

Wszystkie udostępniane materiały, skrypty, notatniki są wyłącznie przeznaczone do użytku prywatnego w celu łatwiejszego opanowania wiedzy.

Nie wolno ich rozprzestrzeniać, ani umieszczać w Internecie.

# Optymalizacja

# Optymalizacja

W zagadnieniach optymalizacyjnych podobnie jak w przypadku poszukiwania pierwiastków, będziemy poszukiwali argumentów, tylko w tym przypadku będących minimum lub maksimum.

## **Przypomnienie:** Warunki istnienia ekstremum

Niech funkcja  $f(x)$  określona w pewnym przedziale  $[a, b]$  osiąga w punkcie wewnętrznym  $x_0$  tego przedziału ekstremum (maksimum lub minimum). Jeśli istnieje w tym punkcie pochodna skończona  $f'(x_0)$ , to:  $f'(x_0) = 0$

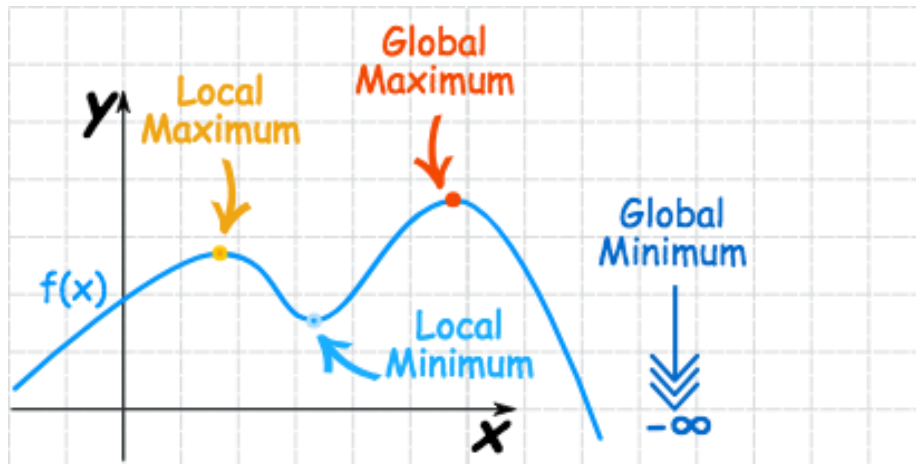
Jeżeli funkcja  $f(x)$  spełnia następujące założenia:

- ma pochodną  $f''(x)$  w pewnym otoczeniu  $x_0$ ,
- $f''(x)$  jest ciągła w punkcie  $x_0$ ,
- $f'(x_0) = 0$  i  $f''(x_0) \neq 0$ ,

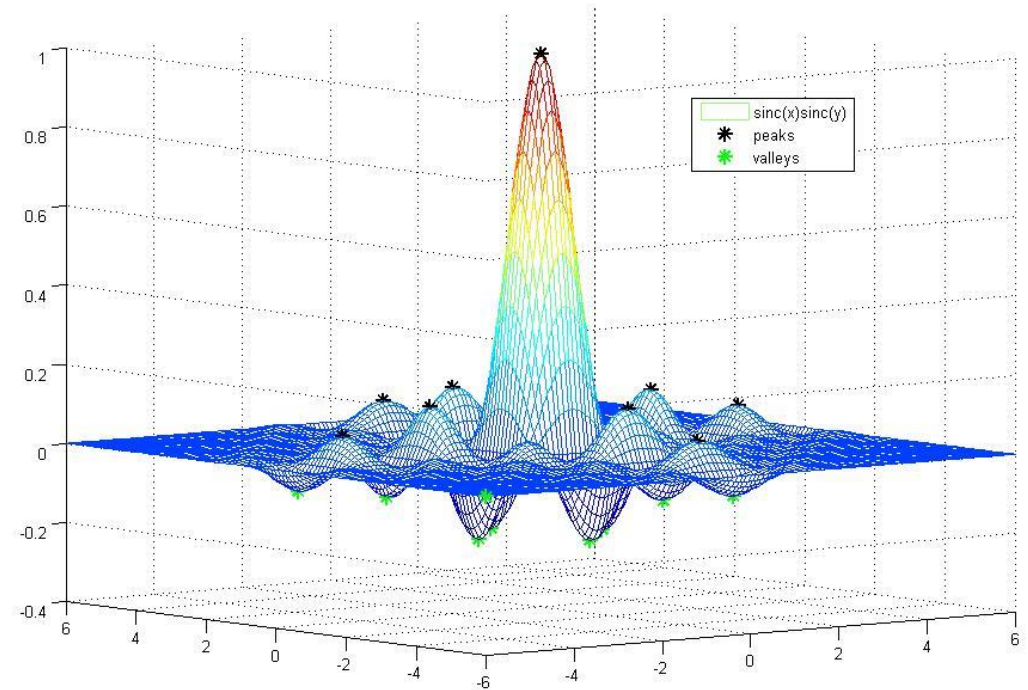
to funkcja  $f(x)$  ma w punkcie  $x_0$ :

