has a local minimum at $x = 0$.

C. f has neither a local maximum nor a local minimum at $x = 2$.

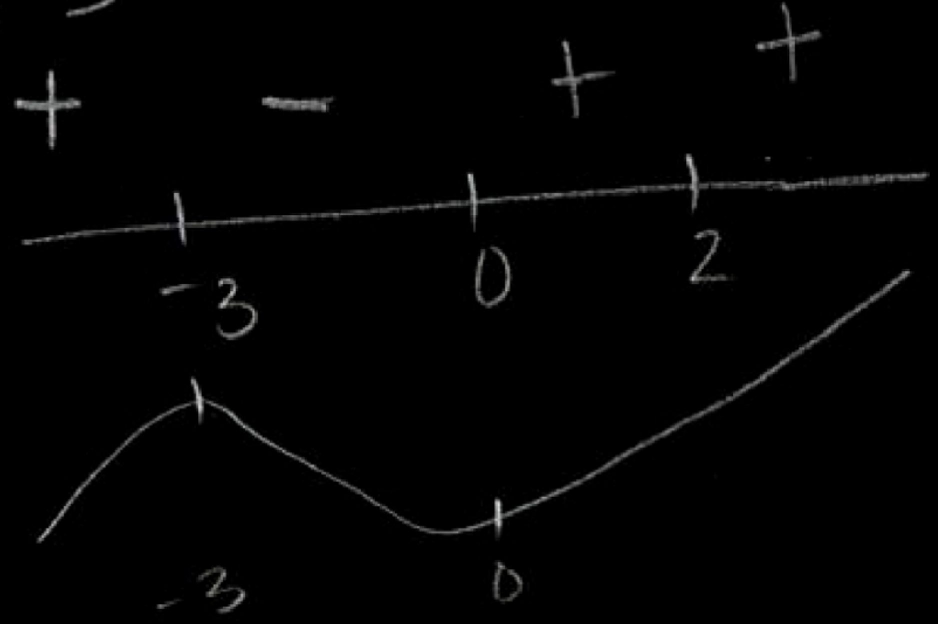
D. None of these



https://beckypatterson.weebly.com/uploads/5/2/3/6/5236342/test_review_ws.pdf

Answer: A,B,C

0, 2, -3



Second Derivative Test

Graphical intuition

Suppose that f is twice differentiable at the point a.

(a) If $f'(a) = 0$ and $f''(a) > 0$, then f has a local minimum at a.

(b) If $f'(a) = 0$ and $f''(a) < 0$, then f has a local maximum at a.

Question:

Using the second derivative test, determine the nature of the critical points for the function $f(x) = x^3 + 3x^2 - 9x$.

A. $x = 1$ is a local minimum and $x = -3$ is a local maximum.

B. $x = 1$ is a local maximum and $x = -3$ is a local minimum.

C. Both $x = 1$ and $x = -3$ are local minima.

D. Both $x = 1$ and $x = -3$ are local maxima.

Answer: A

To determine the nature of the critical points for $f(x) = x^3 + 3x^2 - 9x$:

1. Find the first derivative: $f'(x) = 3(x-1)(x+3)$.
2. Solve $f'(x) = 0$: $x = 1$ and $x = -3$.
3. Find the second derivative: $f''(x) = 6x + 6$.
4. Evaluate $f''(x)$ at the critical points:
    - $f''(1) = 12 > 0$: local minimum at $x = 1$.
    - $f''(-3) = -12 < 0$: local maximum at $x = -3$.

Correct option: A. $x = 1$ is a local minimum and $x = -3$ is a local maximum.

Proof of Second Derivative test

Taylor Series around x_0:

$$f(x_0 + h) = f(x_0) + f'(x_0)h + \frac{1}{2}f''(x_0)h^2 + \cdots$$

https://www2.math.upenn.edu/~kazdan/210/LectureNotes/2-deriv/2-deriv.pdf

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \tfrac{1}{2} f''(x_0)(x - x_0)^2 + \cdots \text{ higher order terms}$$
$$= f(x_0) + \tfrac{1}{2} f''(x_0)(x - x_0)^2 + \cdots \text{ higher order terms.}$$

$x = x_0 + h$

$$f(x_0+h) = f(x_0) + f'(x_0)\,h + \frac{1}{2} f''(x_0)\,h^2 + \ldots$$

Taylor Series around x_0:

$$f(x_0 + h) = f(x_0) + f'(x_0)h + \frac{1}{2}f''(x_0)h^2 + \cdots$$

(Ignoring higher order terms like $h^3, h^4, \ldots etc$ as $h \to 0$)

$$f(x_0 + h) = f(x_0) + f'(x_0)h + \frac{1}{2}f''(x_0)h^2$$

x_0 is a critical point (stationary point)

Taylor Series around x_0:

$$f(x_0+h) = f(x_0) + f'(x_0)h + \frac{1}{2}f''(x_0)h^2 + \cdots$$

(Ignoring higher order terms like $h^3, h^4, \ldots etc$ as $h \to 0$)

$$f(x_0+h) = f(x_0) + f'(x_0)h + \frac{1}{2}f''(x_0)h^2$$

x_0 is a critical point (stationary point)

Taylor Series around x_0:

$$f(x_0+h)=f(x_0)+f'(x_0)h+\frac{1}{2}f''(x_0)h^2+\cdots$$

(Ignoring higher order terms like $h^3, h^4, \ldots etc$ as $h \to 0$)

$$f(x_0+h)=f(x_0)+f'(x_0)h+\frac{1}{2}f''(x_0)h^2$$

Let x_0 be critical point

$$f(x_0+h)=f(x_0)+\frac{1}{2}f''(x_0)h^2$$

Taylor Series around x_0:

$$f(x_0+h) = f(x_0) + f'(x_0)h + \frac{1}{2}f''(x_0)h^2 + \cdots$$

(Ignoring higher order terms like $h^3, h^4, \ldots etc$ as $h \to 0$)

$$f(x_0+h) = f(x_0) + f'(x_0)h + \frac{1}{2}f''(x_0)h^2$$

Let x_0 be critical point

$$f(x_0+h) = f(x_0) + \frac{1}{2}f''(x_0)h^2$$

What should be sign of $f''(x_0)$ for x_0 to be local minima ?

Hint: We want $f(x_0)$ to be least in the neighborhood of x_0.

$$f(x_0+h) = f(x_0) + \frac{1}{2} f''(x_0) h^2$$

$f(x_0+h) > f(x_0)$

because i want x_0 to be local minima

$\Rightarrow f''(x_0) > 0$

Taylor Series around x_0:

$$f(x_0+h) = f(x_0) + f'(x_0)h + \frac{1}{2}f''(x_0)h^2 + \cdots$$

(Ignoring higher order terms like $h^3, h^4, \ldots etc$ as $h \to 0$)

$$f(x_0+h) = f(x_0) + f'(x_0)h + \frac{1}{2}f''(x_0)h^2$$

Let x_0 be critical point

$$f(x_0+h) = f(x_0) + \frac{1}{2}f''(x_0)h^2$$

We want $f(x_0)$ to be least in the neighborhood of x_0.

Taylor Series around x_0:

$$f(x_0+h) = f(x_0) + f'(x_0)h + \frac{1}{2}f''(x_0)h^2 + \cdots$$

(Ignoring higher order terms like $h^3, h^4, \dots etc$ as $h \to 0$)

$$f(x_0+h) = f(x_0) + f'(x_0)h + \frac{1}{2}f''(x_0)h^2$$

Let x_0 be critical point

$$f(x_0+h) = f(x_0) + \frac{1}{2}f''(x_0)h^2$$

We $f(x_0)$ to be least in the neighborhood of x_0.

Hence, We want $f(x_0) < f(x_0+h)$.

Therefore, $f''(x_0) > 0$.

Taylor Series around x_0:

$$f(x_0 + h) = f(x_0) + f'(x_0)h + \frac{1}{2}f''(x_0)h^2 + \cdots$$

(Ignoring higher order terms like $h^3, h^4, \ldots etc$ as $h \to 0$)

$$f(x_0 + h) = f(x_0) + f'(x_0)h + \frac{1}{2}f''(x_0)h^2$$

Let x_0 be critical point

$$f(x_0 + h) = f(x_0) + \frac{1}{2}f''(x_0)h^2$$

We $f(x_0)$ to be least in the neighborhood of x_0.

Hence, We want $f(x_0) < f(x_0 + h)$.

Therefore, $f''(x_0) > 0$.

This completes our proof that for x_0 to be local minimum:

- $f'(x_0) = 0$
- $f''(x_0) > 0$

What about functions of many variables?

Extend theorems that allow us to identify and classifly local minimizers of one variable functions to multivariable cases.

Multivariable Functions

What is meant by a multivariable function?

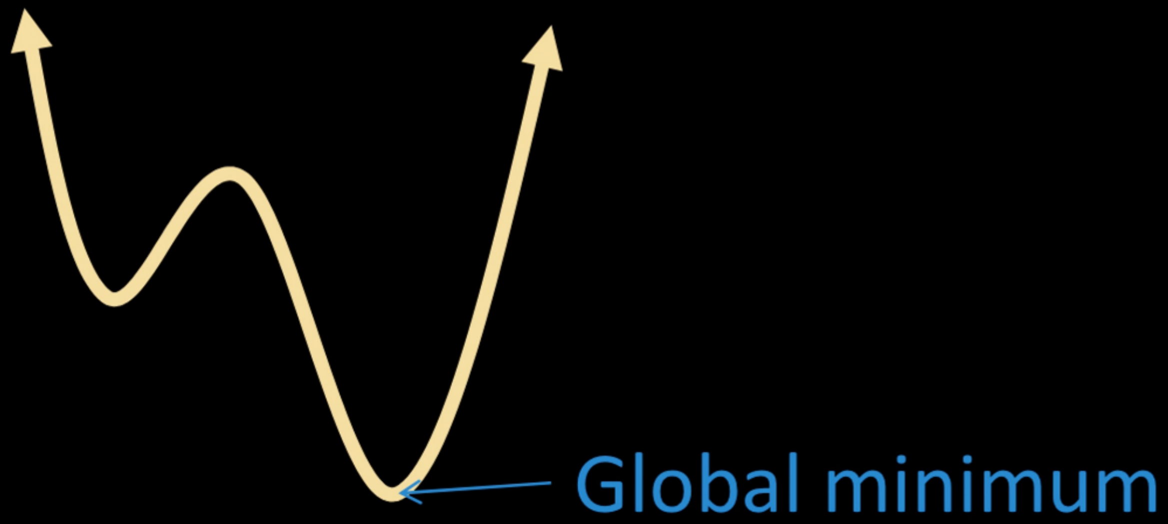
- A multivariable function is a function with several variables.
- Multivariable functions which take more parameters and give one single scalar value as the result.
- These functions are also known as scalar fields.
- The concepts of maxima and minima can be introduced for arbitrary scalar fields defined on subset of $\mathbb{R}^n$.

