Optymalizacja ciągła

5. Metody kierunków poparwy
(metoda Newtona-Raphsona, metoda gradientów sprzężonych)

Wojciech Kotłowski

Instytut Informatyki PP
http://www.cs.put.poznan.pl/wkotlowski/

28.03.2019

Plan wykładu

Minimalizacja **różniczkowalnej** funkcji $f(x)$ **bez ograniczeń**:

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} f(\boldsymbol{x})$$

Często będziemy zakładać **wypukłość** $f(x)$

1. Wprowadzenie (poprzedni wykład)
2. Metoda spadku wzdłuż gradientu (poprzedni wykład)
3. Metoda Newtona-Raphsona
4. Modyfikacje metody Newtona-Rapshona
5. Metoda gradientów sprzężonych

Przypomnienie: metody kierunków poprawy

- Iteracyjne metody generujące punkty $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ zgodnie z:

$$\boldsymbol{x}_{k+1} \; = \; \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k,$$

- Kierunek $\boldsymbol{v}_k$ jest **kierunkiem poprawy**: gwarantuje, że przy dostatecznie małych $\alpha \in \mathbb{R}$:

$$f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k) < f(\boldsymbol{x}_k)$$

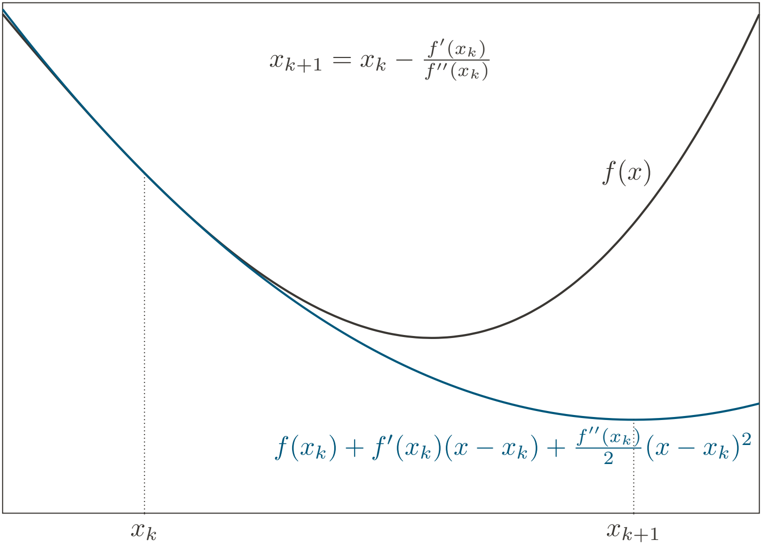
  Równoważnie, **pochodna kierunkowa** wzdłuż $\boldsymbol{v}_k$ musi być **ujemna**:

$$\frac{\partial f(\boldsymbol{x}_k)}{\partial \boldsymbol{v}_k} \; = \; \nabla f(\boldsymbol{x}_k)^\top \boldsymbol{v}_k < 0$$

- **Długość kroku** α_k dobierana jest aby zapewnić odpowiednio duży spadek funkcji celu

Metoda Newtona-Raphsona

Przypomnienie: metoda Newtona



Metoda Newtona-Raphsona

Wielowymiarowa wersja metody Newtona

Generujemy iteracyjnie ciąg punktów $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$

W każdej iteracji $k = 1, 2, \ldots$, przybliżamy funkcję $f(\boldsymbol{x})$ **wielomianem Taylora drugiego rzędu** w aktualnym punkcie $\boldsymbol{x}_k$:

$$P_2(\boldsymbol{x}; \boldsymbol{x}_k) \;=\; f(\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k)^\top (\boldsymbol{x} - \boldsymbol{x}_k) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_k)^\top \nabla^2 f(\boldsymbol{x}_k)(\boldsymbol{x} - \boldsymbol{x}_k)$$

Założenie: hesjan $\nabla^2 f(\boldsymbol{x}_k)$ jest **dodatnio określony**

Jako kolejny punkt $\boldsymbol{x}_{k+1}$ bierzemy punkt minimalizujący $P_2(\boldsymbol{x}; \boldsymbol{x}_k)$

Metoda Newtona-Raphsona

Przypomnienie: Minimum funkcji kwadratowej $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} + c$ z dodatnio określoną macierzą $\boldsymbol{A}$ ma postać $\boldsymbol{x}^* = -\frac{1}{2}\boldsymbol{A}^{-1}\boldsymbol{b}$

Metoda Newtona-Raphsona

Przypomnienie: Minimum funkcji kwadratowej $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} + c$ z dodatnio określoną macierzą $\boldsymbol{A}$ ma postać $\boldsymbol{x}^* = -\frac{1}{2}\boldsymbol{A}^{-1}\boldsymbol{b}$

Wniosek: Minimum funkcji kwadratowej:

$$F(\boldsymbol{v}) = \frac{1}{2}\boldsymbol{v}^\top \nabla^2 f(\boldsymbol{x}_k)\boldsymbol{v} + \nabla f(\boldsymbol{x}_k)^\top \boldsymbol{v} + f(\boldsymbol{x}_k)$$

jest w punkcie $\boldsymbol{v} = -(\nabla^2 f(\boldsymbol{x}_k))^{-1} \nabla f(\boldsymbol{x}_k)$

Metoda Newtona-Raphsona

Przypomnienie: Minimum funkcji kwadratowej $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} + c$ z dodatnio określoną macierzą $\boldsymbol{A}$ ma postać $\boldsymbol{x}^* = -\frac{1}{2}\boldsymbol{A}^{-1}\boldsymbol{b}$

Wniosek: Minimum funkcji kwadratowej:

$$F(\boldsymbol{v}) = \frac{1}{2}\boldsymbol{v}^\top \nabla^2 f(\boldsymbol{x}_k)\boldsymbol{v} + \nabla f(\boldsymbol{x}_k)^\top \boldsymbol{v} + f(\boldsymbol{x}_k)$$

jest w punkcie $\boldsymbol{v} = -(\nabla^2 f(\boldsymbol{x}_k))^{-1}\nabla f(\boldsymbol{x}_k)$
Wniosek: Biorąc $\boldsymbol{v} = \boldsymbol{x} - \boldsymbol{x}_k$, minimum wielomianu:

$$P_2(\boldsymbol{x}; \boldsymbol{x}_k) \; = \; f(\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k)^\top (\boldsymbol{x} - \boldsymbol{x}_k) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_k)^\top \nabla^2 f(\boldsymbol{x}_k)(\boldsymbol{x} - \boldsymbol{x}_k)$$

jest w punkcie $\boldsymbol{x}$ spełniającym:

$$\boldsymbol{x} - \boldsymbol{x}_k \; = \; -(\nabla^2 f(\boldsymbol{x}_k))^{-1}\nabla f(\boldsymbol{x}_k)$$

Metoda Newtona-Raphsona

Przypomnienie: Minimum funkcji kwadratowej $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} + c$ z dodatnio określoną macierzą $\boldsymbol{A}$ ma postać $\boldsymbol{x}^* = -\frac{1}{2}\boldsymbol{A}^{-1}\boldsymbol{b}$

Wniosek: Minimum funkcji kwadratowej:

$$F(\boldsymbol{v}) = \frac{1}{2}\boldsymbol{v}^\top \nabla^2 f(\boldsymbol{x}_k)\boldsymbol{v} + \nabla f(\boldsymbol{x}_k)^\top \boldsymbol{v} + f(\boldsymbol{x}_k)$$

jest w punkcie $\boldsymbol{v} = -(\nabla^2 f(\boldsymbol{x}_k))^{-1}\nabla f(\boldsymbol{x}_k)$

Wniosek: Biorąc $\boldsymbol{v} = \boldsymbol{x} - \boldsymbol{x}_k$, minimum wielomianu:

$$P_2(\boldsymbol{x}; \boldsymbol{x}_k) \;=\; f(\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k)^\top (\boldsymbol{x} - \boldsymbol{x}_k) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_k)^\top \nabla^2 f(\boldsymbol{x}_k)(\boldsymbol{x} - \boldsymbol{x}_k)$$

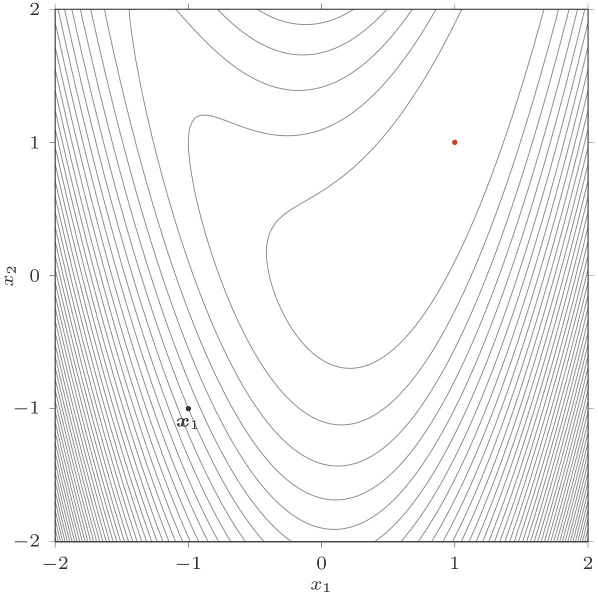
jest w punkcie $\boldsymbol{x}$ spełniającym:

$$\boldsymbol{x} - \boldsymbol{x}_k \;=\; -\,(\nabla^2 f(\boldsymbol{x}_k))^{-1}\nabla f(\boldsymbol{x}_k)$$

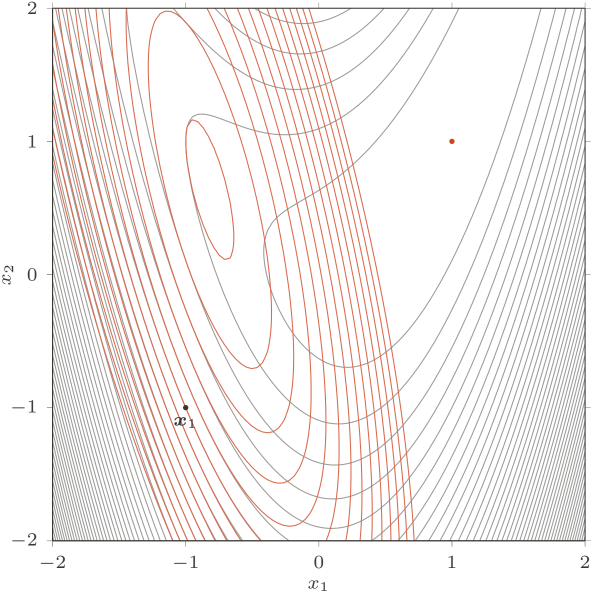
Metoda Newtona-Raphsona:

$$\boldsymbol{x}_{k+1} \;=\; \boldsymbol{x}_k - (\nabla^2 f(\boldsymbol{x}_k))^{-1}\nabla f(\boldsymbol{x}_k)$$

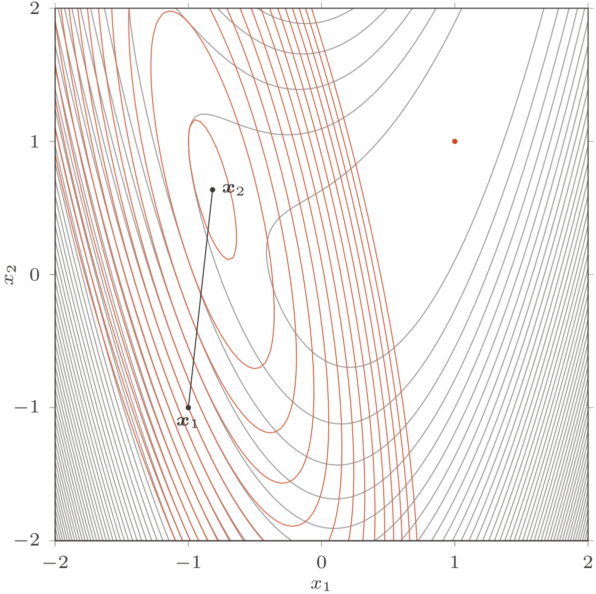
Metoda Newtona-Raphsona



Metoda Newtona-Raphsona



Metoda Newtona-Raphsona



Metoda Newtona-Raphsona

Rozpocznij od punktu $\boldsymbol{x}_1$ i powtarzaj dla $k = 1, 2, \ldots$:

$$\boldsymbol{x}_{k+1} \; = \; \boldsymbol{x}_k - (\nabla^2 f(\boldsymbol{x}_k))^{-1} \nabla f(\boldsymbol{x}_k)$$

