

Optymalizacja (minimalizacja) funkcji

Plan wykładu:

1. Sformułowanie problemu
2. Prosty schemat poszukiwania minimum funkcji
3. Kierunkowe metody poszukiwania minimum funkcji
 - a) Metoda złotego podziału
 - b) Metoda interpolacji kwadratowej Powell'a
 - c) Metoda największego spadku pierwszego rzędu
 - d) Metoda sprzężonego gradientu

Sformułowanie problemu, definicje pomocnicze

Zadaniem optymalizacji jest poszukiwanie minimum lub maksimum funkcji (wielu zmiennych). W praktyce problem sprowadza się do poszukiwania minimum czyli takiego punktu dla którego zachodzi

$$\min f(\mathbf{x}) = f(\mathbf{x}^*) \Leftrightarrow \bigwedge_{\mathbf{x} \in R} f(\mathbf{x}^*) < f(\mathbf{x})$$

$$\mathbf{x} = [x_1, x_2, \dots, x_n]^T$$

z warunkami

$$g_j(\mathbf{x}) \leq 0, \quad j = 1, 2, \dots, m$$

$$h_j(\mathbf{x}) = 0, \quad j = 1, 2, \dots, r$$

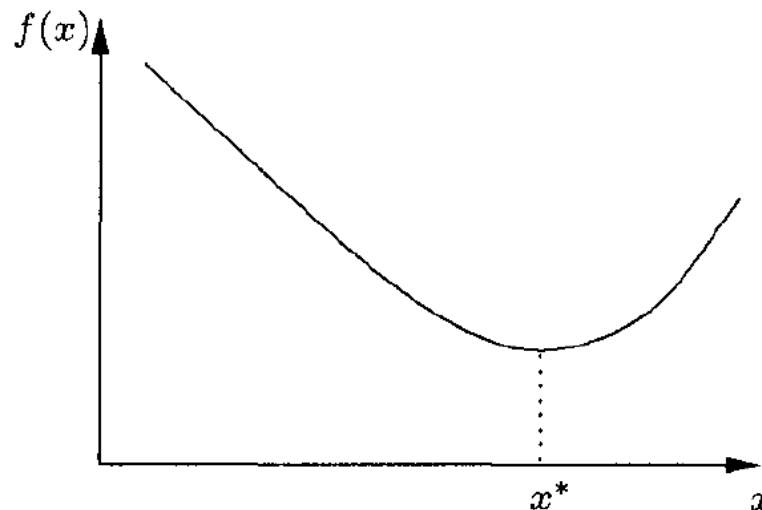
Funkcje: $f(\mathbf{x})$, $g(\mathbf{x})$, $h(\mathbf{x})$ są funkcjami skalarnymi.

$f(\mathbf{x})$ – funkcja celu

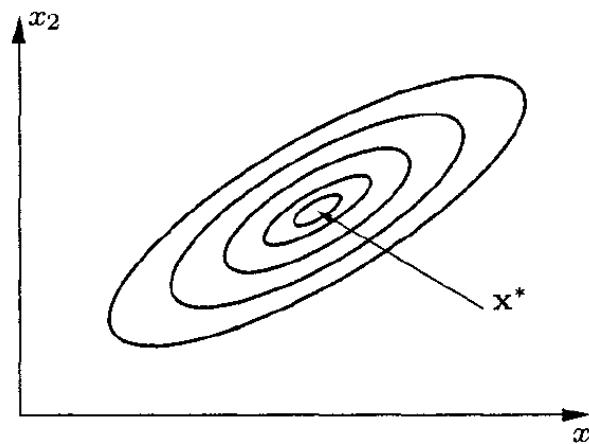
$g(\mathbf{x})$ i $h(\mathbf{x})$ są funkcjami określającymi warunki jakie musi spełniać rozwiązanie (**więzy**)

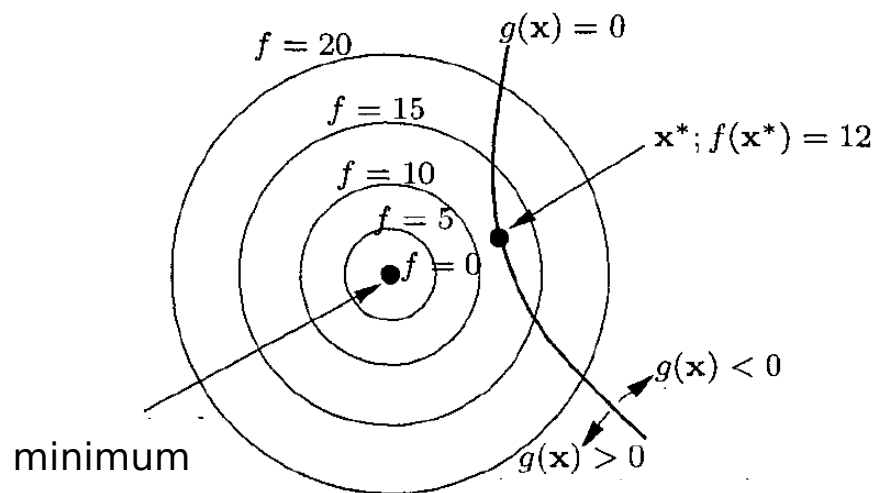
Przykład

Problem jednowymiarowy



Problem dwuwymiarowy





Rys. Przykład poszukiwania minimum z nałożonymi warunkami na rozwiązanie

Trzy przypadki:

1) Problem bez więzów oraz dla $g(x) > 0$

minimum znajdujemy dla $f(x)=0$

2) Jeśli warunkiem jest $g(x)=0$

to minimum znajduje się w punkcie takim że $f(x)=12$

3) Jeśli warunkiem jest $g(x) < 0$

to rozwiązanie znajdziemy w pobliżu punktu w którym $f(x)=12$

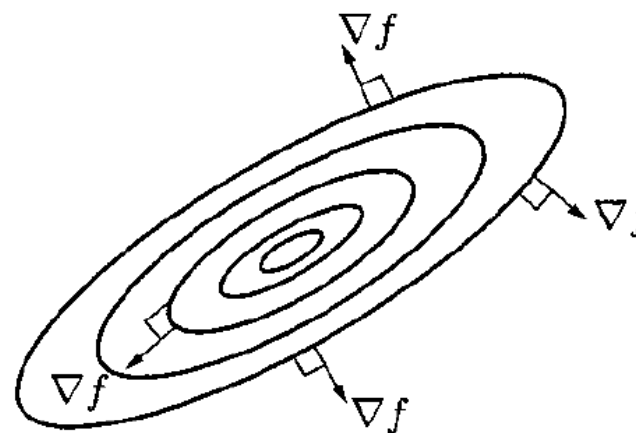
Gradient funkcji - wektor gradientu

Dla funkcji celu

$$f(\mathbf{x}) \in C^2$$

definiujemy funkcję wektorową będącą gradientem funkcji

$$\mathbf{g}(\mathbf{x}) = \nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix}$$



Hesjan (macierz Hessego)

Dla funkcji celu

$$f(\mathbf{x}) \in C^2$$

definiujemy macierz (**hesjan**) której elementami są jej drugie pochodne cząstkowe

$$\begin{aligned} H(\mathbf{x}) &= \left\{ \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} \right\} = \nabla^2 f(\mathbf{x}) \\ &= \begin{bmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2} & \cdots & \cdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_1} & \cdots & \cdots & \cdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} & \cdots & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n^2} \end{bmatrix} \end{aligned}$$

Macierz $H(\mathbf{x})$ jest symetryczna – implikacje numeryczne.

Minimum lokalne, minimum globalne oraz punkt siodłowy funkcji celu

a) Punkt \mathbf{x}^* stanowi **minimum globalne** funkcji jeśli

$$\bigwedge_{\mathbf{x} \in R} f(\mathbf{x}) \geq f(\mathbf{x}^*)$$

b) Punkt \mathbf{x}^* stanowi **minimum lokalne** funkcji jeśli

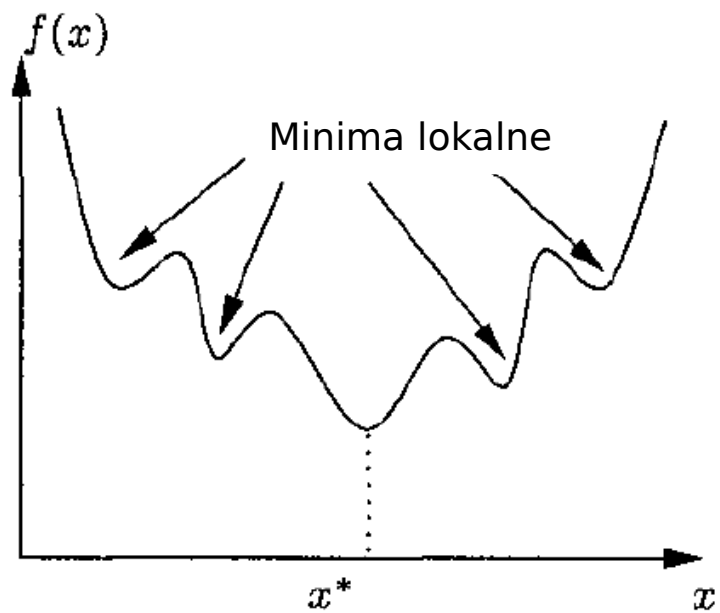
$$\exists \varepsilon : \varepsilon > 0, \varepsilon \in R \quad \bigwedge_{\mathbf{x} : \|\mathbf{x} - \mathbf{x}^*\| < \varepsilon} f(\mathbf{x}) > f(\mathbf{x}^*)$$