- a) minimum właściwe, gdy  $f''(x) > 0$
- b) maksimum właściwe, gdy  $f''(x) < 0$

# Optymalizacja



<https://medium.com/self-study-calculus/extrema-maxima-minima-4e2e6956a44a>



# Optymalizacja

Znajdź  $\mathbf{x}$ , które minimalizuje lub maksymalizuje  $f(\mathbf{x})$  pod warunkiem:

$$\begin{aligned} d_i(\mathbf{x}) &\leq a_i, & i &= 1, 2, \dots, m \\ e_i(\mathbf{x}) &= b_i, & i &= 1, 2, \dots, p \end{aligned}$$

gdzie:

$\mathbf{x}$  jest  $n$ -wymiarowym wektorem

$f(\mathbf{x})$  – funkcja celu

$d_i(\mathbf{x})$  - ograniczenia nierównościowe

$e_i(\mathbf{x})$  - ograniczenia równościowe

$a_i$  i  $b_i$  - wartości stałe

# Optymalizacja

Problemy optymalizacyjne można podzielić ze względu na formę  $f(\mathbf{x})$ :

$f(\mathbf{x})$  i ograniczenia liniowe – **programowanie liniowe**

$f(\mathbf{x})$  jest funkcją kwadratową i/lub co najmniej jedno ograniczenie jest kwadratowe – **programowanie kwadratowe**

$f(\mathbf{x})$  jest funkcją nieliniową i/lub co najmniej jedno ograniczenie jest nieliniowe – **programowanie nieliniowe**

Dodatkowo, w zależności od występowania ograniczeń optymalizacja może być z ograniczeniami lub bez.

Minimalizacja  $f(\mathbf{x})$  jest ekwiwalentem maksymalizacji  $-f(\mathbf{x})$

# Optymalizacja

Problemy optymalizacyjne można podzielić ze względu na formę  $f(\mathbf{x})$ :

$f(\mathbf{x})$  i ograniczenia liniowe – **programowanie liniowe**

$f(\mathbf{x})$  jest funkcją kwadratową i/lub co najmniej jedno ograniczenie jest kwadratowe – **programowanie kwadratowe**

$f(\mathbf{x})$  jest funkcją nieliniową i/lub co najmniej jedno ograniczenie jest nieliniowe – **programowanie nieliniowe**

Dodatkowo, w zależności od występowania ograniczeń optymalizacja może być z ograniczeniami lub bez.

Minimalizacja  $f(\mathbf{x})$  jest ekwiwalentem maksymalizacji  $-f(\mathbf{x})$

# Optymalizacja – metoda złotego podziału

**Metoda złotego podziału:** jest to odpowiednik metody bisekcji poszukiwania miejsc zerowych funkcji.

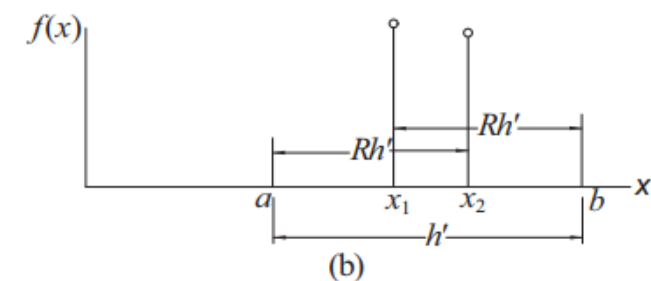
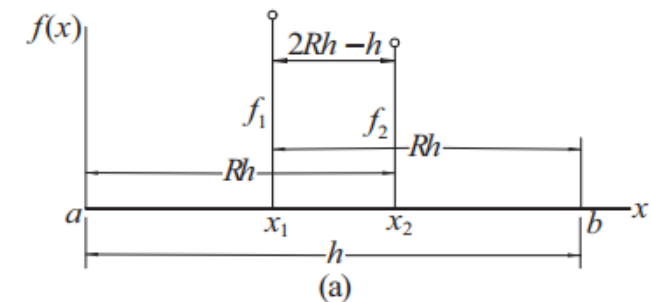
Założmy, że minimum funkcji  $f(x)$  znajduje się w przedziale  $(a, b)$  o szerokości  $h$ .

W celu zawężenia przedziału wyznaczmy wartość funkcji dla  $x_1 = b - Rh$  i  $x_2 = a + Rh$

Jeżeli  $f(x_1) > f(x_2)$ , minimum leży w przedziale  $(x_1, b)$  w przeciwnym przypadku  $(a, x_2)$ .