Metoda Newtona-Raphsona

Rozpocznij od punktu $\boldsymbol{x}_1$ i powtarzaj dla $k = 1, 2, \ldots$:

$$\boldsymbol{x}_{k+1} \ = \ \boldsymbol{x}_k - (\nabla^2 f(\boldsymbol{x}_k))^{-1} \nabla f(\boldsymbol{x}_k)$$

- Porównaj z metodą Newtona dla jednej zmiennej: $x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$

Metoda Newtona-Raphsona

Rozpocznij od punktu $\boldsymbol{x}_1$ i powtarzaj dla $k = 1, 2, \ldots$:

$$\boldsymbol{x}_{k+1} \;=\; \boldsymbol{x}_k - (\nabla^2 f(\boldsymbol{x}_k))^{-1} \nabla f(\boldsymbol{x}_k)$$

- Porównaj z metodą Newtona dla jednej zmiennej: $x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$
- Metoda obliczeniowo **wolniejsza** od spadku wzdłuż gradientu: wymaga odwrócenia macierzy w każdej iteracji
  (w praktyce zamiast odwracać macierz rozwiązuje się układ równań liniowych $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ z $\boldsymbol{A} = \nabla^2 f(\boldsymbol{x}_k)$ i $\boldsymbol{b} = \nabla f(\boldsymbol{x}_k)$)

Metoda Newtona-Raphsona

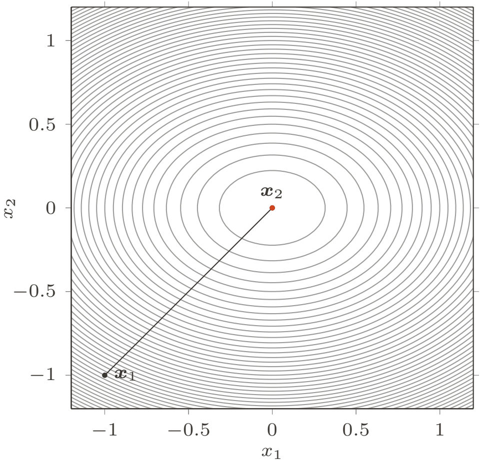
Rozpocznij od punktu $\boldsymbol{x}_1$ i powtarzaj dla $k = 1, 2, \ldots$:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \left(\nabla^2 f(\boldsymbol{x}_k)\right)^{-1} \nabla f(\boldsymbol{x}_k)$$

- Porównaj z metodą Newtona dla jednej zmiennej: $x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$
- Metoda obliczeniowo **wolniejsza** od spadku wzdłuż gradientu: wymaga odwrócenia macierzy w każdej iteracji (w praktyce zamiast odwracać macierz rozwiązuje się układ równań liniowych $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ z $\boldsymbol{A} = \nabla^2 f(\boldsymbol{x}_k)$ i $\boldsymbol{b} = \nabla f(\boldsymbol{x}_k)$)
- Zainicjalizowana dostatecznie blisko minimum zbiega bardzo szybko, ale źle zainicjalizowana może się rozbiec!

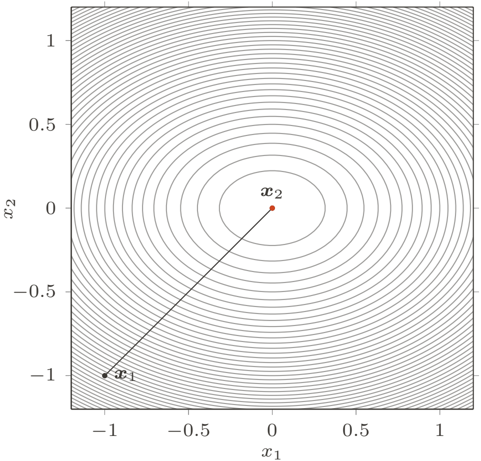
Metoda Newtona-Raphsona: przykład

$$f(x_1, x_2) = x_1^2 + 2x_2^2$$



Metoda Newtona-Raphsona: przykład

$$f(x_1, x_2) = x_1^2 + 2x_2^2$$



Metoda Newtona-Rapshona znajduje minimum funkcji kwadratowej:

$$f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} + c$$

w **jednym kroku**, niezależnie od punktu inicjalizacji!

Metoda Newtona-Raphsona dla funkcji kwadratowej

Rozważmy minimalizację funkcji funkcji kwadratowej z dodatnio określoną macierzą $\boldsymbol{A}$:

$$f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{x}^\top \boldsymbol{b} + c,$$

Metoda Newtona-Raphsona dla funkcji kwadratowej

Rozważmy minimalizację funkcji funkcji kwadratowej z dodatnio określoną macierzą $\boldsymbol{A}$:

$$f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{x}^\top \boldsymbol{b} + c,$$

$$\nabla f(\boldsymbol{x}) = 2\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}, \qquad \nabla^2 f(\boldsymbol{x}) = 2\boldsymbol{A}, \qquad (\nabla^2 f(\boldsymbol{x}))^{-1} = \frac{1}{2}\boldsymbol{A}^{-1}$$

Metoda Newtona-Raphsona dla funkcji kwadratowej

Rozważmy minimalizację funkcji funkcji kwadratowej z dodatnio określoną macierzą $\boldsymbol{A}$:

$$f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{x}^\top \boldsymbol{b} + c,$$

$$\nabla f(\boldsymbol{x}) = 2\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}, \qquad \nabla^2 f(\boldsymbol{x}) = 2\boldsymbol{A}, \qquad (\nabla^2 f(\boldsymbol{x}))^{-1} = \frac{1}{2}\boldsymbol{A}^{-1}$$

Rozpoczynamy od dowolnego punktu $\boldsymbol{x}_1$ i wybieramy punkt $\boldsymbol{x}_2$ jako:

$$\boldsymbol{x}_2 \; = \; \boldsymbol{x}_1 - (\nabla^2 f(\boldsymbol{x}_1))^{-1} \nabla f(\boldsymbol{x}_1)$$

Metoda Newtona-Raphsona dla funkcji kwadratowej

Rozważmy minimalizację funkcji funkcji kwadratowej z dodatnio określoną macierzą $\boldsymbol{A}$:

$$f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{x}^\top \boldsymbol{b} + c,$$

$$\nabla f(\boldsymbol{x}) = 2\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}, \qquad \nabla^2 f(\boldsymbol{x}) = 2\boldsymbol{A}, \qquad (\nabla^2 f(\boldsymbol{x}))^{-1} = \frac{1}{2}\boldsymbol{A}^{-1}$$

Rozpoczynamy od dowolnego punktu $\boldsymbol{x}_1$ i wybieramy punkt $\boldsymbol{x}_2$ jako:

$$\boldsymbol{x}_2 \ = \ \boldsymbol{x}_1 - \frac{1}{2}\boldsymbol{A}^{-1}(2\boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{b})$$

Metoda Newtona-Raphsona dla funkcji kwadratowej

Rozważmy minimalizację funkcji funkcji kwadratowej z dodatnio określoną macierzą $\boldsymbol{A}$:

$$f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{x}^\top \boldsymbol{b} + c,$$

$$\nabla f(\boldsymbol{x}) = 2\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}, \qquad \nabla^2 f(\boldsymbol{x}) = 2\boldsymbol{A}, \qquad (\nabla^2 f(\boldsymbol{x}))^{-1} = \frac{1}{2}\boldsymbol{A}^{-1}$$

Rozpoczynamy od dowolnego punktu $\boldsymbol{x}_1$ i wybieramy punkt $\boldsymbol{x}_2$ jako:

$$\boldsymbol{x}_2 \;=\; \boldsymbol{x}_1 - \underbrace{\boldsymbol{A}^{-1}\boldsymbol{A}}_{\boldsymbol{I}}\boldsymbol{x}_1 - \frac{1}{2}\boldsymbol{A}^{-1}\boldsymbol{b}$$

Metoda Newtona-Raphsona dla funkcji kwadratowej

Rozważmy minimalizację funkcji funkcji kwadratowej z dodatnio określoną macierzą $\boldsymbol{A}$:

$$f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{x}^\top \boldsymbol{b} + c,$$

$$\nabla f(\boldsymbol{x}) = 2\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}, \qquad \nabla^2 f(\boldsymbol{x}) = 2\boldsymbol{A}, \qquad (\nabla^2 f(\boldsymbol{x}))^{-1} = \frac{1}{2}\boldsymbol{A}^{-1}$$

Rozpoczynamy od dowolnego punktu $\boldsymbol{x}_1$ i wybieramy punkt $\boldsymbol{x}_2$ jako:

$$\boldsymbol{x}_2 \; = \; \boldsymbol{x}_1 - \boldsymbol{x}_1 - \frac{1}{2}\boldsymbol{A}^{-1}\boldsymbol{b}$$

Metoda Newtona-Raphsona dla funkcji kwadratowej

Rozważmy minimalizację funkcji funkcji kwadratowej z dodatnio określoną macierzą $\boldsymbol{A}$:

$$f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{x}^\top \boldsymbol{b} + c,$$

$$\nabla f(\boldsymbol{x}) = 2\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}, \qquad \nabla^2 f(\boldsymbol{x}) = 2\boldsymbol{A}, \qquad (\nabla^2 f(\boldsymbol{x}))^{-1} = \frac{1}{2}\boldsymbol{A}^{-1}$$

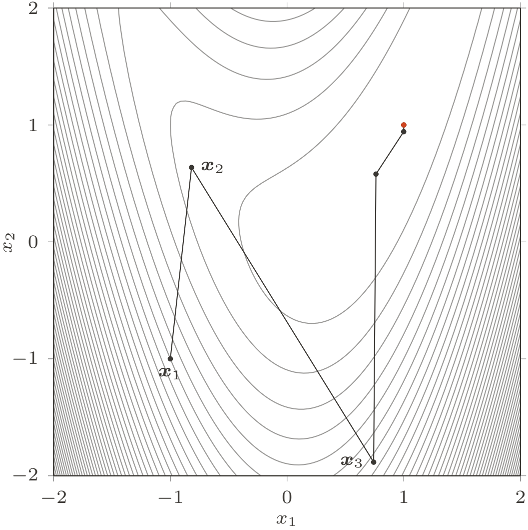
Rozpoczynamy od dowolnego punktu $\boldsymbol{x}_1$ i wybieramy punkt $\boldsymbol{x}_2$ jako:

$$\boldsymbol{x}_2 \;=\; -\frac{1}{2}\boldsymbol{A}^{-1}\boldsymbol{b}$$

Ale jest to punkt minimum $\boldsymbol{x}^*$ funkcji kwadratowej!

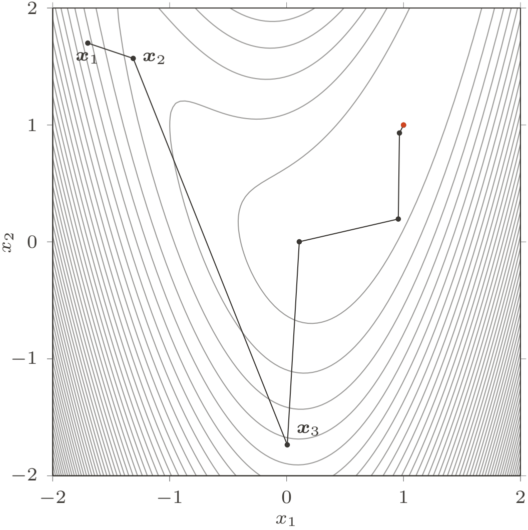
Metoda Newtona-Raphsona: przykład

$$f(x_1, x_2) = 2.5(x_1^2 - x_2)^2 + (1 - x_1)^2$$



Metoda Newtona-Raphsona: przykład

$$f(x_1, x_2) = 2.5(x_1^2 - x_2)^2 + (1 - x_1)^2$$



Metoda Newtona-Raphsona jako metoda spadkowa

Rozpocznij od punktu $\boldsymbol{x}_1$ i powtarzaj dla $k = 1, 2, \ldots$:

$$\boldsymbol{x}_{k+1} \; = \; \boldsymbol{x}_k + \boldsymbol{v}_k, \qquad \boldsymbol{v}_k = -(\nabla^2 f(\boldsymbol{x}_k))^{-1} \nabla f(\boldsymbol{x}_k)$$

Z założenia $\nabla^2 f(\boldsymbol{x}_k)$ jest dodatnio określona, czyli ma wszystkie wartości własne dodatnie

