MASD

Lecture 5
21.09.2021

Pawel Winter

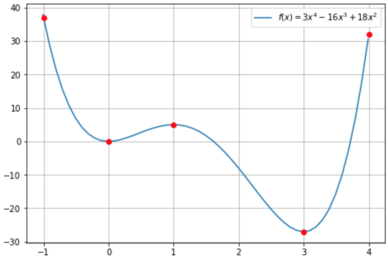
Objectives

- To give you an idea what optimization is and why it is important.
- To become familiar with definitions of minima, maxima and saddle points for functions $f : \mathbb{R}^d \to \mathbb{R}$ both for $d = 1$ and for $d \geq 2$.
- To get familiar with the notion of gradient and why it is important.
- To be able to understand gradient descent for finding minima of differentiable functions.
- To get familiar with the Newton's method
- Sections 4.1-4.3, 4.8, 14.6-14.7

Optimization

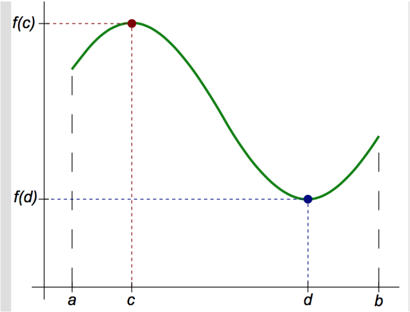
- Models of complex systems (e.g., transportation systems, nuclear power plants, economical models) involve large number of variables and parameters.
  - The performance of such systems can be specified by functions involving many variables and parameters.
  - Constraints can restrict or interrelate variables.
  - Optimization is about finding optimal or at least good solutions subject to the constraints.
  - Optimization can also be used to fix the parameters so the model mirrors the reality.
  - We will today look at how differentiation can be used to analyze the behavior of functions and how to find or approximate their optima.
  - Optimization is a very broad subject and today's lecture is just the tip of a huge iceberg.

Maxima and Minima



- Let $f : D \to \mathbb{R}$ be a one-variable function, and let $c \in D$.
- $f(c)$ is an *absolute* or *global* maximum of f on D if $f(c) \geq f(x)$ for all $x \in D$.
- $f(c)$ is a *local* maximum of f on D if $f(c) \geq f(x)$ when $x \in D$ and x is "near" c. How does one define "near"? Arbitrarily small open interval in D containing c.
- Absolute and local minima are defined analogously.

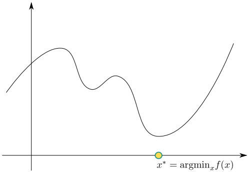
The Extreme Value Theorem (EVT)

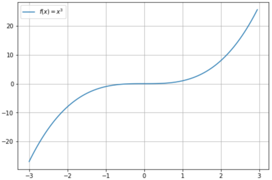


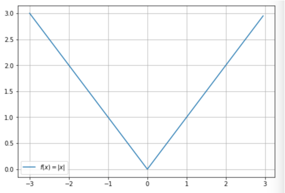
- Let $f : [a, b] \to \mathbb{R}$ be a continuous function on a closed interval $[a, b]$.
- f has a global maximum and a global minimum at some $c \in [a, b]$ and $d \in [a, b]$.
- Intuitively obvious but difficult to prove.
- Why do we require f to be continuous?
- What about $f :]a, b[\to \mathbb{R}$?

source: wikipedia

Min/Max Problems for Functions of One Variable







- Let $f \colon D \to \mathbb{R}$ be a function, $D \subseteq \mathbb{R}$
- How do we find x^* where $f(x^*)$ attains absolute or local optimum (minimum or maximum)? Solve $f'(x) = 0$?
- If $f'(a) = 0$ for some $a \in D$, then the tangent of f at a is horizontal.
- Does that give you an absolute or local optimum? **No**.
- There are functions with $f'(a) = 0$ for some $a \in D$ which are neither absolute nor local optimum.
- There are functions where $f'(a)$ does not exist but the absolute or local optimum is in a.

Fermat's Theorem (FT): If a function f has a local optimum at $c \in D$, and $f'(c)$ exists, then $f'(c) = 0$.