Założmy  $f(x_1) > f(x_2)$ , wtedy podstawiamy  $x_1 \rightarrow a$  i  $x_2 \rightarrow x_1$ , generując nowy przedział  $(a, b)$  o szerokości  $h' = Rh$ .

Kolejne zawężenie przedziału przeprowadzamy wyznaczając wartość funkcji dla  $x_2 = a + Rh'$  i powtarzamy analogiczne procedurę.





# Optymalizacja – metoda złotego podziału

**Metoda złotego podziału:**

Z rysunków wynika, że  $x_2 - x_1 = 2Rh - h$  co odpowiada  $x_1 - a = h' - Rh'$

W konsekwencji otrzymujemy  $2Rh - h = h' - Rh'$

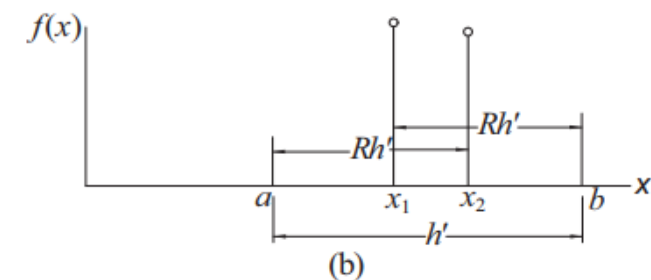
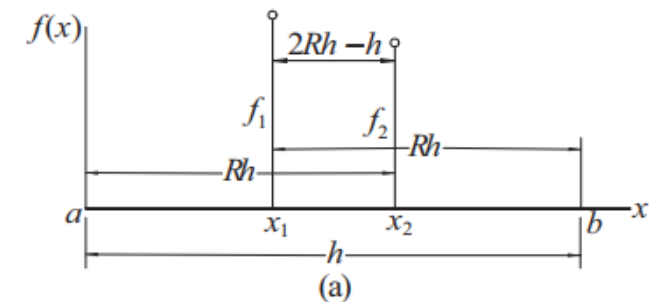
Podstawiając  $h' = Rh$  i dzieląc przez  $h$  otrzymujemy:

$$2R - 1 = R(1 - R)$$

Rozwiązaniem jest złoty podział wynoszący

$$R = \frac{-1 + \sqrt{5}}{2} = 0.618033989 \dots$$

Zmniejszenie przedziału o  $R$  nie jest tak dobre jak o wartość 0.5 jak w przypadku bisekcji, ale za to redukujemy o jeden liczbę wyznaczania wartości w każdej iteracji.



# Optymalizacja – metoda złotego podziału

## Metoda złotego podziału:

Zmniejszenie przedziału o  $R$  nie jest tak dobre jak o wartość 0.5 jak w przypadku bisekcji, ale za to redukujemy o jeden liczbę wyznaczania wartości w każdej iteracji.

Liczba operacji potrzebna do zredukowania przedziału początkowego  $|b - a|$  do poziomu  $\varepsilon$  Jest wyrażona wzorem:

$$\varepsilon = R^n |b - a|$$

czyli  $n$ :

$$n = \frac{\ln\left(\frac{\varepsilon}{|b - a|}\right)}{\ln R} = -2.078087 \ln\left(\frac{\varepsilon}{|b - a|}\right)$$

Złoty podział jest i był wykorzystywany w wielu celach: architektura, sztuka.

# Optymalizacja – wykorzystanie interpolacji

## Interpolacja paraboliczna:

Interpolacja z wykorzystaniem funkcji kwadratowej często jest dobrym przybliżeniem kształtu  $f(x)$  blisko ekstremum.

Idea jest bardzo prosta, jeżeli mamy 3 punkty pomiędzy którymi znajduje się ekstremum możemy dopasować parabolę.

Następnie, liczymy pochodną, którą przyrównujemy do zera, po przekształceniach:

$$x_3 = \frac{f(x_0)(x_1^2 - x_2^2) + f(x_1)(x_2^2 - x_0^2) + f(x_2)(x_0^2 - x_1^2)}{2f(x_0)(x_1 - x_2) + 2f(x_1)(x_2 - x_0) + 2f(x_2)(x_0 - x_1)}$$

Gdzie  $x_0, x_1, x_2$  są punktami początkowymi, a  $x_3$  wartością maksimum odpowiadającą tym punktom.