Metoda Newtona-Raphsona jako metoda spadkowa

Rozpocznij od punktu $\boldsymbol{x}_1$ i powtarzaj dla $k = 1, 2, \ldots$:

$$\boldsymbol{x}_{k+1} \; = \; \boldsymbol{x}_k + \boldsymbol{v}_k, \qquad \boldsymbol{v}_k = -(\nabla^2 f(\boldsymbol{x}_k))^{-1} \nabla f(\boldsymbol{x}_k)$$

Z założenia $\nabla^2 f(\boldsymbol{x}_k)$ jest dodatnio określona, czyli ma wszystkie wartości własne dodatnie

Wtedy $(\nabla^2 f(\boldsymbol{x}_k))^{-1}$ jest też dodatnio określona, gdyż wartości własne $\boldsymbol{A}^{-1}$ są odwrotnościami wartości własnych $\boldsymbol{A}$

Metoda Newtona-Raphsona jako metoda spadkowa

Rozpocznij od punktu $\boldsymbol{x}_1$ i powtarzaj dla $k = 1, 2, \ldots$:

$$\boldsymbol{x}_{k+1} \;=\; \boldsymbol{x}_k + \boldsymbol{v}_k, \qquad \boldsymbol{v}_k = -(\nabla^2 f(\boldsymbol{x}_k))^{-1} \nabla f(\boldsymbol{x}_k)$$

Z założenia $\nabla^2 f(\boldsymbol{x}_k)$ jest dodatnio określona, czyli ma wszystkie wartości własne dodatnie

Wtedy $(\nabla^2 f(\boldsymbol{x}_k))^{-1}$ jest też dodatnio określona, gdyż wartości własne $\boldsymbol{A}^{-1}$ są odwrotnościami wartości własnych $\boldsymbol{A}$

Czyli $\boldsymbol{x}^\top (\nabla^2 f(\boldsymbol{x}_k))^{-1} \boldsymbol{x} > 0$ dla dowolnego niezerowego $\boldsymbol{x}$

Metoda Newtona-Raphsona jako metoda spadkowa

Rozpocznij od punktu $\boldsymbol{x}_1$ i powtarzaj dla $k = 1, 2, \ldots$:

$$\boldsymbol{x}_{k+1} \; = \; \boldsymbol{x}_k + \boldsymbol{v}_k, \qquad \boldsymbol{v}_k = -(\nabla^2 f(\boldsymbol{x}_k))^{-1} \nabla f(\boldsymbol{x}_k)$$

Z założenia $\nabla^2 f(\boldsymbol{x}_k)$ jest dodatnio określona, czyli ma wszystkie wartości własne dodatnie

Wtedy $(\nabla^2 f(\boldsymbol{x}_k))^{-1}$ jest też dodatnio określona, gdyż wartości własne $\boldsymbol{A}^{-1}$ są odwrotnościami wartości własnych $\boldsymbol{A}$

Czyli $\boldsymbol{x}^\top (\nabla^2 f(\boldsymbol{x}_k))^{-1} \boldsymbol{x} > 0$ dla dowolnego niezerowego $\boldsymbol{x}$

Wniosek: $\boldsymbol{v}_k = -(\nabla^2 f(\boldsymbol{x}_k))^{-1} \nabla f(\boldsymbol{x}_k)$ jest **kierunkiem poprawy**

$$\nabla f(\boldsymbol{x}_k)^\top \boldsymbol{v}_k \; = \; - \nabla f(\boldsymbol{x}_k)(\nabla^2 f(\boldsymbol{x}_k))^{-1} \nabla f(\boldsymbol{x}_k) < 0$$

Metoda Newtona-Raphsona jako metoda spadkowa

Rozpocznij od punktu $\boldsymbol{x}_1$ i powtarzaj dla $k = 1, 2, \ldots$:

$$\boldsymbol{x}_{k+1} \;=\; \boldsymbol{x}_k + \boldsymbol{v}_k, \qquad \boldsymbol{v}_k = -(\nabla^2 f(\boldsymbol{x}_k))^{-1} \nabla f(\boldsymbol{x}_k)$$

Z założenia $\nabla^2 f(\boldsymbol{x}_k)$ jest dodatnio określona, czyli ma wszystkie wartości własne dodatnie

Wtedy $(\nabla^2 f(\boldsymbol{x}_k))^{-1}$ jest też dodatnio określona, gdyż wartości własne $\boldsymbol{A}^{-1}$ są odwrotnościami wartości własnych $\boldsymbol{A}$

Czyli $\boldsymbol{x}^\top (\nabla^2 f(\boldsymbol{x}_k))^{-1} \boldsymbol{x} > 0$ dla dowolnego niezerowego $\boldsymbol{x}$

Wniosek: $\boldsymbol{v}_k = -(\nabla^2 f(\boldsymbol{x}_k))^{-1} \nabla f(\boldsymbol{x}_k)$ jest **kierunkiem poprawy**

$$\nabla f(\boldsymbol{x}_k)^\top \boldsymbol{v}_k \;=\; -\nabla f(\boldsymbol{x}_k)(\nabla^2 f(\boldsymbol{x}_k))^{-1} \nabla f(\boldsymbol{x}_k) < 0$$

Uwaga: Długość kroku α_k jest **stała** i równa 1

Zbieżność metody Newtona-Raphsona

Jeśli funkcja $f(\boldsymbol{x})$ jest trzykrotnie różniczkowalna w sposób ciągły, algorytm Newtona zainicjalizowany **odpowiednio blisko** minimum lokalnego $\boldsymbol{x}^*$ z dodatnio określonym hesjanem $\nabla^2 f(\boldsymbol{x}^*)$ zbiega z **kwadratowym** rzędem zbieżności

Dowód koncepcyjnie bardzo podobny jak w przypadku jednowymiarowym, ale skomplikowany przez wektorowo-macierzowe przybliżenie Taylora

Zbieżność metody Newtona-Raphsona

Jeśli funkcja $f(\boldsymbol{x})$ jest trzykrotnie różniczkowalna w sposób ciągły, algorytm Newtona zainicjalizowany **odpowiednio blisko** minimum lokalnego $\boldsymbol{x}^*$ z dodatnio określonym hesjanem $\nabla^2 f(\boldsymbol{x}^*)$ zbiega z **kwadratowym** rzędem zbieżności

Dowód koncepcyjnie bardzo podobny jak w przypadku jednowymiarowym, ale skomplikowany przez wektorowo-macierzowe przybliżenie Taylora

Źle zainicjalizowana metoda Newtona-Raphsona może się nawet rozbiec!

Z tego powodu stosuje się **modyfikacje** metody, które mają zapobiec problemom ze zbieżnością

Modyfikacje metody Newtona-Raphsona

Zmienna długość kroku (*damped Newton-Raphson*)

Metoda Newtona-Raphsona (oryginalna):
Rozpocznij od punktu $\boldsymbol{x}_1$ i powtarzaj dla $k = 1, 2, \ldots$:

$$\boldsymbol{x}_{k+1} \; = \; \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k, \qquad \boldsymbol{v}_k = -(\nabla^2 f(\boldsymbol{x}_k))^{-1} \nabla f(\boldsymbol{x}_k), \quad \alpha_k = 1$$

Metoda nie gwarantuje spadku wartości funkcji $f(\boldsymbol{x}_{k+1}) < f(\boldsymbol{x}_k)$!

Zmienna długość kroku (*damped Newton-Raphson*)

Metoda Newtona-Raphsona (oryginalna):
Rozpocznij od punktu $\boldsymbol{x}_1$ i powtarzaj dla $k = 1, 2, \ldots$:

$$\boldsymbol{x}_{k+1} \;=\; \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k, \qquad \boldsymbol{v}_k = -(\nabla^2 f(\boldsymbol{x}_k))^{-1} \nabla f(\boldsymbol{x}_k), \quad \alpha_k = 1$$

Metoda nie gwarantuje spadku wartości funkcji $f(\boldsymbol{x}_{k+1}) < f(\boldsymbol{x}_k)$!

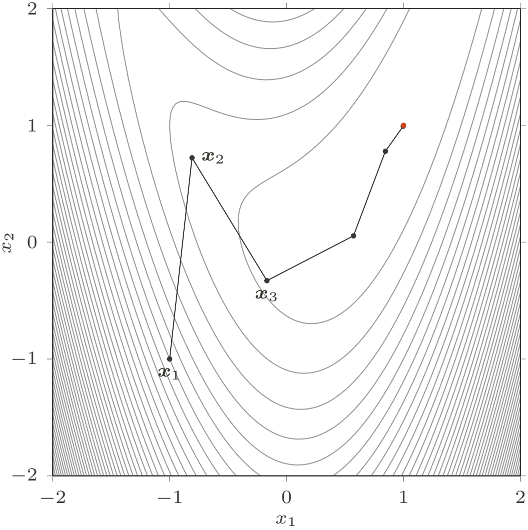
Jeśli hesjan $\nabla^2 f(\boldsymbol{x}_k)$ jest **dodatnio określony**, $\boldsymbol{v}_k$ jest **kierunkiem poprawy** i możemy zmodyfikować metodę dobierając α_k:

- optymalnie: $\alpha_k = \operatorname{argmin}_{\alpha \geqslant 0} f(\boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k)$,
- za pomocą algorytmu *backtracking*,

co zagwarantuje nam $f(\boldsymbol{x}_{k+1}) < f(\boldsymbol{x}_k)$

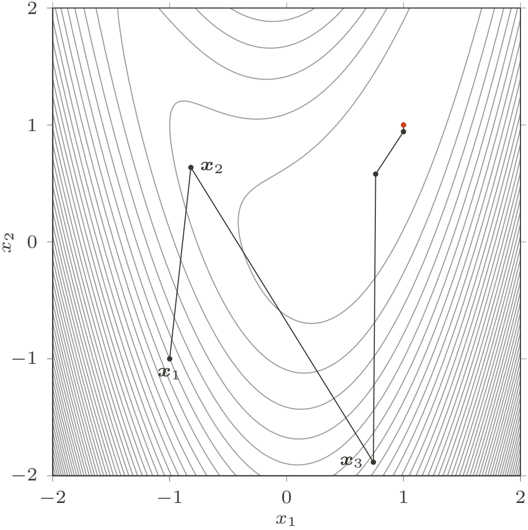
Metoda Newtona Raphsona z optymalnym doborem kroku

$$f(x_1, x_2) = 2.5(x_1^2 - x_2)^2 + (1 - x_1)^2$$



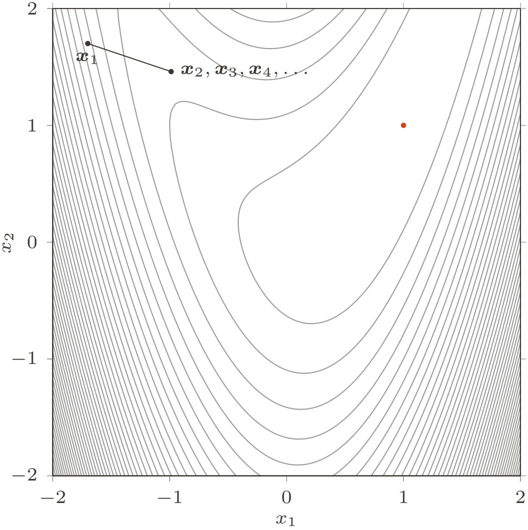
Dla porównania: klasyczna metoda Newtona Raphsona

$$f(x_1, x_2) = 2.5(x_1^2 - x_2)^2 + (1 - x_1)^2$$



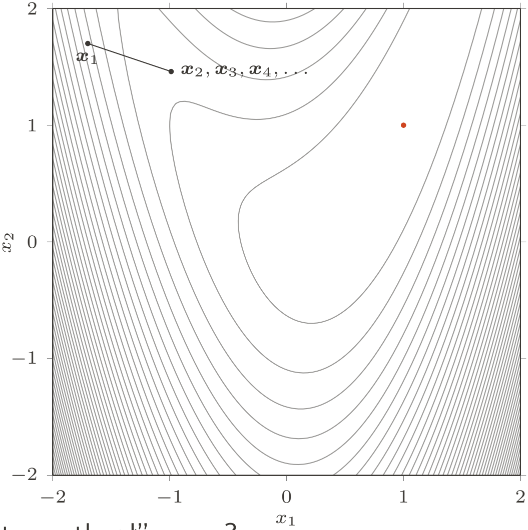
Metoda Newtona Raphsona z optymalnym doborem kroku

$$f(x_1, x_2) = 2.5(x_1^2 - x_2)^2 + (1 - x_1)^2$$



Metoda Newtona Raphsona z optymalnym doborem kroku

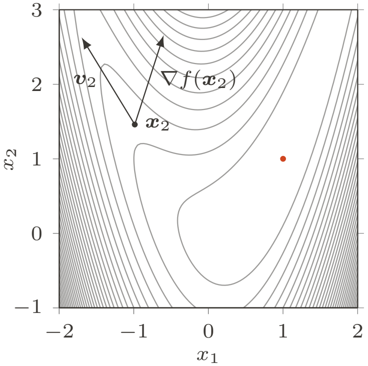
$$f(x_1, x_2) = 2.5(x_1^2 - x_2)^2 + (1 - x_1)^2$$



Dlaczego algorytm „utknął" w $\boldsymbol{x}_2$?