- PROOF: f has a local maximum at c. Then $f(c) \geq f(c+h)$ for every h sufficiently close to 0.
- $f(c) \geq f(c+h) \iff f(c+h) - f(c) \leq 0$
- If $h > 0$ then $\frac{f(c+h)-f(c)}{h} \leq 0$
- As $h \to 0^+$, we get

$$\lim_{h \to 0^+} \frac{f(c+h) - f(c)}{h} \leq \lim_{h \to 0^+} 0 = 0$$

- Since $f'(c)$ exists,

$$f'(c) = \lim_{h \to 0} \frac{f(c+h) - f(c)}{h} = \lim_{h \to 0^+} \frac{f(c+h) - f(c)}{h} \leq 0$$

Fermat's Theorem (FT) continued

- Next we show that $f'(c) \geq 0$. As before $f(c+h) - f(c) \leq 0$
- If $h < 0$ then $\frac{f(c+h)-f(c)}{h} \geq 0$
- As $h \to 0^-$, we get

$$\lim_{h \to 0^-} \frac{f(c+h) - f(c)}{h} \geq \lim_{h \to 0^-} 0 = 0$$

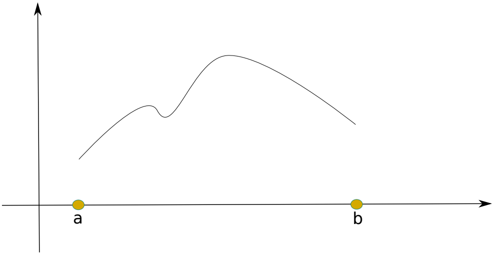
- Since $f'(c)$ exists,

$$f'(c) = \lim_{h \to 0} \frac{f(c+h) - f(c)}{h} = \lim_{h \to 0^-} \frac{f(c+h) - f(c)}{h} \geq 0$$

Critical Numbers

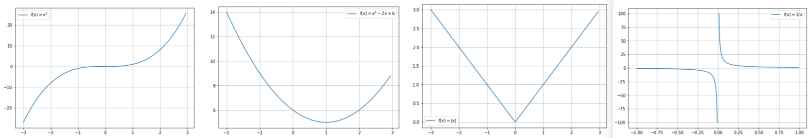
- Let $f : D \to \mathbb{R}$. A number $c \in D$ is *critical* if $f'(c)$ does not exist or $f'(c) = 0$.
- Note that $c \in D$ implies that $f(c)$ is defined.
- FT rephrased: If $f : D \to \mathbb{R}$ has a local optimum at $c \in D$, then c is a critical number.

Critical Numbers for $f : [a, b] \to \mathbb{R}$



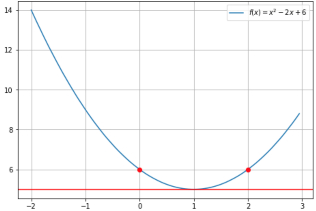
- Neither $f'(a)$ nor $f'(b)$ exist. Therefore a and b are critical numbers.
- There are three x-values with $f'(x) = 0$

Finding Critical Numbers



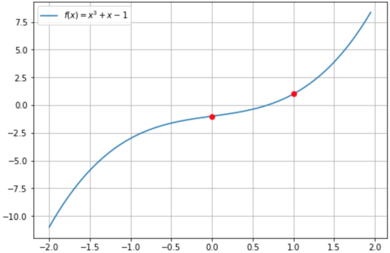
- $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^3$. $\frac{d}{dx}(x^3) = 3x^2$ and $f'(0) = 0$. Hence, 0 is the only critical number.
- $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2 - 2x + 6$. $\frac{d}{dx}(x^2 - 2x + 6) = 2x - 2$ and $f'(x) = 0$ only when $x = 1$. Hence, 1 is the only critical number.
- $f : \mathbb{R} \to \mathbb{R}$, $f(x) = |x|$. $f'(0)$ does not exist so $x = 0$ is a critical number. Since $f'(x) = -1$ for all $x < 0$ and $f'(x) = 1$ for $x > 0$, there is no other critical numbers.
- $f : \mathbb{R} \setminus \{0\} \to \mathbb{R}$, $f(x) = \frac{1}{x}$. $f'(x) = -\frac{1}{x^2}$. $f(x)$ is not defined for $x = 0$. So $x = 0$ is not a critical number. Since $f'(x) < 0$ for all $x \neq 0$, there are no other critical numbers.

Rolle's Theorem (RT): If $f : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$, differentiable on $]a, b[$ and $f(a) = f(b)$, then $\exists c \in]a, b[$ such that $f'(c) = 0$.