c) Punkt

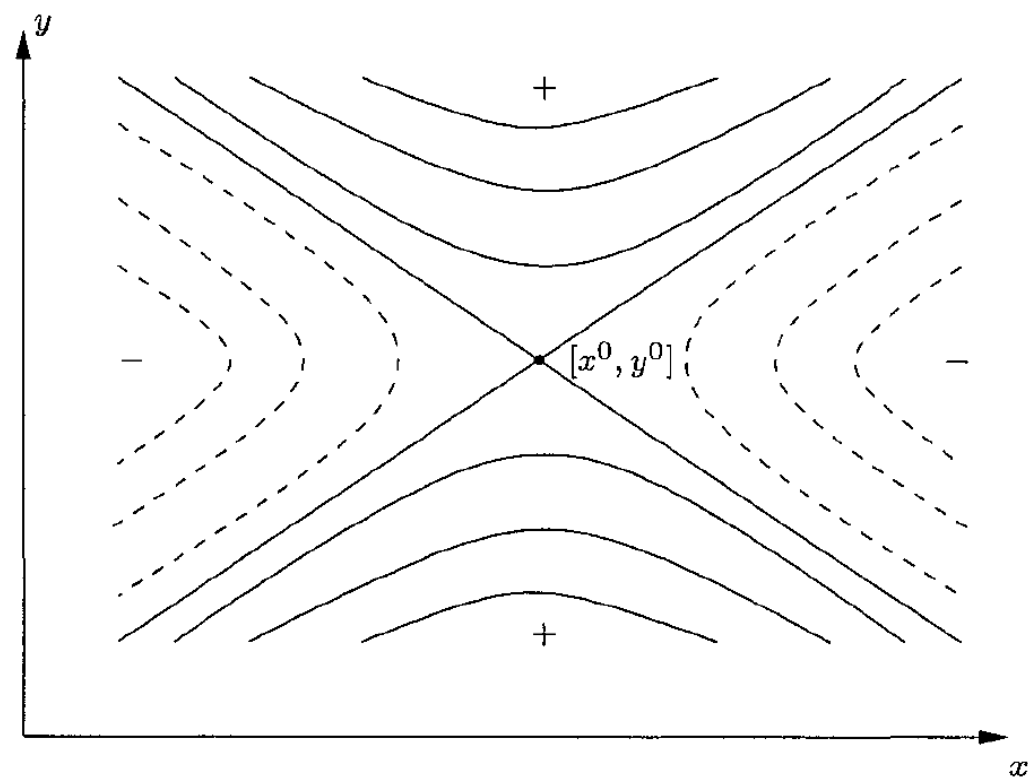
$$\mathbf{x}^* = \begin{bmatrix} x^0 \\ y^0 \end{bmatrix}$$

jest **punktem siodłowym** funkcji jeśli

$$\exists \varepsilon : \varepsilon > 0, \varepsilon \in R \quad \bigwedge_{\substack{\mathbf{x} : \|\mathbf{x} - \mathbf{x}^0\| < \varepsilon \\ \mathbf{y} : \|\mathbf{y} - \mathbf{y}^0\| < \varepsilon}} f(\mathbf{x}, \mathbf{y}^0) \leq f(\mathbf{x}^0, \mathbf{y}^0) \leq f(\mathbf{x}^0, \mathbf{y})$$



Rys. Minima lokalne i minimum globalne



Rys. Punkt siodłowy funkcji

Metoda Newtona poszukiwania minimum funkcji kwadratowej w \mathbb{R}^n

Funkcję kwadratową definiujemy następująco

$$f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T A \mathbf{x} + \mathbf{b}^T \mathbf{x} + c$$

gdzie: A jest pewną macierzą kwadratową oraz

$$\mathbf{x}, \mathbf{b} \in \mathbb{R}^n$$

$$c \in \mathbb{R}$$

Jeśli macierz A jest symetryczna to wówczas zachodzi

$$\nabla f(\mathbf{x}) = A \mathbf{x} + \mathbf{b}$$

Oraz

$$\nabla^2 f(\mathbf{x}) = A$$

$$H(\mathbf{x}) = A$$

Przykład.

Optymalizacja parametrów nieliniowych funkcji falowej cząstki. Funkcja celu jest energią cząstki, która osiąga minimum dla pewnego zestawu wartości tych parametrów.