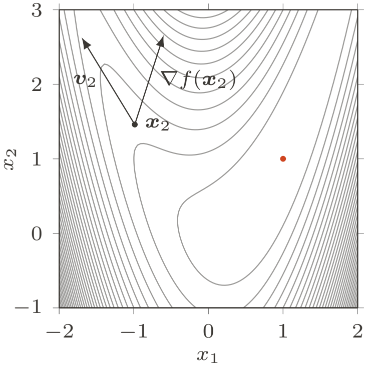
Analiza

$$\begin{aligned}
\boldsymbol{x}_2 &= (0.99, 1.46) \\
\nabla f(\boldsymbol{x}_2) &= (0.77, 2.4) \\
\nabla^2 f(\boldsymbol{x}_2) &= \begin{bmatrix} 16.8 & 9.9 \\ 9.9 & 5.0 \end{bmatrix} \\
\boldsymbol{v}_2 &= -(\nabla^2(\boldsymbol{x}_2))^{-1}\nabla(\boldsymbol{x}_2) \\
&= (-1.42, -2.33) \\
\nabla f(\boldsymbol{x}_2)^\top \boldsymbol{v}_2 &= 4.5 > 0
\end{aligned}$$



Analiza

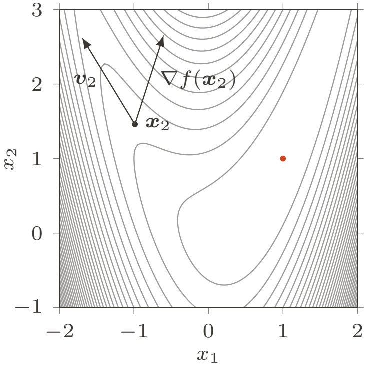
$$
\begin{aligned}
\boldsymbol{x}_2 &= (0.99, 1.46) \\
\nabla f(\boldsymbol{x}_2) &= (0.77, 2.4) \\
\nabla^2 f(\boldsymbol{x}_2) &= \begin{bmatrix} 16.8 & 9.9 \\ 9.9 & 5.0 \end{bmatrix} \\
\boldsymbol{v}_2 &= -(\nabla^2(\boldsymbol{x}_2))^{-1}\nabla(\boldsymbol{x}_2) \\
&= (-1.42, -2.33) \\
\nabla f(\boldsymbol{x}_2)^\top \boldsymbol{v}_2 &= 4.5 > 0
\end{aligned}
$$



Kierunek $\boldsymbol{v}_2$ **nie jest kierunkiem poprawy**

Analiza

$$\begin{aligned}
\boldsymbol{x}_2 &= (0.99, 1.46) \\
\nabla f(\boldsymbol{x}_2) &= (0.77, 2.4) \\
\nabla^2 f(\boldsymbol{x}_2) &= \begin{bmatrix} 16.8 & 9.9 \\ 9.9 & 5.0 \end{bmatrix} \\
\boldsymbol{v}_2 &= -(\nabla^2(\boldsymbol{x}_2))^{-1}\nabla(\boldsymbol{x}_2) \\
&= (-1.42, -2.33) \\
\nabla f(\boldsymbol{x}_2)^\top \boldsymbol{v}_2 &= 4.5 > 0
\end{aligned}$$



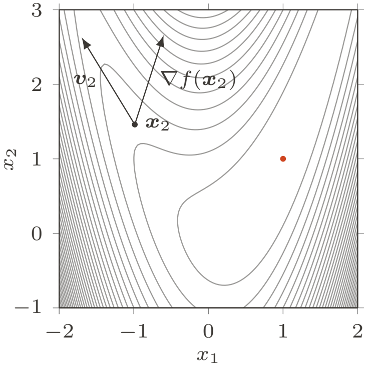
Kierunek $\boldsymbol{v}_2$ **nie jest kierunkiem poprawy**

Wartości własne hesjanu: $(22.4, -0.62) \Longrightarrow$ **nie jest dodatnio określony**

Optymalna długość kroku to $\alpha_2 = 0$, ponieważ nie możemy polepszyć funkcji celu wzdłuż kierunku $\boldsymbol{v}_2$

Analiza

$$
\begin{aligned}
\boldsymbol{x}_2 &= (0.99, 1.46) \\
\nabla f(\boldsymbol{x}_2) &= (0.77, 2.4) \\
\nabla^2 f(\boldsymbol{x}_2) &= \begin{bmatrix} 16.8 & 9.9 \\ 9.9 & 5.0 \end{bmatrix} \\
\boldsymbol{v}_2 &= -(\nabla^2(\boldsymbol{x}_2))^{-1}\nabla(\boldsymbol{x}_2) \\
&= (-1.42, -2.33) \\
\nabla f(\boldsymbol{x}_2)^\top \boldsymbol{v}_2 &= 4.5 > 0
\end{aligned}
$$



Kierunek $\boldsymbol{v}_2$ **nie jest kierunkiem poprawy**

Wartości własne hesjanu: $(22.4, -0.62) \Longrightarrow$ **nie jest dodatnio określony**

Optymalna długość kroku to $\alpha_2 = 0$, ponieważ nie możemy polepszyć funkcji celu wzdłuż kierunku $\boldsymbol{v}_2$

Wniosek: Optymalizacja długości kroku może zawieść, jeśli hesjan nie jest dodatnio określony!

Wartości własne hesjanu

Dotychczas zakładaliśmy, że hesjan $\nabla^2 f(\boldsymbol{x}_k)$ jest **dodatnio określony**, czyli ma wszystkie wartości własne **dodatnie** i stąd jest **odwracalny**.

W praktyce:

- Nawet dla funkcji wypukłych, hesjan może być tylko dodatnio półokreślony i jedna lub więcej wartości własnych może być równa zero; wtedy odwrotność $(\nabla^2 f(\boldsymbol{x}_k))^{-1}$ nie istnieje
- Dla funkcji **niewypukłych**, wartości własne mogą być ujemne, a wtedy kierunek $\boldsymbol{v}_k = -(\nabla^2 f(\boldsymbol{x}_k))^{-1} \nabla f(\boldsymbol{x}_k)$ nie musi być już kierunkiem poprawy!

Przykład

Rozważmy minimalizację funkcji postaci:

$$f(\boldsymbol{x}) = x_1^4 + x_2^2 - x_1 - x_2$$

Przykład

Rozważmy minimalizację funkcji postaci:

$$f(\boldsymbol{x}) = x_1^4 + x_2^2 - x_1 - x_2$$

$$\frac{\partial f}{\partial x_1} = 4x_1^3 - 1, \quad \frac{\partial f}{\partial x_2} = 2x_2 - 1$$

Przykład

Rozważmy minimalizację funkcji postaci:

$$f(\boldsymbol{x}) = x_1^4 + x_2^2 - x_1 - x_2$$

$$\frac{\partial f}{\partial x_1} = 4x_1^3 - 1, \quad \frac{\partial f}{\partial x_2} = 2x_2 - 1$$

$$\frac{\partial^2 f}{\partial x_1^2} = 12x_1^2, \quad \frac{\partial^2 f}{\partial x_2^2} = 2, \quad \frac{\partial^2 f}{\partial x_1 \partial x_2} = 0, \quad \nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} 12x_1^2 & 0 \\ 0 & 2 \end{bmatrix}$$

Przykład

Rozważmy minimalizację funkcji postaci:

$$f(\boldsymbol{x}) = x_1^4 + x_2^2 - x_1 - x_2$$

$$\frac{\partial f}{\partial x_1} = 4x_1^3 - 1, \quad \frac{\partial f}{\partial x_2} = 2x_2 - 1$$

$$\frac{\partial^2 f}{\partial x_1^2} = 12x_1^2, \quad \frac{\partial^2 f}{\partial x_2^2} = 2, \quad \frac{\partial^2 f}{\partial x_1 \partial x_2} = 0, \quad \nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} 12x_1^2 & 0 \\ 0 & 2 \end{bmatrix}$$

Ponieważ hesjan w $\boldsymbol{x}$ ma wartości własne $(12x_1^2, 2)$, **nieujemne** dla dowolnego x_1, jest macierzą **dodatnio półokreśloną dla każdego $\boldsymbol{x}$**, a więc $f(\boldsymbol{x})$ jest funkcją **wypukłą**

Przykład

Rozważmy minimalizację funkcji postaci:

$$f(\boldsymbol{x}) = x_1^4 + x_2^2 - x_1 - x_2$$

$$\frac{\partial f}{\partial x_1} = 4x_1^3 - 1, \quad \frac{\partial f}{\partial x_2} = 2x_2 - 1$$

$$\frac{\partial^2 f}{\partial x_1^2} = 12x_1^2, \quad \frac{\partial^2 f}{\partial x_2^2} = 2, \quad \frac{\partial^2 f}{\partial x_1 \partial x_2} = 0, \quad \nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} 12x_1^2 & 0 \\ 0 & 2 \end{bmatrix}$$

Ponieważ hesjan w $\boldsymbol{x}$ ma wartości własne $(12x_1^2, 2)$, **nieujemne** dla dowolnego x_1, jest macierzą **dodatnio półokreśloną dla każdego $\boldsymbol{x}$**, a więc $f(\boldsymbol{x})$ jest funkcją **wypukłą**

Rozpoczynając metodę Newtona-Raphsona od $\boldsymbol{x}_1 = \mathbf{0}$ mamy hesjan:

$$\nabla^2 f(\boldsymbol{x}_1) = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix},$$

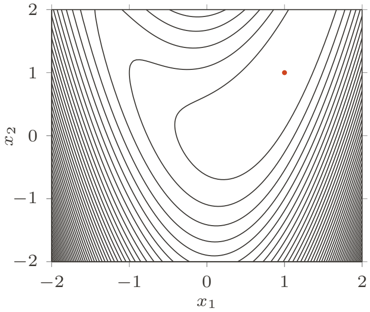
o wartościach własnych $(0, 2)$, nie dający się odwrócić!

Przykład

Rozważmy minimalizację funkcji:

$$f(\boldsymbol{x}) = 2.5(x_1^2 - x_2)^2 + (1 - x_1)^2$$

$$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} 30x_1^2 - 10x_2 + 2 & -10x_1 \\ -10x_1 & 5 \end{bmatrix}$$

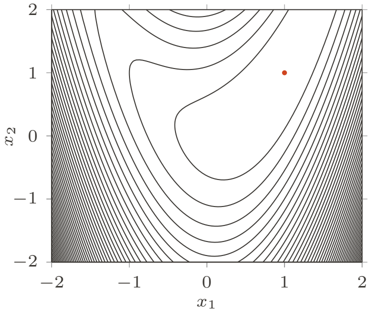


Przykład

Rozważmy minimalizację funkcji:

$$f(\boldsymbol{x}) = 2.5(x_1^2 - x_2)^2 + (1 - x_1)^2$$

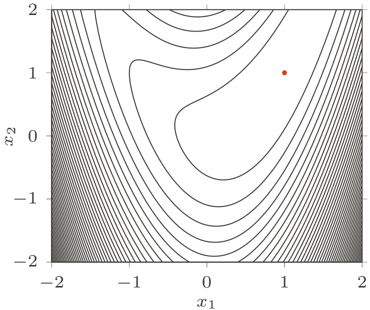
$$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} 30x_1^2 - 10x_2 + 2 & -10x_1 \\ -10x_1 & 5 \end{bmatrix}$$



Rozpoczynając metodę Newtona-Raphsona od $\boldsymbol{x}_1 = (0, 0.3)$ dostajemy:

$$\nabla^2 f(\boldsymbol{x}_1) = \begin{bmatrix} -1 & 0 \\ 0 & 5 \end{bmatrix} \qquad \text{o wartościach własnych } \ (-1, 5)$$

Przykład



Rozważmy minimalizację funkcji:

$$f(\boldsymbol{x}) = 2.5(x_1^2 - x_2)^2 + (1 - x_1)^2$$

$$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} 30x_1^2 - 10x_2 + 2 & -10x_1 \\ -10x_1 & 5 \end{bmatrix}$$

Rozpoczynając metodę Newtona-Raphsona od $\boldsymbol{x}_1 = (0, 0.3)$ dostajemy:

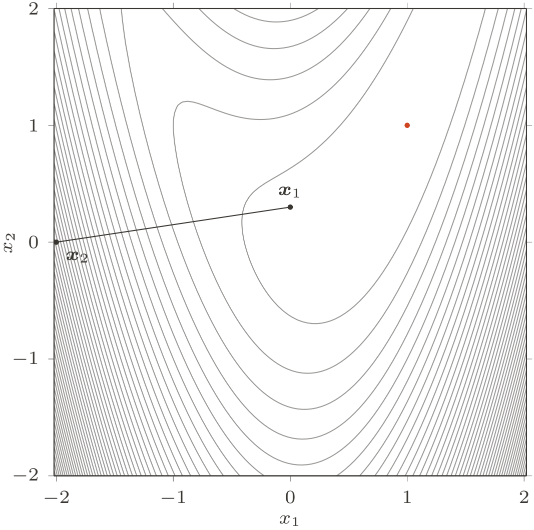
$$\nabla^2 f(\boldsymbol{x}_1) = \begin{bmatrix} -1 & 0 \\ 0 & 5 \end{bmatrix} \qquad \text{o wartościach własnych } (-1, 5)$$

Wielomian Taylora w $\boldsymbol{x}_1$ nie jest funkcją wypukłą, a kolejny punkt $\boldsymbol{x}_2$ jest jego **punktem siodłowym**, a nie minimum!