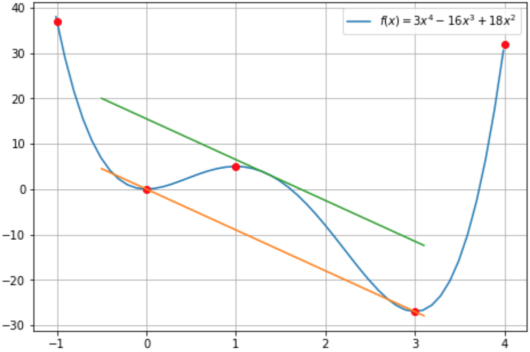
- PROOF: CASE I: f is a constant function. Then $f'(c) = 0$ for all $c \in]a, b[$.
- CASE II: $f(x) < f(a)$ for some $x \in]a, b[$. By EVT, every continuous function f attains minimum at some $c \in [a, b]$. Since $f(a) = f(b)$, c must be in $]a, b[$. Since f is differentiable on $]a, b[$, $f'(c)$ exists, and $f'(c) = 0$ by FT.
- CASE III: $f(x) > f(a)$ is similar to CASE II - omitted.

Applying RT: $f(x) = x^3 + x - 1$ has one real root



- $f(0) = -1$ and $f(1) = 1$.
- f is a polynomium and therefore continuous everywhere.
- By Intermediate Value Theorem (see 2.5.10), there must be $c \in]0, 1[$ such that $f(c) = 0$. c is the root of f.
- Suppose that f has at least 2 roots a and b. Hence, $f(a) = 0$ and $f(b) = 0$.
- f is a polynomium. Hence it is differentiable on $]a, b[$.
- RT implies that there is $c \in]a, b[$ such that $f'(c) = 0$.
- But $f'(x) = 3x^2 + 1 > 0$ for all x, a contradiction.

Mean Value Theorem (MVT): If $f : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$, differentiable on $]a, b[$, then $\exists c \in]a, b[$ such that $f'(c) = \frac{f(b)-f(a)}{b-a}$.



Proof of Mean Value Theorem (MVT): If $f : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$, differentiable on $]a, b[$, then $\exists c \in]a, b[$ such that $f'(c) = \frac{f(b)-f(a)}{b-a}$.

- Let $A = (a, f(a))$ and $B = (b, f(b))$. Consider the line $y - f(a) = \frac{f(b)-f(a)}{b-a}(x - a)$ through A and B.
- Let $h(x) = f(x) - f(a) - \frac{f(b)-f(a)}{b-a}(x - a)$
- $h(a) = f(a) - f(a) - \frac{f(b)-f(a)}{b-a}(a - a) = 0$
- $h(b) = f(b) - f(a) - \frac{f(b)-f(a)}{b-a}(b - a) = 0$
- h is continuous on $[a, b]$ since it is a difference of f and a first-degree polynomium, both continuous on $[a, b]$.
- h is differentiable on $]a, b[$ since f and the first-degree polynomials are differentiable on $]a, b[$.
- By RT, $\exists c \in]a, b[$ such that $h'(c) = 0$
- $0 = h'(c) = f'(c) - \frac{f(b)-f(a)}{b-a}$

Increasing Test: If $f : [a, b] \to \mathbb{R}$ is continuous on some interval $[a, b]$, differentiable on $]a, b[$, with $f'(x) > 0$ for all $x \in]a, b[$, then f is an increasing function in $]a, b[$.

- PROOF. We need to show that

$$x_1 < x_2 \implies f(x_1) < f(x_2)$$

  for all $x_1, x_2 \in]a, b[$.
- Let $x_1, x_2 \in]a, b[$, $x_1 < x_2$, be given.
- All assumptions of MVT are satisfied in $[x_1, x_2]$. Hence, there is $c \in]x_1, x_2[$ such that

$$0 < f'(c) = \frac{f(x_2) - f(x_1)}{x_2 - x_1}$$

- Since $x_2 > x_1$, it follows that $f(x_2) > f(x_1)$.

Decreasing Test: If $f : [a, b] \to \mathbb{R}$ is continuous on some interval $[a, b]$, differentiable on $]a, b[$, with $f'(x) < 0$ for all $x \in]a, b[$, then f is a decreasing function in $]a, b[$.

