

# Analiza matematyczna dla informatyków.

Mieczysław Cichoń, ver. 2.4/2021

**Mieczysław Cichoń - WMI UAM**

# Plan wykładów 9, 10, 11+12 ...

Rachunek różniczkowy funkcji jednej zmiennej rzeczywistej.

Pochodna i jej sens geometryczny.

Zastosowania w informatyce (m.in. podstawy interpolacji, funkcje spline).

Interpretacja geometryczna pochodnej. Liniowe przybliżanie funkcji (lokalne).

Podstawowe twierdzenia rachunku różniczkowego i wnioski z nich.

Metoda Newtona.

Rola wzoru Taylora w szacowaniu błędów.

Badanie przebiegu zmienności funkcji (na ćwiczeniach: obliczanie prostych pochodnych, sprawdzanie monotoniczności funkcji. ....)

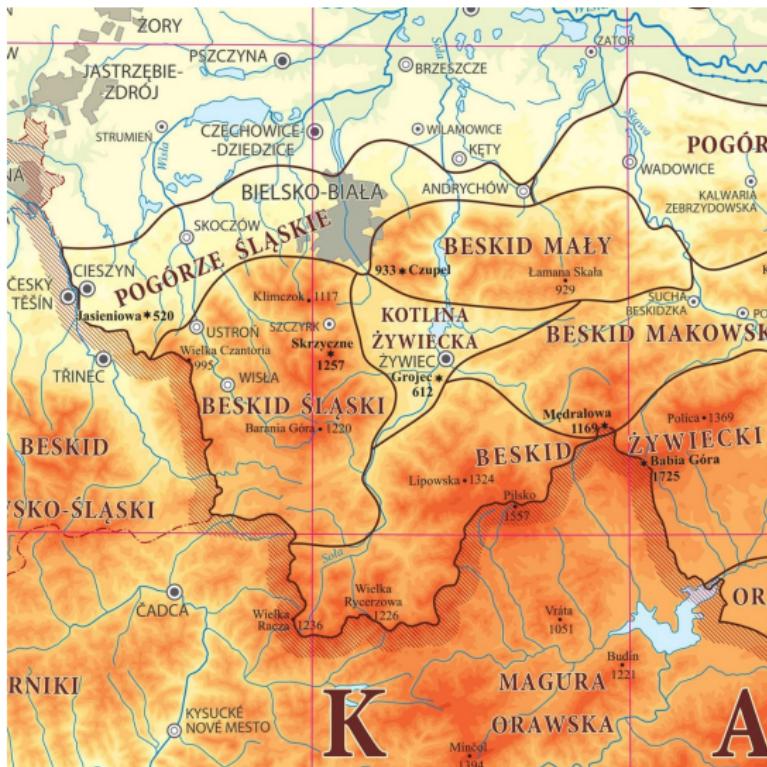
# Dla zainteresowanych.

Tak naprawdę, to plan tych wykładów jest tylko podstawą i w wielu przypadkach trzeba to uogólnić na tzw. *funkcje wielu zmiennych* - np. w nauczaniu maszynowym. Podstawowe idee z naszego wykładu pozostają bez zmian, ale może **warto** sprawdzić jak to się da poszerzyć do przypadku takich funkcji.

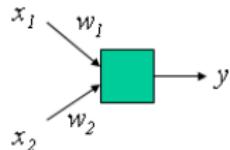
**Zainteresowani** mogą ode mnie otrzymać skrócone wersje takich materiałów z motywacjami w informatyce (będzie o nie - oczywiście - łatwiej niż dla funkcji jednej zmiennej).

Ale: podkreślam, że jednak podstawowe idee zastosowań będą widoczne już w naszym wykładzie...

Ale przecież funkcje wielu zmiennych były już w szkole np. ... na geografii. To prosty wykres funkcji wielu zmiennych...

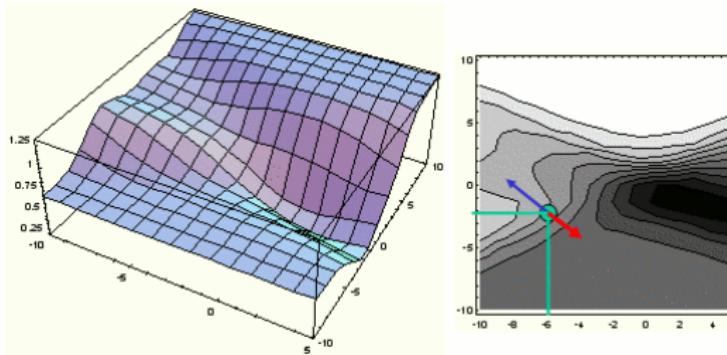
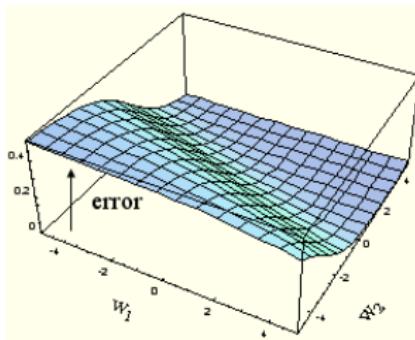


# A to już funkcje wielu zmiennych w sieciach neuronowych...



$$y = f\left(w_0 + \sum_{i=1}^n w_i x_i\right)$$

$$f(s) = \frac{1}{1 + e^{-s}}$$



## Strony do lektury na wykłady 9, 10, 11+12...

To 3 długie i ważne wykłady - naprawdę namawiam do przeczytania (w ciągu 3 tygodni!! - to 3,5 wykładu).

Czytamy najpierw motywacje:

[K] : motywacje - strony 31-32 oraz 35-36

teraz wstępne materiały

[K] : strony 219-255 - zwłaszcza paragraf 15.2 !!! Jest co czytać. Ale polecam lekturę "kolorowych" fragmentów. dla informatyków - BARDZO przydatne!!

oraz

[W] : strony 106-143

(lub alternatywnie: z tego wykładu strony 75-102).

# Pochodna funkcji w punkcie.

Rozpatrzmy funkcję  $f : A \rightarrow \mathbb{R}$ ,  $A \subset \mathbb{R}$ . Niech  $x_0 \in A$  będzie punktem skupienia zbioru  $A$ . Niech  $h \in \mathbb{R}$  będzie takie, że  $x_0 + h \in A$  (oraz dla dowolnego  $h_1 \leq h$ ,  $x_0 + h_1 \in A$ ).

Iloraz postaci  $\frac{f(x_0+h)-f(x_0)}{h}$  nazywamy ilorazem różnicowym funkcji  $f$  w punkcie  $x_0$ .

Można zauważyć, że jest to współczynnik kierunkowy tzw. prostej siecznej czyli prostej przechodzącej przez punkty  $(x_0, f(x_0))$  oraz  $(x_0 + h, f(x_0 + h))$ . Mogą istnieć różne przypadki, gdy obliczamy granicę

$$\lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h}.$$

Mamy kilka możliwości...

(1<sup>0</sup>) granica ta istnieje i jest liczbą skończoną. Np.  
 $f(x) = x^2$  ,  $x_0 \in \mathbb{R}$

$$\lim_{h \rightarrow 0} \frac{(x_0 + h)^2 - x_0^2}{h} = \lim_{h \rightarrow 0} \frac{x_0^2 + 2x_0 h + h^2 - x_0^2}{h} = \lim_{h \rightarrow 0} (2x_0 + h) = 2x_0$$

(2<sup>0</sup>) granica ta istnieje i jest równa  $\pm\infty$  (jest granicą niewłaściwą). Np.  $f(x) = \sqrt[3]{x}$  ,  $x_0 = 0$

$$\lim_{h \rightarrow 0} \frac{\sqrt[3]{h} - \sqrt[3]{0}}{h} = \lim_{h \rightarrow 0} h^{-\frac{2}{3}} = +\infty$$

(3<sup>0</sup>) granica ta nie istnieje, ale istnieją granice jednostronne (oczywiście w tym przypadku różne)

Np.  $f(x) = |x|$ ,  $x_0 = 0$

$$\lim_{h \rightarrow 0^+} \frac{|0+h| - |0|}{h} = \lim_{h \rightarrow 0^+} \frac{h}{h} = 1$$

$$\lim_{h \rightarrow 0^-} \frac{|0+h| - |0|}{h} = \lim_{h \rightarrow 0^-} \frac{-h}{h} = -1$$

(4<sup>0</sup>) nie istnieją nawet granice jednostronne ilorazu różnicowego funkcji  $f$  w punkcie  $x_0$ . Np.

$$f(x) = \begin{cases} x \cdot \sin \frac{1}{x} & , x \neq 0 , \\ 0 & , x = 0 , \end{cases}$$

$$\lim_{h \rightarrow 0^+} \frac{h \cdot \sin \frac{1}{h}}{h} = \lim_{h \rightarrow 0^+} \sin \frac{1}{h} \quad - \text{ nie istnieje}$$

(granica lewostronna - analogicznie).

## Definicja.

**Definicja.** Pochodną funkcji  $f$  w punkcie  $x_0$  nazywamy granicę (skończoną lub nieskończoną) ilorazu różnicowego funkcji  $f$  w punkcie  $x_0$ .

Sytuacja ta dotyczy więc przykładów z (1<sup>0</sup>) i (2<sup>0</sup>). W przypadku (3<sup>0</sup>) granice jednostronne nazywa się odpowiednio pochodną prawo- i lewostronną funkcji  $f$  w punkcie  $x_0$ .

Jak widać nawet funkcja ciągła może nie mieć pochodnej w punkcie ciągłości (nawet jednostronnej) - por. przykład (4<sup>0</sup>).

## Przykład obliczania pochodnej z definicji.

Niech  $f(x) = \cos x$  i  $x_0 \in \mathbb{R}$ .

$$\begin{aligned}\lim_{h \rightarrow 0} \frac{\cos(x_0 + h) - \cos(x_0)}{h} &= \lim_{h \rightarrow 0} \frac{-2 \sin\left(\frac{2x_0 + h}{2}\right) \cdot \sin\frac{h}{2}}{h} = \\ &= \lim_{h \rightarrow 0} \left[ -\sin\left(x_0 + \frac{h}{2}\right) \cdot \frac{\sin\frac{h}{2}}{\frac{h}{2}} \right] = -\sin x_0\end{aligned}$$

gdzie  $\lim_{h \rightarrow 0} \frac{\sin \frac{h}{2}}{\frac{h}{2}} = 1$ , a funkcja  $\sin x$  jest funkcją ciągłą.

Oczywiście można teraz obliczyć pochodne np. z funkcji  $g(x) = x + \cos x$  itd, ale obliczanie z definicji nie jest najlepszą z metod...

# Pochodne.

Najprostsze przykłady zastosowań:

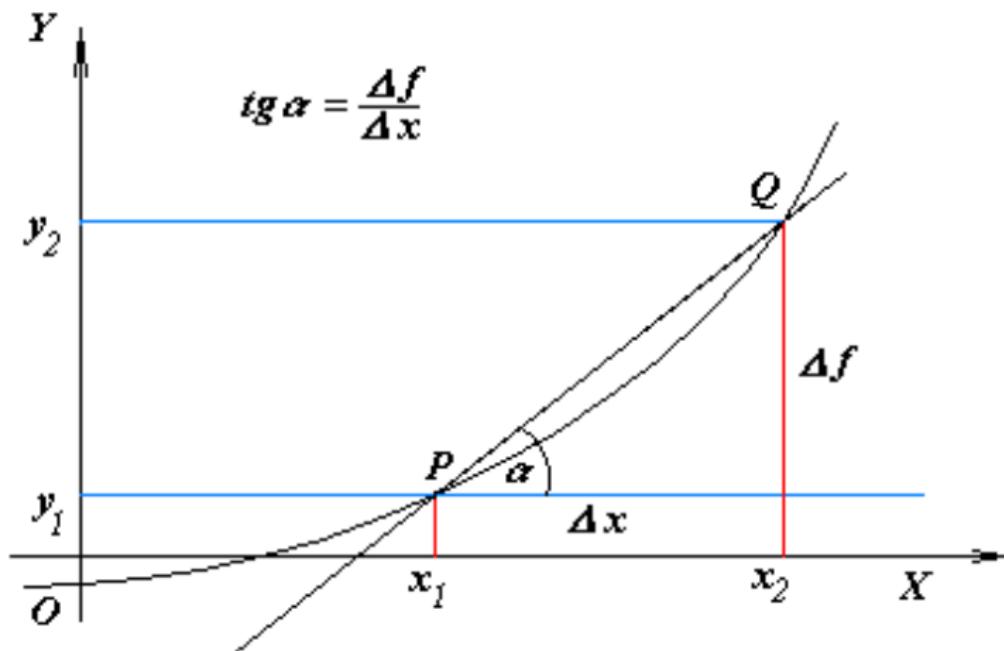
- ▶ metoda stycznych,
- ▶ szacowanie błędów wzorów interpolacyjnych (np. Lagrange'a),
- ▶ AI i automatyka : modelowanie dynamiki bardziej złożonych układów,
- ▶ przy korzystaniu z funkcji tworzących,
- ▶ grafika komputerowa i wizualizacja (w tym metody numeryczne),
- ▶ .....

# Styczna.

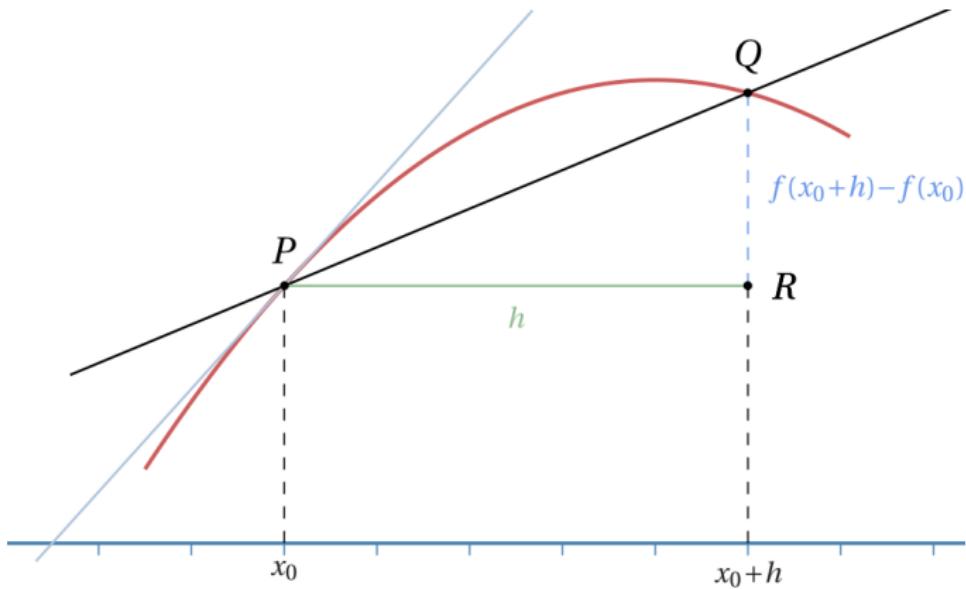
Prostą  $y = mx + n$  o współczynniku  $m = f'(x_0)$  przechodzącą przez punkt  $(x_0, f(x_0))$  nazywamy **styczną**, tak więc pochodna funkcji  $f$  w punkcie  $x_0$  to tangens kąta  $\alpha$  nachylenia stycznej do osi OX.

Uwaga: w tej definicji nie ma żadnego związku pomiędzy ilością punktów wspólnych stycznej i wykresu funkcji (tylko: punkt  $(x_0, f(x_0))$  musi być punktem wspólnym). Proszę nie czytać "internetowych" definicji...

Sieczna.



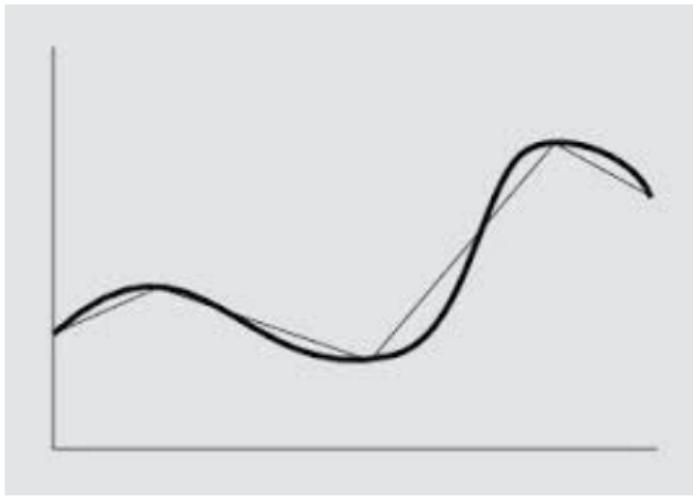
Styczna.



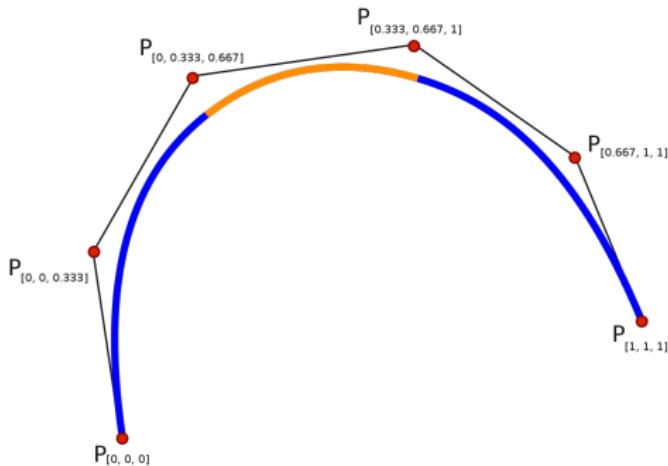
A tu prezentacja: [skrypt ilustracyjny w programie "Mathematica" \(.CDF\)](#).

## Styczne w informatyce...

Ponieważ istnienie stycznej do wykresu funkcji  $f$  w punkcie  $(x_0, f(x_0))$  jest w pewnym sensie równoważne istnieniu pochodnej funkcji  $f$  w  $x_0$ , to nie może być zaskoczeniem, że w wielu zastosowaniach pochodnej pojawi się styczna.



Klasyczny przypadek to przy obrazowaniu ruchu (np. gry, ale nie tylko) mamy skomplikowany kształt krzywej w ruchu. *W dużym uproszczeniu*: zastępujemy zamiany układu współrzędnych (co daje skomplikowane wzory na krzywą) przez skończone układy punktów na krzywej i styczne do tej krzywej. Powstaje wielokąt (wielościan). Teraz kontrolujemy w ruchu tylko skończone układy punktów (lepsze będą **funkcje spline** zamiast stycznych, ale to inna historia...):

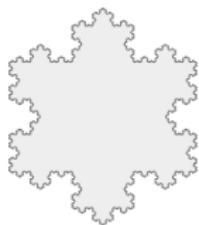


## Styczne w grafice.

To będzie możliwe, gdy krzywa ma styczne (najlepiej w każdym punkcie). A co jeśli nie? To pytanie: czy są krzywe nie posiadające w **żadnym** punkcie stycznych?

**Tak!** To np. wykresy funkcji singularnych (ciągły bez pochodnej w żadnym punkcie). Ogólniej to krzywe zadane parametrycznie, których funkcje parametryzujące są singularne.

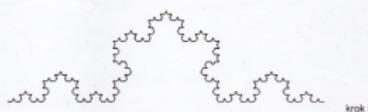
No to czas na fraktale!