Kierunek $-(\nabla^2 f(\boldsymbol{x}_1))^{-1} \nabla f(\boldsymbol{x}_1)$ nie jest kierunkiem poprawy!

Przykład

$$f(\boldsymbol{x}) = 2.5(x_1^2 - x_2)^2 + (1 - x_1)^2, \qquad \boldsymbol{x}_1 = (0, 0.3)$$



Przesunięcie wartości własnych

Fakt: jeśli symetryczna macierz $\boldsymbol{A}$ ma wartości własne $(\lambda_1, \ldots, \lambda_n)$ to macierz $\boldsymbol{A} + \epsilon \boldsymbol{I}$ ma wartości własne $(\lambda_1 + \epsilon, \ldots, \lambda_n + \epsilon)$

Przesunięcie wartości własnych

Fakt: jeśli symetryczna macierz $\boldsymbol{A}$ ma wartości własne $(\lambda_1, \ldots, \lambda_n)$ to macierz $\boldsymbol{A} + \epsilon \boldsymbol{I}$ ma wartości własne $(\lambda_1 + \epsilon, \ldots, \lambda_n + \epsilon)$

Dowód: Jeśli $\boldsymbol{v}_k$ jest wektorem własnym $\boldsymbol{A}$ z wartością własną λ_k:

$$\boldsymbol{A}\boldsymbol{v}_k = \lambda_k \boldsymbol{v}_k,$$

to $\boldsymbol{v}_k$ jest też wektorem własnym $\boldsymbol{A} + \epsilon \boldsymbol{I}$ z wartością własną $\lambda_k + \epsilon$:

$$(\boldsymbol{A} + \epsilon \boldsymbol{I})\boldsymbol{v}_k = \boldsymbol{A}\boldsymbol{v}_k + \epsilon \boldsymbol{I}\boldsymbol{v}_k = \lambda_k \boldsymbol{v}_k + \epsilon \boldsymbol{v}_k = (\lambda_k + \epsilon)\boldsymbol{v}_k$$

Przesunięcie wartości własnych

Fakt: jeśli symetryczna macierz $\boldsymbol{A}$ ma wartości własne $(\lambda_1, \ldots, \lambda_n)$ to macierz $\boldsymbol{A} + \epsilon \boldsymbol{I}$ ma wartości własne $(\lambda_1 + \epsilon, \ldots, \lambda_n + \epsilon)$

Dowód: Jeśli $\boldsymbol{v}_k$ jest wektorem własnym $\boldsymbol{A}$ z wartością własną λ_k:

$$\boldsymbol{A}\boldsymbol{v}_k \;=\; \lambda_k \boldsymbol{v}_k,$$

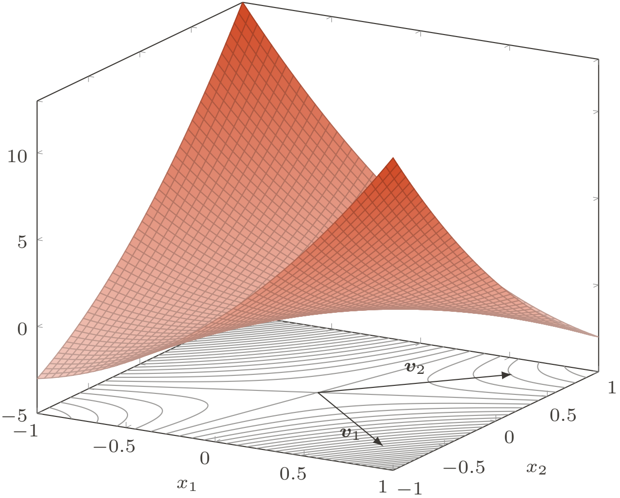
to $\boldsymbol{v}_k$ jest też wektorem własnym $\boldsymbol{A} + \epsilon \boldsymbol{I}$ z wartością własną $\lambda_k + \epsilon$:

$$(\boldsymbol{A} + \epsilon \boldsymbol{I})\boldsymbol{v}_k \;=\; \boldsymbol{A}\boldsymbol{v}_k + \epsilon \boldsymbol{I}\boldsymbol{v}_k \;=\; \lambda_k \boldsymbol{v}_k + \epsilon \boldsymbol{v}_k \;=\; (\lambda_k + \epsilon)\boldsymbol{v}_k$$

Wniosek: Przy odpowiednim doborze ϵ, możemy zagwarantować, że wszystkie wartości własne macierzy $\boldsymbol{A} + \epsilon \boldsymbol{I}$ będą **dodatnie** (tzn. macierz będzie **dodatnio określona**

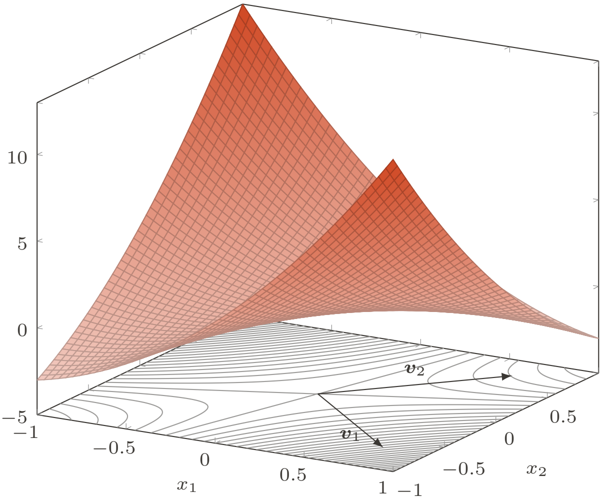
Przykład

$$f(\boldsymbol{x}) = 3x_1^2 - 8x_1x_2 + 2x_2^2, \qquad \boldsymbol{x} = (x_1, x_2)$$



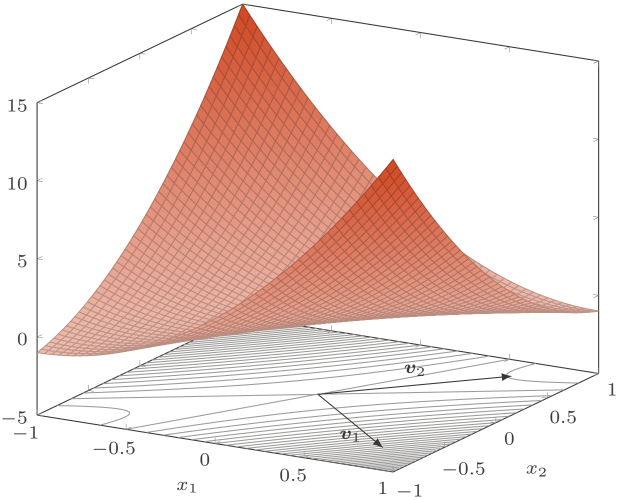
Przykład

$$f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}, \qquad \boldsymbol{A} = \begin{bmatrix} 3 & -4 \\ -4 & 2 \end{bmatrix}, \quad \lambda_1 = 6.53, \lambda_2 = -1.53$$



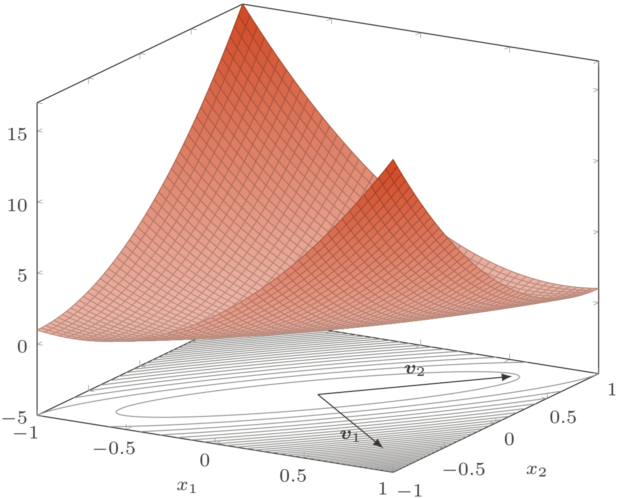
Przykład

$$f(\boldsymbol{x}) = \boldsymbol{x}^\top(\boldsymbol{A}+\boldsymbol{I})\boldsymbol{x}, \qquad \boldsymbol{A}+\boldsymbol{I} = \begin{bmatrix} 4 & -4 \\ -4 & 3 \end{bmatrix}, \quad \lambda_1 = 7.53, \lambda_2 = -0.53$$



Przykład

$$f(\boldsymbol{x}) = \boldsymbol{x}^\top(\boldsymbol{A} + 2\boldsymbol{I})\boldsymbol{x}, \qquad \boldsymbol{A} + 2\boldsymbol{I} = \begin{bmatrix} 5 & -4 \\ -4 & 4 \end{bmatrix}, \quad \lambda_1 = 8.53, \lambda_2 = 0.47$$



Metoda Levenberga-Marquardta

W każdej iteracji $k = 1, 2, \ldots$ dodajemy do hesjanu macierz $\epsilon_k \boldsymbol{I}$ aby zapewnić jego dodatnią określoność:

$$\boldsymbol{x}_{k+1} \;=\; \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k, \qquad \boldsymbol{v}_k = -(\nabla^2 f(\boldsymbol{x}_k) + \epsilon_k \boldsymbol{I})^{-1} \nabla f(\boldsymbol{x}_k)$$

Metoda Levenberga-Marquardta

W każdej iteracji $k = 1, 2, \ldots$ dodajemy do hesjanu macierz $\epsilon_k \boldsymbol{I}$ aby zapewnić jego dodatnią określoność:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k, \qquad \boldsymbol{v}_k = -(\nabla^2 f(\boldsymbol{x}_k) + \epsilon_k \boldsymbol{I})^{-1} \nabla f(\boldsymbol{x}_k)$$

Dobór współczynnika ϵ_k:

- Dla funkcji **wypukłych** wystarczy stała, niewielka wartość ϵ, np. $\epsilon = 10^{-5}$. Długość kroku można dobrać jako $\alpha_k = 1$ (jak w metodzie Newtona-Raphsona) lub optymalizować
- Dla funkcji **niewypukłych**, w każdej iteracji trzeba wyznaczyć wartości własne hesjanu i wybrać ϵ_k tak, aby wszystkie przesunąć powyżej zera

  W tym przypadku zaleca się zawsze używanie zmiennej (optymalizowanej) długości kroku α_k!