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

Unconstrained optimization

Find values of the variable $\mathbf{x}$ to give minimum of an objective function $f(x)$

$$\min_{x} \quad f(\mathbf{x})$$


Global minimum

Constrained optimization

Find values of the variable **x** to give minimum of an objective function $f(x)$

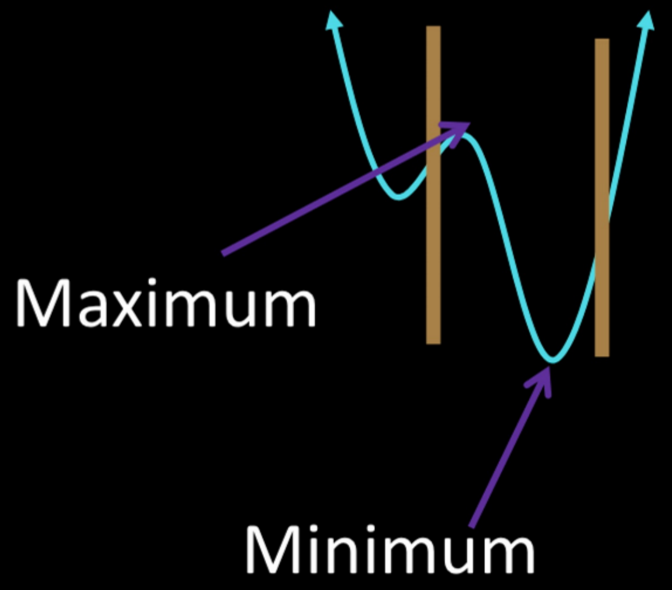
$$\min_{x} \quad f(\mathbf{x})$$

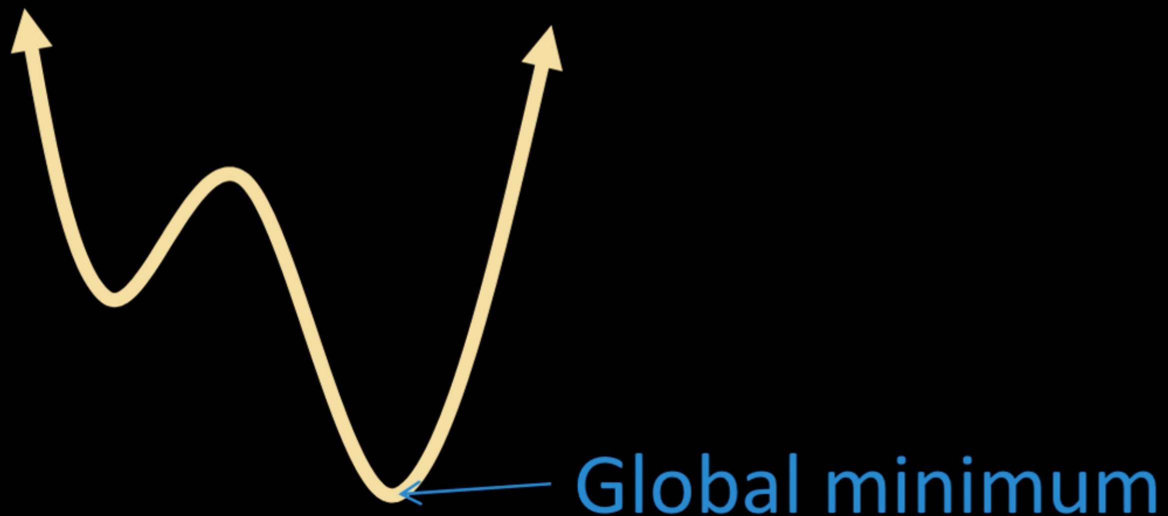
$$\boxed{(x+2)^2}$$

subject to $g(\mathbf{x}) \geq 0$

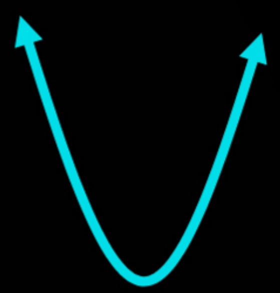
$h(\mathbf{x}) = 0$

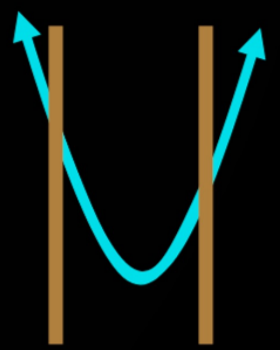
$$\boxed{x \geq 3}$$


Maximum
Minimum


Global minimum

Un Constrained vs Constrained Optimization





Unconstrained optimization

$$\min \underbrace{(x_1 - 2)^2 + (x_2 - 1)^2}_{f(x)}$$

Constrained optimization

$$\min \underbrace{(x_1 - 2)^2 + (x_2 - 1)^2}_{f(x)} \quad \text{s.t.}$$

$$\begin{cases} -x_1^2 + x_2 \geqslant 0 \\ -x_1 - x_2 + 2 \geqslant 0 \end{cases}$$

UnConstrained Optimization

Notation:

If $f(\mathbf{x})$ is a function of n variables with continuous first and second partial derivatives on $\mathbb{R}^n$, then the *gradient* of $f(\mathbf{x})$ is the vector

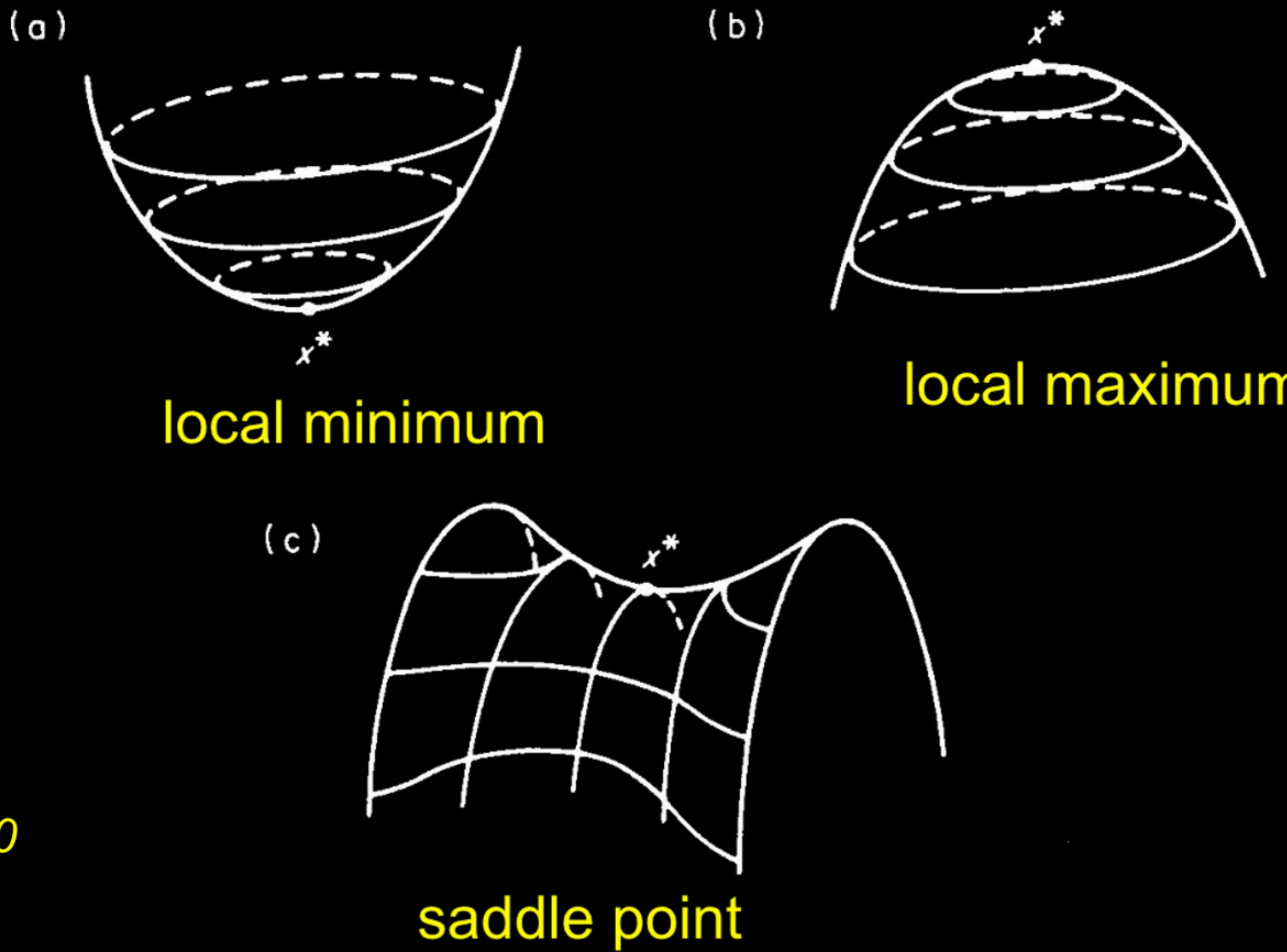
$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \frac{\partial f}{\partial x_3} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

Stationary point (or critical point)
Definition

- Assume f is differentiable at $\mathbf{a}$. If $\nabla f(\mathbf{a}) = \mathbf{0}$ the point $\mathbf{a}$ is called a **stationary point** of f.

Types of stationary points

$\nabla f(x^*) = 0$



(a)-(c) x^* is stationary: $\nabla f(x^*)=0$

In one dimension..