Po wyznaczeniu  $x_3$

1. W nowej iteracji podstawiamy  $z_0 = z_1, z_1 = z_2, z_2 = z_3$  (podobnie jak w metodzie siecznej)
2. Lub stosujemy podejście przedziałowe (podobnie jak w m. złotego podziału lub bisekcji)

# Optymalizacja – wykorzystanie interpolacji

**Interpolacja paraboliczna, przykład, wariant przedziałowy:**

Wykorzystajmy interpolację paraboliczną do wyznaczenia maximum funkcji

$$f(x) = 2 \sin x - \frac{x^2}{10}, \text{ dla punktów początkowych: } x_0 = 0, x_1 = 1, x_2 = 4$$

Podstawiamy, a wyznaczyć wartości funkcji w tych punktach:

$f(x_0) = 0, f(x_1) = 1, f(x_2) = -3.1136$ , podstawiając do równania na  $x_3$ :

$$x_3 = \frac{0(1^2 - 4^2) + 1.5829(4^2 - 0^2) + (-3.1136)(0^2 - 1^2)}{2(0)(1 - 4) + 2(1.5829)(4 - 0) + 2(-3.1136)(0 - 1)} = 1.5055$$

$$f(x_3) = f(1.5055) = 1.7691$$

Następnie, analogicznie jak w metodzie złotego podziału określamy punkt, który powinien być wyeliminowany. Ponieważ wartość funkcji dla nowego punktu jest wyższa niż dla  $x_1$ , a nowy  $x$  jest większy od  $x_1$  eliminujemy  $x_0$ . Proces powtarzamy.

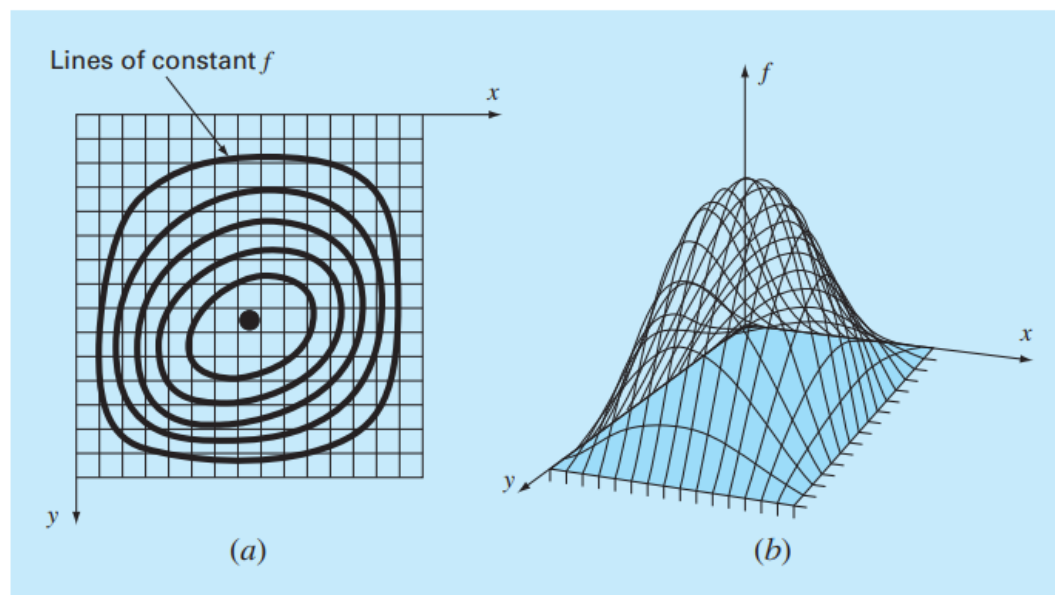
# Optymalizacja wielowymiarowa bez ograniczeń

Podział metod:

Metody optymalizacji wielowymiarowej bez ograniczeń mogą być klasyfikowane na różne sposoby.

Przyjmijmy podział na metody:

- nie wykorzystujące gradientu (ang. *nongradient, direct methods*)
- gradientowe (ang. *gradient, descent (ascent) methods*)



Chapra, Steven C., and Raymond P. Canale. *Numerical methods for engineers*.