Każdy punkt krzywej Kocha przypomina (w pewnym sensie) w charakterze punkt 0 na wykresie  $f(x) = |x|...$

Newton i Leibniz (niezależnie od siebie) tworzą **rachunek różniczkowy** – doskonałe narzędzie (miedzy innym) do badania przebiegu funkcji. W epoce, kiedy maszyny liczące nie posiadały możliwości wykonywania wielu obliczeń w ciągu utartka sekundy, można było stosunkowo łatwo (dzięki Newtonowi i Leibnizowi) przedstawić kształt nawet bardzo skomplikowanej funkcji, jeśli tylko w dostatecznie dużych obszarach była różniczkowalna. Różniczkowalność w interpretacji geometrycznej oznaczała posiadanie stycznej określonej jednoznacznie. Pod koniec XIX i na początku XX wieku przy próbach dokładnego zrozumienia pojęcia podstawowych (takich jak np. „ciągłość” czy „krzywa”) zauważono istnienie struktur, które obecnie nazywamy fraktalami. Wśród matematyków związanych z fraktalami możemy wymienić na przykład Georga Cantora (1872), Giuseppe Peana (1890), Davida Hilberta (1901), Helge'a von Kocha (1904), Wacława Sierpińskiego (1916), Gastona Julia (1918), czy Felixa Hausdorffa (1919). Helge von Koch był szwedzkim matematykiem, który w roku 1904 wprowadził krzywą nazwaną obecnie **Krzywą Kocha**. Linię ta w każdym punkcie jest różniczkowalna. Pojęcie nachylenia krzywej jest zgodne z intuicją i ma związek z pojęciem stycznej. Jeśli jednak krzywa ma zaklinań to pojawi się problem. Nie możemy dopasować jednoznacznie stycznej. Otoż krzywa Kocha jest przykładem krzywej, która w każdym punkcie ma zaklinanie, co prowadzi do tego, że **niemogliśmy jest określić stycznej w sposób jednoznaczny**.

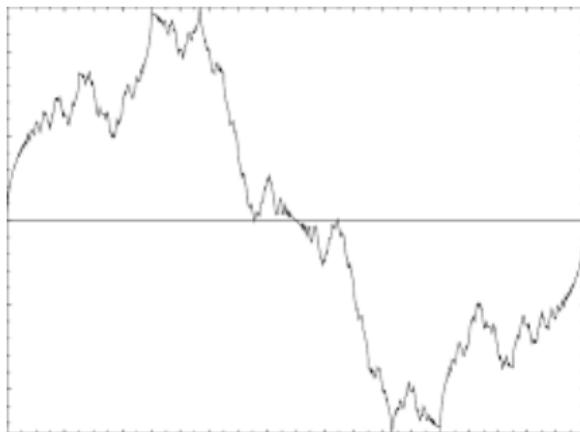
Konstrukcja krzywej Kocha przedstawia się następująco. Zaczyniemy od odcinka (obiekt ten nazwujemy inicjalnym), dzielimy na trzy równe części, a w miejsce środkowej wstawiamy trójkąt równoboczny i usuwamy jego podstawę. Kontrarymy w ten sposób podstawnego krok konstrukcji. Otrzymana figura nosi nazwę generatora. Powtarzamy konstrukcję w ten sposób, że w miejscu każdego odcinka wstawiamy odpowiednio zmniejszony generator. Kilka początkowych kroków przedstawiają rysunki:



# Pochodne i fraktale...

**Brak stycznych** w przypadku krzywych fraktalnych powoduje, że w informatyce (o ile są potrzebne) nie można ich badać, czy manipulować nimi za pomocą rachunku różniczkowego.

Każda krzywa fraktalna ma własność braku stycznych, ale nie odwrotnie - poniżej wykres funkcji Riemanna (ale nie fraktal):



# Prezentacja kilku nieróżniczkowalnych funkcji.

Proszę w skrypcie wybrać po kolej funkcje, zwłaszcza funkcję Weirstrassa (i użyć "zoom"):

[Skrypt ilustracyjny granic ciągów w "Mathematica"](#) -  
potrzebny darmowy *CDF Player* lub *Mathematica*

# Różniczkowalność.

Odróżnimy istnienie pochodnej (skończonej lub nie) od różniczkowalności:

**Definicja.** Mówimy, że funkcja  $f : A \rightarrow \mathbb{R}$  ( $A \subset \mathbb{R}$ ) jest w punkcie  $x_0 \in A$  (punkcie skupienia zbioru  $A$ ) różniczkowalna, jeśli posiada w tym punkcie skończoną pochodną  $f'(x_0)$ .

**Definicja.** Mówimy, że funkcja  $f : A \rightarrow \mathbb{R}$  jest różniczkowalna w zbiorze  $A$  jeżeli jest różniczkowalna w każdym punkcie zbioru  $A$ .

Funkcja różniczkowalna jest lokalnie przybliżana funkcją liniową. Pokażemy, że to różniczka funkcji w punkcie jest tym odwzorowaniem liniowym.

**Twierdzenie.** Jeżeli  $f$  jest różniczkowalna w punkcie  $x_0$  to dla dostatecznie małych  $h$  prawdziwy jest wzór

$$f(x_0 + h) - f(x_0) = f'(x_0) \cdot h + h \cdot \varphi(h)$$

przy czym  $\lim_{h \rightarrow 0} \varphi(h) = 0$ . (czyli "o małe" od  $h$ )

$f(x_0 + h) - f(x_0)$  nazywamy przyrostem funkcji  $f$  w punkcie  $x_0$  o kroku  $h$ , zaś iloczyn  $f'(x_0) \cdot h$  różniczką z funkcji  $f$  w punkcie  $x_0$  na przyroście  $h$  i oznaczamy odpowiednio  $\Delta f(x_0, h)$  i  $df(x_0, h)$ .

Teraz rozpatrzymy sytuację odwrotną.

**Twierdzenie.** Niech  $f : A \subset \mathbb{R} \longrightarrow \mathbb{R}$ , i  $x_0 \in A$  będzie punktem skupienia zbioru  $A$ . Jeżeli przyrost funkcji  $\Delta f(x_0, h)$  daje się zapisać w postaci:

$$\Delta f(x_0, h) = I \cdot h + r(x_0, h) \quad (\Delta)$$

gdzie  $I \in \mathbb{R}$ , funkcja  $r(x_0, h)$  jest określona dla tych  $h \in \mathbb{R}$ , dla których  $x_0 + h \in A$  oraz  $\lim_{h \rightarrow 0} \frac{r(x_0, h)}{h} = 0$  to  $f$  jest różniczkowalna w punkcie  $x_0$  oraz

$$f'(x_0) = I .$$

## Różniczki zupełne.

Przez  $df(x_0, h)$  oznaczać będziemy iloczyn  $f'(x_0) \cdot h$  i nazywać będziemy różniczką z funkcji  $f$  w punkcie  $x_0$  na przyrostie  $h$ . Mamy więc dla funkcji różniczkowalnej

$$f(x_0 + h) - f(x_0) = df(x_0, h) + r(x_0, h)$$

gdzie  $\lim_{h \rightarrow 0} \frac{r(x_0, h)}{h} = 0$ .

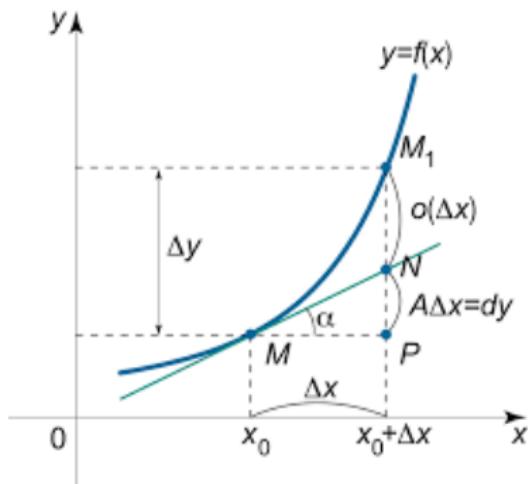
### Interpretacja geometryczna różniczki.

Zauważmy, że jeśli  $f$  jest różniczkowalna w  $x_0$  to dla małych przyrostów  $h$  reszta  $r(x_0, h)$  jest mała, a więc

$$f(x_0 + h) - f(x_0) \approx df(x_0, h) .$$

Zazwyczaj łatwiej jest obliczyć różniczkę funkcji w punkcie, niż jej przyrost - wykorzystamy to.

# Pochodna a różniczka.



Na rysunku: odcinek  $PM_1$  reprezentuje przyrost funkcji  $f$  w punkcie  $x_0$ :  $f(x_0 + h) - f(x_0)$ . Odcinek  $PN$  reprezentuje różniczkę  $df(x_0, h)$ , a odcinek  $NM_1$  resztę  $r(x_0, h)$ .

Widoczne:  $f(x_0 + h) - f(x_0) = df(x_0, h) + r(x_0, h)$  oraz szybki spadek długości odcinków pomiędzy styczną a krzywą  $o(\Delta x)$ !

# Aproksymacja liniowa.

## 15.2.1 Aproksymacja liniowa

Interpretacja pochodnej jako współczynnika kierunkowego nachylenia stycznej w danym punkcie do wykresu nie jest jedyną użyteczną intuicją w tym kontekście. Pojęcie pochodnej wiąże się również z ideą przybliżania funkcji funkcją liniową.

**Definicja 15.2.1** *Funkcja liniowa*  $\mathbb{R} \rightarrow \mathbb{R}^n$  jednej zmiennej to funkcja postaci

$$\mathbb{R} \ni t \mapsto tv \in \mathbb{R}^n$$

dla ustalonego  $v \in \mathbb{R}^n$ . Wektor  $v$  (liczbę w przypadku  $n = 1$ ) nazywać będziemy *współczynnikiem* funkcji liniowej jednej zmiennej.

Funkcję taką oznaczać będziemy  $L_v$ , lub krótko  $L$ , jeśli współczynnik  $v$  jest nieistotny. Latwo sprawdzić, że funkcja liniowa jest ciągła i w zerze ma wartość zero.

**Uwaga 15.2.2** Jeśli  $L : \mathbb{R} \rightarrow \mathbb{R}^m$  jest funkcją liniową, to  $L = L_v$  dla  $v = L(1)$ . Innymi słowy wartość funkcji liniowej dla argumentu jeden jest równa jej współczynnikowi.

Funkcja liniowa daje się bardzo prosto wyliczać, więc jest dobrym kandydatem na przybliżanie (lokalne) innych funkcji. Idea jest następująca. Wiemy, że pewna funkcja  $f$  ma w punkcie  $x_0$  wartość  $y_0 := f(x_0)$ . Interesuje nas wartość funkcji  $f$  w punkcie  $x_1$  bliskim  $x_0$ , czyli  $y_1 := f(x_1)$ . Czy dysponując przyrostem zmiennej  $x$ , czyli różnicą  $x_1 - x_0$ , możemy przy pomocy pewnej funkcji liniowej wyliczyć przyrost zmiennej  $y$ , czyli różnicę  $y_1 - y_0 = f(x_1) - f(x_0)$ ? Podkreślimy, że chodzi nam o funkcję liniową w układzie współrzędnych przesuniętym do punktu  $(x_0, y_0)$ , a więc w układzie  $(\bar{x}, \bar{y})$ , gdzie  $\bar{x} := x - x_0$ ,  $\bar{y} := y - y_0$ , zwanym lokalnym układem współrzędnych względem punktu  $(x_0, y_0)$ . Zwrócić uwagę, że punkt o współrzędnych  $(x_0, y_0)$  w lokalnym układzie współrzędnych ma współrzędne  $(0, 0)$ . Nasze pytanie można więc sformułować następująco (rys. 15.2):

Czy istnieje funkcja liniowa  $L$  w lokalnym układzie współrzędnych  $(\bar{x}, \bar{y})$ , taka że

$$y_1 - y_0 \approx \bar{y} = L(\bar{x}) = L(x_1 - x_0)?$$

Zwróćmy uwagę, że to pytanie należy uściślić. Jeśli przybliżenie rozumieć będziemy jako wymóg, by w granicy, przy  $x \rightarrow x_0$ , popełniony błąd, czyli wyrażenie

$$y_1 - y_0 - L(x_1 - x_0) = f(x_1) - f(x_0) - L(x_1 - x_0)$$

zmierzało do zera, to wymóg ten spełni każda funkcja liniowa, bo jako ciągła ma w zerze granicę równą wartości, czyli zero. Naturalne jest jednak oczekiwanie, że im bliżej znajdziemy się  $x_0$ , tym przybliżenie będzie lepsze.

Miarą jakości przybliżenia jest błąd względem przyrostu  $\bar{x} = x_1 - x_0$  mierzony ilorazem

$$\frac{y_1 - y_0 - L(x_1 - x_0)}{x_1 - x_0} = \frac{f(x_0 + \bar{x}) - f(x_0) - L(\bar{x})}{\bar{x}}.$$

Chcemy, by ten błąd zmierzał do zera przy  $\bar{x} \rightarrow 0$ . Niech  $L = L_v$ . Przy użyciu bardziej tradycyjnego oznaczenia  $h$  w miejscu  $\bar{x}$ , nasz wymóg można zapisać jako

$$\lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0) - hv}{h} = 0,$$

co jak łatwo sprawdzić, sprowadza się do żądania, by

$$\lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h} = v.$$

Tak więc, jeśli stosowne przybliżenie liniowe istnieje, to współczynnik  $v$  jest pochodną funkcji w punkcie  $x_0$ .

# I pierwszy symbol Landaua...

**Definicja 15.2.3** Mówimy, że funkcja  $f$  jest *o male* od funkcji  $g$  w otoczeniu punktu  $x_0$  i piszemy

$$f(x) = o(g(x)) \text{ przy } x \rightarrow x_0,$$

jeżeli

$$\forall \epsilon > 0 \exists \delta > 0 \quad |x - x_0| < \delta \Rightarrow ||f(x)|| < \epsilon |g(x)|.$$

Następującą uwagę można wywnioskować wprost z definicji granicy.

**Uwaga 15.2.4** Relacja  $f(x) = o(g(x))$  przy  $x \rightarrow x_0$  zachodzi wtedy i tylko wtedy gdy

$$f(x_0) = 0 \text{ i } \lim_{x \rightarrow x_0} \frac{||f(x)||}{|g(x)|} = 0.$$

□

Wróćmy teraz do idei przybliżania przyrostu funkcji funkcją liniową. Wprowadźmy tradycyjne oznaczenia na przyrost wartości argumentu i przyrost wartości funkcji (różnice) w punkcie  $x_0$ :

$$\begin{aligned}\Delta x &:= x - x_0, \\ \Delta y &:= f(x) - f(x_0).\end{aligned}$$

Niech  $L = L_{f'(x_0)}$  będzie funkcją liniową przybliżającą w lokalnym układzie współrzędnych przyrost funkcji  $f$ . Wtedy

$$\Delta y \approx L(\Delta x) = f'(x_0)\Delta x, \tag{15.5}$$

przy czym popelniony błąd względny jest tym mniejszy im mniejsze jest  $\Delta x$ . Leibniz nie dysponował precyzyjnym aparatem i mówił, że jeśli  $\Delta x$  jest nieskończenie małe, to przybliżona równość (15.5) staje się równością. Takie nieskończenie małe różnice  $\Delta x$  i  $\Delta y$  nazywały różniczkami (malusinkimi różnicami) i oznaczały  $dy$  i  $dx$ . Używając jego języka napisalibyśmy, że różniczki  $dx$  i  $dy$  spełniają równanie

$$dy = f'(x)dx. \tag{15.6}$$

Problem w tym, że różniczki jako wielkości nieskończenie małe trudno jest precyzyjnie zdefiniować. Pojęcie różniczki odczarował A. Cauchy, traktując równanie (15.6) jako równanie odwzorowania liniowego przybliżającego przyrost funkcji, a samą różniczkę definiując jako to przybliżające odwzorowanie liniowe. I tak do dziś rozumiemy różniczkę. Natomiast termin różniczkowalność, ze względu na bezpośredni związek pomiędzy różniczką i pochodną, jest odnoszony zarówno do istnienia pochodnej jak i różniczki, bo, jak zobaczymy, są to fakty równoważne.

## Przykład 1.

Obliczyć wartość wyrażenia  $e^{0,03}$ .

Rozpatrzmy funkcję  $f(x) = e^x$ , wówczas jest to wartość tej funkcji w punkcie  $x = 0,03$  tj.  $f(0,03)$ . Nie jest to łatwe, ale łatwo jest obliczyć wartość  $f$  w „bliskim” punkcie  $x_0 = 0$ . Mamy więc

$$f(x_0 + h) \approx f(x_0) + df(x_0, h)$$

gdzie  $h = 0,03$  ,  $x_0 = 0$ . Ponieważ  $f'(x) = e^x$  (co pokażemy za chwilę), więc  $f'(x_0) = e^0 = 1$ , czyli

$$f(0,03) \approx 1 + 1 \cdot 0,03 = 1,03 .$$

## Przykład 2.

Obliczyć wartość wyrażenia  $A = \sin 40^0$ .

Wykorzystamy miarę łukową kąta, a więc

$$A = \sin \frac{40 \cdot \pi}{180} = \sin \frac{2\pi}{9} .$$

Niech

$$f(x) = \sin x, \quad x_0 = \frac{\pi}{4}, \quad h = \frac{2\pi}{9} - \frac{\pi}{4} = -\frac{\pi}{36} \approx -0,09 .$$

Obliczamy (skorzystamy z pochodnych funkcji sinus - co można sprawdzić z definicji) :

$$f(x_0) = \sin \frac{\pi}{4} = \frac{\sqrt{2}}{2} \approx 0,71 \quad f'(x) = \cos x, \quad \cos \frac{\pi}{4} = \frac{\sqrt{2}}{2} \approx 0,71$$

a więc

$$A = f(x_0 + h) \approx f(x_0) + df(x_0, h) = 0,71 + (0,71) \cdot (-0,09) \approx 0,65.$$

(tu wszystkie obliczenia prowadzimy z dokładnością 2 miejsc po przecinku).

# Różniczkowalność a ciągłość.

**Twierdzenie.** Jeżeli  $f$  jest różniczkowalna w punkcie  $x_0$ , to jest w tym punkcie ciągła.

Jak już wspomnieliśmy ciągłość nie jest wystarczająca nawet do istnienia pochodnych jednostronnych...

# Funkcja pochodna.

A teraz pojęcie funkcji pochodnej - funkcję przyporządkowującą punktom  $x_0 \in A^d$  liczby  $f'(x_0)$  nazywamy **funkcją pochodną** funkcji  $f$ . Funkcję tę oznaczać będziemy przez  $f'$ .

Na ogół dziedzina funkcji  $f'$  jest zawarta w dziedzinie funkcji  $f$ .

Można wyprowadzić wzory na najważniejsze z funkcji (m.in. elementarne).