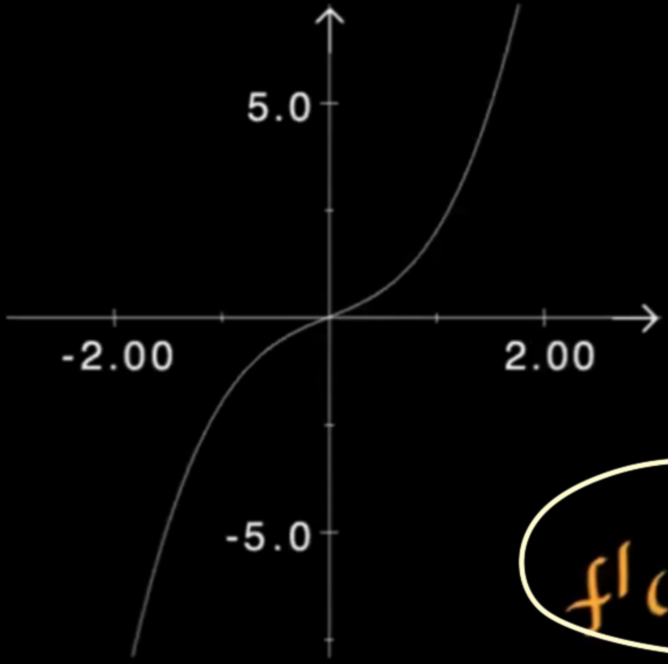
In one dimension, we have seen point of inflection.

If $f''(x) = 0$ and $f''(x)$ changes the sign then we call it point of inflection.

At point of inflection, $f'(x)$ may or may not be zero.

In one dimension..

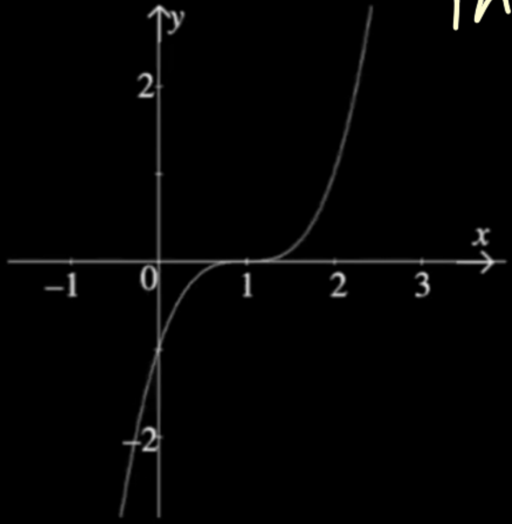
$f''(x) = 0$ for point of inflection.



The graph of $y = x^3 + x$.

$f'(x) \neq 0$

at point of inflection

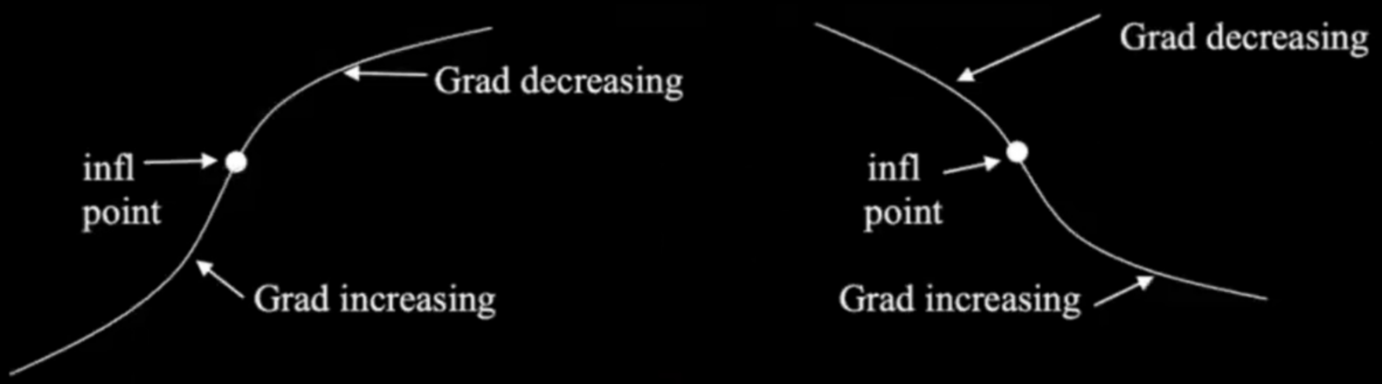


The graph of $y = x^3 - 3x^2 + 3x - 1$.

$f'(x) = 0$

In one dimension..

These are **ordinary** inflection points:

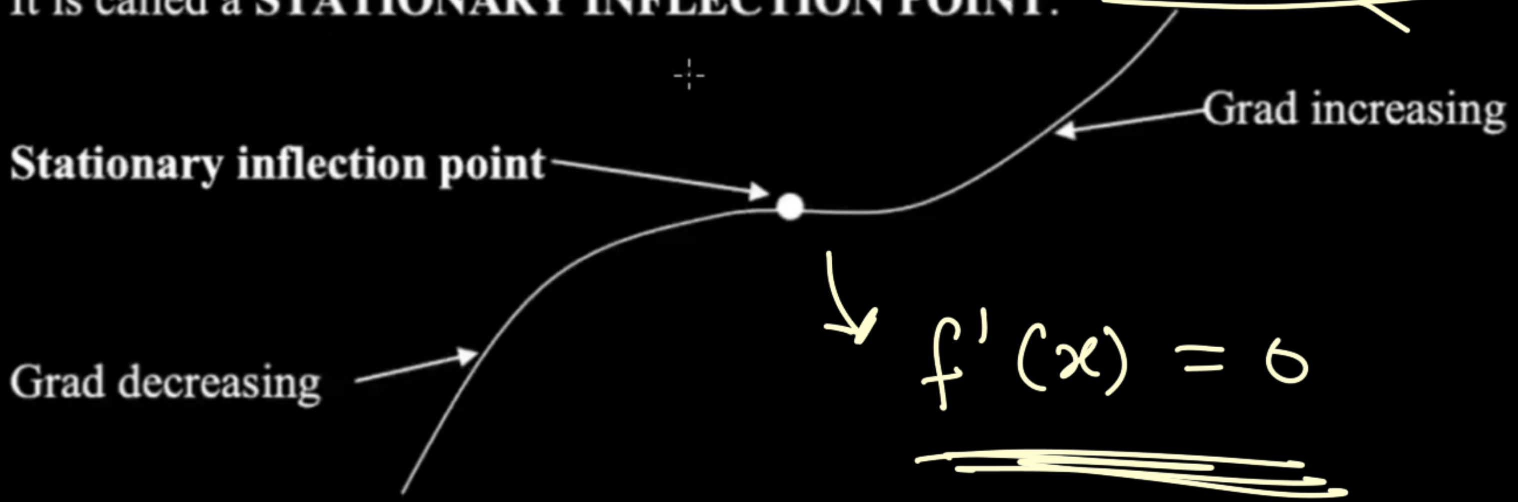


$f'(x) \neq 0$ $\qquad$ $f'(x) \neq 0$

In one dimension..

This is a special inflection point because the gradient is also zero.
It is called a **STATIONARY INFLECTION POINT**.

saddle point



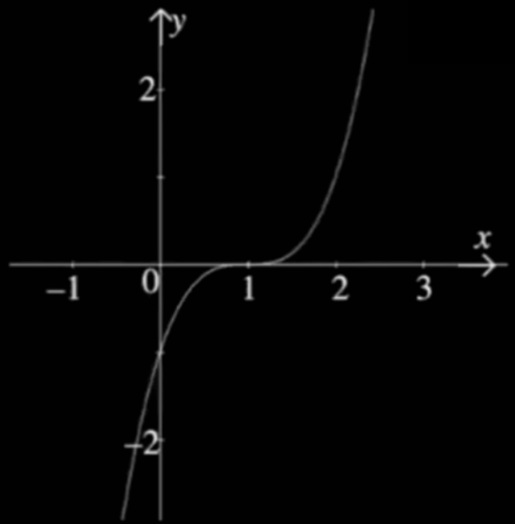
$f'(x) = 0$

(A typical example of this is the curve $y = x^3$)

Saddle point

$f'(x) = 0$ and it is neither minima nor maxima.

2 Conditions of saddle point



The graph of $y = x^3 - 3x^2 + 3x - 1$.

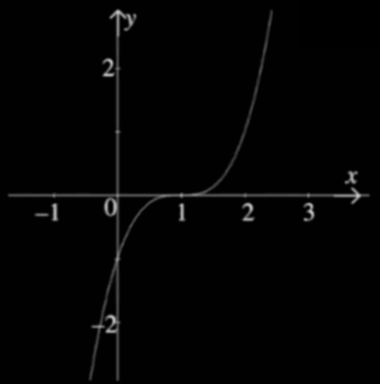
$f'(x) = 0$

in case of <u>one - D</u>

saddle point is just inflection point where $f'(x) = 0$

the kind of points satisfy the two conditions of saddle point.

GO CLASSES



The graph of $y = x^3 - 3x^2 + 3x - 1$.

$f'(x) = 0$

in case of one - D

saddle point is just

inflection point where $f'(x) = 0$

the kind of points satisfy

the two conditions of saddle

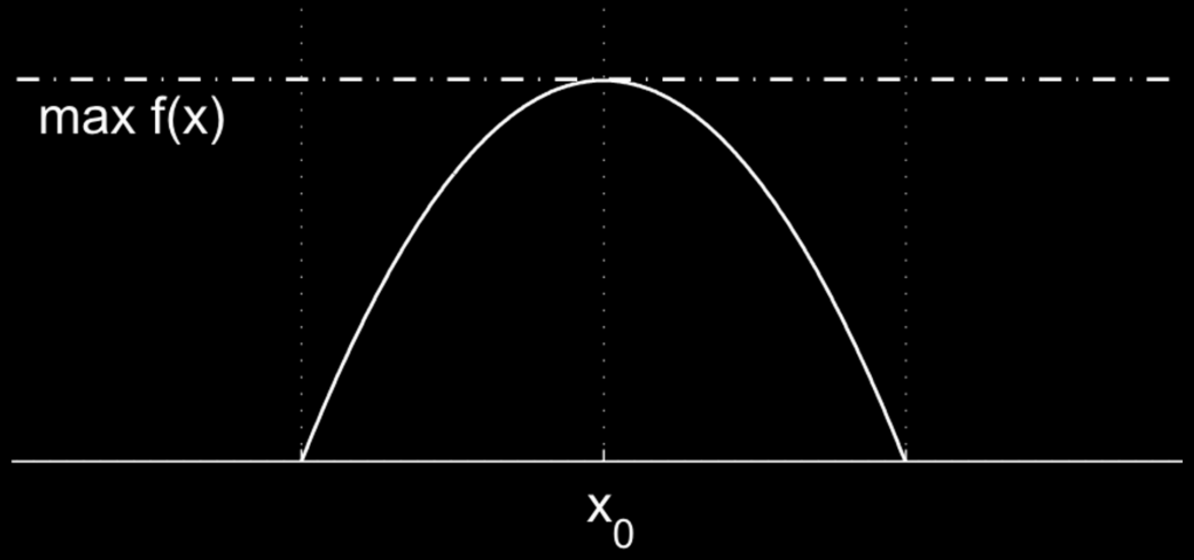
point.

in case of one-D saddle point

is not of much interest.