Metoda Levenberga-Marquardta: interpretacja

Rozważmy ogólną metodę spadkową postaci $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k$

Metoda Levenberga-Marquardta: interpretacja

Rozważmy ogólną metodę spadkową postaci $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k$

W metodzie Newtona-Rapshona $\boldsymbol{v}_k = -(\nabla^2 f(\boldsymbol{x}_k))^{-1} \nabla f(\boldsymbol{x}_k)$ możemy przedstawić jako rozwiązanie problemu minimalizacji funkcji kwadratowej:

$$g(\boldsymbol{v}) \;=\; \nabla f(\boldsymbol{x}_k)^\top \boldsymbol{v} + \frac{1}{2} \boldsymbol{v}^\top \nabla^2 f(\boldsymbol{x}_k) \boldsymbol{v}$$

Metoda Levenberga-Marquardta: interpretacja

Minimum funkcji kwadratowej $\boldsymbol{x}^\top \boldsymbol{b} + \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$
to $\boldsymbol{x}^* = -\boldsymbol{A}^{-1}\boldsymbol{b}$

P… $_{1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k$

W metodzie Newtona-Rapshona $\boldsymbol{v}_k = -(\nabla^2 f(\boldsymbol{x}_k))^{-1}\nabla f(\boldsymbol{x}_k)$ możemy przedstawić jako rozwiązanie problemu minimalizacji funkcji kwadratowej:

$$g(\boldsymbol{v}) \;=\; \nabla f(\boldsymbol{x}_k)^\top \boldsymbol{v} + \frac{1}{2}\boldsymbol{v}^\top \nabla^2 f(\boldsymbol{x}_k)\boldsymbol{v}$$

Metoda Levenberga-Marquardta: interpretacja

Rozważmy ogólną metodę spadkową postaci $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k$

W metodzie Newtona-Rapshona $\boldsymbol{v}_k = -(\nabla^2 f(\boldsymbol{x}_k))^{-1} \nabla f(\boldsymbol{x}_k)$ możemy przedstawić jako rozwiązanie problemu minimalizacji funkcji kwadratowej:

$$g(\boldsymbol{v}) \ = \ \nabla f(\boldsymbol{x}_k)^\top \boldsymbol{v} + \frac{1}{2} \boldsymbol{v}^\top \nabla^2 f(\boldsymbol{x}_k) \boldsymbol{v}$$

W metodzie spadku wzdłuż gradientu $\boldsymbol{v}_k = -\nabla f(\boldsymbol{x}_k)$ możemy przedstawić jako rozwiązanie problemu minimalizacji funkcji:

$$g(\boldsymbol{v}) \ = \ \nabla f(\boldsymbol{x}_k)^\top \boldsymbol{v} + \frac{1}{2} \boldsymbol{v}^\top \boldsymbol{I} \boldsymbol{v}$$

Metoda Levenberga-Marquardta: interpretacja

Rozważmy ogólną metodę spadkową postaci $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k$

W metodzie Newtona-Rapshona $\boldsymbol{v}_k = -(\nabla^2 f(\boldsymbol{x}_k))^{-1} \nabla f(\boldsymbol{x}_k)$ możemy przedstawić jako rozwiązanie problemu minimalizacji funkcji kwadratowej:

$$g(\boldsymbol{v}) \; = \; \nabla f(\boldsymbol{x}_k)^\top \boldsymbol{v} + \frac{1}{2} \boldsymbol{v}^\top \nabla^2 f(\boldsymbol{x}_k) \boldsymbol{v}$$

W metodzie spadku wzdłuż gradientu $\boldsymbol{v}_k = -\nabla f(\boldsymbol{x}_k)$ możemy przedstawić jako rozwiązanie problemu minimalizacji funkcji:

$$g(\boldsymbol{v}) \; = \; \nabla f(\boldsymbol{x}_k)^\top \boldsymbol{v} + \frac{1}{2} \boldsymbol{v}^\top \boldsymbol{I} \boldsymbol{v}$$

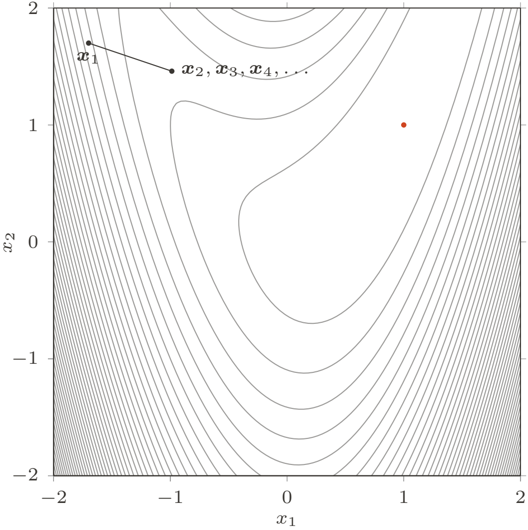
Metoda Levenberga-Marquardta jest **interpolacją** obu metod dobierając kierunek spadku jako rozwiązanie problemu minimalizacji funkcji:

$$g(\boldsymbol{v}) \; = \; \nabla f(\boldsymbol{x}_k)^\top \boldsymbol{v} + \frac{1}{2} \boldsymbol{v}^\top \Big(\epsilon_k \boldsymbol{I} + \nabla^2 f(\boldsymbol{x}_k) \Big) \boldsymbol{v}$$

Jest to szczególny przypadek **metod quasi-newtonowskich**, gdzie w funkcji kwadratowej dobiera się dowolną macierz dodatnio określoną $\boldsymbol{A}_k$

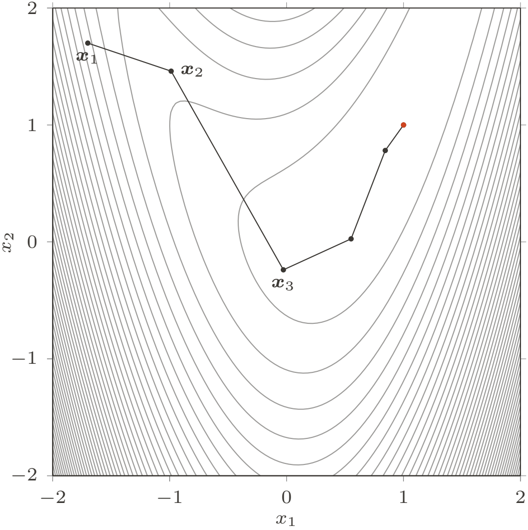
Metoda Newtona Raphsona z optymalnym doborem kroku

$$f(x_1, x_2) = 2.5(x_1^2 - x_2)^2 + (1 - x_1)^2$$

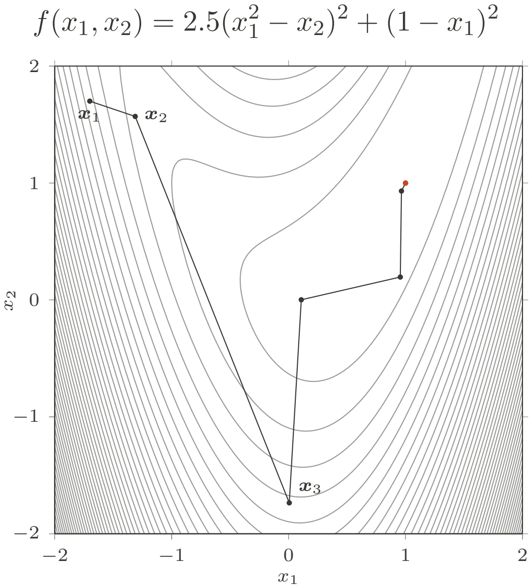


M. Levenberga-Marquardta z optymalnym doborem kroku

$$f(x_1, x_2) = 2.5(x_1^2 - x_2)^2 + (1 - x_1)^2$$



Dla porównania: klasyczna metoda Newtona Raphsona



Podsumowanie

Dwie modyfikacje:

1. Optymalizacja długości kroku α_k
2. Zapewnienie dodatniej określoności hesjanu (Levenberg-Marquardt)

Podsumowanie

Dwie modyfikacje:

1. Optymalizacja długości kroku α_k
2. Zapewnienie dodatniej określoności hesjanu (Levenberg-Marquardt)

Jeśli stosujemy (1), powinniśmy również stosować (2), inaczej algorytm może się zatrzymać w nieoptymalnym rozwiązaniu!

Podsumowanie

Dwie modyfikacje:

1. Optymalizacja długości kroku α_k
2. Zapewnienie dodatniej określoności hesjanu (Levenberg-Marquardt)

Jeśli stosujemy (1), powinniśmy również stosować (2), inaczej algorytm może się zatrzymać w nieoptymalnym rozwiązaniu!

Optymalizacja długości kroku zapewnia zbieżność algortymu, ale niekoniecznie polepsza szybkość jego działania, a równocześnie zwiększa złożoność obliczeniową

Podsumowanie

Dwie modyfikacje:

1. Optymalizacja długości kroku α_k
2. Zapewnienie dodatniej określoności hesjanu (Levenberg-Marquardt)

Jeśli stosujemy (1), powinniśmy również stosować (2), inaczej algorytm może się zatrzymać w nieoptymalnym rozwiązaniu!

Optymalizacja długości kroku zapewnia zbieżność algortymu, ale niekoniecznie polepsza szybkość jego działania, a równocześnie zwiększa złożoność obliczeniową

Dobry kompromis: Stosuj klasyczną metodę Newtona-Raphsona (z $\alpha_k = 1$), ale jeśli jej krok nie daje spadku funkcji celu (lub hesjan jest nieodwracalny), zastosuj krok Levenberga-Marquardta z optymalizacją α_k.

Metoda gradientów sprzężonych

Przypomnienie z algebry: baza przestrzeni wektorowej

Zbiór wektorów $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$ nazywamy **liniowo niezależnym**, jeśli żaden wektor z tego zbioru nie da się przedstawić w postaci kombinacji liniowej pozostałych

W przestrzeni $\mathbb{R}^n$ zbiór wektorów liniowo niezależnych może składać się maksymalnie z n wektorów, i zbiór taki nazywamy **bazą przestrzeni wektorowej**

Wniosek: Jeśli $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ jest bazą to **dowolny inny wektor** $\boldsymbol{x} \in \mathbb{R}^n$ musi być kombinacją liniową wektorów z bazy, tzn. dla pewnych współczynników $c_1, \ldots, c_n \in \mathbb{R}$

$$\boldsymbol{x} = \sum_{k=1}^{n} c_k \boldsymbol{v}_k$$

Wektory sprzężone

Niech $\boldsymbol{A}$ będzie macierzą **dodatnio określoną**
Zbiór **niezerowych** wektorów $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ nazywamy **sprzężonymi** (ze względu na macierz $\boldsymbol{A}$), jeśli

$$\boldsymbol{v}_i^\top \boldsymbol{A} \boldsymbol{v}_j = 0, \qquad \text{dla dowolnych} \quad i \neq j$$

Wektory sprzężone

Niech $\boldsymbol{A}$ będzie macierzą **dodatnio określoną**
Zbiór **niezerowych** wektorów $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ nazywamy **sprzężonymi** (ze względu na macierz $\boldsymbol{A}$), jeśli

$$\boldsymbol{v}_i^\top \boldsymbol{A} \boldsymbol{v}_j = 0, \qquad \text{dla dowolnych} \quad i \neq j$$

Fakt: Zbiór wektorów sprzężonych jest **bazą**

Wektory sprzężone

Niech $\boldsymbol{A}$ będzie macierzą **dodatnio określoną**
Zbiór **niezerowych** wektorów $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ nazywamy **sprzężonymi** (ze względu na macierz $\boldsymbol{A}$), jeśli

$$\boldsymbol{v}_i^\top \boldsymbol{A} \boldsymbol{v}_j = 0, \qquad \text{dla dowolnych} \quad i \neq j$$

Fakt: Zbiór wektorów sprzężonych jest **bazą**

Dowód: Wystarczy pokazać ich liniową niezależność. Załóżmy przeciwnie, że istnieje wektor $\boldsymbol{v}_k$ liniowo zależny od pozostałych, tzn.

$$\boldsymbol{v}_k = \sum_{i \neq k} \alpha_i \boldsymbol{v}_i$$

Wektory sprzężone

Niech $\boldsymbol{A}$ będzie macierzą **dodatnio określoną**
Zbiór **niezerowych** wektorów $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ nazywamy **sprzężonymi** (ze względu na macierz $\boldsymbol{A}$), jeśli

$$\boldsymbol{v}_i^\top \boldsymbol{A} \boldsymbol{v}_j = 0, \qquad \text{dla dowolnych} \quad i \neq j$$

Fakt: Zbiór wektorów sprzężonych jest **bazą**

Dowód: Wystarczy pokazać ich liniową niezależność. Załóżmy przeciwnie, że istnieje wektor $\boldsymbol{v}_k$ liniowo zależny od pozostałych, tzn.

$$\boldsymbol{v}_k = \sum_{i \neq k} \alpha_i \boldsymbol{v}_i$$

Ale oznacza to, że:

$$\boldsymbol{A}\boldsymbol{v}_k = \sum_{i \neq k} \alpha_i \boldsymbol{A}\boldsymbol{v}_i$$

Wektory sprzężone

Niech $\boldsymbol{A}$ będzie macierzą **dodatnio określoną**
Zbiór **niezerowych** wektorów $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ nazywamy **sprzężonymi** (ze względu na macierz $\boldsymbol{A}$), jeśli

$$\boldsymbol{v}_i^\top \boldsymbol{A} \boldsymbol{v}_j = 0, \qquad \text{dla dowolnych} \quad i \neq j$$