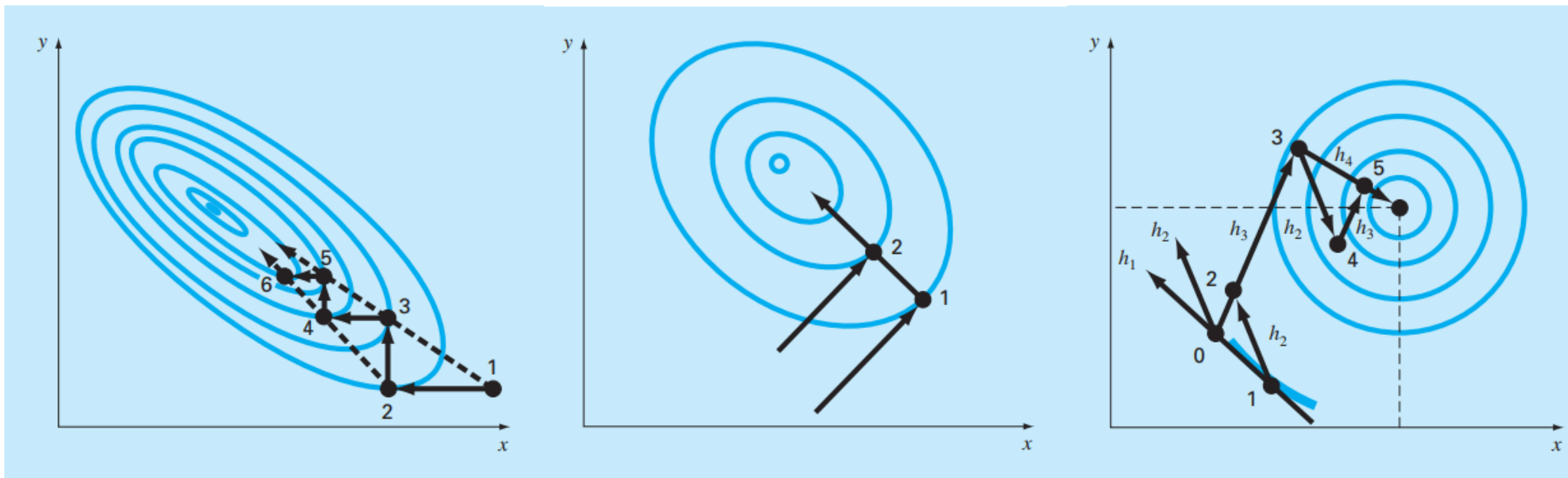
Boston: McGraw-Hill Higher Education, 2010.

# Optymalizacja wielowymiarowa bez ograniczeń

Metody nie wykorzystujące gradientu:

**Poszukiwanie losowe** – rozwiązanie siłowe (ang. *brute force*). Powtarzalne wyznaczanie wartości funkcji dla losowo wybranych argumentów. Ekstremalnie nieefektywne podejście.

**wyszukiwanie jednowymiarowe, wzorzec wyszukiwania** – wyszukiwanie dla jednego wymiaru, gdy pozostałe są stałe. Redukcja do sekwencji poszukiwań jednowymiarowych.



# Optymalizacja wielowymiarowa bez ograniczeń

Metody nie wykorzystujące gradientu:

## Metoda Powella

Wybierz punkt startowy  $\mathbf{x}_0$  w przestrzeni

Wybierz wektory początkowe  $\mathbf{v}_i, i = 1, \dots, n$

(najczęściej  $\mathbf{v}_i = \mathbf{e}_i$  gdzie  $\mathbf{e}_i$  - wektory jednostkowe)

**While:**

**For**  $i = 1, \dots, n$

minimalizuj  $F(\mathbf{x})$  wzdłuż prostej przechodzącej przez  $\mathbf{x}_{i-1}$  w kierunku  $\mathbf{v}_i$ .

Niech minimum będzie punkt  $\mathbf{x}_i$

**End For**

$\mathbf{v}_{n+1} \leftarrow \mathbf{x}_0 - \mathbf{x}_n$

minimalizuj  $F(\mathbf{x})$  wzdłuż prostej przechodzącej przez  $\mathbf{x}_0$  w kierunku  $\mathbf{v}_{n+1}$ .

Niech minimum będzie punkt  $\mathbf{x}_{n+1}$

**if**  $|\mathbf{x}_{n+1} - \mathbf{x}_0| < \varepsilon$  **break**

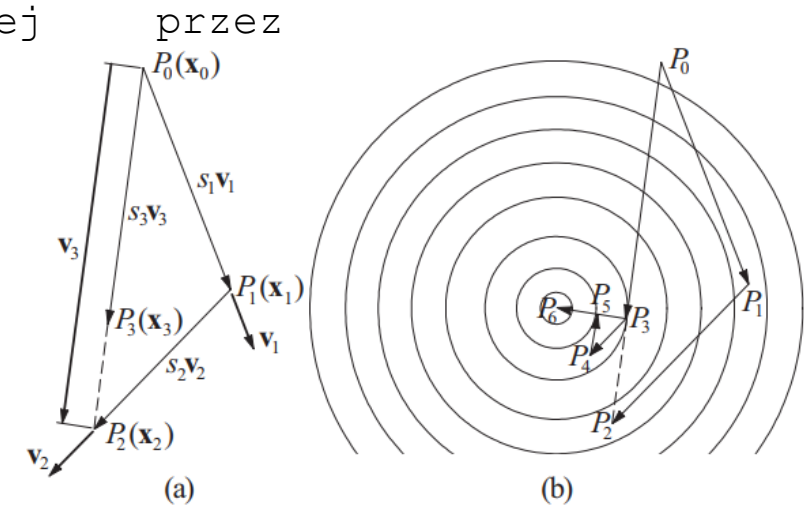
**For**  $i = 1, \dots, n$

$\mathbf{v}_i \leftarrow \mathbf{v}_{i+1}$  ( $\mathbf{v}_1$  jest odrzucane, pozostałe wektory są ponownie wykorzystywane)