- Maximizing the objective function is the same as minimizing the negative of the objective function.

$$\max_{x \in \mathbb{R}^n} f(x) = -\min_{x \in \mathbb{R}^n} [-f(x)]$$


max f(x)
x_0

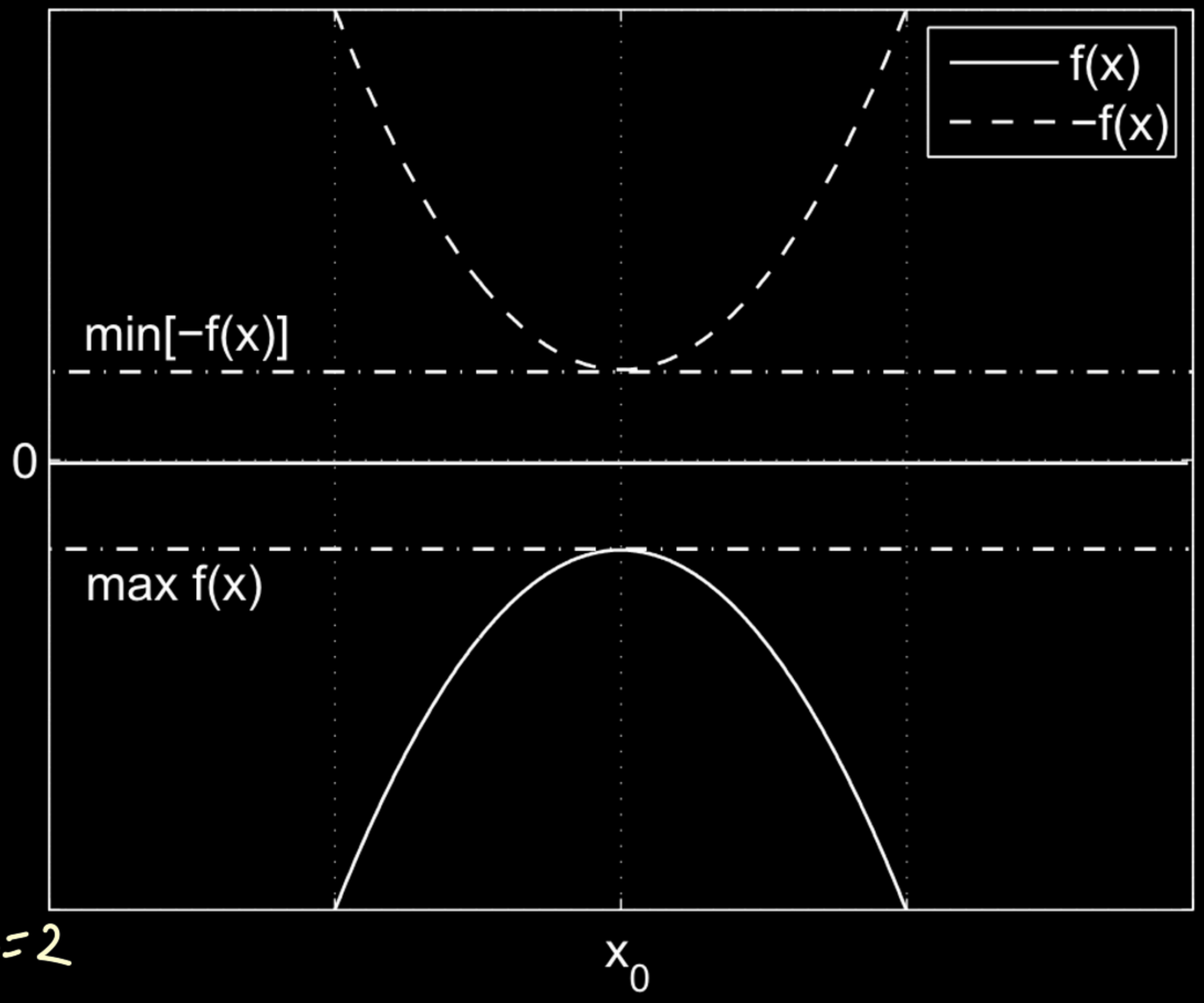
Loss functions

$x=0$ $x=1$ $x=2$

1, 3, 9

-1 -3 -9

$x=0$ $x=1$ $x=2$



Question:

Example: Find all critical points for each given function:

1. $f(x, y) = x^2 + y^2.$
2. $g(x, y) = x^2 + 2x - y^2 - 6y + 4.$
3. $p(x, y) = x^3 + y^3 - 3xy.$
4. $q(x, y) = x^2 + 4xy + 2y^2 + 6x - 4y + 3.$
5. $j(x, y) = x^2y^2 - x^2 - y^2.$

$$\begin{bmatrix} 2x \\ 2y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$x = 0, y = 0$

https://web.northeastern.edu/dummit/teaching_sp21_2321/2321_lecture_10_minima_maxima_classification.pdf

Example: Find all critical points for each given function:

1. $f(x, y) = x^2 + y^2$.

- We have $f_x = 2x$ and $f_y = 2y$.
- Clearly $f_x = 0$ requires $x = 0$ while $f_y = 0$ requires $y = 0$.
- Therefore, we get one critical point: $(x, y) = (0, 0)$.

2. $g(x, y) = x^2 + 2x - y^2 - 6y + 4$.

- We have $g_x = 2x + 2$ and $g_y = -2y - 6$.
- Then $g_x = 0$ requires $x = -1$ while $g_y = 0$ requires $2y = -6$ so $y = -3$.
- Therefore, we get one critical point: $(x, y) = (-1, -3)$.

Example: Find all critical points for each given function:

3. $p(x, y) = x^3 + y^3 - 3xy$.

- We have $p_x = 3x^2 - 3y$ and $p_y = 3y^2 - 3x$.
- So we get the equations $3x^2 - 3y = 0$ and $3y^2 - 3x = 0$.
- Neither equation gives us a value for x or y directly.
- But we can solve the first equation for y in terms of x:this gives $y = x^2$.
- Now plugging into the second equation yields $3(x^2)^2 - 3x = 0$, so that $3x^4 - 3x = 0$.
- Factoring gives $3x(x^3 - 1) = 0$, which has the solutions $x = 0$ and $x = 1$.
- If $x = 0$, then $y = x^2 = 0$, so we get the point $(0, 0)$.
- If $x = 1$, then $y = x^2 = 1$, so we get the point $(1, 1)$.
- In total, there are two critical points: $(0, 0)$ and $(1, 1)$.

Example: Find all critical points for each given function:

4. $q(x, y) = x^2 + 4xy + 2y^2 + 6x - 4y + 3.$

- We have $q_x = 2x + 4y + 6$ and $q_y = 4x + 4y - 4$.
- So we get the equations $2x + 4y + 6 = 0$ and $4x + 4y - 4 = 0$.
- Neither equation gives us a value for x or y directly.
- But we can solve the second equation for y in terms of x:this gives $y = 1 - x$.
- Now plugging into the first equation yields $2x + 4(1 - x) + 6 = 0$, so that $10 - 2x = 0$ and thus $x = 5$.
- Then $y = 1 - x = -4$.
- Therefore, there is one critical point: $(x, y) = (5, -4)$.

There are many other ways to solve this system: for example, we could have solved the first equation for y in terms of x, or for x in terms of y, or we could have subtracted the two equations.

Example: Find all critical points for each given function:

5. $j(x,y) = x^2y^2 - x^2 - y^2$.

- We have $j_x = 2xy^2 - 2x$ and $j_y = 2x^2y - 2y$.
- So we get the equations $2xy^2 - 2x = 0$ and $2x^2y - 2y = 0$.
- We can factor the first equation: $2x(y^2 - 1) = 0$.
- This means either $2x = 0$ (so that $x = 0$) or $y^2 - 1 = 0$ (so that $y = 1$ or $y = -1$).
- If $x = 0$, then the second equation gives $-2y = 0$ so that $y = 0$. We get the point $(0,0)$.
- If $y = 1$, then the second equation gives $2x^2 - 2 = 0$ so that $x = 1$ or $x = -1$. We get points $(1,1)$ and $(-1,1)$.
- If $y = -1$, then the second equation gives $-2x^2 + 2 = 0$ so that $x = 1$ or $x = -1$. We get points $(1,-1)$ and $(-1,-1)$.
- In total we get five critical points: $(0,0)$, $(1,1)$, $(-1,1)$, $(1,-1)$, $(-1,-1)$.

Definition

A <u>local minimum</u> is a point where f is nearby always bigger.
A <u>local maximum</u> is a point where f is nearby always smaller.
A <u>saddle point</u> is a critical point where f nearby is bigger in some directions and smaller in others.

Question:

Find and classify the critical points of the function
$f(x, y) = x^2 + 10x + 3y^2 - 18y + 5.$

(1) find the critical points.

(2) Classify the critical points.

unconstrained optimisation in multivariate

Question:

Find and classify the critical points of the function
$f(x, y) = x^2 + 10x + 3y^2 - 18y + 5.$

(1) find the critical points.

(2) Classify the critical points.

Next, We will see how to solve such questions.