Jeśli A jest dodatniookreślona to rozwiązanie można łatwo znaleźć, ponieważ

$$\nabla f(\mathbf{x}) = A \mathbf{x} + \mathbf{b} = 0$$

$$\mathbf{x}^* = -A^{-1} \mathbf{b}$$

W metodzie Newtona zakładamy

$$\mathbf{x}^* = \mathbf{x}^i + \delta$$

gdzie: \mathbf{x}^i – przybliżone rozwiązanie w i -tej iteracji

Korzystając z rozwinięcia funkcji w szereg Taylora możemy zapisać

$$\begin{aligned} 0 &= \nabla f(\mathbf{x}^*) = \nabla f(\mathbf{x}^i + \delta) \\ &= \nabla f(\mathbf{x}^i) + H(\mathbf{x}^i) \delta + O(\|\delta\|^2) \end{aligned}$$

Jeśli pominiemy wyrazy rzędu $\|\delta\|^2$ to

$$\nabla f(\mathbf{x}^i) + H(\mathbf{x}^i) \delta = 0$$

W i -tej iteracji poprawiamy rozwiązanie, tj.

$$\mathbf{x}^{i+1} = \mathbf{x}^i + \delta$$

i ostatecznie

$$\mathbf{x}^{i+1} = \mathbf{x}^i - H^{-1}(\mathbf{x}^i) \nabla f(\mathbf{x}^i)$$

Zastosujmy schemat do funkcji kwadratowej

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}A^T\mathbf{x} + \mathbf{b}^T\mathbf{x} + c$$

W pierwszej iteracji dostajemy

$$\begin{aligned}\mathbf{x}^1 &= \mathbf{x}^0 - A^{-1}(A\mathbf{x}^0 + \mathbf{b}) \\ &= \mathbf{x}^0 - \mathbf{x}^0 - A^{-1}\mathbf{b} = -A^{-1}\mathbf{b}\end{aligned}$$

Czyli

$$\mathbf{x}^1 = \mathbf{x}^*$$

Zbieżność do minimum uzyskujemy w jednym kroku.

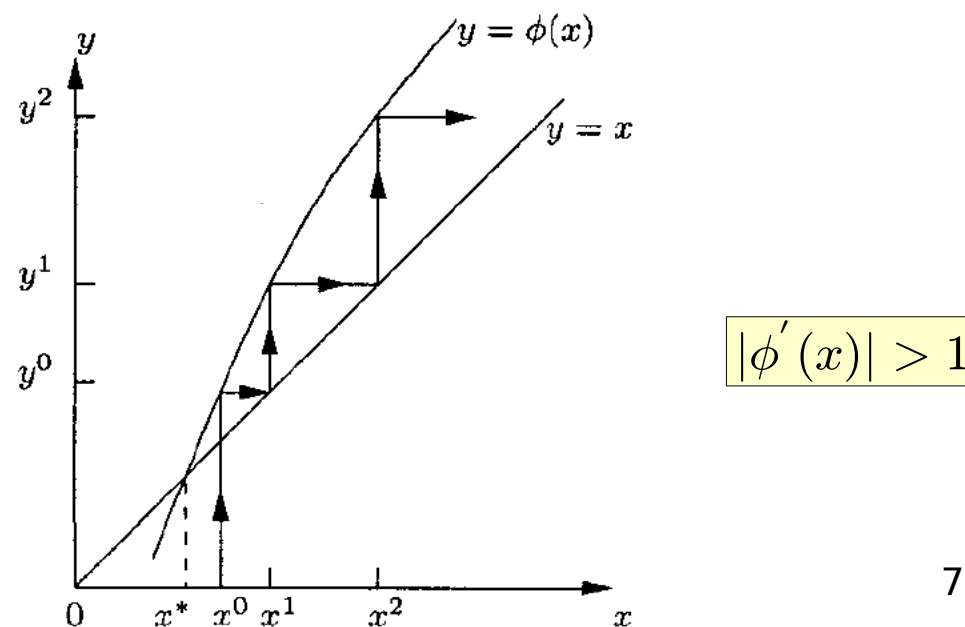
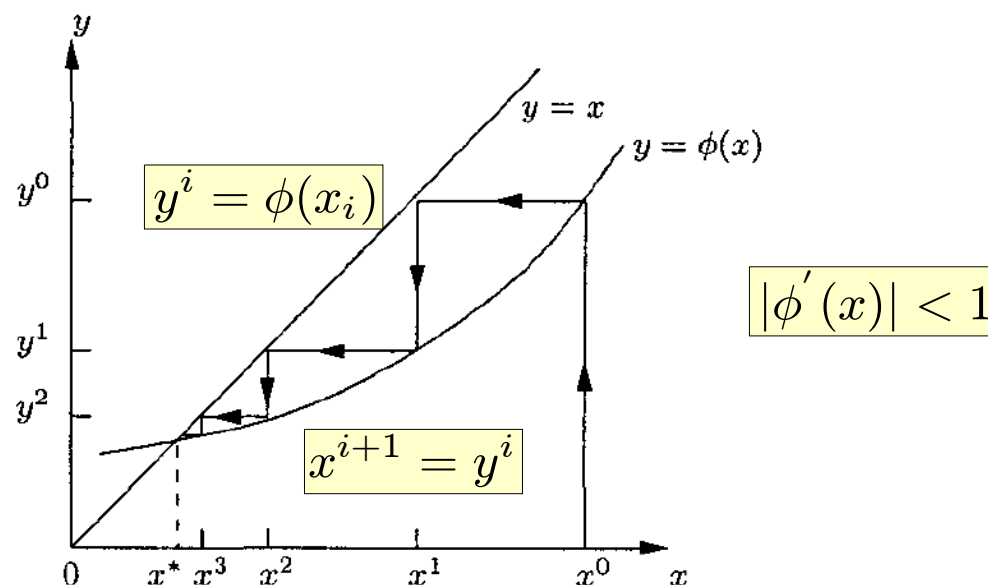
Ale metoda ma wady:

- 1) Metoda nie zawsze jest zbieżna nawet w pobliżu minimum
- 2) Wymaga znalezienia A^{-1} w każdej iteracji

Przykład. Należy znaleźć minimum funkcji

$$f'(x) = 0$$

$$x^{i+1} = x^i - \frac{f^{(1)}(x^i)}{f^{(2)}(x^i)} = \phi(x_i)$$



Kierunkowe metody poszukiwania minimum funkcji

Pochodna kierunkowa funkcji celu

Różniczka zupełna funkcji celu

$$df = \frac{\partial f}{\partial x_1} dx_1 + \dots + \frac{\partial f}{\partial x_n} dx_n = \nabla f(\mathbf{x}) d\mathbf{x}$$

Jeśli wektor \mathbf{u} wyznacza kierunek prostej łączącej punkty \mathbf{x} i \mathbf{x}' to są one ze sobą powiązane

$$\mathbf{x}(\lambda) = \mathbf{x}' + \lambda \mathbf{u}$$

Dla bardzo małych zmian wartości λ możemy zapisać

$$d\mathbf{x} = \mathbf{u} d\lambda$$

Na prostej łączącej (**ustalone punkty**) \mathbf{x} i \mathbf{x}' wartość funkcji celu zależna będzie od nowej zmiennej

$$F(\lambda) = f(\mathbf{x}' + \lambda \mathbf{u}) = f(\mathbf{x})$$

Liczmy różniczkę zupełną dla funkcji celu zależnej od λ

$$dF = df = \nabla^T f(\mathbf{x}) \mathbf{u} d\lambda$$

Możemy teraz wyrazić pochodną kierunkową funkcji celu w punkcie \mathbf{x} dla kierunku \mathbf{u} następująco

$$\frac{dF(\lambda)}{d\lambda} = \left. \frac{df(\mathbf{x})}{d\lambda} \right|_{\mathbf{u}} = \nabla^T f(\mathbf{x}) \mathbf{u}$$

Jak wykorzystać pochodną kierunkową do znalezienia minimum funkcji?

Startując z określonego punktu \mathbf{x}_0 poszukujemy kolejnego przybliżenia tj. \mathbf{x}_1 w kierunku spadku wartości funkcji. W ten sposób wyznaczamy ciąg kolejnych przybliżeń

$$\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \dots$$

poszukiwanego minimum. Procedurę iteracyjną przerywamy, gdy spełniony jest jeden z warunków:

a) $\|\mathbf{x}^{i+1} - \mathbf{x}^i\| < \varepsilon$

b) $\nabla f(\mathbf{x}) = 0$

c) w kolejnych iteracjach rośnie wartość

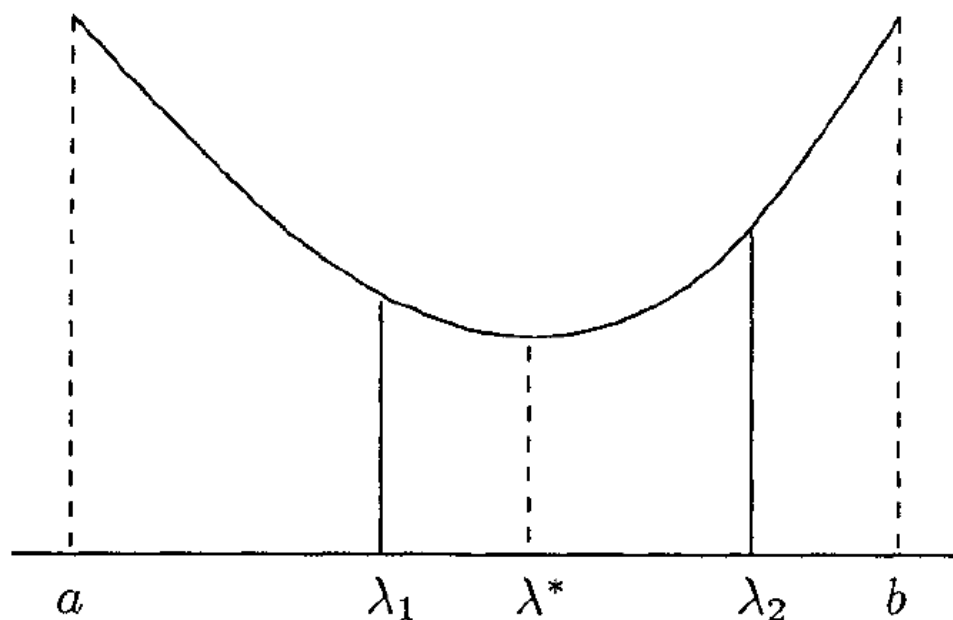
$$\|\mathbf{x}^{i+1} - \mathbf{x}^i\|$$

co oznacza brak zbieżności

Jeśli korzystamy z pochodnej kierunkowej, to musimy ją wyliczać w każdej iteracji. Ponadto, warunkiem na to aby kolejny wyznaczony punkt \mathbf{x}^{i+1} znajdował się bliżej rzeczywistego rozwiązania jest