# Tabela pochodnych.

Dziedzina $f$	Funkcja $f$	Dziedzina $f'$	Funkcja $f'$
$\mathbb{R}$	$x^\alpha$	$\mathbb{R}$	$\alpha \cdot x^{\alpha-1}$
$\mathbb{R}$	$a$	$\mathbb{R}$	0
$\mathbb{R}$	$\sin x$	$\mathbb{R}$	$\cos x$
$\mathbb{R}$	$\cos x$	$\mathbb{R}$	$-\sin x$
$\mathbb{R}$	$e^x$	$\mathbb{R}$	$e^x$
$(0, +\infty)$	$\ln x$	$(0, +\infty)$	$\frac{1}{x}$
$\mathbb{R}$	$a^x (a > 0)$	$\mathbb{R}$	$a^x \cdot \ln a$
$< -1, 1 >$	$\arcsin x$	$(-1, 1)$	$\frac{1}{\sqrt{1-x^2}}$
$\mathbb{R}$	$\arctan x$	$\mathbb{R}$	$\frac{1}{1+x^2}$
$< -1, 1 >$	$\arccos x$	$(-1, 1)$	$\frac{-1}{\sqrt{1-x^2}}$
...	...	...	...

Prezentacja: Skrypt ilustracyjny możliwości obliczeniowych pochodnych programu "Mathematica".

Pochodne można obliczać korzystając z bardziej ogólnych reguł niż definicja...

**Twierdzenie.** *Dane są 2 funkcje  $f, g : A \rightarrow \mathbb{R}$ . Założymy, że  $f$  i  $g$  są różniczkowalne w ustalonym punkcie  $x \in A$ . Wówczas w tym samym punkcie  $x \in A$  różniczkowalne są funkcje  $f + g$ ,  $f - g$ ,  $f \cdot g$ ,  $\lambda \cdot g$  ( $\lambda \in \mathbb{R}$ ), a jeśli ponadto  $g(x) \neq 0$  to także  $\frac{f}{g}$ .*

Prawdziwe są ponadto wzory:

$$(a) \quad (f + g)'(x) = f'(x) + g'(x),$$

$$(b) \quad (f - g)'(x) = f'(x) - g'(x),$$

$$(c) \quad (f \cdot g)'(x) = f'(x) \cdot g(x) + f(x) \cdot g'(x),$$

$$(d) \quad (\lambda \cdot g)'(x) = \lambda \cdot g'(x),$$

$$(e) \quad \left(\frac{f}{g}\right)'(x) = \frac{f'(x) \cdot g(x)' - f(x) \cdot g'(x)}{g^2(x)} .$$

**Twierdzenie.** (pochodna superpozycji) Niech  $f : A \rightarrow B \subset \mathbb{R}$ ,  $g : B \rightarrow \mathbb{R}$ . Jeżeli  $f$  jest różniczkowalna w punkcie  $x \in A$ , a  $g$  jest różniczkowalna w  $y = f(x) \in B$  to funkcja złożona (superpozycja)  $g \circ f : A \rightarrow \mathbb{R}$  dana wzorem  $(g \circ f)(x) = g(f(x))$  jest różniczkowalna w punkcie  $x \in A$ , oraz prawdziwy jest wzór

$$(g \circ f)'(x) = g'(y) \cdot f'(x) \quad (\text{gdzie } y = f(x)).$$

**Twierdzenie.** (pochodna funkcji odwrotnej) Niech  $g$  będzie funkcją ciągłą i ściśle monotoniczną w przestrzeni  $P$ . Przez  $f$  oznaczmy funkcję odwrotną do  $g$ . Jeżeli funkcja  $g$  jest różniczkowalna w punkcie  $y \in P$  oraz pochodna tej funkcji  $g'(y) \neq 0$  to  $f$  jest różniczkowalna w punkcie  $x = g(y)$  oraz

$$f'(x) = \frac{1}{g'(y)} \quad (\text{czyli} \quad (g^{-1})'(x) = \frac{1}{g'(y)}).$$

## Przykłady.

(1) Niech  $f(x) = x^3 + e^{\sin x}$ ,  $x \in \mathbb{R}$ .

Wówczas  $f(x) = f_1(x) + f_2(x)$ , gdzie  $f_1(x) = x^3$ ,  $f_2(x) = e^{\sin x}$ .  
Stąd  $(f_1)'(x) = 3x^2$ , zaś  $f_2$  jest funkcją złożoną

$$g(u) = e^u, \quad h(v) = \sin v.$$

Z działań na pochodnych mamy  $(f_2)'(x) = \cos x \cdot e^{\sin x}$ .  
Ostatecznie  $f'(x) = 3x^2 + \cos x \cdot e^{\sin x}$ .

(2) Weźmy teraz  $f(x) = \ln x$ . Jest to funkcja odwrotna do  
 $g(y) = e^y$

$$(x = e^y) \iff (y = \ln x)$$

$$g : \mathbb{R} \xrightarrow{\text{na}} (0, +\infty) \quad \text{oraz} \quad f(0, +\infty) \xrightarrow{\text{na}} \mathbb{R}$$

$g$  - ścisłe rosnąca na  $P = (-\infty, \infty)$  (oraz ciągła), co więcej

$$g'(y) = e^y \neq 0 \quad \text{dla dowolnych } y \in \mathbb{R}.$$

Mamy więc

$$(x = e^y) \quad f'(x) = \frac{1}{g'(y)} = \frac{1}{e^y} = \frac{1}{e^{\ln x}} = \frac{1}{x}$$

(kolejny wynik do tabeli pochodnych ...).

# Ekstrema lokalne.

Mówimy, że funkcja  $f : A \rightarrow \mathbb{R}$  ma w tym punkcie maksimum lokalne ( $x_0$  jest punktem wewnętrznym zbioru  $A$ ) jeżeli istnieje  $\delta > 0$  taka, że dla każdego  $x \in (x_0 - \delta, x_0 + \delta)$  zachodzi nierówność

$$f(x) \leq f(x_0) .$$

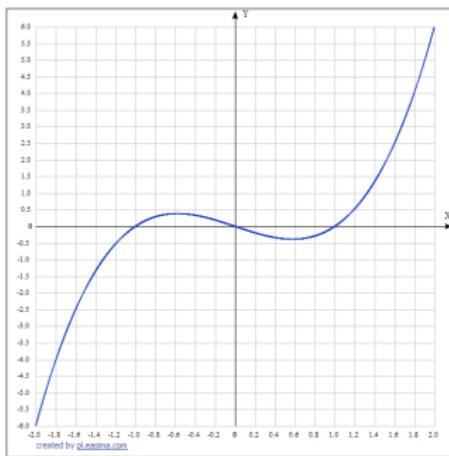
Funkcja ta ma minimum lokalne w  $x_0$  jeżeli

$$\exists_{\delta>0} \quad \forall_{x \in (x_0 - \delta, x_0 + \delta)} \quad f(x) \geq f(x_0) .$$

Minima i maksima lokalne noszą łączną nazwę ekstremów lokalnych.

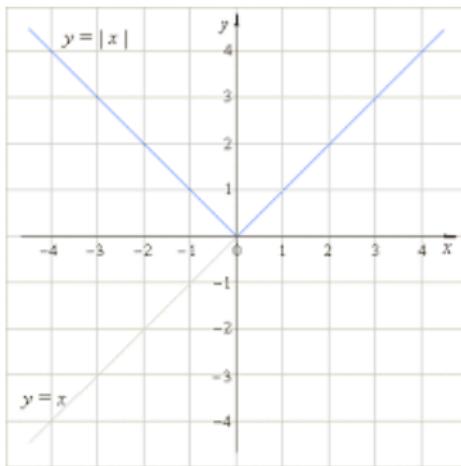
$$f(x) = x \cdot (x - 1) \cdot (x + 1)$$

funkcja ma w punkcie  $x_0 = -\frac{\sqrt{3}}{3}$  maksimum lokalne, a w punkcie  $x_1 = \frac{\sqrt{3}}{3}$  minimum lokalne, ale funkcja ta jest nieograniczona z góry i z dołu).

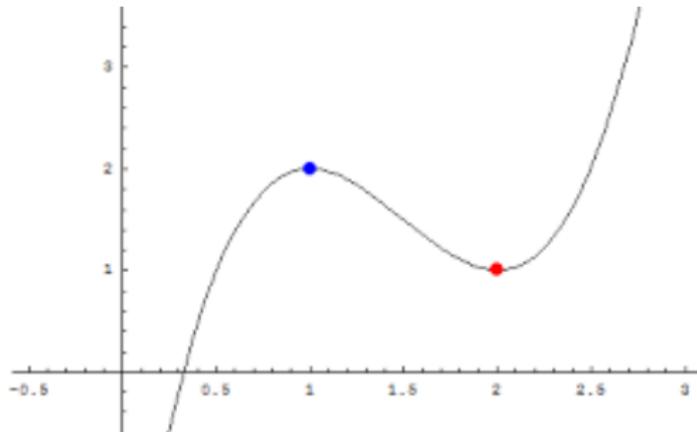


Rozpatrzmy kolejny przykład:  $f(x) = |x|$ .

Łatwo zauważyć, że w punkcie  $x_0 = 0$  jest minimum lokalne ( $f(x_0) = 0$  a dla każdego innego  $x \in \mathbb{R}$   $f(x) > 0$  - jest to tzw. minimum globalne). Niemniej nie istnieje (jak już wiemy ...)  $f'(x_0)$ .



# Ekstrema lokalne.



W maksimum lokalnym nie musi być wartością największą funkcji (niebieski punkt na wykresie), a w minimum lokalnym nie musi być wartością najmniejszą funkcji (czerwony punkt na wykresie).

# Zasadnicze twierdzenia rachunku różniczkowego.

**Twierdzenie.** (tw. Fermata). Jeżeli  $f : A \rightarrow \mathbb{R}$  jest różniczkowalna w  $x_0$  i ma w nim ekstremum lokalne, to  $f'(x_0) = 0$ .

Odwrotna zależność (jak już wiemy) nie zachodzi ...

Nie można stąd oczywiście wyciągać wniosków, że warunek  $f'(x_0) = 0$  jest warunkiem koniecznym istnienia ekstremum.

Dla sformułowania warunków wystarczających istnienia ekstremum lokalnego musimy wprowadzić nowe pojęcia i zbadać je. Do tematu oczywiście wróćmy...

A tu prezentacja: [twierdzenia Fermata w formacie .CDF](#).

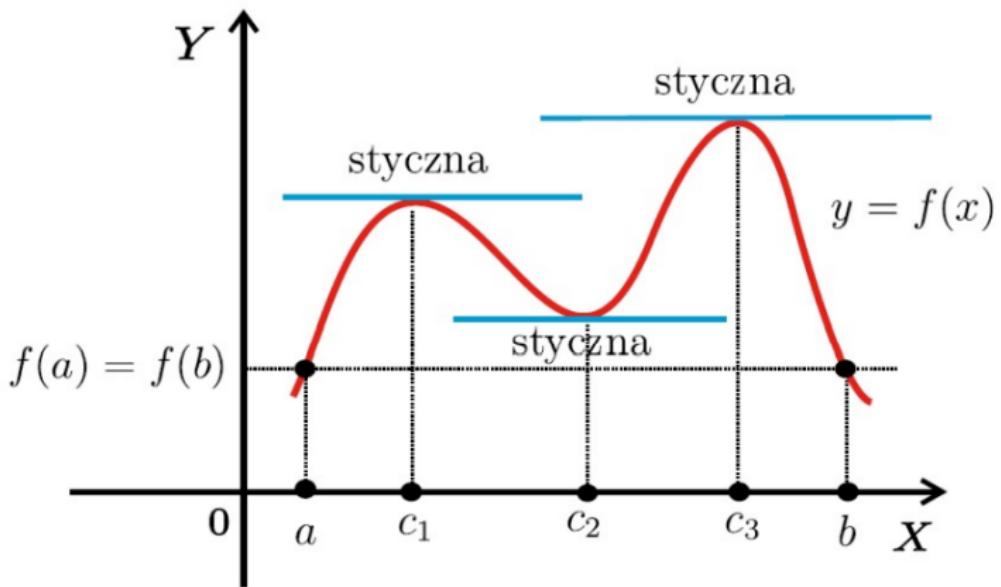
## Twierdzenie Rolle'a.

**Twierdzenie.** (Rolle'a). Założmy, że funkcja  $f$  określona w przedziale  $\langle a, b \rangle$  spełnia następujące warunki:

- (a)  $f$  jest ciągła w  $\langle a, b \rangle$ ,
- (b)  $f$  jest różniczkowalna w  $(a, b)$ ,
- (c)  $f(a) = f(b)$ .

Wówczas istnieje taki punkt  $\xi \in (a, b)$ , że  $f'(\xi) = 0$ .

Sens geometryczny twierdzenia Rolle'a: na łuku krzywej będącej wykresem funkcji  $f$ , którego końce mają tę samą rzędną znajduje się co najmniej 1 punkt, w którym styczna jest równoległa do osi OX.



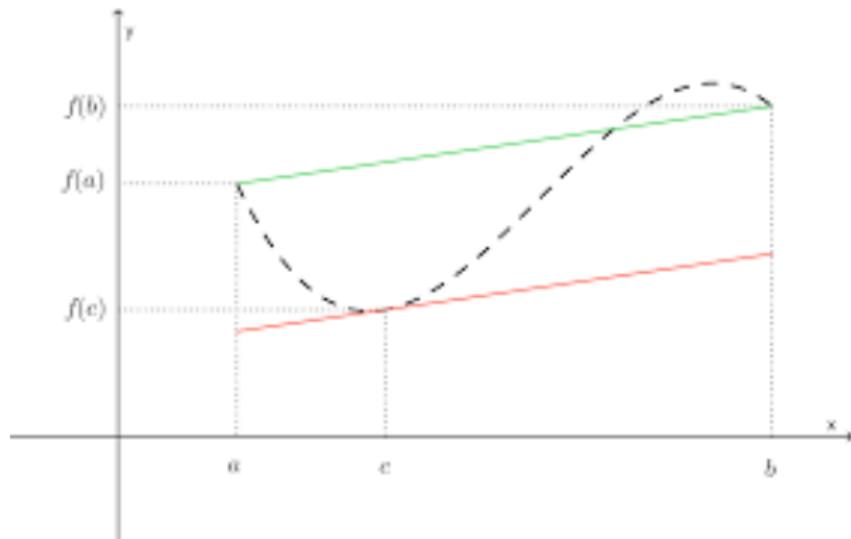
# Twierdzenie Lagrange'a o wartości średniej.

**Twierdzenie.** (Lagrange'a o wartości średniej). Jeżeli funkcja  $f$  jest ciągła w przedziale  $\langle a, b \rangle$  i różniczkowalna w  $(a, b)$  to istnieje punkt  $c \in (a, b)$  taki, że

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

Geometrycznie możemy to zilustrować następująco: istnieje więc co najmniej jeden punkt  $c \in (a, b)$  w którym styczna jest równoległa do siecznej wykresu przechodzącej przez  $(a, f(a))$  i  $(b, f(b))$ .

# Interpretacja graficzna.



I jeszcze ilustracja graficzna tego twierdzenia w postaci skryptu: [skrypt ilustracyjny w programie "Mathematica" \(.CDF\)](#).

## Twierdzenie Cauchy'ego.

**Twierdzenie.** (Cauchy'ego). Założymy, że 2 funkcje  $f$  i  $g$  są ciągłe w  $a, b >$  i różniczkowalne w  $(a, b)$  oraz  $g'(x) \neq 0$  dla  $x \in (a, b)$ .

Wówczas istnieje  $\xi \in (a, b)$  takie, że

$$\frac{f'(\xi)}{g'(\xi)} = \frac{f(b) - f(a)}{g(b) - g(a)} .$$

## Wnioski.

**Wniosek 1.** Jeżeli  $f$  jest ciągła w  $(a, b)$  i jeżeli  $f'(x) = 0$  dla każdego  $x \in (a, b)$  to  $f(x) = c = \text{const}$  dla  $x \in (a, b)$  (funkcja jest stała).

**Wniosek 2.** Jeżeli dane są 2 funkcje  $f$  i  $g$  określone w  $(a, b)$  (o wartościach w  $\mathbb{R}$ ) i  $f'(x) = g'(x)$  dla  $x \in (a, b)$  to funkcje różnią się co najwyżej o stałą tj.  $f(x) = g(x) + C$  ( $C = \text{const}, x \in (a, b)$ )

**Przykład zastosowania:**  $f(x) = \arcsin x$

$$g(x) = -\arccos x, \quad x \in (-1, 1).$$

Wówczas

$$f'(x) = \frac{1}{\sqrt{1-x^2}}, \quad g'(x) = \frac{1}{\sqrt{1-x^2}},$$

czyli  $f'(x) = g'(x)$  dla  $x \in (-1, 1)$ .

Stąd  $(f(x) - g(x))' = f'(x) - g'(x) = 0$ ,

a więc  $f(x) - g(x) = \text{const}$ .

$$\arcsin x + \arccos x = \text{const.} \quad \text{dla } x \in (-1, 1)$$

Ponieważ

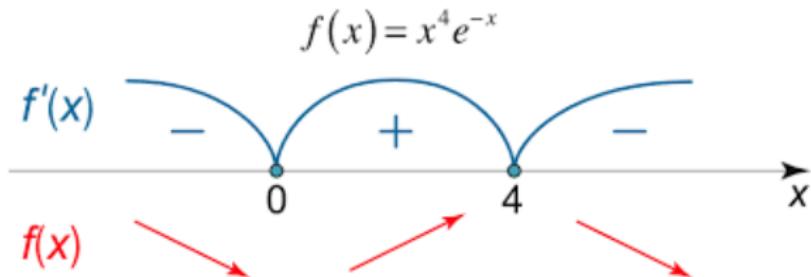
$$\arcsin 0 = 0, \quad \arccos 0 = \frac{\pi}{2},$$

więc ostatecznie

$$\arcsin x + \arccos x = \frac{\pi}{2} \quad \text{dla } x \in (-1, 1).$$

**Wniosek 3.** Jeżeli  $f$  jest ciągła w  $(a, b)$  i  $f'(x) \neq 0$  dla każdego  $x \in (a, b)$  to  $f$  jest różniczkowalna.

**Wniosek 4.** Jeżeli  $f$  jest ciągła w  $(a, b)$  i różniczkowalna w  $(a, b)$  oraz  $f'(x) > 0$  dla  $x \in (a, b)$  [ $f'(x) < 0$  dla  $x \in (a, b)$ ] to  $f$  jest ścisłe rosnąca [malejąca] w  $(a, b)$ .



## Wniosek 5. (reguła de l'Hôpitala).

Rozpatrzmy kilka przypadków:

(A) Założymy, że

$$\lim_{x \rightarrow x_0} f(x) = \lim_{x \rightarrow x_0} g(x) = 0 , \quad (*)$$

gdzie funkcje  $f$  i  $g$  są określone i różniczkowalne w pewnym otoczeniu punktu  $x_0$ . Jeżeli  $g'(x) \neq 0$  w tym otoczeniu, a ponadto istnieje granica  $\lim_{x \rightarrow x_0} \frac{f'(x)}{g'(x)}$  to istnieje granica  $\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)}$  oraz zachodzi wzór

$$\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = \lim_{x \rightarrow x_0} \frac{f'(x)}{g'(x)}$$

Uwagi:

(1<sup>0</sup>) regułę tę można stosować do granic jednostronnych,

(2<sup>0</sup>) również w przypadku granic niewłaściwych ( $x_0 = \pm\infty$ ) reguła jest prawdziwa.

(B) Przy założeniach reguły (A), gdzie w miejsce (\*) zakładamy

$$\lim_{x \rightarrow x_0} f(x) = \lim_{x \rightarrow x_0} g(x) = \infty \quad (**)$$

mamy

$$\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = \lim_{x \rightarrow x_0} \frac{f'(x)}{g'(x)}$$

Tu również możemy stosować tę regułę dla granic niewłaściwych ( $x_0 = \pm\infty$ ) oraz w przypadku  $\lim_{x \rightarrow x_0} \frac{f'(x)}{g'(x)} = \pm\infty$ .