Gradient and Hessian

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

differentiation of a vector w.r.t. a vector

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \\ \frac{\partial f}{\partial y \partial x} & \end{bmatrix}$$

$$\frac{\partial}{\partial x} \qquad \frac{\partial}{\partial y}$$

Gradient and Hessian

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \quad H = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}$$

← Symmetric matrix

derivative of gradient

Question:

Find the hessian matrix for the given function:

$$f(x, y) = 8x - 2x^2y^2$$

https://web.stanford.edu/group/sisl/k12/optimization/MO-unit4-pdfs/4.9findingHessians.pdf

Question:

Find the hessian matrix for the given function:

$$f(x, y) = 8x - 2x^2y^2$$

$$\nabla f = \begin{bmatrix} 8 - 4xy^2 \\ -4yx^2 \end{bmatrix} \quad \begin{bmatrix} -4y^2 & -8xy \\ -8xy & -4x^2 \end{bmatrix}$$

https://web.stanford.edu/group/sisl/k12/optimization/MO-unit4-pdfs/4.9findingHessians.pdf

Given the function $f(x,y) = 8x - 2x^2y^2$

Gradient:

$$\nabla f(x,y) = \begin{bmatrix} 8 - 4xy^2 \\ -4x^2y \end{bmatrix}$$

1. **First column**

$$\begin{bmatrix} \frac{\partial^2 f}{\partial x^2} \\ \frac{\partial^2 f}{\partial y \partial x} \end{bmatrix} = \begin{bmatrix} -4y^2 \\ -8xy \end{bmatrix}$$

2. **Second column**

$$\begin{bmatrix} \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y^2} \end{bmatrix} = \begin{bmatrix} -8xy \\ -4x^2 \end{bmatrix}$$

Hessian:

$$H = \begin{bmatrix} -4y^2 & -8xy \\ -8xy & -4x^2 \end{bmatrix}$$

Objective:

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \; f(x)$$

Necessary & Sufficinet Conditions for Optimality

x^* is a local minimum of $f(x)$ iff:

1. Zero gradient at x^*:

$$\nabla_x f(x^*) = 0$$

2. Hessian at x^* is positive definite:

$$d^\top \nabla^2 f(x^*) d > 0$$

$$d^T H(x^*) d > 0$$

Question:

Example: $f(x, y) = x^2 + y^2$.

1. Find all critical points.
2. Classify the type of critical points.

Question:

Example: $f(x, y) = x^2 + y^2$.

1. Find all critical points.
2. Classify the type of critical points.

$\nabla = \begin{bmatrix} 2x \\ 2y \end{bmatrix}$ $x = 0$ $y = 0$

$d^T H d$ $d = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$

$H = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ PD

local minima

Solution:

Example: Classify the type of critical point that $f(x, y) = x^2 + y^2$ has at the origin $(0, 0)$.

- We saw earlier that $(0, 0)$ is a critical point of this function.
- To classify it, we compute $f_{xx} = 2$, $f_{xy} = 0$, and $f_{yy} = 2$.
- Then $D = f_{xx} f_{yy} - (f_{xy})^2 = 2 \cdot 2 - 0^2 = 4$. (Here, D is constant, but normally we would need to evaluate it at our point.)
- So, by the second derivatives test, since $D > 0$ and $f_{xx} > 0$ at $(0, 0)$, we see that $(0, 0)$ is a local minimum.

Question:

Find and classify the critical points of the function
$f(x, y) = x^2 + 10x + 3y^2 - 18y + 5.$

(1) find the critical points.

(2) Classify the critical points.

https://www.math.ucdavis.edu/~deloera/TEACHING/LOWDIVISION/MATH16C/lec7.3-7.5.pdf

Question:

Find and classify the critical points of the function
$f(x, y) = x^2 + 10x + 3y^2 - 18y + 5.$

(1) find the critical points.

(2) Classify the critical points.

$$\nabla = \begin{bmatrix} 2x+10 \\ 6y-18 \end{bmatrix} = 0$$

$$H = \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix}$$

$x = -5 \quad , \quad y = 3$

$\boxed{\text{minima at } x = -5, \ y = 3}$

Solution:

Find and classify the critical points of the function
$f(x, y) = x^2 + 10x + 3y^2 - 18y + 5$.

Answer: **(1)** find the critical points.

$f_x(x, y) = 2x+10 = 0 \Leftrightarrow x = -5 \quad f_y(x, y) = 6y-18 = 0 \Leftrightarrow y = 3.$

So, there is one and only one critical point: $(-5, 3)$.

(2) Classify the critical point: Compute the 2nd order partial derivatives:

$$f_{xx}(x, y) = 2, f_{yy}(x, y) = 6, f_{xy}(x, y) = 0 = f_{yx}(x, y).$$

Then,

$d(-5, 3) = f_{xx}(-5, 3)\cdot f_{yy}(-5, 3)-(f_{xy}(-5, 3))^2 = (2)(6)-(0)^2 = 12 > 0.$

Since $f_{xx}(-5, 3) = 2 > 0$, the critical point is a relative minimum.

Proof of Second Derivative test for 1D Function

Taylor Series around x_0:

$$f(x_0+h) = f(x_0) + f'(x_0)h + \frac{1}{2}f''(x_0)h^2 + \cdots$$

(Ignoring higher order terms like $h^3, h^4, \ldots etc$ as $h \to 0$)

$$f(x_0+h) = f(x_0) + f'(x_0)h + \frac{1}{2}f''(x_0)h^2$$

Let x_0 be critical point

$$f(x_0+h) = f(x_0) + \frac{1}{2}f''(x_0)h^2$$

We $f(x_0)$ to be least in the neighborhood of x_0.

Hence, We want $f(x_0) < f(x_0+h)$.

Therefore, $f''(x_0) > 0$.

This completes our proof that for x_0 to be local minimum:

- $f'(x_0) = 0$
- $f''(x_0) > 0$

Proof of Second Derivative test for multivariate function

minima cond^{n}s:

$$\left.\begin{array}{l} \nabla f(x^*) = 0 \\ H(x^*) \quad PD \end{array}\right\}$$

Taylor Series around x:

vector

$$f(\mathbf{x}+h) = f(\mathbf{x}) + \nabla f(\mathbf{x})h + \frac{1}{2}h^T\nabla^2 f(\mathbf{x})h + \ldots$$

$\frac{1}{2}h^2 f''(x)$

https://scottmccracken.weebly.com/uploads/9/0/6/6/9066859
/unconstrained_optimization-print_version.pdf

Taylor Series around x:

for $x = x^*$

$$f(\mathbf{x}+h) = f(\mathbf{x}) + \underbrace{\cancel{\nabla f(\mathbf{x})}}_{0} h + \frac{1}{2} h^T \nabla^2 f(\mathbf{x}) h + \dots$$

Let x^* be critical point

Taylor Series around x:

$$f(\mathbf{x}+h) = f(\mathbf{x}) + \nabla f(\mathbf{x})h + \underbrace{\frac{1}{2}h^T\nabla^2 f(\mathbf{x})h + \dots}_{}$$

ignoring higher order terms

Let x^* be critical point

$$f(\mathbf{x}^*+h) = f(\mathbf{x}^*) + \frac{1}{2}h^T\nabla^2 f(\mathbf{x}^*)h$$

Taylor Series around x:

$$f(\mathbf{x}+h) = f(\mathbf{x}) + \nabla f(\mathbf{x})h + \frac{1}{2}h^T\nabla^2 f(\mathbf{x})h + \underbrace{\ldots}_{\text{ignoring higher order terms}}$$

Let x^* be critical point

$$f(\mathbf{x}^*+h) = f(\mathbf{x}^*) + \frac{1}{2}h^T\nabla^2 f(\mathbf{x}^*)h$$

for x^* to be minima $f(x^*+h) > f(x^*)$

$\Rightarrow h^T \nabla^2 f(x^*)\, h > 0 \Rightarrow$ Hessian is PD

Question:

$$f(x_1, x_2) = \frac{1}{3}x_1^3 + \frac{1}{2}x_1^2 + 2x_1x_2 + \frac{1}{2}x_2^2 - x_2 + 1.$$

Find all the local minima or maxima of above function.

https://members.cbio.mines-paristech.fr/~jvert/teaching/2006insead/slides/3_optimality/optimality.pdf

$$\nabla f(x_1, x_2) = \begin{pmatrix} x_1^2 + x_1 + 2x_2 \\ 2x_1 + x_2 - 1 \end{pmatrix}, \Rightarrow \quad \nabla = 0 \quad \Rightarrow \quad \begin{matrix} (1, -1) \\ (2, -3) \end{matrix}$$

$$\nabla^2 f(x_1, x_2) = \begin{pmatrix} 2x_1 + 1 & 2 \\ 2 & 1 \end{pmatrix}.$$

Saddle point

There are two stationary points: $x_a = (1, -1)^\top$ and $x_b = (2, -3)^\top$. The corresponding Hessian are:

$$\nabla^2_{(1,-1)} = \begin{pmatrix} 3 & 2 \\ 2 & 1 \end{pmatrix}$$

$$\underbrace{|H| < 0 \;\&\; f_{xx} > 0}$$

$$\Downarrow$$

Neither PD nor Negitive definite

$$\nabla^2_{(2,-3)} = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}$$

$|H| > 0$
$f_{xx} > 0$

$$\nabla f(x_1, x_2) = \begin{pmatrix} x_1^2 + x_1 + 2x_2 \\ 2x_1 + x_2 - 1 \end{pmatrix},$$

$$\nabla^2 f(x_1, x_2) = \begin{pmatrix} 2x_1 + 1 & 2 \\ 2 & 1 \end{pmatrix}.$$

There are two stationary points: $x_a = (1, -1)^\top$ and $x_b = (2, -3)^\top$. The corresponding Hessian are:

$$\nabla^2 f(x_a) = \begin{pmatrix} 3 & 2 \\ 2 & 1 \end{pmatrix} \quad \text{and} \quad \nabla^2 f(x_b) = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}$$

- $\det\left(\nabla^2 f(x_a)\right) = -1$ so the Hessian has a negative and a positive eigenvalue: x_a is neither a local maximum nor a local minimum } Saddle point
- $\nabla^2 f(x_b) \succ 0$ so x_b is a local minimum.

We have various different types of critical points:

Definition

A local minimum is a point where f is nearby always bigger.
A local maximum is a point where f is nearby always smaller.
A saddle point is a critical point where f nearby is bigger in some directions and smaller in others.

Question:

Example: For $f(x, y) = 3x^2 + 2y^3 - 6xy$, find the critical points of f and classify them as minima, maxima, or saddle points.