**End For**

$\mathbf{x}_0 \leftarrow \mathbf{x}_{n+1}$

**End While**



# Optymalizacja wielowymiarowa bez ograniczeń

**Metody gradientowe:** wykorzystują gradient w celu poszukiwania lokalnych ekstremów.

## Gradient

Założmy, że mamy dwuwymiarową funkcję  $f(x, y)$ . Wartość funkcji interpretujemy jako wysokość. Założmy, że znajdujemy się w punkcie  $(a, b)$  i chcemy ustalić jakie jest nachylenie w dowolnym kierunku.

Jednym z rozwiązań jest zdefiniowanie kierunku wzdłuż nowej osi  $h$ , która tworzy z osią  $x$  kąt  $\theta$ .

Wysokość może być zdefiniowana jako nowa funkcja  $g(h)$ . Jeżeli teraz nasze położenie zdefiniujemy jako początek osi ( $h = 0$ ) to nachylenie w tym kierunku będzie  $g'(0)$  (pochodna kierunkowa):

$$g'(0) = \frac{\partial f}{\partial x} \cos \theta + \frac{\partial f}{\partial y} \sin \theta$$

dla pochodnych cząstkowych wyznaczanych dla  $x = a$  i  $y = b$



# Optymalizacja wielowymiarowa bez ograniczeń

**Gradient** wskazuje w którym kierunku będzie najbardziej strome podejście (największa zmiana wartości)

$$\nabla f = \frac{\partial f}{\partial x} \mathbf{i} + \frac{\partial f}{\partial y} \mathbf{j}$$

Dla przypadku  $n$  wymiarowego gradient ma postać:

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\mathbf{x}) \end{bmatrix}$$

# Optymalizacja wielowymiarowa bez ograniczeń

Hessian:

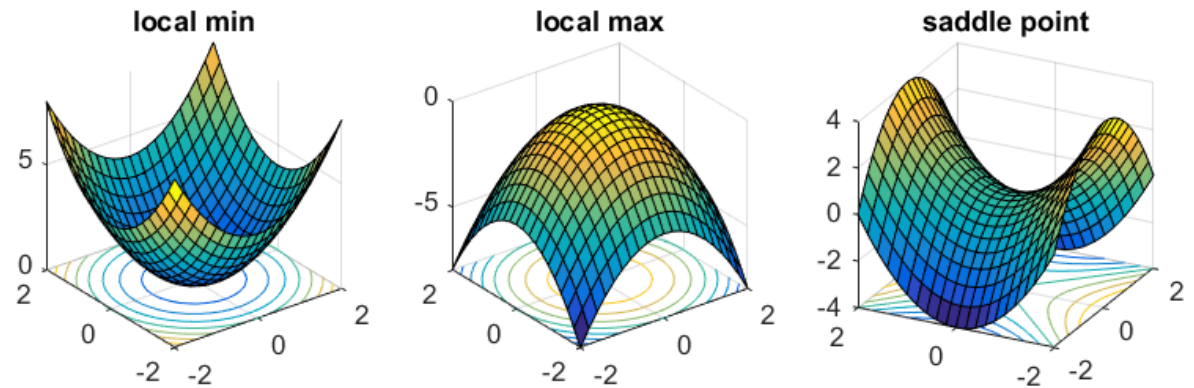
$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}$$

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

Jeżeli  $|H| > 0$  i  $\frac{\partial^2 f}{\partial x^2} > 0$  to  $f(x, y)$  ma lokalne minimum

Jeżeli  $|H| > 0$  i  $\frac{\partial^2 f}{\partial x^2} < 0$  to  $f(x, y)$  ma lokalne maximum

Jeżeli  $|H| < 0$  to  $f(x, y)$  ma punkt siodłowy



<https://www.offconvex.org/2016/03/22/saddlepoints/>

# Optymalizacja wielowymiarowa bez ograniczeń

**W sytuacji gdy obliczenie pochodnych analitycznie jest bardzo trudne lub niewygodne Gradient i Hesjan mogą być policzone numeryczne (np. metoda różnic skończonych)**

$$\frac{\partial f}{\partial x} = \frac{f(x + \delta x, y) - f(x - \delta x, y)}{2\delta x}$$

$$\frac{\partial f}{\partial y} = \frac{f(x, y + \delta y) - f(x, y - \delta y)}{2\delta y}$$