- PROOF. We need to show that

  $$x_1 < x_2 \implies f(x_1) > f(x_2)$$

  for all $x_1, x_2 \in]a, b[$.
- Let $x_1, x_2 \in]a, b[$, $x_1 < x_2$, be given.
- All assumptions of MVT are satisfied in $[x_1, x_2]$. Hence, there is $c \in]x_1, x_2[$ such that

  $$0 > f'(c) = \frac{f(x_2) - f(x_1)}{x_2 - x_1}$$

- Since $x_2 > x_1$, it follows that $f(x_2) < f(x_1)$.

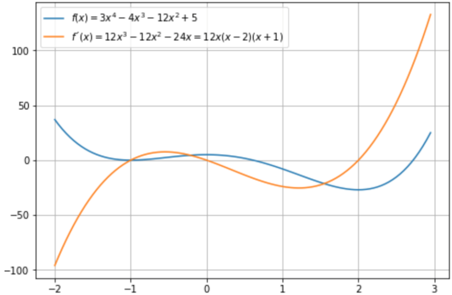
First Derivative Test

Let $f : [a, b] \to \mathbb{R}$ be a continuous function which is differentiable on $]a, b[$. Let $c \in]a, b[$

- If f' changes from positive to negative at c, then f has a local maximum at c.
- If f' changes from negative to positive at c, then f has a local minimum at c.
- If f' has the same sign on both sides of c, then f has neither local maximum nor local minimum at c.

First Derivative Test also works when f is continuous but not differentiable.

Applying 1-st Derivative Test: $f(x) = 3x^4 - 4x^3 - 12x^2 + 5$

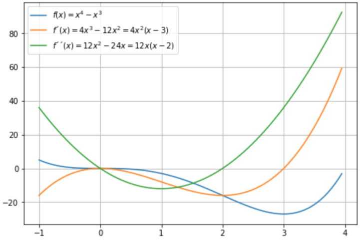


- $f'(x) = 12x^3 - 12x^2 - 24x = 12x(x^2 - x - 2) = 12x(x-2)(x+1)$
- $f'(-1) = f'(0) = f'(2) = 0$
- For $x \in]-\infty, -1[$, $f'(x) < 0$. f is decreasing in this interval.
- For $x \in]-1, 0[$, $f'(x) > 0$. f is increasing in this interval.
- For $x \in]0, 2[$, $f'(x) < 0$. f is decreasing in this interval.
- For $x \in]2, \infty[$, $f'(x) > 0$. f is increasing in this interval.
- f has local minimum at -1, local maximum at 0, local minimum at 2.

Second Derivative Test

- Suppose that f' is differentiable and f'' is continuous in some interval [a,b] containing c. Assume that $f'(c) = 0$.
- If $f''(c) > 0$, then f has a local minimum at c.
- If $f''(c) < 0$, then f has a local maximum at c.
- If $f''(c) = 0$ or $f''(c)$ does not exist, then inconclusive.
- Intuition: Tangent at $f(c)$ is horizontal.
  - $f''(c) > 0$ means that the slope of the tangent is increasing. We say that f is *concave upward*.
  - $f''(c) < 0$ means that the slope of the tangent is decreasing. We say that f is *concave downward*.

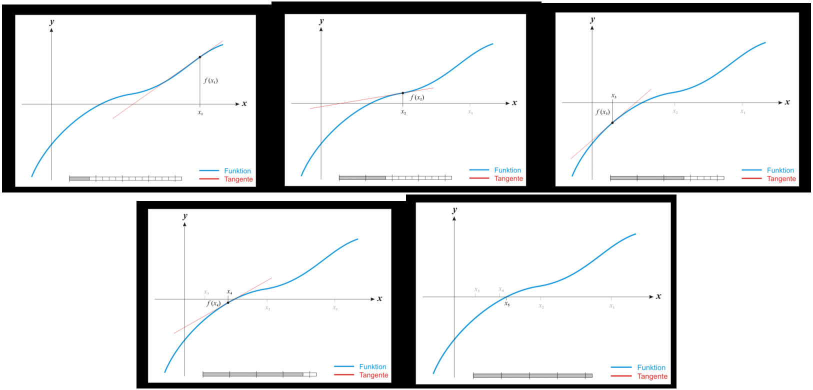
Applying 2-nd Derivative Test: $f(x) = x^4 - 4x^3$



- $f'(x) = 4x^3 - 12x^2 = 4x^2(x-3)$
- $f''(x) = 12x^2 - 24x = 12x(x-2)$
- f, f', f'' are polynomials $\implies$ continuous and differentiable everywhere. Critical numbers are only those satisfying $f'(c) = 0$.
- $f'(0) = f'(3) = 0$, $f''(0) = 0$, $f''(3) = 36$
- $f(3) = -27$ is a local minimum. Could also be decided by First Derivative Test.
- Since $f''(0) = 0$, nothing can be said about 0.