Question:

Example: For $f(x,y) = 3x^2 + 2y^3 - 6xy$, find the critical points of f and classify them as minima, maxima, or saddle points.

$$\nabla = \begin{bmatrix} 6x - 6y \\ 6y^2 - 6x \end{bmatrix} \Rightarrow (x, y) = 0, 0 \quad (1, 1)$$

$$H = \begin{bmatrix} 6 & -6 \\ -6 & 12y \end{bmatrix}$$

$H_{(0,0)}$ $f_{xx} > 0$

$|H_{(0,0)}| < 0$

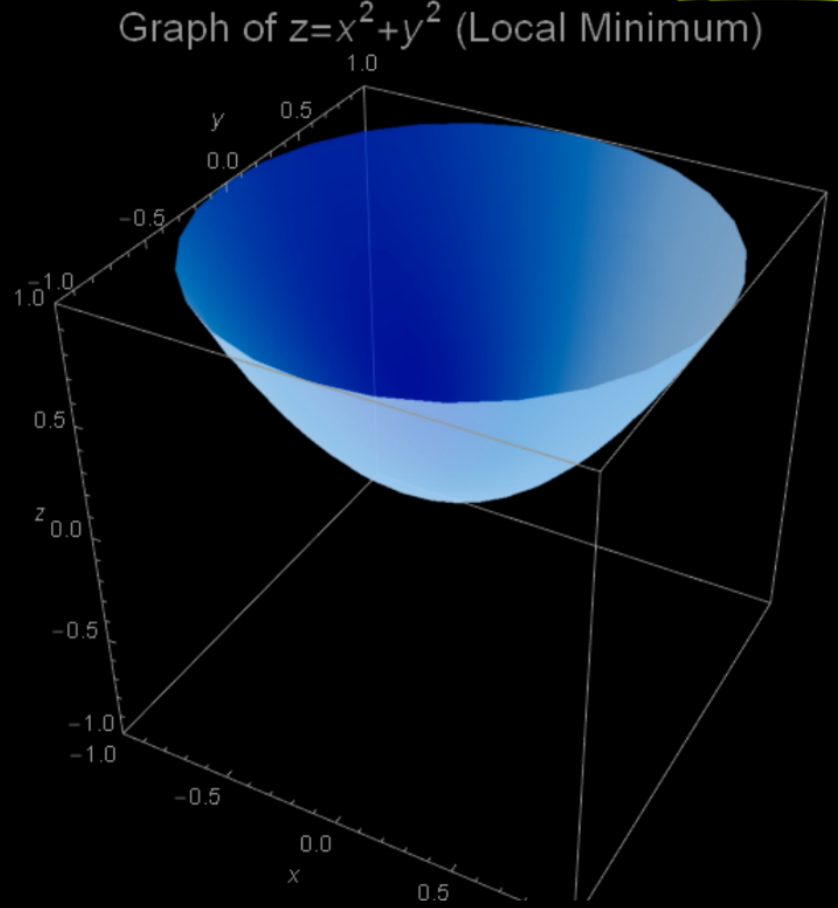
$|H_{(1,1)}| > 0$ $f_{xx} > 0$

Solution:

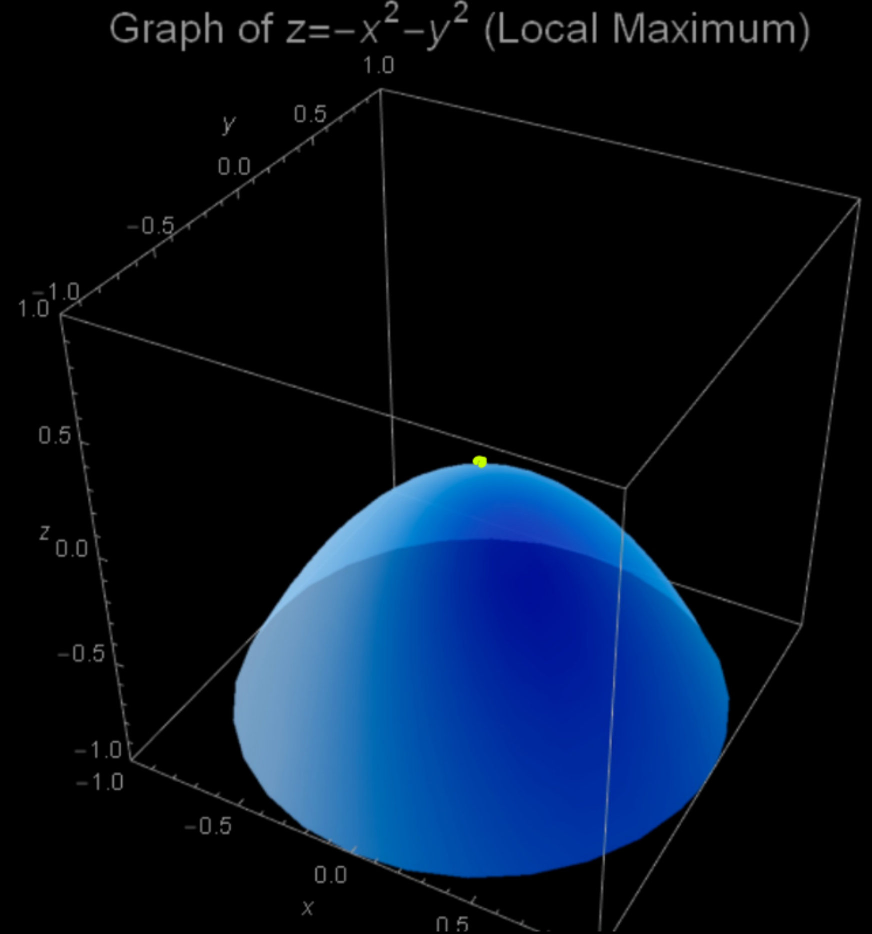
Example: For $f(x, y) = 3x^2 + 2y^3 - 6xy$, find the critical points of f and classify them as minima, maxima, or saddle points.

- First, $f_x = 6x - 6y$ and $f_y = 6y^2 - 6x$. They are both defined everywhere so we need only find where they are both zero.
- Next, we can see that f_x is zero only when $y = x$.
- Then the equation $f_y = 0$ becomes $6x^2 - 6x = 0$, which by factoring we can see has solutions $x = 0$ or $x = 1$.
- Since $y = x$, we see $(0, 0)$ and $(1, 1)$ are the critical points.
- To classify them, we compute $f_{xx} = 6$, $f_{xy} = -6$, and $f_{yy} = 12y$. Then $D(0, 0) = 6 \cdot 0 - (-6)^2 < 0$ and $D(1, 1) = 6 \cdot 12 - (-6)^2 > 0$. Also, $f_{xx} > 0$ at $(1, 1)$.
- So, by the second derivatives test, $(0, 0)$ is a saddle point and $(1, 1)$ is a local minimum.

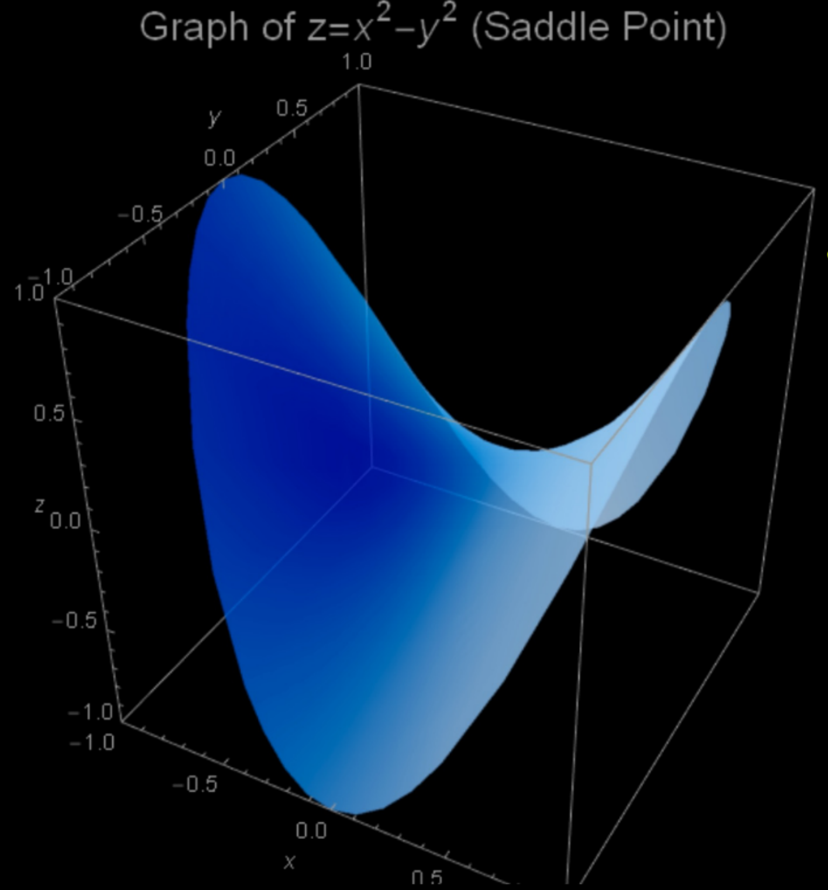
Example: $f(x, y) = x^2 + y^2$ has a local minimum at $(0, 0)$:



Example: $f(x, y) = -x^2 - y^2$ has a local maximum at $(0, 0)$:



Example: $f(x, y) = x^2 - y^2$ has a saddle point at $(0, 0)$:



$\nabla = \begin{bmatrix} 2x \\ -2y \end{bmatrix}$

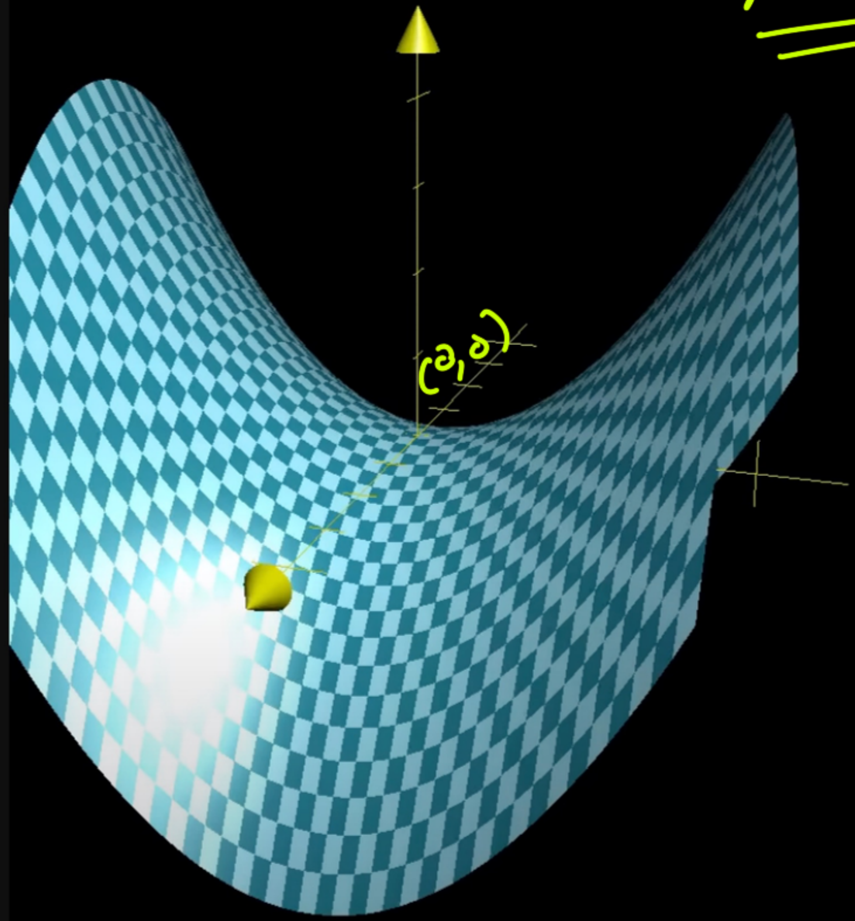
indefinite

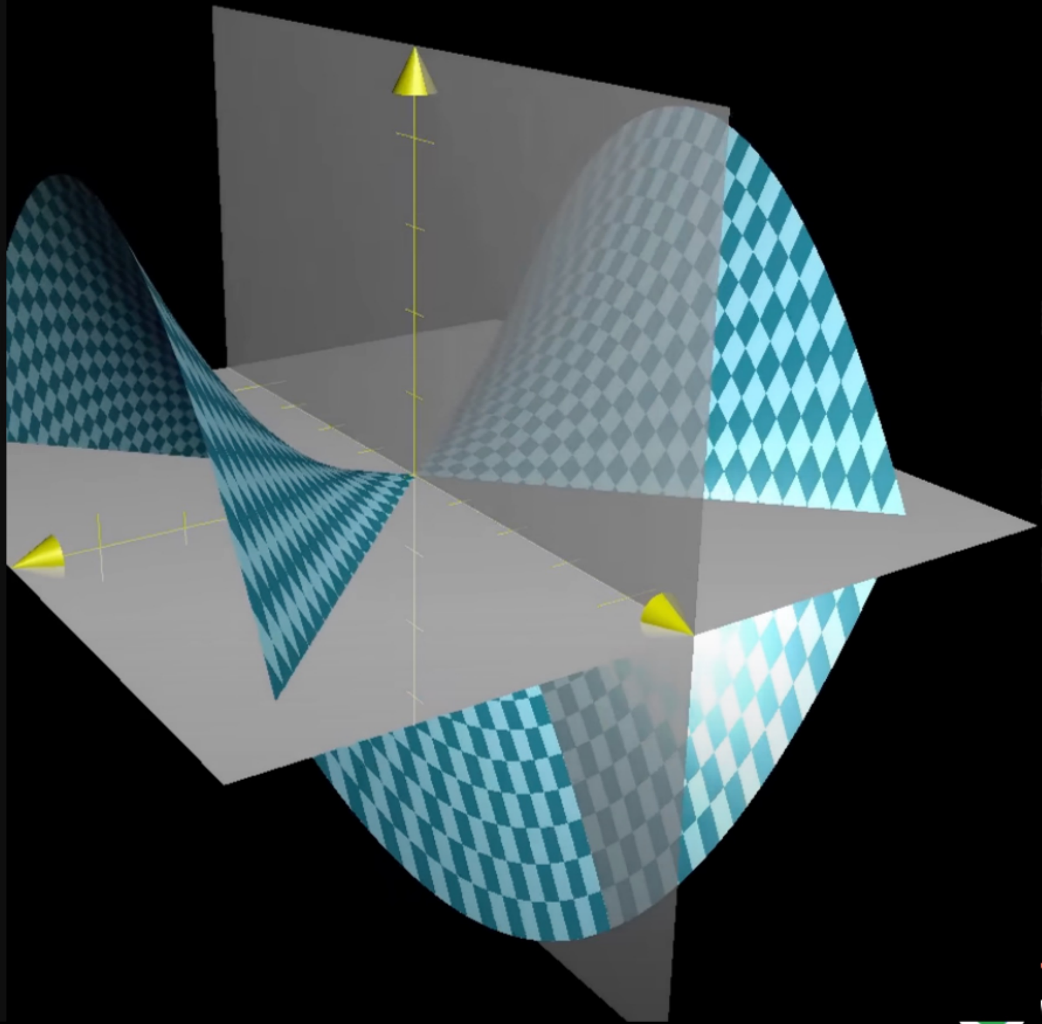
$H = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$

Local minima and local maxima are (presumably) familiar from the one-variable setting. Saddle points, however, are a new kind of critical point.

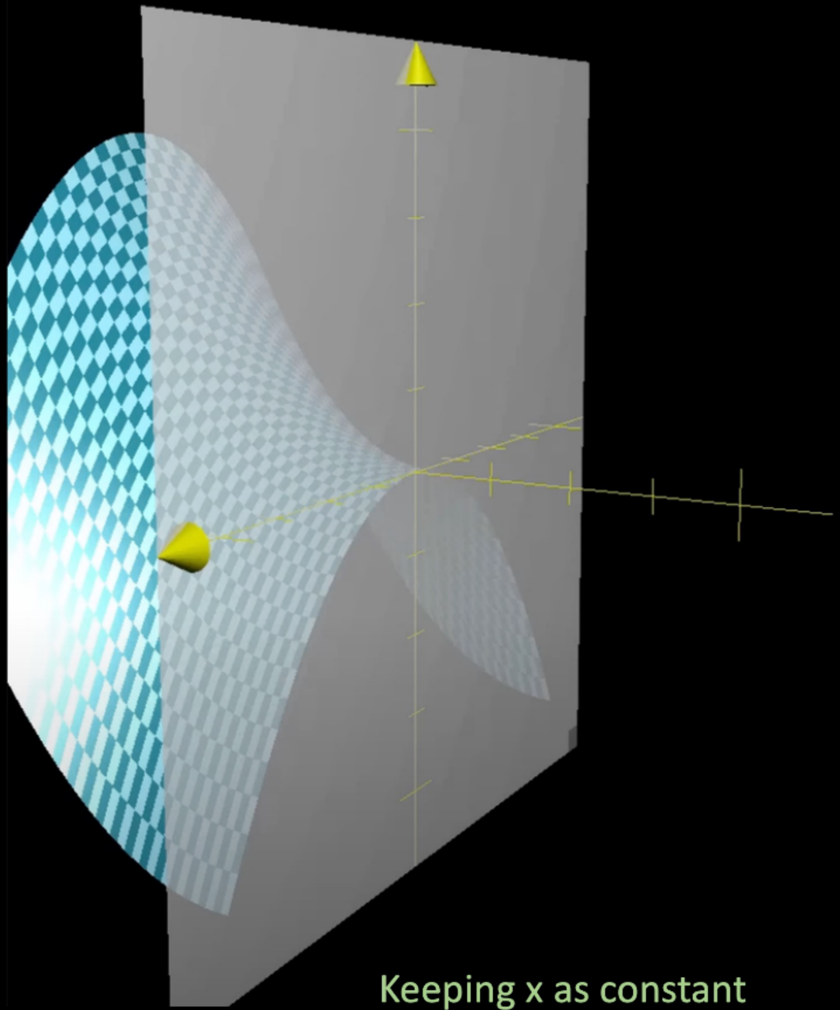
- A saddle point will look like a local minimum along some directions and a local maximum along other directions.
- For example, $f(x, y) = x^2 - y^2$ looks like a minimum in the x-direction (as x varies and y is held fixed at 0, the function is $f(x, 0) = x^2$) but a maximum in the y-direction (as y varies and x is held fixed at 0, the function is $f(0, y) = -y^2$).
- In contrast, a local minimum looks like a minimum in every direction, while a local maximum looks like a maximum in every direction.