$$\left. \frac{df(\mathbf{x}^i)}{d\lambda} \right|_{\mathbf{u}^{i+1}} = \nabla^T f(\mathbf{x}^i) \mathbf{u}^{i+1} < 0$$

Metoda złotego podziału (metoda jednowymiarowa)



Rys. Wyznaczanie kolejnych przybliżeń w metodzie złotego podziału

Jak wyznaczyć kolejne przybliżenia w metodzie złotego podziału?

- 1) Wstępnie wyznaczamy przedział $[a, b]$ w którym spodziewamy się że zlokalizowane jest minimum
- 2) W przedziale $[a, b]$ wyznaczamy dwa punkty λ_1 i λ_2
- 3) Jeśli

$$F(\lambda_2) > F(\lambda_1)$$

to zmieniamy granice przedziału na $[a, \lambda_2]$

- 4) Jeśli

$$F(\lambda_2) < F(\lambda_1)$$

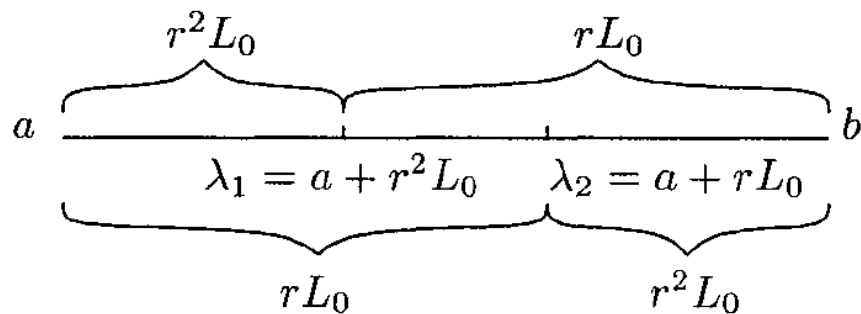
to zmieniamy granice przedziału na $[\lambda_1, b]$

- 5) Proces podziału prowadzimy iteracyjnie aż do spełnienia warunku

$$|a^i - b^i| < \varepsilon$$

$$\lambda^* = \frac{b^i - a^i}{2}$$

Pozostaje tylko kwestia jak wyznaczyć punkty tak aby wybór był optymalny tzn. chcemy wykonać jak najmniejszą ilość podziałów.



Punktem wyjścia jest zależność

$$\frac{(\lambda_1 - a) + (b - \lambda_1)}{b - \lambda_1} = \frac{b - \lambda_1}{\lambda_1 - a} = \varphi$$

$$b - a = L \Rightarrow b = L + a$$

$$\frac{L}{L + a - \lambda_1} = \frac{L + a - \lambda_1}{\lambda_1 - a}$$

$$L(\lambda_1 - a) = (L - (\lambda_1 - a))^2$$

$$(\lambda_1 - a) = L \left(1 - \frac{(\lambda_1 - a)}{L} \right)^2 = Lr^2$$

$$b - \lambda_1 = L \left(1 - \frac{(\lambda_1 - a)}{L} \right) = Lr$$

$$\frac{Lr^2 + Lr}{Lr} = \frac{Lr}{Lr^2} = \frac{1}{r} \Rightarrow r^2 + r - 1 = 0$$

Pierwiastki równania

$$r_1 = \frac{\sqrt{5} - 1}{2} = 0.618034 > 0$$

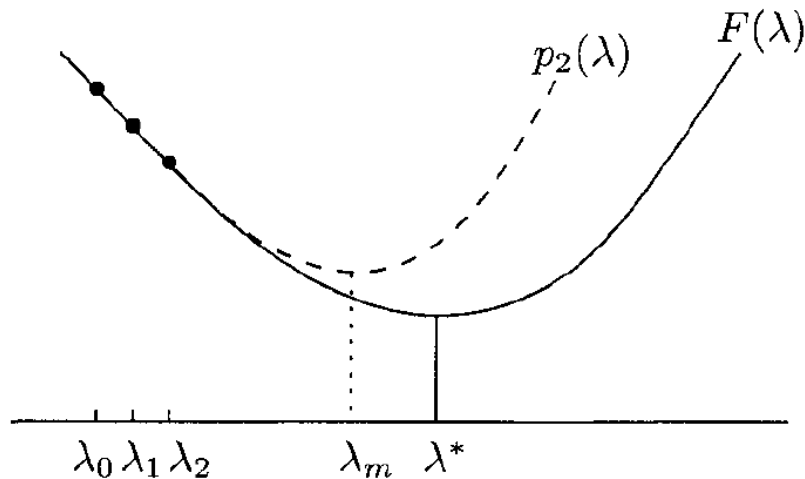
$$r_2 = \frac{-\sqrt{5} - 1}{2} < 0$$

Po wyborze $r=r_1$ możemy określić wartości λ_1 i λ_2 zakładając ponadto, że oba punkty powinny być symetryczne względem krańców przedziału