W tej sytuacji będziemy mówić o wyrażeniu nieoznaczonym typu  $[\frac{\infty}{\infty}]$ .

(C) Jeżeli (\*) zastąpimy przez

$$\lim_{x \rightarrow x_0} f(x) = 0 \quad \text{oraz} \quad \lim_{x \rightarrow x_0} g(x) = \infty$$

to

$$\lim_{x \rightarrow x_0} f(x) \cdot g(x) = \lim_{x \rightarrow x_0} \frac{f(x)}{\frac{1}{g(x)}}$$

i możemy stosować regułę (A).

Jest to tzw. symbol nieoznaczony typu  $[0 \cdot \infty]$ .

(D) Przy założeniu (\*\*):

$$[f(x) - g(x)] = \frac{1}{\frac{1}{f(x)}} - \frac{1}{\frac{1}{g(x)}} = \frac{\frac{1}{g(x)} - \frac{1}{f(x)}}{\frac{1}{f(x) \cdot g(x)}}$$

i do ostatniego wyrażenia możemy stosować regułę (A) (tzw. symbol nieoznaczony typu  $[\infty - \infty]$ ).

A tu prezentacja: [reguły de l'Hôpitala w formacie .CDF](#).

Uwaga: do symboli nieoznaczonych zaliczamy też  $[0^0]$ ,  $[\infty^\infty]$  oraz  $[1^\infty]$  (mamy nadzieję, że Czytelnik „rozszыfruje” przez analogię jak one wyglądają – stosują się oczywiście do wyrażeń  $f(x)^{g(x)}$  ...).

Jeżeli również można sprowadzić do postaci występującej w regule (A) lub (B) np:

$$\lim_{x \rightarrow x_0} f(x) = \lim_{x \rightarrow x_0} g(x) = 0 \quad (\text{tj. } [0^0]) .$$

Niech  $F(x) = f(x)^{g(x)}$ . Przy założeniach gwarantujących istnienie funkcji  $F(x)$  ( $f(x), g(x) > 0$ ,  $f(x) \neq 1$ ) możemy obliczyć  $G(x) = \ln F(x)$ . Wtedy

$$G(x) = \ln F(x) = \ln(f(x)^{g(x)}) = g(x) \cdot \ln f(x) ,$$

a to ostatnie wyrażenie jest typu  $[0 \cdot \infty]$ . Możemy więc (zgodnie z (C)) obliczyć  $\lim_{x \rightarrow x_0} G(x)$ . Z ciągłości funkcji logarytmicznej i wykładniczej:

$$\lim_{x \rightarrow x_0} f(x)^{g(x)} = \lim_{x \rightarrow x_0} F(x) = \lim_{x \rightarrow x_0} e^{G(x)} = e^{\lim_{x \rightarrow x_0} G(x)}$$

czyli obliczymy granicę wyjściowego wyrażenia.

## Uwagi dla informatyków.

Reguła de l'Hôpitala dotyczy **funkcji i punktów skupienia dziedziny**. **Nie wolno jej stosować do ciągów!!** Takie błędny komentarz to duży skrót myślowy! Przy ciągach (jak widać nawet tu przydadzą się pochodne!) postępujemy tak:

(1) Rozszerzamy dziedzinę: ciąg  $a_n = f(n)$  określony jest na liczbach naturalnych, czyli rozpatrujemy **funkcję**  $\tilde{f} : [1, \infty) \rightarrow \mathbb{R}$  taką, że  $\tilde{f}(n) = f(n)$  np. dla ciągu  $a_n = n^2 \cdot \ln n$  ta funkcja jest postaci  $\tilde{f}(x) = x^2 \cdot \ln x$  - określona dla dziedziny  $x > 0$ .

(2) Do funkcji  $\tilde{f}(x)$  stosujemy regułę de l'Hôpitala (o ile spełnione są założenia - w tym przypadku jest to granica w nieskończoności) uzyskując np. granicę  $g$ .

(3) Skoro znana jest granica funkcji  $\tilde{f}$ , to z definicji Heinego granic dla każdego ciągu  $(x_n)$  z dziedziny zbieżnego do  $\infty$  mamy  $\lim_{n \rightarrow \infty} \tilde{f}(x_n) = g$ , a więc w szczególności dla  $x_n = n$  tak jest. **Mamy szukaną granicę ciągu!**

## Pochodne wyższych rzędów.

**Definicja.** Jeżeli  $f'$  jest różniczkowalna w zbiorze  $A$  to funkcję  $f'' : A \rightarrow \mathbb{R}$  określoną wzorem  $f''(x) = (f')'(x)$  nazywać będziemy pochodną drugiego rzędu z funkcji  $f$ .

Podobnie możemy zdefiniować pochodne wyższych rzędów:

jeżeli  $f^{(n)}$  jest różniczkowalna w  $A$  to pochodną rzędu  $(n+1)$  definiujemy następująco:

$$f^{(n+1)}(x) = (f^{(n)})'(x)$$

O funkcji mającej pochodne rzędu od 1 do  $n$  w  $x_0$  będziemy krótko mówić, że jest  $n$ -krotnie różniczkowalna w  $x_0$ .

Będziemy mówili, że funkcja  $f$  jest klasy  $C^n$  w zbiorze  $A$ , o ile  $f$  jest  $n$ -krotnie różniczkowalna w  $A$  i jeżeli  $f^{(n)}$  jest funkcją ciągłą w  $A$ . Podobnie powiemy, że funkcja  $f$  jest klasy  $C^\infty$  w zbiorze  $A$ , jeżeli ma w  $A$  pochodne dowolnego rzędu. Fakty te zapisywać będziemy odpowiednio:  $f \in C^n$  [ $f \in C^\infty$ ].

## Wzór Taylora.

**Twierdzenie 12.12.** (wzór Taylora). Założmy, że funkcja  $f$  określona w przedziale  $P$  jest w nim  $n$ -krotnie różniczkowalna.

Niech  $x_0$  i  $h$  będą takie, że  $x_0, x_0 + h \in P$ . Wówczas istnieje taka liczba  $0 < \nu < 1$ , że

$$f(x_0+h) = f(x_0) + \frac{f'(x_0)}{1!} \cdot h + \frac{f''(x_0)}{2!} \cdot h^2 + \dots + \frac{f^{(n-1)}(x_0)}{(n-1)!} \cdot h^{n-1} + \nu_n(h),$$

gdzie  $\nu_n(h) = \frac{f^{(n)}(x_0 + \nu h)}{n!} \cdot h^n$  (i jest nazywana resztą Lagrange'a we wzorze Taylora).

Prezentacja: [Skrypt ilustracyjny szeregu Taylora w "Mathematica"](#)  
- oraz taki dodatkowy [skrypt ilustracyjny...](#)

I jeszcze prezentacja tych [wielomianów Taylora \(.CDF\)](#).

Można także zapisać uzyskany wzór w następującej postaci:  
 niech  $x, x_0 \in P$ , kładziemy  $h = x - x_0$  , a stąd

$$\begin{aligned} f(x) &= f(x_0) + \frac{f'(x_0)}{1!}(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \dots \\ &+ \frac{f^{(n-1)}(x_0)}{(n-1)!}(x - x_0)^{n-1} + \nu_n(x), \end{aligned}$$

gdzie

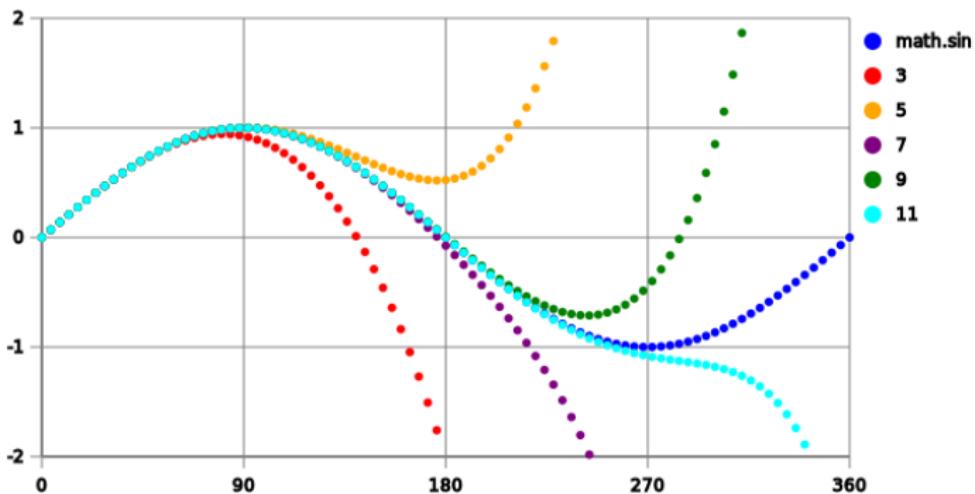
$$\nu_n(x) = \frac{f^{(n)}(x_0 + \nu(x - x_0))}{n!}(x - x_0)^n,$$

a w szczególności jeżeli  $x_0 = 0 \in P$ , to

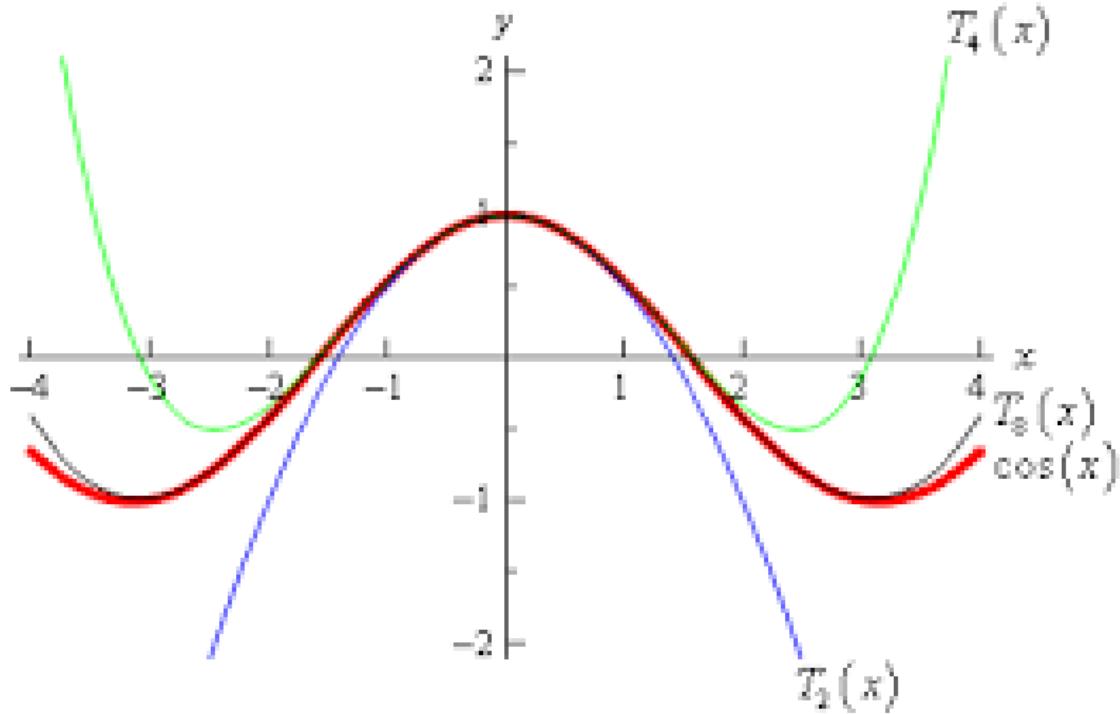
$$f(x) = f(0) + \frac{f'(0)}{1!} \cdot x + \frac{f''(0)}{2!} \cdot x^2 + \dots + \frac{f^{(n-1)}(0)}{(n-1)!} \cdot x^{n-1} + \frac{f^{(n)}(\nu x)}{n!} \cdot x^n.$$

Ten ostatni wzór nazywamy **wzorem MacLaurina**.

### Sines calculated with Taylor Polynomials to degree 3, 5, 7, 9 and 11



A tu applet ilustrujący wielomiany Taylora: niskich stopni.  
Zwracam uwagę: **błąd (oszacowanie reszty) rośnie** wyraźnie z  
odległością  $x$  od  $x_0$ ! To lokalne przybliżanie funkcji...



I tutaj obliczenia szeregów Taylora w Python'ie - może kogoś zainteresuje...

## Uwagi.

**Jak rozumieć wzór Taylora?** Zauważmy, że kolejne pochodne danej funkcji, wyznaczone w tym samym punkcie, "kodują" coraz "większą i bardziej dokładną" informację o kształcie funkcji w "szerszym" przedziale. Już wykorzystanie kilku lub kilkunastu liczb wystarcza do osiągnięcia wystarczającej jakości aproksymacji. Skojarzone pojęcie z informatyki: "**błąd obcięcia**".

Jest to więc coś w rodzaju standaryzacji **zapisu funkcji i algorytmu ich kompresji** w jednym.

Przypominam - **te slajdy to szkic wykładu.** Proszę doczytać ze źródeł ważne dla informatyków informacje, np. własność Darboux dla pochodnej (str 122, Twierdzenie 6.47 [W]): proszę przeanalizować czym się różni od klasycznego wyniku (por. str 241 [K], Twierdzenie 15.10.2).

Poza tym np. **postacie reszt we wzorze Taylora** - od str. 125 [W], a ze źródła [K] funkcje klasy  $C^k$  - od str. 234 w [K].

# Aproksymacja pochodnych.

Wielomian Taylor'a aproksymuje wartość funkcji poprzez **aproksymowanie wartości jej kolejnych pochodnych**. Wszystko staje się znacznie jaśniejsze po wyznaczeniu kolejnych **pochodnych wielomianu!**

Analizując poniższy schemat należy zwrócić uwagę jak człony wzoru przesuwają się "w lewo":

$$T_n(x) = \underbrace{f(x_0) + f^{(1)}(x_0)(x - x_0)}_{\text{liniowa aproksymacja } f(x) \text{ na bazie różniczki}} + \underbrace{\frac{1}{2!} f^{(2)}(x_0)(x - x_0)^2 + \cdots + \frac{1}{n!} f^{(n)}(x_0)(x - x_0)^n}_{\text{korekcja aproksymacji } f(x) \text{ na bazie pochodnych wyższych rzędów}}$$

$$T_n^{(1)}(x) = \underbrace{f^{(1)}(x_0) + f^{(2)}(x_0)(x - x_0)}_{\text{liniowa aproksymacja } f^{(1)}(x) \text{ na bazie różniczki}} + \underbrace{\frac{3}{3!} f^{(3)}(x_0)(x - x_0)^2 + \cdots + \frac{n}{n!} f^{(n)}(x_0)(x - x_0)^{n-1}}_{\text{korekcja aproksymacji } f^{(1)}(x) \text{ na bazie pochodnych wyższych rzędów}}$$

$$T_n^{(2)}(x) = \underbrace{f^{(2)}(x_0) + f^{(3)}(x_0)(x - x_0)}_{\text{liniowa aproksymacja } f^{(2)}(x) \text{ na bazie różniczki}} + \underbrace{\frac{3 \cdot 4}{4!} f^{(4)}(x_0)(x - x_0)^2 + \cdots + \frac{(n-1)n}{n!} f^{(n)}(x_0)(x - x_0)^{n-2}}_{\text{korekcja aproksymacji } f^{(2)}(x) \text{ na bazie pochodnych wyższych rzędów}}$$

Czyli

$$T_n^{(k)}(x_0) = f^{(k)}(x_0).$$

# Zastosowania wzoru Taylora.

Rola wzoru Taylora w informatyce jest nie do przecenienia! Proszę sprawdzić w dowolnej książce z metod numerycznych - większość szacowań błędów obliczeń wartości funkcji jest oparta o szacowanie **reszty we wzorze Taylora**. Patrz np. ([po angielsku](#)).

Proszę zwrócić uwagę i **starannie przeczytać materiał [W]: strony 124-131** i zwrócić uwagę na różne reszty. Pytanie po lekturze: po co tyle różnych postaci reszt (łatwiej vs. dokładniej...)? ("**błędy obcięcia**") Np.

**Twierdzenie 15.13.5 (wzór Taylora z resztą Lagrange'a)** Niech  $f : (c, d) \rightarrow \mathbb{R}$  będzie funkcją  $n + 1$ -różniczkowalną. Niech  $a, b \in (c, d)$  spełniają  $a < b$ . Wtedy istnieje  $\xi \in (a, b)$  takie, że

$$f(b) = T_a^n f(b-a) + \frac{f^{(n+1)}(\xi)}{(n+1)!} (b-a)^{n+1}. \quad (15.32)$$

Przyjmując oznaczenie  $h := b - a$  oraz rozpisując wielomian Taylora możemy też powyższą tezę zapisać jako

$$f(a+h) = f(a) + \frac{f'(a)}{1!} h + \frac{f''(a)}{2!} h^2 + \dots + \frac{f^{(n)}(a)}{n!} h^n + \frac{f^{(n+1)}(\xi)}{(n+1)!} h^{n+1}.$$

Podamy teraz ogólne twierdzenie, pozwalające dokładniej szacować resztę we wzorze Taylora.

**Twierdzenie 6.59 (wzór Taylora z resztą Schlömilcha–Roche'a).** *Niech  $f: (a, b) \rightarrow \mathbb{R}$  ma w przedziale  $(a, b)$  pochodne do rzędu  $(k+1)$  włącznie. Założymy, że  $x_0, x_0 + d \in (a, b)$ ,  $d > 0$ . Oznaczmy*

$$r_k(x) = f(x) - \sum_{j=0}^k \frac{f^{(j)}(x_0)}{j!} (x - x_0)^j.$$

*Wówczas dla każdego  $x \in (x_0, x_0 + d)$  i każdego  $p > 0$  istnieje liczba  $\theta \in (0, 1)$  taka, że*

$$r_k(x) = \frac{f^{(k+1)}(x_0 + \theta(x - x_0))}{k! \cdot p} (1 - \theta)^{k+1-p} (x - x_0)^{k+1}. \quad (6.18)$$

A ta poniższa postać reszty jest ważna, wróćmy do temtu, gdy już "oficjalnie" wprowadzimy całki!

#### 9.2.4 Wzór Taylora z resztą w postaci całkowej

W poprzednich podrozdziałach wykorzystaliśmy wzór na wielokrotne całkowanie przez części do dowodów niewymierności  $\pi$  i przestępności  $e$ . Teraz wskażemy jeszcze jedno zastosowanie tego wzoru, bardzo przydatne (także do analizowania funkcji wielu zmiennych, z którymi Czytelnik wielokrotnie zetknie się w późniejszych swoich studiach).

**Twierdzenie 9.49 (wzór Taylora z resztą całkową).** *Niech  $g \in C^{k+1}([a, b])$ ,  $a < x < b$ . Wówczas*

$$g(x) = g(a) + \sum_{j=1}^k \frac{g^{(j)}(a)}{j!} (x - a)^j + \int_a^x \frac{(x - t)^k}{k!} g^{(k+1)}(t) dt. \quad (9.54)$$

# Zwiększamy dokładność.

Kolejny raz zwracam uwagę: postęp w zastosowaniu w informatyce (tu: dokładność oszacowań) jest skutkiem użycia bardziej zaawansowanego aparatu matematycznego...