$x^2 - y^2$

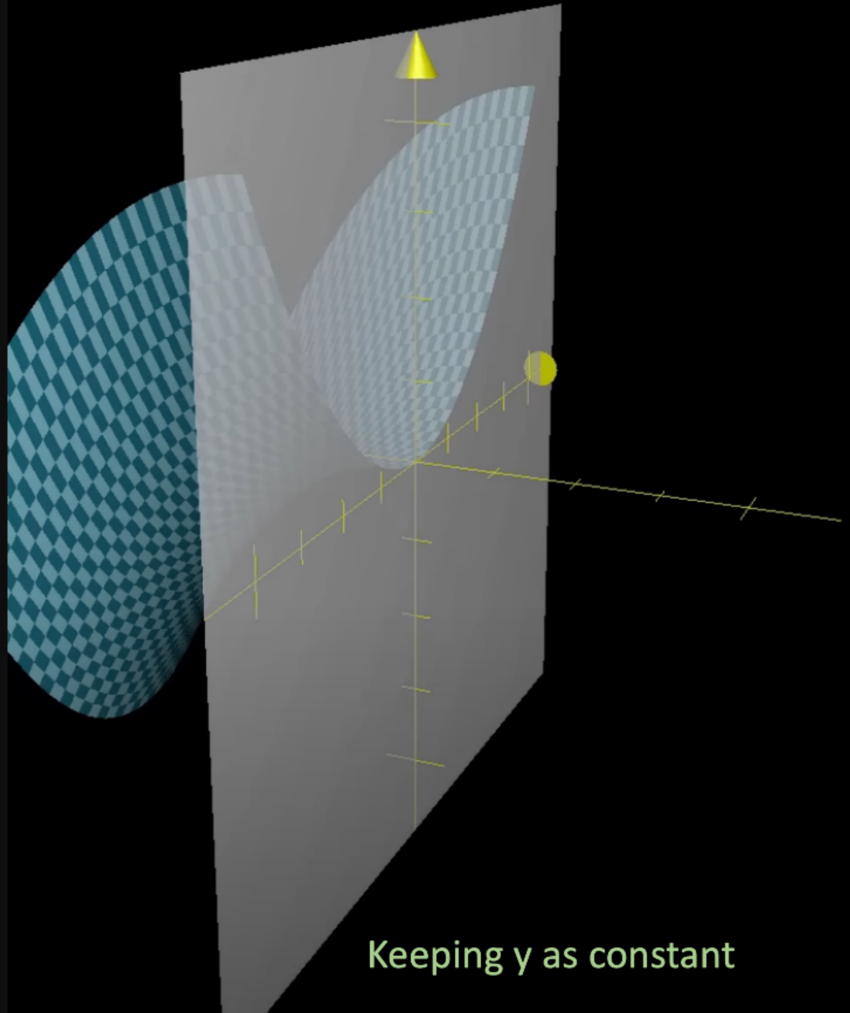




GO CLASSES



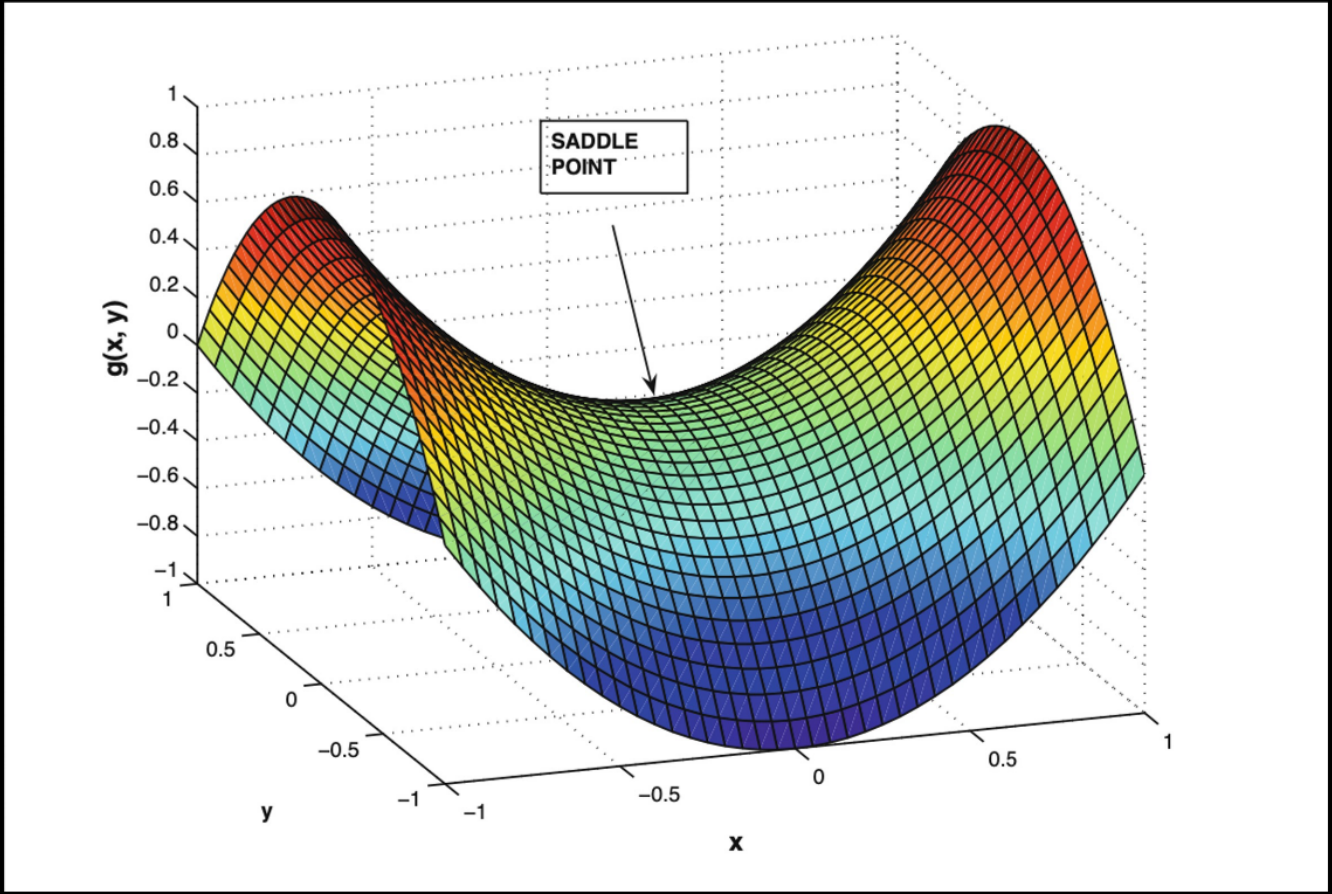
Keeping x as constant

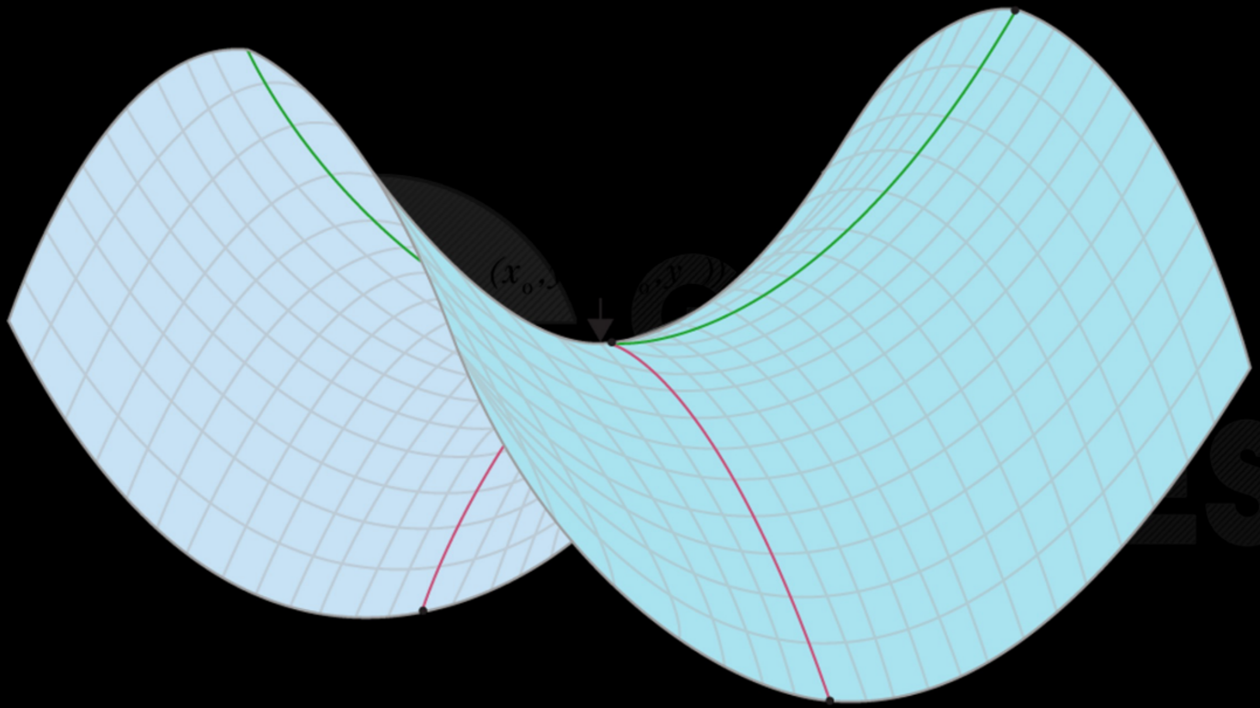


Keeping y as constant


PERFECT FLAVOUR IN EVERY BITE
Pringles
CHEESE & ONION

Calculus





- At a point $\mathbf{x} = \mathbf{x}^* + \mathbf{d}$ in the neighborhood of a saddle point $\mathbf{x}^*$, the Taylor series gives

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \frac{1}{2}\mathbf{d}^T\mathbf{H}(\mathbf{x}^*)\mathbf{d} + \ldots$$

since $\mathbf{g}(\mathbf{x}^*) = 0$. As $\mathbf{x}^*$ is neither a maximizer nor a minimizer, there must be directions $\mathbf{d}_1$, $\mathbf{d}_2$ (or $\mathbf{x}_1$, $\mathbf{x}_2$) such that

$$\begin{array}{rcl} f(\overbrace{\mathbf{x}^* + \mathbf{d}_1}^{\mathbf{x}_1}) < f(\mathbf{x}^*) & \Rightarrow & \mathbf{d}_1^T\mathbf{H}(\mathbf{x}^*)\mathbf{d}_1 < 0 \\ f(\underbrace{\mathbf{x}^* + \mathbf{d}_2}_{\mathbf{x}_2}) > f(\mathbf{x}^*) & \Rightarrow & \mathbf{d}_2^T\mathbf{H}(\mathbf{x}^*)\mathbf{d}_2 > 0 \end{array}$$

Then, $\mathbf{H}(\mathbf{x}^*)$ is indefinite

http://personal.cimat.mx:8181/~dalmau/courses/2018A/clase5.pdf

Question:

Example: Classify the two critical points $(0, 0)$ and $(1, 1)$ for $p(x, y) = x^3 + y^3 - 3xy$.

https://web.northeastern.edu/dummit/teaching_sp21_2321/2321_lecture_10_minima_maxima_classification.pdf

Solution:

Remember that $D = f_{xx} \cdot f_{yy} - (f_{xy})^2 = \begin{vmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{vmatrix}$.

- We have $f_x = 3x^2 - 3y$, $f_y = 3y^2 - 3x$, so $f_{xx} = 6x$, $f_{xy} = -3$, and $f_{yy} = 6y$.
- Therefore, $D = (6x)(6y) - (-3)^2$.
- Thus, at $(x, y) = (0, 0)$ we get $D = (0)(0) - 9 = -9$.
- Also, at $(x, y) = (1, 1)$ we get $D = (6)(6) - 9 = 27$.

- We saw earlier that $(0, 0)$ and $(1, 1)$ are the critical points of this function and that $D = (6x)(6y) - (-3)^2$.
- At $(0, 0)$, we have $D = -9$, so $(0, 0)$ is a saddle point.
- At $(1, 1)$, we have $D = (6)(6) - (-3)^2 = 27$, and also $f_{xx} = 6x = 6 > 0$, so $(1, 1)$ is a local minimum.

Gradient descent

Constrained optimisiation.