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

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

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{f(x + \delta x, y + \delta y) - f(x + \delta x, y - \delta y) - f(x - \delta x, y + \delta y) + f(x - \delta x, y - \delta y)}{4\delta x \delta y}$$

# Optymalizacja wielowymiarowa bez ograniczeń

## Metoda najszybszego wzrostu/spadku (ang. *Steepest Ascent/Descent Method*)

Naturalną strategią wspinaczki na wzgórze wydaje się określenie maksymalnego nachylenia z punktu widzenia pozycji wyjściowej, a następnie pójście w tym kierunku.

Niestety skuteczność takiego podejścia zależy od wyboru punktu startowego. Jeżeli nie będziemy na stoku prowadzącym bezpośrednio do maksimum, bardzo szybko odejdziemy od najbardziej stromego kierunku wznoszenia.

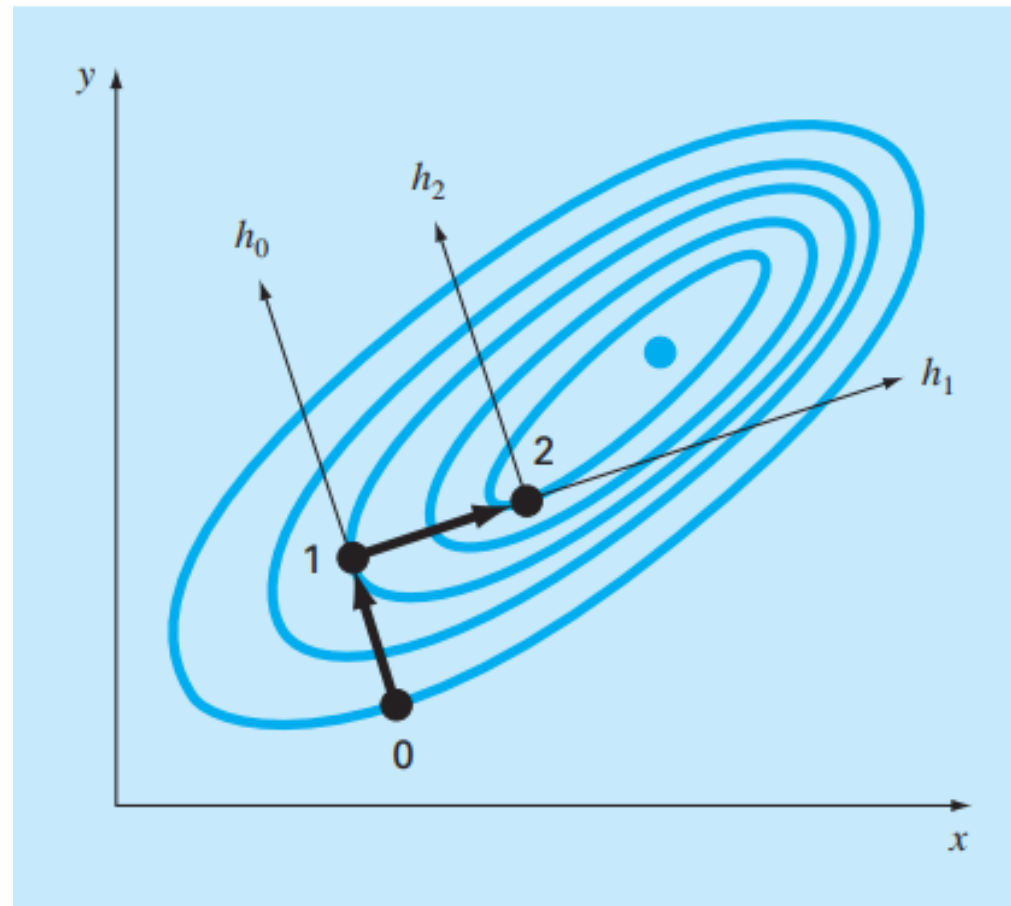
**Rozwiązanie:** idziemy niewielką odległość wzdłuż kierunku nachylenia. Następnie zatrzymujemy się i ponownie wyznaczamy gradient i idziemy w nowym kierunku. Powtarzamy proces. Niestety duże koszty obliczeniowe.

## Rozwiązanie preferowane:

Przemieszczamy się po ustalonej ścieżce wzdłuż początkowego gradientu, aż  $f(x, y)$  przestaje rosnąć. Ten punkt zatrzymania staje się punktem wyjścia, w którym ponownie wyznaczamy gradient. Proces powtarza się aż do osiągnięcia szczytu.