$$\lambda_1 = a + r^2 L$$

$$\lambda_2 = a + rL$$

Metoda interpolacji kwadratowej Powell'a



Rys. Wyznaczanie przybliżonego rozwiązania w metodzie Powell'a.

Przez trzy punkty: $\lambda_1, \lambda_2, \lambda_3$ przechodzi wielomian 2 stopnia

$$\begin{aligned} p_2(\lambda) &= F(\lambda_0) + F[\lambda_0, \lambda_1](\lambda - \lambda_0) \\ &+ F[\lambda_0, \lambda_1, \lambda_2](\lambda - \lambda_0)(\lambda - \lambda_1) \end{aligned}$$

Z warunkiem na minimum

$$\begin{aligned} \frac{dp_2}{d\lambda} &= F[\lambda_0, \lambda_1] + 2\lambda F[\lambda_0, \lambda_1, \lambda_2] \\ &- F[\lambda_0, \lambda_1, \lambda_2](\lambda_0 + \lambda_1) \end{aligned}$$

gdzie: $F[\lambda_0, \lambda_1]$ – iloraz różnicowy 1 rzędu,
 $F[\lambda_0, \lambda_1, \lambda_2]$ – iloraz różnicowy 2 rzędu

Punkt λ_m jest kolejnym przybliżeniem

$$\lambda_m = \frac{F[\lambda_0, \lambda_1, \lambda_2](\lambda_0 + \lambda_1) - F[\lambda_0, \lambda_1]}{2F[\lambda_0, \lambda_1, \lambda_2]} \approx \lambda^*$$

Aby znaleziony punkt był rzeczywistym minimum, druga pochodna ($F[\lambda_0, \lambda_1, \lambda_2]$) musi spełniać warunek

$$F[\lambda_0, \lambda_1, \lambda_2] > 0$$

Algorytm:

1) Wybierz λ_0 i oblicz $F[\lambda_0 + h] < F[\lambda_0]$, $F[\lambda_0 + 2h] < F[\lambda_0 + h]$ (ewentualnie zmień znak: $-h$, jeśli nierówności nie są spełnione)

2) Wyznacz λ_m i sprawdź czy jest minimum

3) Jeśli

$$|\lambda_m - \lambda_n| > h$$

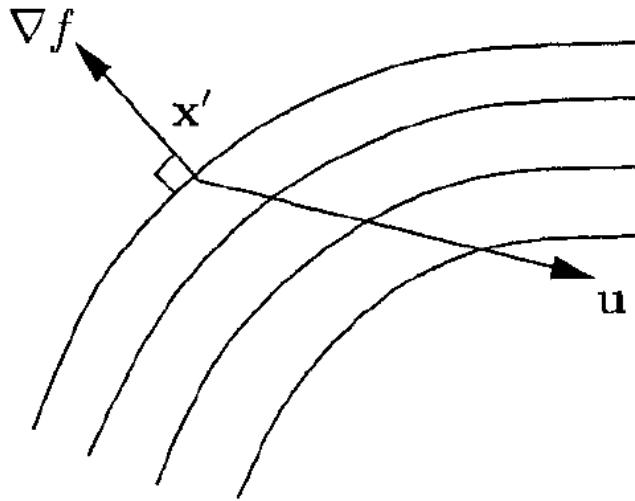
odrzuć najdalej położony od λ_m punkt i ponownie wykonaj obliczenia z pkt. 2.

λ_n – najbliższy położony punkt względem λ_m

Punkt λ_m akceptujemy jako minimum jeśli

$$|\lambda_m - \lambda_n| < \varepsilon$$

Metoda największego spadku (pierwszego rzędu)



Rys. Gradient funkcji i kierunek poszukiwań w metodzie największego spadku.

Korzystamy z pochodnej kierunkowej funkcji w punkcie \mathbf{x}'

$$\left. \frac{df(\mathbf{x}')}{d\lambda} \right|_{\mathbf{u}} = \frac{dF(0)}{d\lambda} = \nabla^T f(\mathbf{x}') \mathbf{u}$$

Wektor kierunkowy \mathbf{u} ma długość równą 1

$$\|\mathbf{u}\| = 1$$

Korzystamy z nierówności Schwartza

$$\nabla^T f(\mathbf{x}') \mathbf{u} \geq -\|\nabla^T f(\mathbf{x}')\| \cdot \|\mathbf{u}\| = -\|\nabla^T f(\mathbf{x}')\| = \min$$

Jeśli wektor kierunkowy wybierzemy w postaci

$$\mathbf{u} = \frac{-\nabla f(\mathbf{x}')}{\|\nabla f(\mathbf{x}')\|}$$

to będzie on wskazywał kierunek największego spadku.