Aproxymacja liniowa poprzez różniczkę jest prosta, ale często nie jest wystarczająco dokładna. Możemy wtedy rozważać aproksymację krzywą drugiego stopnia (parabolą) zadaną wzorem  $y = a_0 + a_1h + a_2h^2$  gdzie  $h = x - x_0$  jest przyrostem zmiennej niezależnej (rys. 15.7). Możemy też przybliżać wielomianem  $n$ -ego stopnia postaci

$$W(h) := a_0 + a_1h + a_2h^2 + \dots + a_nh^n.$$

Różnica

$$f(x) - W[h] = f(x_0 + h) - (a_0 + a_1h + a_2h^2 + \dots + a_nh^n)$$

powinna być  $o(h^n)$ , bo oczekujemy błędu, który szybciej zmiera się do zera niż wielomian aproksymacyjny. Pokażemy, że jeśli funkcja  $r$  jest  $n$ -różniczkowalna w zerze, to własność  $r(h) = o(h^n)$  jest równoważna zerowaniu się pochodnych  $r$  do  $n$ -tej włącznie, tzn. warunkowi

$$r(0) = r'(0) = r''(0) = \dots = r^{(n)}(0) = 0.$$

W odniesieniu do naszego problemu aproksymacyjnego oznacza to, że funkcja

$$h \mapsto f(x_0 + h) - W(h)$$

powinna mieć zerowe pochodne w zerze do  $n$ -tej włącznie co jest równoważne postulatowi by funkcje  $h \mapsto f(x_0 + h)$  i  $h \mapsto W(h)$  miały te same pochodne do  $n$ -tej włącznie. Jak zobaczymy pochodne wielomianu jest stosunkowo łatwo policzyć:  $W^{(i)}(0) = i!a_i$ . Zatem tyle muszą wynosić pochodne funkcji  $h \mapsto f(x_0 + h)$  w zerze. Są one równe pochodnym  $x \mapsto f(x)$  w  $x_0$ , zatem poszukiwana aproksymacja ma postać

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

gdzie  $r(h) = o(h^n)$ . Otrzymany w ten sposób wielomian aproksymacyjny nazywany jest wielomianem Taylora funkcji  $f$  w  $x_0$ , bo pochodzi od angielskiego matematyka Brook'a Taylora (rys. 15.8).

## Niektóre zalety wzoru Taylora link: (źródło)

- ▶ Wzór Taylora definiuje wielomian Taylora, który tym lepiej (zwykle) aproksymuje funkcję im wyższy jest jego stopień.
- ▶ Twierdzenie Taylora (w wersji powyższej z resztą w postaci Peano), daje informację o błędzie aproksymacji (notacja „o-małe” w zagadnieniach asymptotycznego tempa wzrostu). Istnieją inne dodatkowe nierówności pozwalające szacować resztę - w innych podanych postaciach reszty.
- ▶ Wielomiany to sumy częściowe szeregu Taylora. Bardzo łatwo wyznaczać je numerycznie (tylko operacje mnożenia i dodawania). Z tego względu szereg kalkulatorów, programów i języków programowania implementuje inne funkcje (choć nie wszystkie - o czym mówiliśmy), poprzez odpowiednie wielomiany Taylora.
- ▶ Wielomiany z łatwością oddają się różniczkowaniu i całkowaniu. Można je zapisać w postaci iloczynowej, ułatwiając rozwiązywanie szeregu równań i nierówności. Wielomiany są określone na całej prostej rzeczywistej, co umożliwia analityczne „przedłużanie” wartości funkcji na dziedzinę, w której wyjściowa funkcja jest nieokreślona.

## Funkcje wypukłe - powtórka.

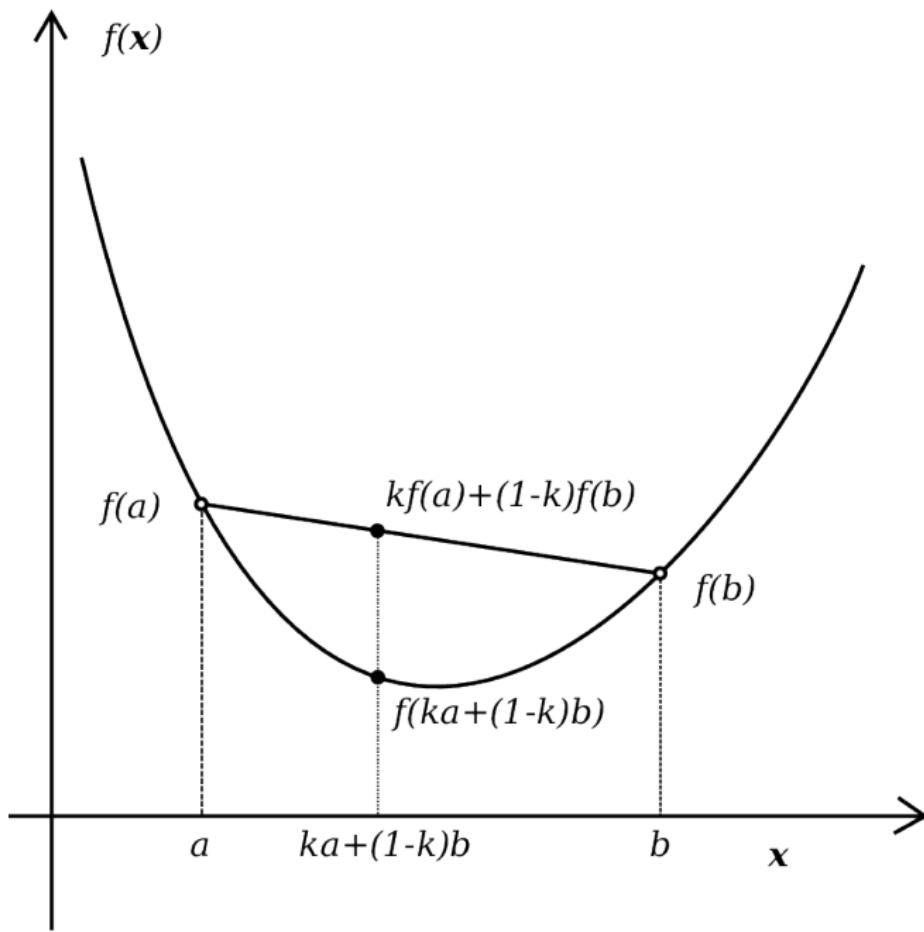
**Definicja.** Niech  $A \subset \mathbb{R}$  będzie przedziałem. Funkcję  $f : A \rightarrow \mathbb{R}$  nazywamy **wypukłą** w  $A$  gdy dla dowolnych  $x_1, x_2 \in A$  oraz dowolnych  $s, t \in \mathbb{R}$ ,  $s, t \geq 0$ ,  $s + t = 1$ , zachodzi nierówność

$$f(s \cdot x_1 + t \cdot x_2) \leq s \cdot f(x_1) + t \cdot f(x_2).$$

W przypadku, gdy nierówność zachodzi w przeciwnym kierunku funkcję nazywamy **wklesłą** w  $A$ .

Ponownie zwracamy uwagę, że ta własność także zależy od zbioru, a nierówność jest na ogół bardzo dobrym oszacowaniem dla wartości funkcji  $f$  często wykorzystywanym w różnych zastosowaniach. Nieco później podamy inną metodę badania wypukłości funkcji  $f$ . Ilustracją graficzną tej cechy jest fakt, iż odcinek łączący dowolne dwa punkty wykresu  $\{(x, y) : x \in A, y = f(x)\}$  „leży nad” wykresem funkcji (dokładnie to stwierdza nierówność z definicji!! - zrobić odpowiedni rysunek).

To pojęcie niezbędne do zrozumienia jednej z najważniejszych z metod optymalizacyjnych (np. w nauczaniu maszynowym) – optymalizacja wypukła! Np. informacje podstawowe.



Przykładami funkcji wypukłych są np.  $f(x) = x^2$ ,  $x \in \mathbb{R}$  czy  $f(x) = e^x$ ,  $x \in \mathbb{R}$ , natomiast funkcja  $f(x) = \sin x$  jest wypukła w  $A = < \pi, 2\pi >$ , ale nie jest wypukła w swojej dziedzinie. Funkcje wklęsłe to np.  $f(x) = -x^2$ ,  $x \in \mathbb{R}$  czy  $f(x) = \log x$ ,  $x \in (0, \infty)$ .

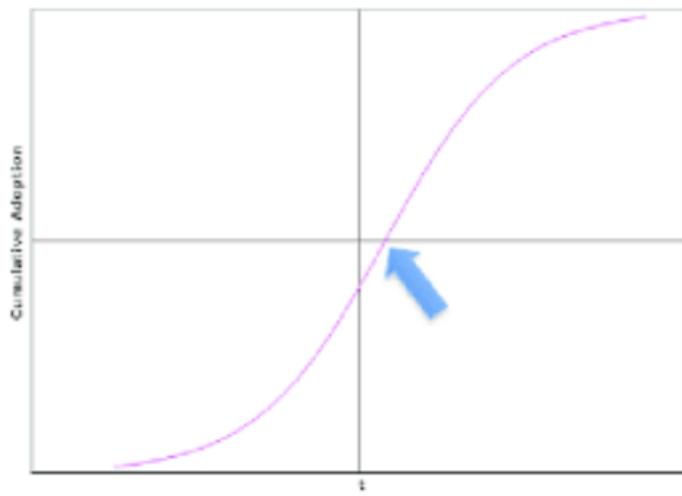
**Przykład.** Ponieważ  $2 = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 3$ , a funkcja  $f(x) = \sqrt{x}$ ,  $x \in < 0, \infty >$  jest wklęsła (sprawdzić !), to m.in. (!) wstawiając  $x_1 = 1$  oraz  $x_2 = 3$  do definicji uzyskamy

$$\sqrt{2} \geq \frac{1}{2} \cdot \sqrt{1} + \frac{1}{2} \sqrt{3},$$

czyli  $2\sqrt{2} - \sqrt{3} \geq 1$ , a ta nierówność nie dla wszystkich jest oczywista...

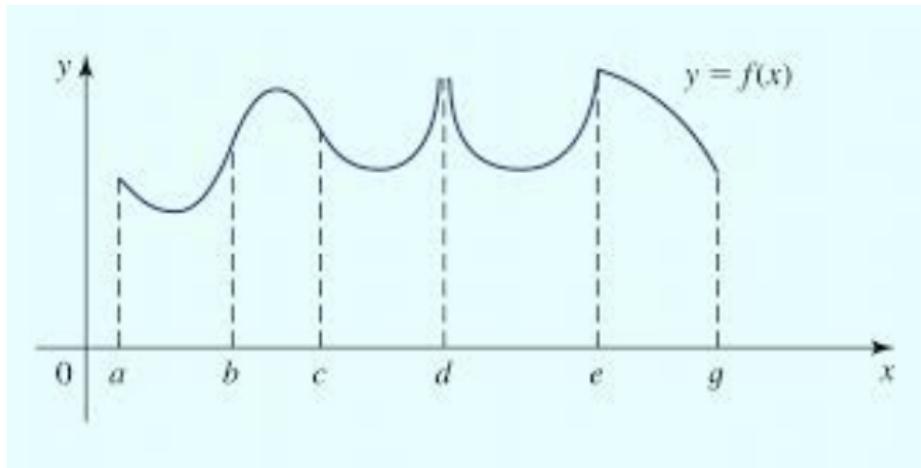
Podobnie natychmiast mamy przydatne oszacowanie pierwiastka:  $\sqrt{2} \leq \frac{3}{2}$  (tu:  $\frac{1}{2} = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 0$ , gdyż  $f(x) = 2^x$  - wypukła).

# Punkt przegięcia - zmiana wypukłości funkcji.



Zmiana wypukłości funkcji: na rysunku z wypukłej na wklęsłą  
(w definicji dopuszczalna też odwrotna zmiana)...

Punkty, w których zmienia się własność funkcji z wypukłej na wklęsłą lub odwrotnie nazywamy **punktami przegięcia** - za chwilę formalna definicja.



Funkcja  $f$  ma punkty przegięcia w  $b, c$  oraz  $e$ !

# Punkty przegięcia.

**Definicja.** Założymy, że  $f : P \rightarrow \mathbb{R}$  jest ciągła w przedziale  $P$ , oraz, że  $x_0$  jest punktem wewnętrznym tego przedziału.

Mówimy, że  $x_0$  jest punktem przegięcia funkcji  $f$ , jeżeli ten punkt jest jednocześnie końcem przedziału ścisłej wypukłości funkcji  $f$  i końcem przedziału ścisłej wklesłości tej funkcji.

**Twierdzenie.** Jeżeli  $x_0$  jest punktem przegięcia funkcji  $f$  oraz  $f$  jest klasy  $C^2$  w otoczeniu tego punktu, to  $f''(x_0) = 0$ .

Zwracam uwagę, że nie jest to warunek wystarczający.

Np.  $f(x) = x^4$  nie ma w punkcie  $x_0 = 0$  punktu przegięcia (ma w nim minimum lokalne), ale  $f''(0) = 0$ .

**Uwaga:** dla funkcji wielu zmiennych mamy też inne punkty charakterystyczne wykresu (np. punkt siodłowy) i ich lokalizacja będzie pewnym problemem w stosowaniu algorytmów gradientowych.

# Badanie wypukłości funkcji.

Teraz pokażemy jeszcze jedno zastosowanie pochodnych wyższych rzędów do badania funkcji. Przypomnijmy, że definicja wypukłości funkcji była podana wcześniej.

**Twierdzenie.** (badanie wypukłości funkcji różniczkowalnej).  
*Założymy, że funkcja  $f$  określona w przedziale  $P$  jest w nim różniczkowalna. Na to by  $f$  była wypukła w  $P$  potrzeba i wystarcza by pochodna  $f'$  była funkcją rosnącą.*

Wniosek. Jeżeli  $f$  jest dwukrotnie różniczkowalna w  $P$  i  $f'' \geq 0$  (czyli  $f'$  jest rosnąca) to  $f$  jest wypukła.

**Twierdzenie.** Jeżeli  $x_0$  jest punktem przegięcia funkcji  $f$  oraz  $f$  jest klasy  $C^2$  w otoczeniu tego punktu, to  $f''(x_0) = 0$ .

Zwracamy uwagę, że nie jest to warunek wystarczający.

Np.  $f(x) = x^4$  nie ma w punkcie  $x_0 = 0$  punktu przegięcia (ma w nim minimum lokalne), ale  $f''(0) = 0$ .

Podamy więc warunek wystarczający:

**Twierdzenie.** Jeżeli  $f$  jest klasy  $C^2$  w otoczeniu punktu  $x_0$  oraz  $f''$  ma następujące własności:

$$(a) \quad f''(x_0) = 0$$

$$(b) \quad f''(x) > 0 \text{ dla } x > x_0 \text{ (lub } x < x_0\text{) oraz } f''(x) < 0 \text{ dla } x < x_0 \text{ (}x > x_0 \text{ odpowiednio}),$$

to  $x_0$  jest punktem przegięcia funkcji  $f$ .

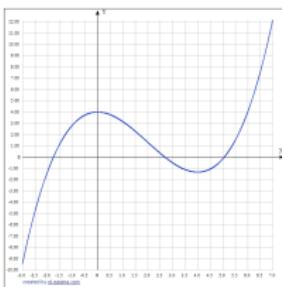
# Przykład.

Rozpatrzmy funkcję:

$$f(x) = \frac{1}{6}x^3 - x^2 + 4 , \quad f : \mathbb{R} \longrightarrow \mathbb{R}$$

Mamy  $f'(x) = \frac{1}{2}x^2 - 2x$  oraz  $f''(x) = x - 2$   $f''(x) > 0$  dla  $x > 2$

oraz  $f''(x) < 0$  dla  $x < 2$  (i oczywiście  $f''(2) = 0$ ). Stąd  $x_0 = 2$  jest punktem przegięcia funkcji  $f$ .



# Uwaga.

Zwracam uwagę, że zachodzi pewna analogia:

$f'(x_0) = 0$  jest warunkiem koniecznym istnienia ekstremum lokalnego funkcji różniczkowalnej (ale nie wystarczającym),

$f''(x_0) = 0$  jest warunkiem koniecznym istnienia punktu przegięcia takiej funkcji (i również nie jest wystarczającym ...).

# Warunek wystarczający istnienia ekstremum lokalnego,

Teraz czas na kolejne ważne twierdzenie:

**Twierdzenie.** (warunek wystarczający istnienia ekstremum lokalnego). Jeżeli  $f \in C^n$  w otoczeniu punktu  $x_0$  ( $n \geq 2$ ) i jeżeli

$$f'(x_0) = f''(x_0) = \dots = f^{(n-1)}(x_0) = 0 ,$$

oraz  $f^{(n)}(x_0) \neq 0$  , to mamy 2 możliwości:

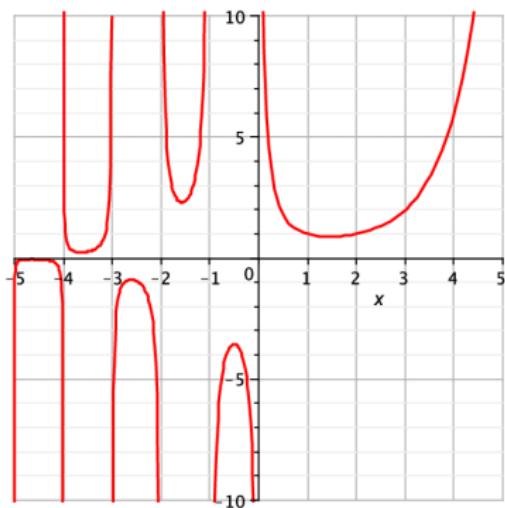
(1<sup>0</sup>) gdy  $n$  jest liczbą parzystą, to  $f$  ma w punkcie  $x_0$  ekstremum lokalne, przy czym jeśli  $f^{(n)}(x_0) > 0$  jest to minimum, a jeśli  $f^{(n)}(x_0) < 0$  to jest to maksimum,

(2<sup>0</sup>) gdy  $n$  jest liczbą nieparzystą to  $f$  ma w  $x_0$  punkt przegięcia ( $f$  jest monotoniczna w otoczeniu punktu  $x_0$ ).

# Wykresy.

Aby narysować wykres funkcji będziemy potrzebować **wszystkich** podanych wcześniej informacji!