# Optymalizacja wielowymiarowa bez ograniczeń

Metoda najszybszego wzrostu/spadku (ang. *Steepest Ascent/Descent Method*)



# Optymalizacja wielowymiarowa bez ograniczeń

Metoda najszybszego wzrostu/spadku (ang. *Steepest Ascent/Descent Method*)

**Przykład:** Znajdź maksimum funkcji  $f(x, y) = 2xy + 3x - x^2 - 2y^2$  dla w.p.  $x = -1, y = 1$

$$\frac{\partial f}{\partial x} = 2y + 2 - 2x = 0$$

$$\frac{\partial f}{\partial y} = 2x - 4y = 0$$

$$\frac{\partial^2 f}{\partial x^2} = -2$$

$$\frac{\partial^2 f}{\partial y^2} = -4$$

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} = 2$$

Hesjan jest równy  $|H| = -2(-4) - 2^2 = 4$

$|H| > 0$  i  $\frac{\partial^2 f}{\partial x^2} < 0 \rightarrow$  maksimum

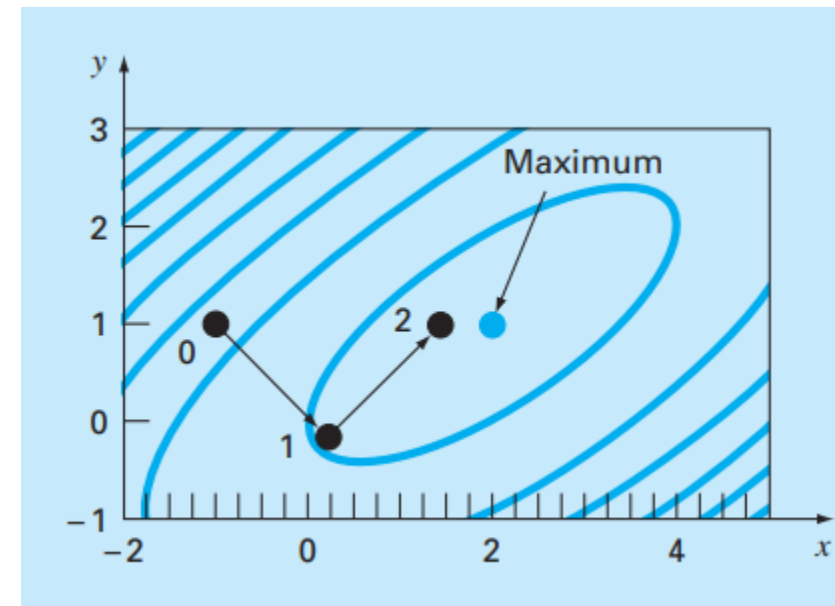
Wyznaczając krok początkowy:  $g(h) = -180h^2 + 72h - 7$

Wyznaczając wprost maksimum ( $h = h^*$ ):

$$g'(h^*) = 0$$

$$-360h^* + 72 = 0$$

$$h^* = 0.2$$



# Optymalizacja wielowymiarowa bez ograniczeń

## Metoda Newtona

Metoda Newtona dla funkcji jednej zmiennej może być uogólniona dla funkcji wielu zmiennych:

$$f(\mathbf{x}) = f(\mathbf{x}_i) + \nabla f^T(\mathbf{x}_i)(\mathbf{x} - \mathbf{x}_i) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_i)^T H_i(\mathbf{x} - \mathbf{x}_i)$$

gdzie  $H_i$  jest macierzą Hessego (Hesjan). W minimum:

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}_j} = 0 \text{ dla } j = 1, 2, \dots, n$$

Zatem:

$$\nabla f = \nabla f(\mathbf{x}_i) + H_i(\mathbf{x} - \mathbf{x}_i) = 0$$

Jeżeli  $H$  jest macierzą nieosobliwą:

$$\mathbf{x}_{i+1} = \mathbf{x}_i - H_i^{-1} \nabla f$$

# Optymalizacja wielowymiarowa bez ograniczeń

## **Metoda Marquardt**

największy spadek gdy daleko od optimum, a gdy blisko metoda Newtona (modyfikacja diagonalu Hessianu)

## **Metody quasi-Newtonowskie (metody zmiennej metryki)**

bazują na oszacowaniu bezpośredniej drogi do optimum w sposób podobny do metody Newtona. Jednak zamiast wykorzystywać macierz Hessego, wykorzystują aproksymację tej macierzy przy użyciu tylko pierwszych pochodnych cząstkowych.

Główne metody:

Davidon-Fletcher-Powell (DFP)

Broyden-Fletcher-Goldfarb-Shanno (BFGS)