Newton's method

- Let $f(x) = 21x^{10} + x^7 - 3x^2 + 10$. To find roots, solve $f(x) = 0$.
- In calculus, Newton's method is an iterative procedure for finding roots of differentiable functions.
- Newton's method can be applied to f' of a twice-differentiable functions to find the roots of f' (critical numbers).
- $f'(x) = 210x^9 + 7x^6 - 6x$. To find critical numbers, solve $f'(x) = 0$.


y
x
f(x1)
x1
Funktion
Tangente
f(x2)
x2
x3
f(x3)
x4
f(x4)
x5

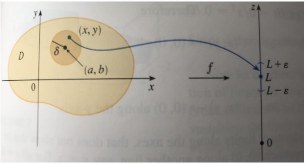
Functions with 2 Variables

- Let $D \subseteq \mathbb{R} \times \mathbb{R}$ and let $f : D \to \mathbb{R}$, $(a, b) \in D$ and $L \in \mathbb{R}$. We say that the limit of $f(x, y)$ as (x, y) approaches (a, b) is L and we write

$$\lim_{(x,y)\to(a,b)} f(x, y) = L$$

if for every $\epsilon > 0$, there is $\delta > 0$ such that

$$(x, y) \in D \wedge \sqrt{(x-a)^2 + (y-b)^2} < \delta \implies |f(x, y) - L| < \epsilon$$



- f is continuous at (a, b) if

$$\lim_{(x,y)\to(a,b)} f(x, y) = f(a, b)$$

Maximum Rate of Change

- Directional derivatives of f at a given point give the rates of change of f in all possible direction.
- What is the direction where f changes fastest (the slope of the corresponding tangent is steepest)?
- Let f be a differentiable function of two variables.
- **Claim**: The maximum value of the directional derivative $D_{\mathbf{u}}f(x,y)$ is $|\nabla f(x,y)|$ and is obtained in the direction of the gradient vector $\nabla f(x,y)$

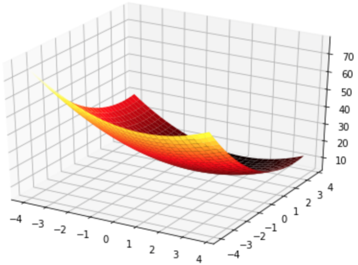
  $$D_{\mathbf{u}}f(x,y) = \nabla f(x,y) \cdot \mathbf{u} = |\nabla f(x,y)||\mathbf{u}|\cos(\theta) = |\nabla f(x,y)|\cos(\theta)$$

  where θ is the angle between $\nabla f(x,y)$ and $\mathbf{u}$. $\cos(\theta)$ is maximum for $\theta = 0$ (and $\cos(0) = 1$)
- Therefore the maximum value of $D_{\mathbf{u}}f(x,y)$ is $|\nabla f(x,y)|$ and occurs when $\mathbf{u}$ and $\nabla f(x,y)$ have the same direction.

Critical Points

- A function of two variables has a *local minimum* at (a, b) if $f(x, y) \geq f(a, b)$ when (x, y) is "near" (a, b). Definitions of *local maximum*, *global minimum* and *maximum* should be obvious.
- 1. Derivative Test: If f has a local minimum or maximum at (a, b) and first partial derivatives of f exist at (a, b), then $f_x(a, b) = f_y(a, b) = 0$. (Proof in the book). Note that this is equivalent to $\nabla f(x, y) = \mathbf{0}$.
- A point (a, b) is called a *critical point* of f if $f_x(a, b) = f_y(a, b) = 0$, or if one of these partial derivatives does not exist.
- Generalizes to functions with 3 or more variables.