Fakt: Zbiór wektorów sprzężonych jest **bazą**

Dowód: Wystarczy pokazać ich liniową niezależność. Załóżmy przeciwnie, że istnieje wektor $\boldsymbol{v}_k$ liniowo zależny od pozostałych, tzn.

$$\boldsymbol{v}_k = \sum_{i \neq k} \alpha_i \boldsymbol{v}_i$$

Ale oznacza to, że:

$$\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k = \sum_{i \neq k} \alpha_i \boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_i$$

Wektory sprzężone

Niech $\boldsymbol{A}$ będzie macierzą **dodatnio określoną**
Zbiór **niezerowych** wektorów $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ nazywamy **sprzężonymi** (ze względu na macierz $\boldsymbol{A}$), jeśli

$$\boldsymbol{v}_i^\top \boldsymbol{A} \boldsymbol{v}_j = 0, \qquad \text{dla dowolnych} \quad i \neq j$$

Fakt: Zbiór wektorów sprzężonych jest **bazą**

Dowód: Wystarczy pokazać ich liniową niezależność. Załóżmy przeciwnie, że istnieje wektor $\boldsymbol{v}_k$ liniowo zależny od pozostałych, tzn.

$$\boldsymbol{v}_k = \sum_{i \neq k} \alpha_i \boldsymbol{v}_i$$

Ale oznacza to, że:

$$\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k = \sum_{i \neq k} \alpha_i \underbrace{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_i}_{=0}$$

Wektory sprzężone

Niech $\boldsymbol{A}$ będzie macierzą **dodatnio określoną**
Zbiór **niezerowych** wektorów $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ nazywamy **sprzężonymi** (ze względu na macierz $\boldsymbol{A}$), jeśli

$$\boldsymbol{v}_i^\top \boldsymbol{A} \boldsymbol{v}_j = 0, \qquad \text{dla dowolnych} \quad i \neq j$$

Fakt: Zbiór wektorów sprzężonych jest **bazą**

Dowód: Wystarczy pokazać ich liniową niezależność. Załóżmy przeciwnie, że istnieje wektor $\boldsymbol{v}_k$ liniowo zależny od pozostałych, tzn.

$$\boldsymbol{v}_k = \sum_{i \neq k} \alpha_i \boldsymbol{v}_i$$

Ale oznacza to, że:

$$\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k = 0$$

co prowadzi do sprzeczności, ponieważ $\boldsymbol{A}$ jest dodatnio określona, a $\boldsymbol{v}_k$ jest niezerowy, więc $\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k > 0$

Funkcja kwadratowa

Będziemy rozważali minimalizację funkcji kwadratowej:

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{b}, \qquad \boldsymbol{A} \text{ – dodatnio określona}$$

Pomijamy współczynnik c (nie wpływa na optymalizację)

Pozostałe współczynniki dobieramy w ten sposób, aby:

$$\nabla f(\boldsymbol{x}) \;=\; \mathbf{0} \quad\Longleftrightarrow\quad \boldsymbol{A}\boldsymbol{x} \;=\; \boldsymbol{b},$$

tzn. minimum $\boldsymbol{x}^* = \boldsymbol{A}^{-1}\boldsymbol{b}$ jest rozwiązaniem układu równań $\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{b}$

Metoda gradientów sprzężonych: wprowadzenie

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{b}, \qquad \boldsymbol{A} \text{ – dodatnio określona}$$

Niech $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ będzie zbiorem wektorów sprzężonych ze względu na $\boldsymbol{A}$

Będziemy rozważali metodę spadkową minimalizacji funkcji kwadratowej, w której kolejne kierunki poprawy to **kolejne wektory sprzężone**, tzn.:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k, \qquad \text{tzn.} \;\; \boldsymbol{x}_{k+1} = \boldsymbol{x}_1 + \sum_{i=1}^{k} \alpha_i \boldsymbol{v}_i,$$

natomiast współczynniki α_k wyznaczamy rozwiązując:

$$\alpha_k \;=\; \operatorname*{argmin}_{\alpha} \underbrace{f(\boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k)}_{g(\alpha)}$$

Metoda gradientów sprzężonych: wprowadzenie

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{b}, \qquad \boldsymbol{A} \text{ – dodatnio określona}$$

Niech $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ będzie zbiorem wektorów sprzężonych ze względu na $\boldsymbol{A}$

Będziemy rozważali metodę spadkową minimalizacji funkcji kwadratowej, w której kolejne kierunki poprawy to **kolejne wektory sprzężone**, tzn.:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k, \qquad \text{tzn.} \;\; \boldsymbol{x}_{k+1} = \boldsymbol{x}_1 + \sum_{i=1}^{k} \alpha_i \boldsymbol{v}_i,$$

natomiast współczynniki α_k wyznaczamy rozwiązując:

$$\alpha_k = \operatorname*{argmin}_{\alpha} \underbrace{f(\boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k)}_{g(\alpha)}$$

$$g(\alpha) = \frac{1}{2}(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k)^\top \boldsymbol{A}(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k) - (\boldsymbol{x}_k + \alpha \boldsymbol{v}_k)^\top \boldsymbol{b}$$

Metoda gradientów sprzężonych: wprowadzenie

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{b}, \qquad \boldsymbol{A} \text{ – dodatnio określona}$$

Niech $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ będzie zbiorem wektorów sprzężonych ze względu na $\boldsymbol{A}$

Będziemy rozważali metodę spadkową minimalizacji funkcji kwadratowej, w której kolejne kierunki poprawy to **kolejne wektory sprzężone**, tzn.:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k, \qquad \text{tzn. } \boldsymbol{x}_{k+1} = \boldsymbol{x}_1 + \sum_{i=1}^{k} \alpha_i \boldsymbol{v}_i,$$

natomiast współczynniki α_k wyznaczamy rozwiązując:

$$\alpha_k = \operatorname*{argmin}_{\alpha} \underbrace{f(\boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k)}_{g(\alpha)}$$

$$g(\alpha) = \frac{\alpha^2}{2}\boldsymbol{v}_k^\top \boldsymbol{A}\boldsymbol{v}_k + \alpha\big(\boldsymbol{v}_k^\top \boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{v}_k^\top \boldsymbol{b}\big) + \text{const}$$

Metoda gradientów sprzężonych: wprowadzenie

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{b}, \qquad \boldsymbol{A} \text{ – dodatnio określona}$$

Niech $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ będzie zbiorem wektorów sprzężonych ze względu na $\boldsymbol{A}$

Będziemy rozważali metodę spadkową minimalizacji funkcji kwadratowej, w której kolejne kierunki poprawy to **kolejne wektory sprzężone**, tzn.:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k, \qquad \text{tzn.} \quad \boldsymbol{x}_{k+1} = \boldsymbol{x}_1 + \sum_{i=1}^{k} \alpha_i \boldsymbol{v}_i,$$

gradient: $\boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{b} = \nabla f(\boldsymbol{x}_k)$
oznaczamy przez $\boldsymbol{g}_k$

natomias … rozwiązując:

$$\alpha_k \;=\; \operatorname*{argmin}_{\alpha} \underbrace{f(\boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k)}_{g(\alpha)}$$

$$g(\alpha) = \frac{\alpha^2}{2}\boldsymbol{v}_k^\top \boldsymbol{A}\boldsymbol{v}_k + \alpha \boldsymbol{v}_k^\top (\boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{b}) + \text{const}$$

Metoda gradientów sprzężonych: wprowadzenie

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{b}, \qquad \boldsymbol{A} \text{ – dodatnio określona}$$

Niech $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ będzie zbiorem wektorów sprzężonych ze względu na $\boldsymbol{A}$

Będziemy rozważali metodę spadkową minimalizacji funkcji kwadratowej, w której kolejne kierunki poprawy to **kolejne wektory sprzężone**, tzn.:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k, \qquad \text{tzn. } \boldsymbol{x}_{k+1} = \boldsymbol{x}_1 + \sum_{i=1}^{k} \alpha_i \boldsymbol{v}_i,$$

natomiast współczynniki α_k wyznaczamy rozwiązując:

$$\alpha_k = \operatorname*{argmin}_{\alpha} \underbrace{f(\boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k)}_{g(\alpha)}$$

$$g(\alpha) = \frac{\alpha^2}{2}\boldsymbol{v}_k^\top \boldsymbol{A}\boldsymbol{v}_k + \alpha \boldsymbol{v}_k^\top \boldsymbol{g}_k + \text{const}$$

Metoda gradientów sprzężonych: wprowadzenie

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{b}, \qquad \boldsymbol{A} \text{ – dodatnio określona}$$

Niech $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ będzie zbiorem wektorów sprzężonych ze względu na $\boldsymbol{A}$

Będziemy rozważali metodę spadkową minimalizacji funkcji kwadratowej, w której kolejne kierunki poprawy to **kolejne wektory sprzężone**, tzn.:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k, \qquad \text{tzn.} \;\; \boldsymbol{x}_{k+1} = \boldsymbol{x}_1 + \sum_{i=1}^{k} \alpha_i \boldsymbol{v}_i,$$

natomiast współczynniki α_k wyznaczamy rozwiązując:

$$\alpha_k = \operatorname*{argmin}_{\alpha} \underbrace{f(\boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k)}_{g(\alpha)}$$

$$g'(\alpha) = \alpha \boldsymbol{v}_k^\top \boldsymbol{A}\boldsymbol{v}_k + \boldsymbol{v}_k^\top \boldsymbol{g}_k$$

Metoda gradientów sprzężonych: wprowadzenie

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{b}, \qquad \boldsymbol{A} \text{ – dodatnio określona}$$

Niech $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ będzie zbiorem wektorów sprzężonych ze względu na $\boldsymbol{A}$

Będziemy rozważali metodę spadkową minimalizacji funkcji kwadratowej, w której kolejne kierunki poprawy to **kolejne wektory sprzężone**, tzn.:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k, \qquad \text{tzn. } \boldsymbol{x}_{k+1} = \boldsymbol{x}_1 + \sum_{i=1}^{k} \alpha_i \boldsymbol{v}_i,$$

natomiast współczynniki α_k wyznaczamy rozwiązując:

$$\alpha_k = \operatorname*{argmin}_{\alpha} \underbrace{f(\boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k)}_{g(\alpha)}$$

$$g'(\alpha) = 0 \iff \alpha \boldsymbol{v}_k^\top \boldsymbol{A}\boldsymbol{v}_k = -\boldsymbol{v}_k^\top \boldsymbol{g}_k$$

Metoda gradientów sprzężonych: wprowadzenie

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{b}, \qquad \boldsymbol{A} \text{ – dodatnio określona}$$

Niech $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ będzie zbiorem wektorów sprzężonych ze względu na $\boldsymbol{A}$

Będziemy rozważali metodę spadkową minimalizacji funkcji kwadratowej, w której kolejne kierunki poprawy to **kolejne wektory sprzężone**, tzn.:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k, \qquad \text{tzn.} \;\; \boldsymbol{x}_{k+1} = \boldsymbol{x}_1 + \sum_{i=1}^{k} \alpha_i \boldsymbol{v}_i,$$

natomiast współczynniki α_k wyznaczamy rozwiązując:

$$\alpha_k = \operatorname*{argmin}_{\alpha} \underbrace{f(\boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k)}_{g(\alpha)}$$

$$\alpha_k = -\frac{\boldsymbol{v}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A}\boldsymbol{v}_k}$$

Metoda gradientów sprzężonych: wprowadzenie

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{b}, \qquad \boldsymbol{A} \text{ – dodatnio określona}$$

Niech $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ będzie zbiorem wektorów sprzężonych ze względu na $\boldsymbol{A}$

Będziemy rozważali metodę spadkową minimalizacji funkcji kwadratowej, w której kolejne kierunki poprawy to **kolejne wektory sprzężone**, tzn.:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k, \qquad \text{tzn.} \;\; \boldsymbol{x}_{k+1} = \boldsymbol{x}_1 + \sum_{i=1}^{k} \alpha_i \boldsymbol{v}_i,$$

natomiast współczynniki α_k wyznaczamy rozwiązując:

$$\alpha_k \;=\; \operatorname*{argmin}_{\alpha} \underbrace{f(\boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k)}_{g(\alpha)}$$

$$\alpha_k \;=\; -\frac{\boldsymbol{v}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A}\boldsymbol{v}_k} \;=\; -\frac{\boldsymbol{v}_k^\top (\boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{b})}{\boldsymbol{v}_k^\top \boldsymbol{A}\boldsymbol{v}_k}$$

Metoda gradientów sprzężonych: wprowadzenie

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{b}, \qquad \boldsymbol{A} \text{ – dodatnio określona}$$