Tzw. badanie przebiegu zmienności funkcji (o czym za chwilę) pozwoli narysować nawet takie przypadki jak ten poniżej (funkcja  $\Gamma$  Eulera):

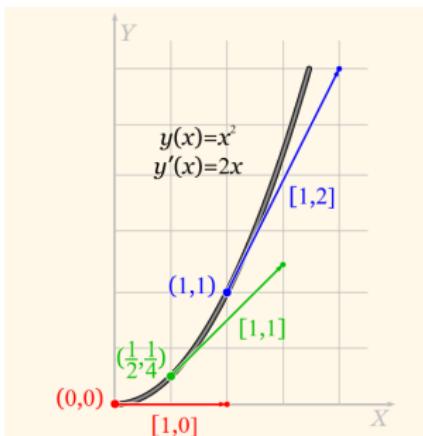


# Grafika komputerowa - krzywe w $\mathbb{R}^2$ .

Najważniejszą postacią zapisu krzywych w grafice wektorowej jest ich postać parametryczna

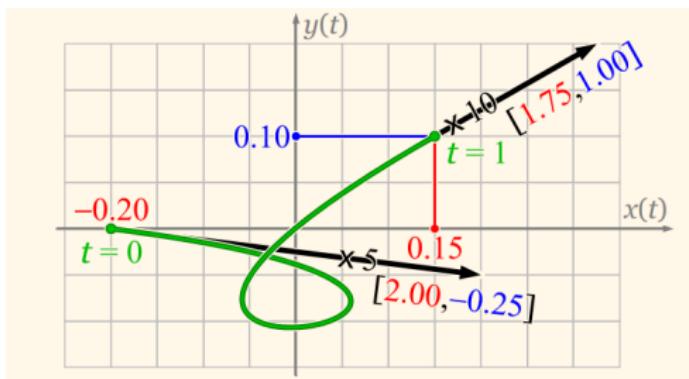
$$x = x(t), y = y(t)$$

dla  $t \in [a, b]$ . Najkorzystniej - ze względu na szybkość obliczeń aproksymować je za pomocą wielomianów (i to najczęściej stopnia co najwyżej trzeciego). Dla krzywych zadanych jawnie  $y = f(x)$  wygodnie jest zaczepić w wybranych punktach wektory o pierwszej składowej równej 1, a drugiej równej  $f'(x_k)$ .



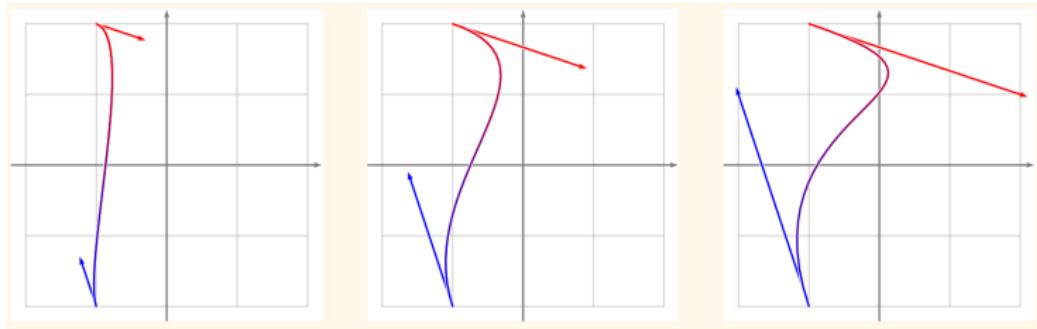
# Krzywe i ich aproksymacja.

W przypadku krzywej parametrycznej postąpimy podobnie: w wybranych punktach krzywej zaczepiamy wektory o składowych  $[x'(t), y'(t)]$ . Na ogół wystarczy zaczepić te wektory w początkowym i końcowym punkcie krzywej (mogliśmy ją podzielić na fragmenty opisane jednym wzorem), co pozwala na ich **gładkie sklejanie**.

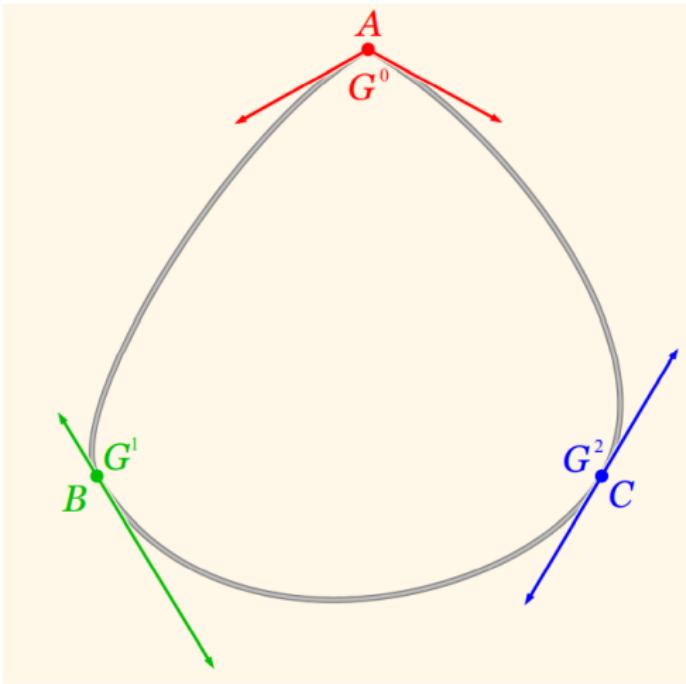


# Długość łuku krzywej.

Jak się okaże - ale w pełni dopiero jak wprowadzimy całki - te wektory oparte o pochodne będą potrzebne również do obliczeń (i oszacowań!!) długości łuku krzywej i jej krzywizny:

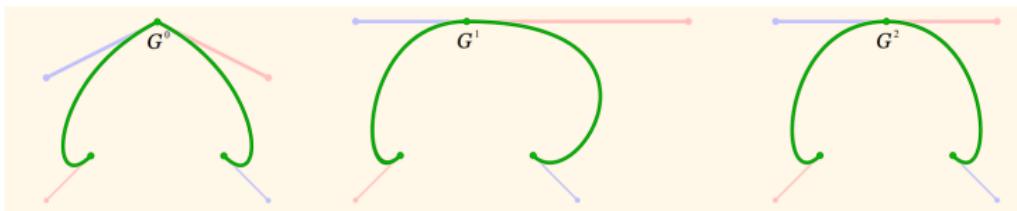


Aby rozpatrywane krzywe, które parametryzujemy kawałkami wyglądaly naturalnie, należy zapewnić odpowiednio gładkie przejście z jednej krzywej w drugą - co zapewniają właśnie te wektory oparte o pochodne. Widać różnice? Który z punktów wygląda najbardziej "dopasowany"?



cdn.

Więcej - na grafice komputerowej, fukcje Hermite'a czy Bezier'a (wielomiany Bernsteina - w matematyce) i b-spline itd...



**Uwaga!** Proszę zauważyć, że przywoałem przykład krzywych - ich parametryzacje są **funkcjami jednej zmiennej**. Podobnie będzie się badać płyty powierzchniowe, ale to wymaga funkcji **wielu zmiennych**, co wykracza poza ten wykład - ale można (i trzeba) to uzupełnić na kolejnych przedmiotach uzupełniających wiedzę matematyczną dla informatyków - **zachęcam!!**

## Metoda Newtona obliczania miejsc zerowych funkcji (różniczkowalnych).

Wybieramy początkowy punkt  $c_1$  aproksymacji miejsca zerowego  $c_0$ . W punkcie  $(c_1, f(c_1))$  prowadzimy styczną do wykresu funkcji  $f$ , która przetnie się z osią OX w punkcie  $(c_2, 0)$ . Teraz prowadzimy styczną do wykresu funkcji w punkcie  $(c_2, f(c_2))$ , a jej punkt wspólny z osią OX oznaczamy przez  $(c_3, 0)$ . Postępując tak dalej możemy oczekiwać, że  $c_n$  dąży do wartości  $c_0$  (co stanowi jej przybliżenie). Teraz konkretnie:

$$y - f(c_1) = f'(c_1) \cdot (x - c_1)$$

jest równaniem stycznej do wykresu funkcji  $f$  w  $(c_1, f(c_1))$ . Stąd:

$$0 - f(c_1) = f'(c_1) \cdot (c_2 - c_1)$$

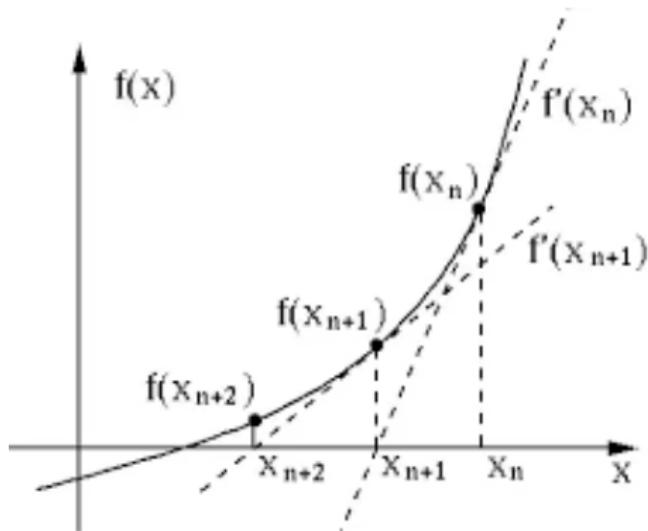
$$c_2 = c_1 - \frac{f(c_1)}{f'(c_1)} .$$

## Metoda Newtona II.

Prowadząc dalej analogiczne rozumowanie otrzymamy:

$$c_{n+1} = c_n - \frac{f(c_n)}{f'(c_n)} .$$

Taki sposób postępowania nazywamy metodą Newtona.



# Zbieżność metody Newtona.

**Twierdzenie.** Jeżeli  $f : \langle a, b \rangle \rightarrow \mathbb{R}$  jest ciągła w  $\langle a, b \rangle$  oraz:

(a) istnieje miejsce zerowe funkcji  $f$  w przedziale  $(a, b)$ ,

(b)  $f$  jest ściśle monotoniczna w  $\langle a, b \rangle$ ,

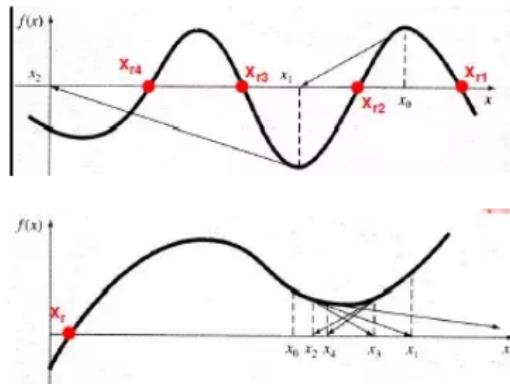
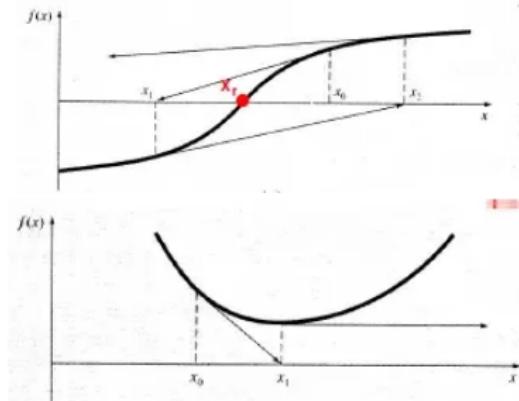
(c)  $f$  nie ma punktów przegięcia w  $\langle a, b \rangle$ ,

(d)  $\left| \frac{f(c)}{f'(c)} \right| \leq b - a$  dla punktu  $c \in \langle a, b \rangle$ , w którym  $|f'|$  ma wartość najmniejszą,

wtedy ciąg aproksymacji metodą Newtona jest zbieżny dla dowolnego wyboru  $c_1 \in \langle a, b \rangle$ .

# Sprawdzajmy założenia...

**Wzór** to nie twierdzenie (ani algorytm), ma swoje założenia.  
Na rysunkach przypadki, w których metoda Newtona nie daje ciągu zbieżnego do rozwiązania (**nie działa**).



Proszę sprawdzić dlaczego?

## Metoda Newtona III.

A teraz **błąd oszacowania metody Newtona w  $n$ -tym kroku:**

**Twierdzenie.** Jeżeli istnieje miejsce zerowe funkcji  $f$  w przedziale  $\langle a, b \rangle$  oraz  $|f''(x)| \leq M$ ,  $|f'(x)| \geq m > 0$  dla  $x \in \langle a, b \rangle$  i pewnych stałych  $m, M \in \mathbb{R}$ , to błąd oszacowania  $B_n$  w  $n$ -tym kroku metody Newtona ma następujące oszacowanie:

$$B_{n+1} \leq B_n^2 \cdot \frac{M}{2m} .$$

Jak widać nie wyliczamy tu błędu, lecz podajemy zależność rekurencyjną błędów z kolejnych kroków.

Proszę zwrócić uwagę na wykorzystanie wielu pojęć wprowadzonych na tym wykładzie...

## Uwaga.

Główny problem metody Newtona - konieczność znajomości pochodnej i jej obliczanie. Można zmodyfikować algorytm (**metoda Steffensa**) eliminując tę konieczność (tak naprawdę stosuje się iloraz różnicowy w miejsce pochodnej, czyli korzystamy z definicji pochodnej), ale nie ma "lepszych" metod - jest ona czasochłonna.

Dobry, polecany przez mnie: [przegląd metod](#). Uwaga na założenia matematyczne!

# Badanie przebiegu zmienności funkcji.

Korzystając z wprowadzonych wcześniej twierdzeń możemy określić własności funkcji na tyle dokładnie, że pozwala to na narysowanie (w miarę precyzyjnie) wykresu tej funkcji. Najpierw na przykładzie pokażemy niezbędne w tym celu czynności, a na zakończenie podsumujemy algorytm postępowania.

Niezbędne przy tym będą wszystkie wprowadzone wcześniej pojęcia i twierdzenia ...

Tu [strona w Geogebrze](#), gdzie można prześledzić badanie przebiegu zmienności funkcji.

# Algorytm postępowania.

W celu ułatwienia zadania naszkicowania wykresu funkcji stosujemy następujące kroki, zwane schematem badania funkcji.  
Należy:

1. wyznaczyć dziedzinę funkcji,
2. zbadać parzystość, nieparzystość i okresowość (tylko wtedy, gdy podejrzewamy, że taka własność zachodzi),
3. wyznaczyć miejsca zerowe funkcji oraz punkty przecięcia z osią  $OY$ ,
4. wyznaczyć granice lub wartości na końcach przedziałów określoności funkcji,
5. wyznaczyć asymptoty pionowe i ukośne,
6. wyznaczyć przedziały monotoniczności i ekstrema lokalne,
7. wyznaczyć przedziały wypukłości i punkty przegięcia,
8. zebrać informacje w tabeli,
9. narysować wykres funkcji.

**Uwaga:** pewne kroki możemy opuścić - o ile nie wpływa to zasadniczo na zdolność wykonania wykresu, a są skomplikowane obliczeniowo (np. czcionką pochyloną powyżej lub przy skomplikowanych obliczeniach drugiej pochodnej).

# Przykład.

## Przykład 1.

$$f(x) = \frac{x^3 + 4}{x^2}$$

Najpierw musimy określić dziedzinę funkcji. Tu oczywiście  $x^2 \neq 0$ , a więc  $D = \mathbb{R} \setminus \{0\}$ . Granice funkcji na końcach przedziału określoności pozwolą w miarę dokładnie szkicować wykres:

$$\lim_{x \rightarrow -\infty} f(x) = \lim_{x \rightarrow -\infty} \frac{x^3 + 4}{x^2} = -\infty ,$$

$$\lim_{x \rightarrow 0^-} f(x) = \lim_{x \rightarrow 0^+} f(x) = +\infty ,$$

$$\lim_{x \rightarrow \infty} f(x) = \lim_{x \rightarrow \infty} \frac{x^3 + 4}{x^2} = +\infty .$$

Następnie warto znaleźć punkty przecięcia wykresu z osiami współrzędnych. Tu, ponieważ  $x \neq 0$ , szukamy tylko wartości argumentu dla którego  $f(x) = 0$ .

Oczywiście  $f(x) = 0 \iff x^3 + 4 = 0 \iff x = -\sqrt[3]{4}$ .

Teraz znajdziemy asymptoty tej funkcji.

Pionową jest oczywiście  $x = 0$ , szukamy asymptot ukośnych. Niech  $y = mx + n$ , wówczas

$$m_1 = \lim_{x \rightarrow \infty} \frac{f(x)}{x} = \lim_{x \rightarrow \infty} \frac{x^3 + 4}{x^3} = 1$$

$$m_2 = \lim_{x \rightarrow -\infty} \frac{f(x)}{x} = 1,$$

funkcja ta ma więc co najwyżej jedną asymptotę.

$$n = \lim_{x \rightarrow \pm\infty} [f(x) - mx] = \lim_{x \rightarrow \pm\infty} \left[ \frac{x^3 + 4}{x^2} - x \right] = \lim_{x \rightarrow \pm\infty} \frac{4}{x^2} = 0$$

Asymptotą ukośną jest więc prosta  $y = x$ .

Czas na badanie pochodnych funkcji  $f$ .

Obliczamy:

$$\begin{aligned}f'(x) &= \left( \frac{x^3 + 4}{x^2} \right)' = \frac{3x^2 \cdot x^2 - 2x(x^3 + 4)}{x^4} = \\&= \frac{3x^4 - 2x^4 - 8x}{x^4} = \frac{x^3 - 8}{x^3}\end{aligned}$$

dla  $x \neq 0$ .

Zbadamy dokładniej tę pochodną.

$$f'(x) = 0 \iff \frac{x^3 - 8}{x^3} = 0 \iff x^3 = 8 \iff x = 2$$

$$\begin{aligned}f'(x) > 0 &\iff \frac{x^3 - 8}{x^3} > 0 \iff \frac{(x-2)(x^2+2x+4)}{x^3} > 0 \iff \\&\iff x \in (-\infty, 0) \cup (2, \infty)\end{aligned}$$

oraz  $f'(x) < 0 \iff x \in (0, 2)$

Funkcja jest więc rosnąca w przedziale  $(-\infty, 0)$  oraz w przedziale  $(2, \infty)$ , a malejąca w przedziale  $(0, 2)$ . Z definicji ekstremum wynika więc, że dla  $x = 2$  funkcja ma minimum lokalne oraz  $f(2) = 3$ .

Zbadamy teraz drugą pochodną  $f''(x)$ :

$$f''(x) = \left( \frac{x^3 - 8}{x^3} \right)' = \frac{3x^2 \cdot x^3 - 3x^2 \cdot (x^3 - 8)}{x^6} = \frac{24x^2}{x^6} = \frac{24}{x^4},$$

$x \neq 0$ . Stąd oczywiście  $f''(x) > 0$  dla  $x \in (-\infty, 0) \cup (0, \infty)$ .

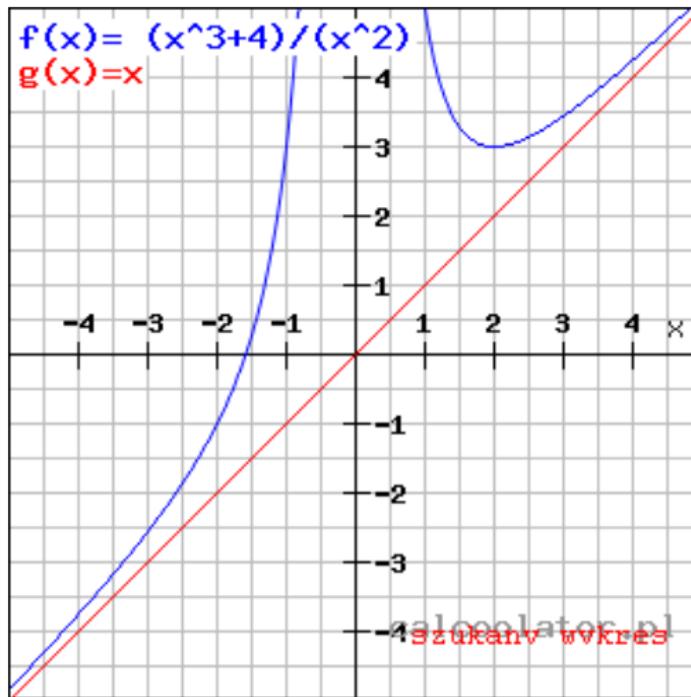
Funkcja jest wypukła w przedziale  $(-\infty, 0)$  oraz w przedziale  $(0, \infty)$ .