$f(x, y) = x^2 + y^2 - 2x - 6y + 14$



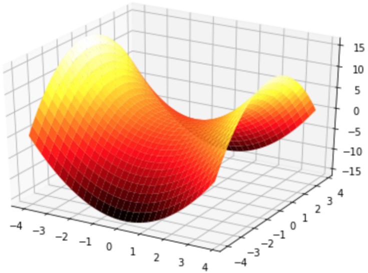
$$f_x(x, y) = 2x - 2 \qquad f_y(x, y) = 2y - 6$$

- These partial derivatives are 0 for $x = 1$ and $y = 3$.
- $(1, 3)$ is the only critical point and $f(1, 3) = 4$

$$f(x, y) = (x - 1)^2 + (y - 3)^2 + 4$$

- Hence $f(x, y) \geq 4$ for all $x \in \mathbb{R}$ and for all $y \in \mathbb{R}$.
- f has not only local but global minimum in $(1, 3)$

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



$$f_x(x, y) = -2x \qquad f_y(x, y) = 2y$$

- These partial derivatives are 0 for $x = 0$ and $y = 0$.
- Move along the x-axis with $y = 0$. Then $f(x, 0) \leq 0$ for all $x \in \mathbb{R}$. $(0, 0)$ cannot be a local minimum.
- Move along the y-axis with $x = 0$. Then $f(0, y) \geq 0$ for all $y \in \mathbb{R}$. $(0, 0)$ cannot be a local maximum.
- Such critical points are called *saddle points*.

Second Derivatives Test

- Suppose that second partial derivatives of some function f are continuous around some point (a, b), and $f_x(a, b) = f_y(a, b) = 0$. Let

$$D(a, b) = f_{xx}(a, b) f_{yy}(a, b) - f_{xy}(a, b)^2$$

- if $D(a, b) > 0$ and $f_{xx}(a, b) > 0$ then $f(a, b)$ is a local minimum.
- if $D(a, b) > 0$ and $f_{xx}(a, b) < 0$ then $f(a, b)$ is a local maximum.
- if $D(a, b) < 0$ then (a, b) is a saddle point.
- if $D(a, b) = 0$, the test is inconclusive.
- Generalizes to higher dimensions.

Does $f(x) = x^2 + y^2 - 3xy$ Have Local Minimum in $(0, 0)$?

- $f_x(x, y) = 2x - 3y$, $f_y(x, y) = 2y - 3x$
- $f_{xx}(x, y) = 2$, $f_{yy}(x, y) = 2$, $f_{xy}(x, y) = -3$
- Does f really have a minimum in $(0, 0)$?
- Let $x = y$. Then $f(x, x) = -x^2$

Multidimensional Second Derivative Test

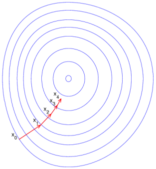
- Suppose that all partial derivatives exist at som $\mathbf{x_0}$ and $\nabla f(\mathbf{x}_0) = 0$. So $\mathbf{x}_0$ is a critical point.
- Compute the $d \times d$ *Hessian* at $\mathbf{x_0}$

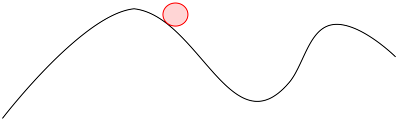
$$H_f = \begin{pmatrix} \frac{\partial^2 f}{\partial^2 x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_d} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial^2 x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_d} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_d \partial x_1} & \frac{\partial^2 f}{\partial x_d \partial x_2} & \cdots & \frac{\partial^2 f}{\partial^2 x_d} \end{pmatrix}$$

  and its eigenvalues.
- If the eigenvalues are all positive, then $\mathbf{x}_0$ is a local *minimum*
- If the eigenvalues are all negative, then $\mathbf{x}_0$ is a local *maximum*
- If the eigenvalues are all nonzero but mixed, then $\mathbf{x}_0$ is a *saddle point*
- Otherwise, the test is inconclusive.

Solving maximization problems numerically: Gradient ascent

- **Idea:** Walk uphill as long as you can. If there are several uphill directions, follow the steepest one.





- Where do I start?
  Common: Random initialization
- How long steps?
  Common: step size $\eta \|\nabla f(x)\|$, η (eta) is called the *learning rate*
- When have I reached a (local) top?
  Common: Threshold on $\sharp$ steps, function value difference or gradient magnitude

Min or max? Descent or ascent?

- If you want to find local maximum of a function (or its approximation), move in the direction of the gradient ∇f. We call this **gradient ascent**.
- If you want to find local minimum of a function (or its approximation), move in the direction opposite of the gradient, $-\nabla f$. We call this **gradient descent**.

Summary

- How to look for optima of functions with one or several variables
- Iterative methods to find optima of functions