Pochodna kierunkowa osiąga wtedy najmniejszą wartość

$$\frac{dF(0)}{d\lambda} = -\nabla^T f(\mathbf{x}') \frac{\nabla f(\mathbf{x}')}{\|\nabla f(\mathbf{x}')\|} = \min$$

Algorytm:

1) Wybierz \mathbf{x}^0

2) Iteracyjnie obliczaj

$$\mathbf{u}^i = \frac{-\nabla f(\mathbf{x}^{i-1})}{\|\nabla f(\mathbf{x}^{i-1})\|}$$

$$\mathbf{x}^i = \mathbf{x}^{i-1} + \lambda \mathbf{u}^i$$

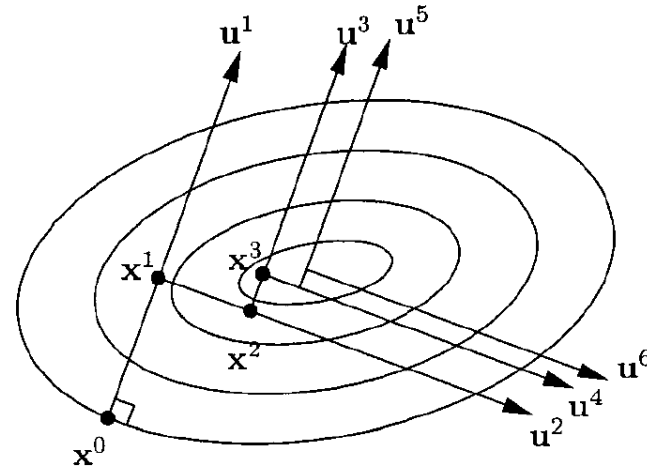
$$F(\lambda_i) = f(\mathbf{x}^{i-1} + \lambda_i \mathbf{u}^i) = \min_{\lambda} f(\mathbf{x}^{i-1} + \lambda \mathbf{u}^i)$$

3) Warunki zakończenia obliczeń

$$\|\mathbf{x}^i - \mathbf{x}^{i-1}\| < \varepsilon_1$$

$$\|\nabla f(\mathbf{x}^i)\| < \varepsilon_2$$

$$|f(\mathbf{x}^i) - f(\mathbf{x}^{i-1})| < \varepsilon_3$$



Metoda największego spadku może być mało wydajna, jeśli kontur wartości funkcji celu jest wydłużony (elipsa). Pojawiają się wówczas częste zmiany kierunków poszukiwań – zigzag.

Metoda sprzężonego gradientu (CG)

Metoda jest wydajna i dostarcza rozwiązania w skończonej liczbie kroków dla problemu w postaci

$$f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T A \mathbf{x} + \mathbf{b}^T \mathbf{x} + c$$

Dwa wektory są wzajemnie sprzężone jeśli

$$\mathbf{u}^T A \mathbf{v} = (\mathbf{u}, A \mathbf{v}) = 0$$

gdzie: A jest macierzą dodatniookreśloną

Jeśli ciąg wektorów

$$\mathbf{u}^i \in R^n, \quad i = 1, 2, \dots, n$$

tworzą wektory wzajemnie sprzężone, to stanowią one bazę w przestrzeni R^n i wówczas każdy wektor

$$\mathbf{x}^i \in R^n$$

można rozwinąć w tej bazie

$$\mathbf{x} = \sum_{k=1}^n \alpha_k \mathbf{u} = \mathbf{x}_0 + \sum_{i=1}^n \lambda_i \mathbf{u}$$

$$\alpha_i = \frac{(\mathbf{u}^i, A \mathbf{x})}{(\mathbf{u}^i, A \mathbf{u}^i)}$$

Algorytm CG Fletchera-Reevsa

- 1) Wybieramy punkt startowy \mathbf{x}^0
- 2) Obliczamy iteracyjnie

$$\mathbf{x}^i = \mathbf{x}^{i-1} + \lambda_i \mathbf{u}_i$$

kierunek poszukiwań:

$$\text{a) } i=0 \quad \mathbf{u}^1 = -\nabla f(\mathbf{x}^0)$$

$$\text{b) } i=1, 2, 3, \dots, n \quad \mathbf{u}^{i+1} = -\nabla f(\mathbf{x}^i) + \beta_i \mathbf{u}^i$$

Wartość λ_i wyznaczamy identycznie jak w metodzie największego spadku.

Parametr β_i

$$\beta_i = \frac{\|\nabla f(\mathbf{x}^i)\|^2}{\|\nabla f(\mathbf{x}^{i-1})\|^2}$$