Wyniki zbierzymy w postaci tabeli.

$x$	$(-\infty; 0)$	0	$(0; 2)$	2	$(2; \infty)$
$f'(x)$	+	X	-	0	+
$f''(x)$	+	X	+	+	+
$f(x)$	↗	X	↘	$3_{min}$	↗

Niezbędny jest komentarz: w pierwszym wierszu umieszczamy wszystkie punkty, które mogą ułatwić przygotowanie wykresu oraz przedziały pomiędzy nimi; w wierszach „ $f'(x)$ ” oraz „ $f''(x)$ ” podajemy jaki znak posiadają te pochodne w danym punkcie lub przedziale. Wreszcie ostatni wiersz to podsumowanie - strzałki mają sygnalizować wzrost (i wypukłość) funkcji, a w odpowiednich miejscach umieszczamy też wyliczone wcześniej granice. Warto tu zauważyć, że jest to także pewna forma kontroli - nie może być sprzeczności pomiędzy wzrostem funkcji, a jej granicami.

I na koniec - wykres...



# Zastosowania...

To nie miejsce na szczegółowe przedstawienie roli pochodnych **w zastosowaniach informatycznych** - z jednej strony wprowadzamy dopiero podstawowe pojęcia matematyczne, a z drugiej potrzebujemy najpierw poznać pewne działy informatyki, aby zobaczyć potrzebę stosowania matematyki.

**Dla chętnych:** proszę np. poczytać o zastosowaniach (UJ) [w grafice komputerowej - strony od 155](#) - pochodne, styczne, ciągłość, funkcje gładkie i klas  $C^{(k)}$ , krzywe Hermite'a i Beziera, funkcje spline itd. (teraz wiadomo, dlaczego na ilustracji wcześniej podanej nie były styczne, tylko inne krzywe...). Szczegóły - oczywiście na innym przedmiocie, ale bez matematyki nie da się tego zrozumieć...

Grafika komputerowa ma **3 podstawowe działy**: **podstawy matematyczne (transformacje, krzywe i powierzchnie, oświetlenie - algebra, geometria i analiza)**, później **stosowanie istniejących bibliotek graficznych** (np. Direct3D, WebGL) i na koniec **aplikacje graficzne**. Proszę pamiętać: bez pierwszej części nie poznamy możliwości i ograniczeń pozostałych dwóch działów.

Proszę też poczytać materiał dla początkujących **grafików komputerowych** por. strona 3 mini-kursu: okaże się po co nam pochodne...

Funkcje wielu zmiennych (a więc i powierzchnie 3D) będą przedstawiane w ramach kolejnego kursu analizy dla informatyków - dla chętnych...

# Tajne/poufne.

- ▶ Styczna do wykresu funkcji. Warunki jej istnienia. Związek z pewnym twierdzeniem o wartości średniej (jakim?). Omów wybrane zastosowanie w informatyce (np. metoda stycznych, grafika komputerowa itp.).
- ▶ Co to jest szereg (potęgowy) Taylora? Podaj definicję, wybrany przez siebie **przykład** oraz omów zastosowanie w obliczeniach na komputerze wartości pewnych funkcji.
- ▶ Jeśli funkcja  $f$ , której wartość należy obliczyć za pomocą komputera, spełnia warunek Lipschitza, to każdy algorytm numeryczny poprawny jest też numerycznie stabilny. Podaj co oznacza spełnianie warunku Lipschitz'a i wybrany warunek wystarczający, aby on zachodził.

- Za pomocą reguły de l'Hôpitala oblicz granicę:

$$\lim_{x \rightarrow \infty} \frac{x^{100} + x - 2}{x^{50} + x - 2}.$$

Sprawdź niezbędne założenia.

- Można sprawdzić, że algorytm Karatsuby mnożenia bardzo długich liczb całkowitych jest rzędu  $O(n^{\log_2 3})$  ("Algorytmy i struktury danych"). Korzystając ze wzoru  $\log_2 3 = \frac{\log_{10} 3}{\log_{10} 2}$  oszacuj tę wielkość i sprawdź, że  $n^{\log_2 3} = o(n^2)$ . Czy jest również rzędu  $o(n^{3/2})$ ?
- Oblicz  $f'''(x)$  dla funkcji  $f(x) = \frac{1}{\ln x}$  (sprawdź dziedzinę!). Podaj definicję pochodnej wyższego rzędu i wykorzystane w obliczeniach twierdzenia lub własności funkcji.
- Omów zależności monotoniczności funkcji i znaku pochodnej. Zwrót uwagę na założenia! Podaj przykład (narysuj) wykres funkcji rosnącej i **nieciągłej** na przedziale  $[0, 1]$ . Co można powiedzieć o pochodnej funkcji z podanego przykładu?
- Co to jest szereg (potęgowy) Taylora? Podaj definicję, wybrany przez siebie **przykład** oraz omów zastosowanie do obliczania na komputerze wartości liczby rzeczywistej e.

- ▶ Oblicz  $f'''(x)$  dla funkcji  $f(x) = \operatorname{arctg} x$ . Podaj definicję pochodnej wyższego rzędu i wykorzystane w obliczeniach twierdzenia lub własności funkcji.
- ▶ W ocenie złożoności obliczeniowej wykorzystywane jest pojęcie "asymptotycznej niewiększości", czyli  $f(n) = O(g(n))$ . Warunkiem wystarczającym na to jest

$$\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} < \infty.$$

**Oblicz**, które z podanych niżej funkcji są asymptotycznie niewiększe niż funkcja (ciąg)  $g(n) = n \ln n$  (np. zastosuj granice ciągów lub rozszerzenie do funkcji na  $\mathbb{R}$  i regułę de l'Hôspitala)?

- [a]  $f_1(n) = n^2$ ,
- [b]  $f_2(n) = n \ln(\ln n)$ ,
- [c]  $f_3(n) = 5n$ .

## Definicja. Pochodne cząstkowe funkcji.

Niech funkcja  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  będzie określona przynajmniej na otoczeniu punktu  $(x_0, y_0)$ .

**Pochodną cząstkową I rzędu funkcji  $f$  w punkcie  $(x_0, y_0)$  względem zmiennej  $x$  nazywamy granicę ilorazu różnicowego:**

$$\frac{\partial f}{\partial x}(x_0, y_0) = \lim_{h \rightarrow 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h}.$$

Pochodną tę oznaczać będziemy również:

$$f'_x(x_0) \quad lub \quad D_1 f(x_0, y_0).$$

## Definicja. Pochodne cząstkowe funkcji.

Niech funkcja  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  będzie określona przynajmniej na otoczeniu punktu  $(x_0, y_0)$ .

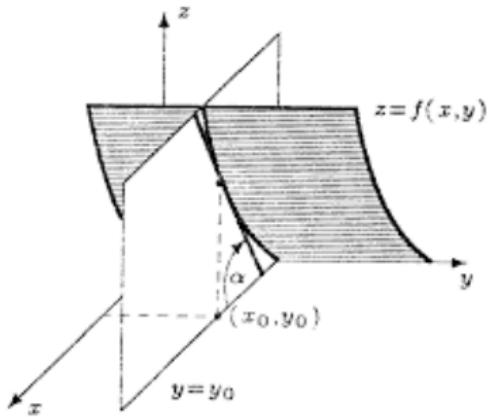
**Pochodną cząstkową I rzędu funkcji  $f$  w punkcie  $(x_0, y_0)$  względem zmiennej  $y$  nazywamy granicę ilorazu różnicowego:**

$$\frac{\partial f}{\partial y}(x_0, y_0) = \lim_{h \rightarrow 0} \frac{f(x_0, y_0 + h) - f(x_0, y_0)}{h}.$$

Pochodną tę oznaczać będziemy również:

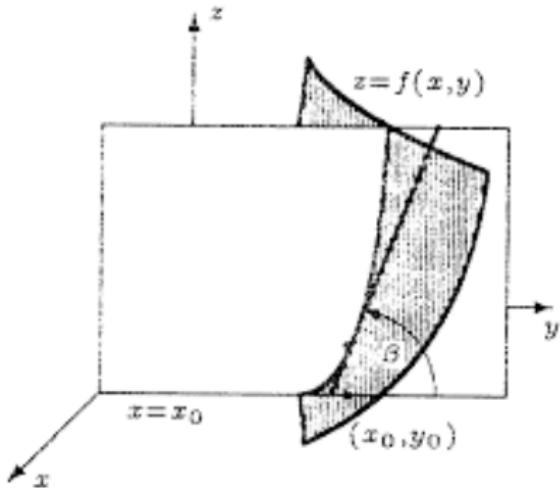
$$f'_y(x_0) \quad lub \quad D_2 f(x_0, y_0).$$

# Interpretacja geometryczna pochodnej cząstkowej $\frac{\partial f}{\partial x}(x_0, y_0)$ .



$\frac{\partial f}{\partial x}(x_0, y_0) = \operatorname{tg} \alpha$ . Pochodna cząstkowa  $\frac{\partial f}{\partial x}(x_0, y_0)$  jest miarą lokalnej szybkości wzrostu funkcji  $f$  względem zmiennej  $x$  przy ustalonej wartości zmiennej  $y$ .

# Interpretacja geometryczna pochodnej cząstkowej $\frac{\partial f}{\partial y}(x_0, y_0)$ .



$$\frac{\partial f}{\partial y}(x_0, y_0) = \operatorname{tg} \beta.$$

# WAŻNE

Podstawowe reguły różniczkowania - dotyczące sumy, różnicy, iloczynu i ilorazu pozostają bez zmian dla pochodnych cząstkowych.

## Przykład

►  $f(x, y) = x^2 \cdot y + e^{x \cdot y};$

$$\frac{\partial f}{\partial x}(x, y) = 2 \cdot x \cdot y + y \cdot e^{x \cdot y}$$

$$\frac{\partial f}{\partial y}(x, y) = x^2 + x e^{x \cdot y}$$

►  $f(x, y) = \sin^2(x - y^2);$

$$\frac{\partial f}{\partial x}(x, y) = 2 \sin(x - y^2) \cdot \cos(x - y^2) \cdot (1 - 0)$$

$$\frac{\partial f}{\partial y}(x, y) = 2 \sin(x - y^2) \cdot \cos(x - y^2) \cdot (0 - 2y)$$

# Gradient.

**Gradientem funkcji wielu zmiennych**  $f$  w punkcie  $(x_1, x_3, \dots, x_n)$ , w którym istnieją wszystkie pochodne cząstkowe, nazywamy wektor

$$\nabla f(x_1, x_3, \dots, x_n) = \left( \frac{\partial f}{\partial x_1}(x_1, x_3, \dots, x_n), \dots, \frac{\partial f}{\partial x_n}(x_1, x_3, \dots, x_n) \right).$$

W przypadku, gdy pochodne cząstkowe istnieją w każdym punkcie,  $\nabla f$  jest po prostu funkcją wielu zmiennych otrzymanych jako pochodne cząstkowe.

Wtedy **symbol**  $\nabla$  oznacza wektorowy operator liniowy  $(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n})$ , który przeprowadza funkcje wielu zmiennych posiadające wszystkie pochodne cząstkowe na wektorową funkcję wielu zmiennych.

Niech funkcja  $f$  ma pochodne cząstkowe  $\frac{\partial f}{\partial x}$ ,  $\frac{\partial f}{\partial y}$  przynajmniej na  $O(x_0, y_0)$ . Pochodne cząstkowe II rzędu funkcji  $f$  w punkcie  $(x_0, y_0)$  określamy wzorami:

$$\frac{\partial^2 f}{\partial x^2}(x_0, y_0) = \frac{\partial}{\partial x} \left( \frac{\partial f}{\partial x} \right)(x_0, y_0)$$

$$\frac{\partial^2 f}{\partial x \partial y}(x_0, y_0) = \frac{\partial}{\partial x} \left( \frac{\partial f}{\partial y} \right)(x_0, y_0)$$

$$\frac{\partial^2 f}{\partial y \partial x}(x_0, y_0) = \frac{\partial}{\partial y} \left( \frac{\partial f}{\partial x} \right)(x_0, y_0)$$

$$\frac{\partial^2 f}{\partial y^2}(x_0, y_0) = \frac{\partial}{\partial y} \left( \frac{\partial f}{\partial y} \right)(x_0, y_0)$$

## Twierdzenie Schwarza.

Niech funkcja dwóch zmiennych  $f : [a, b] \times [c, d] \rightarrow \mathbb{R}$  posiada na otoczeniu punktu  $(x_0, y_0) \in (a, b) \times (c, d)$  pochodne mieszane II rzędu,  $f_{xy}$  i  $f_{yx}$ , które są w tym punkcie ciągłe (jako funkcje dwóch zmiennych).

Wtedy pochodne te w  $(x_0, y_0)$  są sobie równe.

# Różniczki funkcji wielu zmiennych.

Rezygnujemy tu z w pełni formalnego wprowadzenia różniczek jako pewnych odwzorowań wieloliniowych (pozostajemy przy intuicji dla funkcji jednej zmiennej).

Przedstawimy jedynie pewne wprowadzenie oznaczeń.

Wartość różniczki

$$d_a f \in L(\mathbb{R}^n, \mathbb{R})$$

na wektorze  $h = (h_1, h_2, \dots, h_n) \in \mathbb{R}^n$  wyraża sumę

$$d_a f(h) = \frac{\partial f(a)}{\partial x_1} h_1 + \frac{\partial f(a)}{\partial x_2} h_2 + \dots + \frac{\partial f(a)}{\partial x_n} h_n.$$

Sumę tę można także wyrazić bez argumentu  $h$  jako

$$d_a f = \frac{\partial f(a)}{\partial x_1} dx_1 + \frac{\partial f(a)}{\partial x_2} dx_2 + \dots + \frac{\partial f(a)}{\partial x_n} dx_n,$$

gdzie  $dx_i : \mathbb{R}^n \ni h = (h_1, h_2, \dots, h_n) \mapsto dx_i(h) = h_i \in \mathbb{R}$  jest rzutowaniem na  $i$ -tą współrzędną.

Założymy, że  $f \in C^m(U)$ , gdzie  $U \subset \mathbb{R}^n$  jest podzbiorem otwartym przestrzeni skończenie wymiarowej  $\mathbb{R}^n$ .

Wówczas różniczkę rzędu  $m$  można wyrazić efektywnie za pomocą pochodnych cząstkowych rzędu  $m$ .

**Twierdzenie.** Jeśli  $f \in C^m(U)$ , to w dowolnym punkcie  $a \in U$  wartość różniczki rzędu  $m$  na  $m$ -ce jednakowych wektorów  $h = (h_1, h_2, \dots, h_n) \in \mathbb{R}^n$  wyraża sumą

$$d_a^m f \underbrace{(h, h, \dots, h)}_{m \text{ wektorów } h} = \sum_{|\alpha|=m} \binom{m}{\alpha} \frac{\partial^m}{\partial x^\alpha} f(a) h^\alpha,$$

gdzie sumowanie przebiega po wszystkich możliwych wielowskaźnikach ( $n$ -wskaźnikach)  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n) \in \mathbb{N}_0^n$  o długości  $|\alpha| = \alpha_1 + \alpha_2 + \dots + \alpha_n = m$ .

Natomiast  $\binom{m}{\alpha} := \frac{m!}{(m - |\alpha|)! \alpha!}$ , jest uogólnieniem symbolu Newtona, w którym silnię wielowskaźnika  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$  definiujemy za pomocą iloczynu silni jego współrzędnych, tj.  
 $\alpha! = \alpha_1! \alpha_2! \dots \alpha_n!$  oraz  $h^\alpha = h_1^{\alpha_1} h_2^{\alpha_2} \dots h_n^{\alpha_n}$ .

Uwaga.

Wzór

$$d_a^m f(h, h, \dots, h) = \sum_{|\alpha|=m} \binom{m}{\alpha} \frac{\partial^m}{\partial x^\alpha} f(a) h^\alpha,$$

który podaliśmy w tezie twierdzenia czasem zapisuje się bez  
wyszczególniania argumentów w następującej postaci

$$d_a^m f = \sum_{|\alpha|=m} \binom{m}{\alpha} \frac{\partial^m f(a)}{\partial x^\alpha} dx^\alpha$$

lub

$$d_\cdot^m f = \sum_{|\alpha|=m} \binom{m}{\alpha} \frac{\partial^m f}{\partial x^\alpha} dx^\alpha,$$

gdzie  $dx^\alpha : \mathbb{R}^n \mapsto \mathbb{R}$  definiujemy na wektorze  $h \in \mathbb{R}^n$  wzorem  
 $dx^\alpha(h) := h^\alpha = h_1^{\alpha_1} h_2^{\alpha_2} \dots h_n^{\alpha_n} \in \mathbb{R}$ .

# Twierdzenie Taylora.

Dla dowolnych punktów  $a$  oraz  $a + h$  zbioru  $U$  takich, że odcinek  $\{a + th, t \in [0, 1]\} \subset U$ , zachodzi równość

$$\begin{aligned} f(a+h) &= f(a) + d_a f(h) + \frac{1}{2!} d_a^2 f(h, h) + \frac{1}{3!} d_a^3 f(h, h, h) + \dots \\ &+ \frac{1}{m!} d_a^m f \underbrace{(h, h, \dots, h)}_{m \text{ wektorów } h} + R_m f(a, h), \end{aligned}$$

gdzie

$$\|R_m f(a, b)\|_y \leq \frac{1}{(m+1)!} \sup\{|d_{a+th}^{m+1}(h, h, \dots, h)|, t \in [0, 1]\}.$$

# Wielomiany Taylora.

Funkcję

$$\begin{aligned} X \in h \mapsto T_a^m f(h) &= f(a) + d_a f(h) + \frac{1}{2!} d_a^2 f(h, h) + \dots \\ &+ \frac{1}{m!} d_a^m f \underbrace{(h, h, \dots, h)}_{m \text{ razy}} \\ &= \sum_{k=0}^m \frac{1}{k!} d_a^k \underbrace{(h, h, \dots, h)}_{k \text{ razy}} \in Y \end{aligned}$$

nazywamy **wielomianem Taylora rzędu  $m$  funkcji  $f$**  o środku w punkcie  $a$ .

# Gradient a wzrost wartości funkcji.

Gradient wskazuje **kierunek największego wzrostu funkcji, a jego długość odpowiada współczynnikowi kierunkowemu tego wzrostu** (to wynika wprost z własności iloczynu skalarnego: iloczyn skalarny wektorów jest równy iloczynowi ich długości razy cosinus kąta między nimi; cosinus, a zatem pochodna kierunkowa jest największa gdy kąt jest zerowy, a wartość tej pochodnej, czyli iloczyn długości jest długością gradientu, bo drugi wektor ma długość 1).

To **kluczowa własność w zastosowaniach** - na ogólnie mówimy o "metodach gradientowych" gdy określamy najszybszy wzrost wartości pewnych funkcji wielu zmiennych...

# Pojęcie gradientu w analizie obrazu...

## The gradient of an image


$$\nabla f = \left[ \frac{\partial f}{\partial x}, 0 \right]$$

## The gradient of an image


$$\nabla f = \left[ \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right]$$

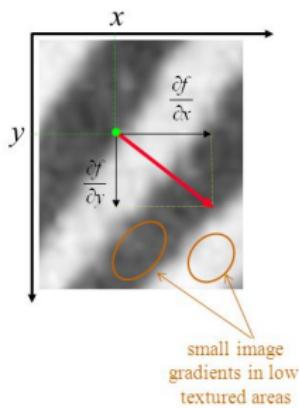
Jak to wyjaśnić? Po prostu: tworzymy funkcję nasycenia wypełnieniem i sprawdzamy jej gradient!

- The **gradient** at a point  $(x,y)$

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \approx \begin{bmatrix} \nabla_x * f \\ \nabla_y * f \end{bmatrix}$$

- **Gradient Magnitude**

$$\| \nabla f \| = \sqrt{\left( \frac{\partial f}{\partial x} \right)^2 + \left( \frac{\partial f}{\partial y} \right)^2} \approx$$



- **Gradient Orientation**

- direction of the “steepest ascend”
- orthogonal to object boundaries in the image

- Edge at a point is orthogonal to the gradient vector at that point.
  - So the direction angle of the edge in this example is  $\alpha - 90^\circ = 45^\circ$
  - All edge points in the figure have the same gradient, so that entire edge segment is in the same direction.
- 



## Ekstrema warunkowe wprowadzenie. Warunek konieczny.

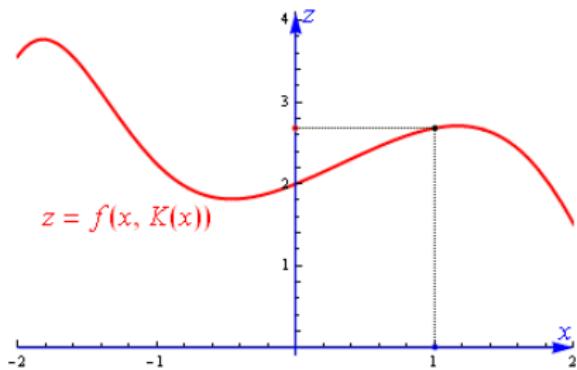
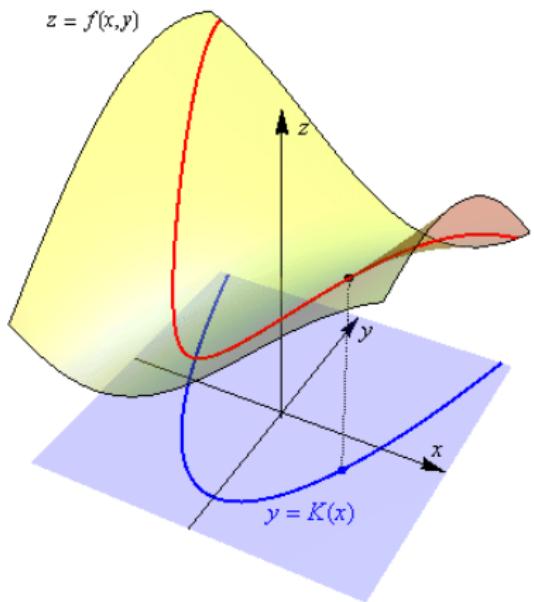
Wektor  $v$  jest styczny do hiperpowierzchni  $H$  w punkcie  $z_0$  (co zapiszemy jako  $v \in T_{z_0} H$ ), jeżeli istnieje krzywa  $\gamma : (-\epsilon, \epsilon) \rightarrow H$  klasy  $C^1$  taka że:

$$\gamma'(0) = v, \quad \gamma(0) = z_0.$$

Jeżeli  $\Phi$  przyjmuje ekstremum lokalne związane przez  $H$  w punkcie  $z_0 \in H$  i istnieje  $\Phi'(z_0)$  to  $\forall v \in T_{z_0} H, \Phi'(z_0)(v) = 0$ ,  $T_{z_0} H \subset \text{Ker } \Phi'(z_0)$  - z opisu hiperpowierzchni przez wektory styczne  $\Phi : U \rightarrow \mathbb{R}$   $U \subset \mathbb{R}^N$  i niech  $H \subset \mathbb{R}^N$  będzie hiperpowierzchnią klasy  $C^k$   $n$ -wymiarową. Mamy określoną funkcję:  $\phi : \Phi|_H : U \cap H \rightarrow \mathbb{R}$ .

Mówimy że  $\Phi$  przyjmuje max (min) lokalne związane przez  $H$  (lub inaczej mówiąc: **warunkowe**) w punkcie  $z_0 \in H$  jeśli istnieje  $\epsilon > 0$  taki że dla dowolnego  $z \in B(z_0, \epsilon) \cap H$  (*i tylko takiego!!*)

$$f(z) \leq f(z_0).$$



# Funkcja Lagrange'a.

Problem optymalizacji z ograniczeniem dla  $f(x)$  pod warunkiem  $h(x) = 0$  można sprowadzić do problemu optymalizacji bez ograniczeń funkcji

$$L(x, \lambda) = f(x) - \lambda \cdot h(x),$$

zwanej funkcją Lagrange'a, bowiem warunki pierwszego rzędu optymalizacji tej funkcji będą dokładnie równe warunkom

$$\nabla f(x) - \lambda h(x) = 0 \quad , \quad h(x) = c.$$

Gradient ograniczenia w punkcie optymalnym musi być różny od wektora zerowego.

## Przykład.

Wyznaczmy najmniejszą i największą wartość funkcji

$$F(x, y, z) = x - y - 2z$$

na przecięciu się dwóch walców

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

(poźniej powiemy sobie, że to przykład **programowania kwadratowego**).

Podany warunek można opisać za pomocą poziomicy zerowej funkcji  $G(x, y, z) = (x^2 + z^2 - 1, y^2 + z^2 - 1)$ . Spośród punktów poziomicy  $G = 0$  tylko dwa nie są regularne:  $(0, 0, 1)$  oraz  $(0, 0, -1)$ .

Poza tymi dwoma punktami możemy zastosować metodę mnożników Lagrange'a, która sprowadza się do wyznaczenia rozwiązań układu równań:

$$\left\{ \begin{array}{l} \frac{\partial F}{\partial x} = \lambda_1 \frac{\partial g_1}{\partial x} + \lambda_2 \frac{\partial g_2}{\partial x} \\ \frac{\partial F}{\partial y} = \lambda_1 \frac{\partial g_1}{\partial y} + \lambda_2 \frac{\partial g_2}{\partial y} \\ \frac{\partial F}{\partial z} = \lambda_1 \frac{\partial g_1}{\partial z} + \lambda_2 \frac{\partial g_2}{\partial z} \\ g_1(x, y, z) = 0 \\ g_2(x, y, z) = 0 \end{array} \right. \quad \text{czyli} \quad \left\{ \begin{array}{l} 1 = 2\lambda_1 x \\ -1 = 2\lambda_2 y \\ -2 = 2(\lambda_1 + \lambda_2)z \\ x^2 + z^2 - 1 = 0 \\ y^2 + z^2 - 1 = 0 \end{array} \right.$$

Układ ten ma dwa rozwiązania

$$-x = y = z = 2 - \sqrt{2},$$

przy czym  $\lambda_1 = \lambda_2 = -2 - \sqrt{2}$ , oraz

$$x = -y = -z = 2 - \sqrt{2}$$

przy czym  $\lambda_1 = \lambda_2 = 2 - \sqrt{2}$ .

Wartość funkcji  $F$  w tych punktach wynosi

$$F(-2 - \sqrt{2}, 2 - \sqrt{2}, 2 - \sqrt{2}) = -2\sqrt{2}$$

oraz

$$F(-2 - \sqrt{2}, 2 - \sqrt{2}, 2 - \sqrt{2}) = 2\sqrt{2}.$$

W obu punktach nieregularnych poziomicy  $G = 0$  mamy  
 $F(0, 0, -1) = 2$  oraz  $F(0, 0, 1) = -2$ .

Po porównaniu tych wartości:  $-2\sqrt{2} < -2 < 2 < 2\sqrt{2}$  stwierdzamy, że największą wartość na poziomicy  $G = 0$  równą  $2\sqrt{2}$  funkcja  $F$  osiąga w punkcie  $(2 - \sqrt{2}, -2 - \sqrt{2}, -2 - \sqrt{2})$ , a najmniejszą, równą  $-2\sqrt{2}$  w punkcie  $(-2 - \sqrt{2}, 2 - \sqrt{2}, 2 - \sqrt{2})$ .

# Przykład zastosowanie ekstremum warunkowego - kodowanie.

**Kody przedrostkowe.** Wartość oczekiwana  $L$  długości słowa kodującego  $\varphi$  jest dana wzorem

$$L = E(\text{lgth}(\varphi)) := \sum_{s \in S} p(s) \cdot \text{lgth}(\varphi(s)) = \sum_i p_i \cdot l_i.$$

Jak dobrać wartości  $l_i$  by minimalizować wartość oczekiwanaą ilości pamięci?

Ponieważ wiemy jakie długości są dopuszczalne, dostajemy  
**problem minimalizacji**

$$L(l_1, \dots, l_n) := \sum_i p_i l_i$$

przy warunku

$$\sum_i d^{-l_i} \leq 1.$$

Zapominamy o tym, że są całkowite (dostaniemy przybliżenie), i wtedy możemy zwiększyć  $L$  zakładając równość!

# Metody optymalizacyjne.

Metody wyznaczania optymalnych rozwiązań, czyli rozwiązań najlepszych ze względu na przyjmowane kryteria. Na ogólny sposób zagadnienie może mieć różne kryteria (minimalny nakład pracy, maksymalny zysk, minimalna złożoność obliczeniowa czy wymagana pamięć itd.), co prowadzi do różnych rozwiązań optymalnych. Podamy ogólne metody, pozwalające je badać w jednolity sposób - niezależnie od konkretnej postaci kryterium.

Na ogólny jest to badanie ekstremum (niekiedy lokalnego, niekiedy globalnego) pewnej **funkcji celu** z uwzględnieniem ograniczeń - na ogólny w postaci równań (warunków brzegowych) lub nierówności. Np.

$$f(x_1, x_2, \dots, x_k) \rightarrow \text{MIN.} \quad \text{o ile} \quad (x_1, x_2, \dots, x_k) \in K,$$

$$f(x_1, x_2, \dots, x_k) \rightarrow \text{MAX.} \quad \text{o ile} \quad g(x_1, x_2, \dots, x_k) = 0 \quad \dots$$

Na ogół są to metody krokowe (iteracyjne), czyli korzysta się z dyskretyzacji pojęć znanych z analizy, a służących do określania ekstremum (z ograniczeniami wynikającymi z ich dyskretnej postaci).

O ile funkcja celu jest zależna od jednej zmiennej, to stosujemy metody z analizy funkcji jednej zmiennej (znane z kursu analizy matematycznej). Opiszemy bardziej realistyczną sytuację funkcji wielu zmiennych.

- Metody gradientowe (ang. *descent methods*)
  - Metoda najszybszego spadku (ang. *steepest descent method*)
  - Metoda Newtona
  - Metoda Marquardta
  - Metoda sprzężonego gradientu Fletcher-Reevesa
  - Metody quasinewtonowskie:
    - \* Metoda Davidon-Fletcher-Powella (DFP)
    - \* Metoda Broyden-Fletcher-Goldfarb-Shannona (BFGS)

# Proste algorytmy.

Metodą “naiwną” jest minimalizacja funkcji wielu zmiennych względem współrzędnych.

1. Ustalamy  $(x_0, y_0)$  - punkt startowy.
2. Ustalamy krok  $k$ .
3. Kontrolujemy wartości funkcji w 4 punktach:  
 $(x_0, y_0 + k), (x_0, y_0 - k), (x_0 + k, y_0), (x_0 - k, y_0)$ .
4. Jeżeli w jednym z punktów wartość funkcji  $f(x, y)$  jest mniejsza niż w punkcie  $(x_0, y_0)$  to kładziemy go za punkt  $(x_0, y_0)$  i powtarzamy procedurę z poprzedniego kroku.
5. Jeśli w poprzednim punkcie algorytmu nie znaleziono takiego punktu kratowego, to zmniejszamy krok i powtarzamy punkt 3.

Metoda jest zawodna, bo kontrolujemy tylko 4 główne kierunki...

# Metoda najszybszego spadku

... jest dość prostą gradientową, iteracyjną metodą optymalizacji bez ograniczeń. Każda iteracja polega na wykonaniu kroku w lokalnie najlepszym kierunku czyli w kierunku wyznaczonym przez **ujemny gradient** (niekiedy nazywany *antygradientem*). Przy czym długość kroku może być stała (metoda gradientu prostego) lub wybrana tak by spadek wartości funkcji celu w nowym punkcie był jak największy (dobra metoda optymalizacji jednowymiarowej). Uwaga: jednak nie musi prowadzić do ekstremum lokalnego!

Oznaczenia:

$x_0$  – pierwsze przybliżenie rozwiązania (punkt startowy)

$x_i$  – i-te przybliżenie rozwiązania

i – numer iteracji (na początku równa 0)

Wybieramy punkt początkowy i obliczamy w nim wartość funkcji i wartość gradientu tej funkcji w tym punkcie.

### Algorytm:

1. sprawdź czy punkt  $x_i$  spełnia warunek stopu tzn.

$$|\nabla f(x_i)|^2 \leq \varepsilon,$$

jeśli tak to zakończ ( $x_i$  jest rozwiązaniem)

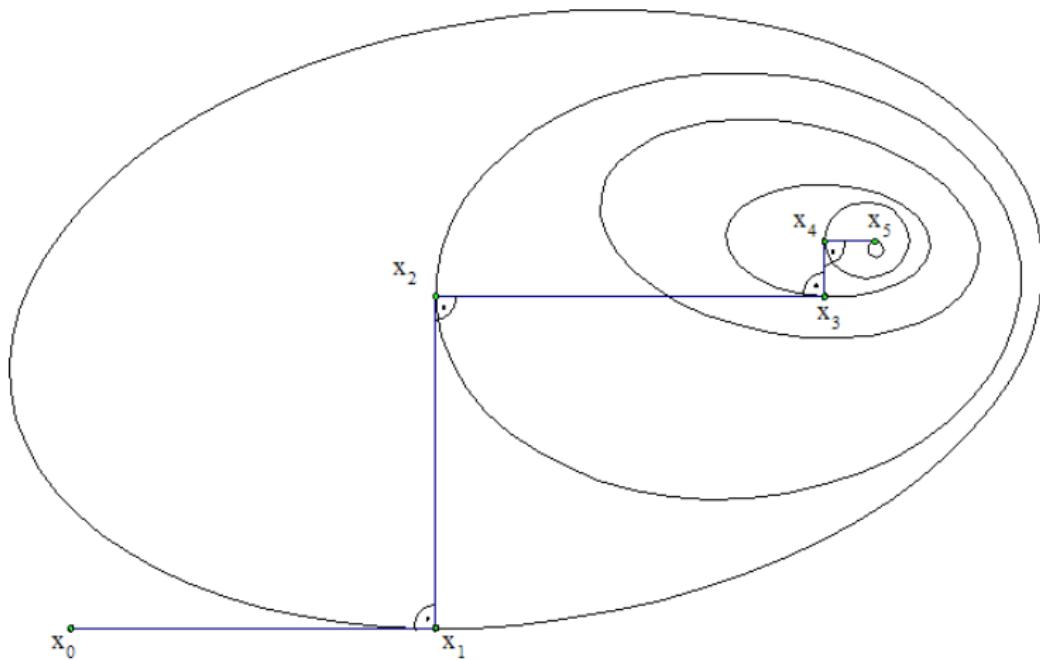
2.  $i := i + 1$ , obliczamy w nowym punkcie wartość funkcji i gradientu,

3.  $x_i := x_{i-1} - \lambda_i \nabla f(x_{i-1})$ , gdzie  $\lambda_i > 0$  – długość kroku minimalizująca jednowymiarową funkcję

$$f(\lambda_i) = f(x_{i-1} - \lambda_i \nabla f(x_{i-1})),$$

4. przejdź do pkt. 1

Przykładowa trajektoria:

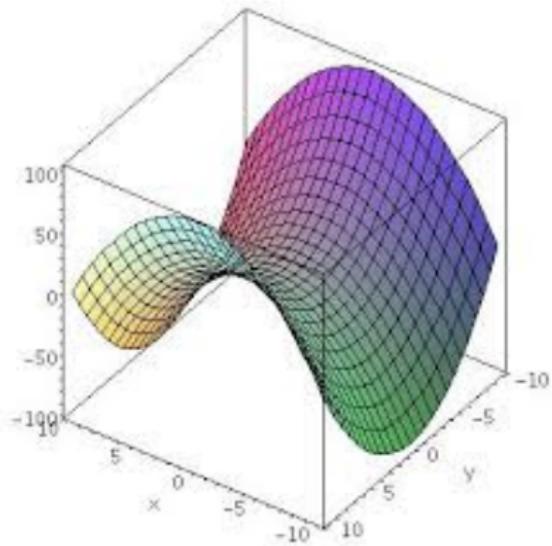


Uwidoczniono poziomice dwuwymiarowej funkcji celu i kolejne punkty eksperymentów.

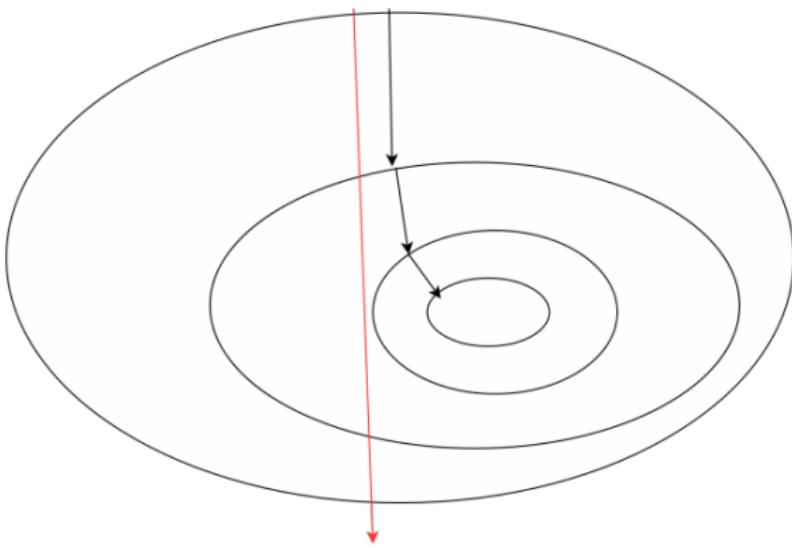
## Organiczenia metody.

- ▶ Jeśli funkcja  $f$  ma więcej niż jedno minimum lokalne, otrzymany wynik może zależeć od punktu początkowego.
- ▶ Wybierając różne punkty początkowe, możemy otrzymać inne rozwiązania, a więc na końcu możemy wybrać najlepsze (najmniejsze) z nich.
- ▶ W sytuacjach gdy istnieje wiele minimów lokalnych wykorzystuje się sposoby dające możliwość wyjścia z optimum lokalnego poprzez rozszerzenie lokalnych poszukiwań.

Uwaga: metody gradientowe mają jednak słabe punkty - te procedury załamują się w otoczeniu punktów siodłowych i takie przypadki trzeba wykluczyć!!



Przypomnienie - w metodach gradientowych oprócz doboru kierunku istotne jest też określenie jak daleko można podążyć w danym kierunku. Na czerwono zaznaczono sytuację dla zbyt dużego współczynnika uczenia.



Zbyt mała jego wartość powodowałaby dużą liczbę drobnych kroków.

Optymalizacja kierunkowa służy do odpowiedniego doboru  $\lambda$ .