Niech $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ będzie zbiorem wektorów sprzężonych ze względu na $\boldsymbol{A}$

Będziemy rozważali metodę spadkową minimalizacji funkcji kwadratowej, w której kolejne kierunki poprawy to **kolejne wektory sprzężone**, tzn.:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k, \qquad \text{tzn. } \boldsymbol{x}_{k+1} = \boldsymbol{x}_1 + \sum_{i=1}^{k} \alpha_i \boldsymbol{v}_i,$$

natomiast współczynniki α_k wyznaczamy rozwiązując:

$$\alpha_k = \operatorname*{argmin}_{\alpha} \underbrace{f(\boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k)}_{g(\alpha)}$$

$$\alpha_k = -\frac{\boldsymbol{v}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A}\boldsymbol{v}_k} = -\frac{\boldsymbol{v}_k^\top (\boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{A}\boldsymbol{x}_1 - \boldsymbol{b})}{\boldsymbol{v}_k^\top \boldsymbol{A}\boldsymbol{v}_k}$$

Metoda gradientów sprzężonych: wprowadzenie

Ponieważ $\boldsymbol{x}_k - \boldsymbol{x}_1 = \sum_{i=1}^{k-1} \alpha_i \boldsymbol{v}_i$, otrzymujemy

$$\boldsymbol{v}_k^\top \boldsymbol{A}(\boldsymbol{x}_k - \boldsymbol{x}_1) = \sum_{i=1}^{k-1} \alpha_i \underbrace{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_i}_{=0} = 0$$

dodatnio określona

sprzężonych ze względu na $\boldsymbol{A}$

alizacji funkcji kwadratowej,

wektory sprzężone, tzn.:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k, \qquad \text{tzn.} \quad \boldsymbol{x}_{k+1} = \boldsymbol{x}_1 + \sum_{i=1}^{k} \alpha_i \boldsymbol{v}_i,$$

natomiast współczynniki α_k wyznaczamy rozwiązując:

$$\alpha_k = \operatorname*{argmin}_{\alpha} \underbrace{f(\boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k)}_{g(\alpha)}$$

$$\alpha_k = -\frac{\boldsymbol{v}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k} = -\frac{\boldsymbol{v}_k^\top (\boldsymbol{A}\boldsymbol{x}_1 - \boldsymbol{b}) + \boldsymbol{v}_k^\top \boldsymbol{A}(\boldsymbol{x}_k - \boldsymbol{x}_1)}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}$$

Metoda gradientów sprzężonych: wprowadzenie

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{b}, \qquad \boldsymbol{A} \text{ – dodatnio określona}$$

Niech $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ będzie zbiorem wektorów sprzężonych ze względu na $\boldsymbol{A}$

Będziemy rozważali metodę spadkową minimalizacji funkcji kwadratowej, w której kolejne kierunki poprawy to **kolejne wektory sprzężone**, tzn.:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k, \qquad \text{tzn. } \boldsymbol{x}_{k+1} = \boldsymbol{x}_1 + \sum_{i=1}^{k} \alpha_i \boldsymbol{v}_i,$$

natomiast współczynniki α_k wyznaczamy rozwiązując:

$$\alpha_k = \operatorname*{argmin}_{\alpha} \underbrace{f(\boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k)}_{g(\alpha)}$$

$$\alpha_k = -\frac{\boldsymbol{v}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A}\boldsymbol{v}_k} = \frac{\boldsymbol{v}_k^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_1)}{\boldsymbol{v}_k^\top \boldsymbol{A}\boldsymbol{v}_k}$$

Optymalna wartość α_k zależy tylko od punktu startowego $\boldsymbol{x}_1$!

Możemy wyznaczyć α_k niezależne od pozostałych współczynników i wcześniejszych kroków algorytmu

Funkcja kwadratowa a wektory sprzężone

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{b}, \qquad \boldsymbol{A} \text{ – dodatnio określona}$$

Niech $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ – zbiór wektorów sprzężonych. Ponieważ jest to baza,

$$\boldsymbol{x}^* - \boldsymbol{x}_1 = \sum_{i=1}^{n} \alpha_i \boldsymbol{v}_i$$

Wyznaczmy optymalne współczynniki α_k

Funkcja kwadratowa a wektory sprzężone

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{b}, \qquad \boldsymbol{A} \text{ – dodatnio określona}$$

Niech $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ – zbiór wektorów sprzężonych. Ponieważ jest to baza,

$$\boldsymbol{x}^* - \boldsymbol{x}_1 = \sum_{i=1}^{n} \alpha_i \boldsymbol{v}_i$$

Wyznaczmy optymalne współczynniki α_k

$$\boldsymbol{A}\boldsymbol{x}^* - \boldsymbol{A}\boldsymbol{x}_1 = \sum_{i=1}^{n} \alpha_i \boldsymbol{A}\boldsymbol{v}_i$$

Funkcja kwadratowa a wektory sprzężone

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{b}, \qquad \boldsymbol{A} \text{ – dodatnio określona}$$

Niech $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ – zbiór wektorów sprzężonych. Ponieważ jest to baza,

$$\boldsymbol{x}^* - \boldsymbol{x}_1 = \sum_{i=1}^{n} \alpha_i \boldsymbol{v}_i$$

Wyznaczmy optymalne współczynniki α_k

$$\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_1 = \sum_{i=1}^{n} \alpha_i \boldsymbol{A}\boldsymbol{v}_i$$

Funkcja kwadratowa a wektory sprzężone

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{b}, \qquad \boldsymbol{A} \text{ – dodatnio określona}$$

Niech $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ – zbiór wektorów sprzężonych. Ponieważ jest to baza,

$$\boldsymbol{x}^* - \boldsymbol{x}_1 = \sum_{i=1}^{n} \alpha_i \boldsymbol{v}_i$$

Wyznaczmy optymalne współczynniki α_k

$$\boldsymbol{v}_k^\top(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_1) = \sum_{i=1}^{n} \alpha_k \underbrace{\boldsymbol{v}_k^\top \boldsymbol{A}\boldsymbol{v}_i}_{=0 \text{ dla } k \neq i}$$

Funkcja kwadratowa a wektory sprzężone

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{b}, \qquad \boldsymbol{A} \text{ – dodatnio określona}$$

Niech $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ – zbiór wektorów sprzężonych. Ponieważ jest to baza,

$$\boldsymbol{x}^* - \boldsymbol{x}_1 = \sum_{i=1}^{n} \alpha_i \boldsymbol{v}_i$$

Wyznaczmy optymalne współczynniki α_k

$$\boldsymbol{v}_k^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_1) = \alpha_k \boldsymbol{v}_k^\top \boldsymbol{A}\boldsymbol{v}_k$$

Funkcja kwadratowa a wektory sprzężone

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{b}, \qquad \boldsymbol{A} \text{ – dodatnio określona}$$

Niech $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ – zbiór wektorów sprzężonych. Ponieważ jest to baza,

$$\boldsymbol{x}^* - \boldsymbol{x}_1 = \sum_{i=1}^{n} \alpha_i \boldsymbol{v}_i$$

Wyznaczmy optymalne współczynniki α_k

$$\alpha_k = \frac{\boldsymbol{v}_k^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_1)}{\boldsymbol{v}_k^\top \boldsymbol{A}\boldsymbol{v}_k},$$

Funkcja kwadratowa a wektory sprzężone

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{b}, \qquad \boldsymbol{A} \text{ – dodatnio określona}$$

Niech $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ – zbiór wektorów sprzężonych. Ponieważ jest to baza,

$$\boldsymbol{x}^* - \boldsymbol{x}_1 = \sum_{i=1}^{n} \alpha_i \boldsymbol{v}_i$$

Wyznaczmy optymalne współczynniki α_k

$$\alpha_k = \frac{\boldsymbol{v}_k^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_1)}{\boldsymbol{v}_k^\top \boldsymbol{A}\boldsymbol{v}_k},$$

Ale są to **te same** współczynniki, które dobiera w kolejnych krokach metoda gradientów sprzężonych, tzn.

$$\boldsymbol{x}_{n+1} \;=\; \boldsymbol{x}_1 + \sum_{k=1}^{n} \alpha_k \boldsymbol{v}_k \;=\; \boldsymbol{x}^*$$

Funkcja kwadratowa a wektory sprzężone

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{b}, \qquad \boldsymbol{A} \text{ – dodatnio określona}$$

Niech $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ – zbiór wektorów sprzężonych. Ponieważ jest to baza,

$$\boldsymbol{x}^* - \boldsymbol{x}_1 = \sum_{i=1}^{n} \alpha_i \boldsymbol{v}_i$$

Wyznaczmy optymalne współczynniki α_k

$$\alpha_k = \frac{\boldsymbol{v}_k^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_1)}{\boldsymbol{v}_k^\top \boldsymbol{A}\boldsymbol{v}_k},$$

Ale są to **te same** współczynniki, które dobiera w kolejnych krokach metoda gradientów sprzężonych, tzn.

$$\boldsymbol{x}_{n+1} \;=\; \boldsymbol{x}_1 + \sum_{k=1}^{n} \alpha_k \boldsymbol{v}_k \;=\; \boldsymbol{x}^*$$

Wniosek: Metoda gradientów sprzężonych w każdym kroku optymalnie dobiera współczynnik α_k, a rozwiązanie $\boldsymbol{x}^*$ otrzymujemy po n iteracjach!

Ortogonalność gradientów

Gradient w danej iteracji $\boldsymbol{g}_k = \nabla f(\boldsymbol{x}_k)$ jest **ortogonalny** do wszystkich poprzednich kierunków poprawy $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{k-1}$, tzn.

$$\boldsymbol{v}_i^\top \boldsymbol{g}_k = 0, \qquad \text{dla} \quad i = 1, \ldots, k-1$$

Ortogonalność gradientów

Gradient w danej iteracji $\boldsymbol{g}_k = \nabla f(\boldsymbol{x}_k)$ jest **ortogonalny** do wszystkich poprzednich kierunków poprawy $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{k-1}$, tzn.

$$\boldsymbol{v}_i^\top \boldsymbol{g}_k = 0, \qquad \text{dla} \quad i = 1, \ldots, k-1$$

Dowód:

$$\boldsymbol{v}_i^\top \boldsymbol{g}_k \quad = \quad \boldsymbol{v}_i^\top (\boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{b})$$

Ortogonalność gradientów

Gradient w danej iteracji $\boldsymbol{g}_k = \nabla f(\boldsymbol{x}_k)$ jest **ortogonalny** do wszystkich poprzednich kierunków poprawy $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{k-1}$, tzn.

$$\boldsymbol{v}_i^\top \boldsymbol{g}_k = 0, \qquad \text{dla} \;\; i = 1, \ldots, k-1$$

Dowód:

$$\begin{aligned}
\boldsymbol{v}_i^\top \boldsymbol{g}_k &= \boldsymbol{v}_i^\top (\boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{b}) \\
&= \boldsymbol{v}_i^\top \left(\boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{A}\Big(\sum_{j=1}^{k-1} \alpha_j \boldsymbol{v}_j \Big) - \boldsymbol{b} \right)
\end{aligned}$$

Ortogonalność gradientów

Gradient w danej iteracji $\boldsymbol{g}_k = \nabla f(\boldsymbol{x}_k)$ jest **ortogonalny** do wszystkich poprzednich kierunków poprawy $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{k-1}$, tzn.

$$\boldsymbol{v}_i^\top \boldsymbol{g}_k = 0, \qquad \text{dla} \;\; i = 1, \ldots, k-1$$

Dowód:

$$\begin{aligned}
\boldsymbol{v}_i^\top \boldsymbol{g}_k &= \boldsymbol{v}_i^\top (\boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{b}) \\
&= \boldsymbol{v}_i^\top \left(\boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{A}\Big(\sum_{j=1}^{k-1} \alpha_j \boldsymbol{v}_j \Big) - \boldsymbol{b} \right) \\
&= \boldsymbol{v}_i^\top (\boldsymbol{A}\boldsymbol{x}_1 - \boldsymbol{b}) + \sum_{j=1}^{k-1} \alpha_j \underbrace{\boldsymbol{v}_i^\top \boldsymbol{A} \boldsymbol{v}_j}_{=0 \text{ dla } i \neq j}
\end{aligned}$$

Ortogonalność gradientów

Gradient w danej iteracji $\boldsymbol{g}_k = \nabla f(\boldsymbol{x}_k)$ jest **ortogonalny** do wszystkich poprzednich kierunków poprawy $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{k-1}$, tzn.

$$\boldsymbol{v}_i^\top \boldsymbol{g}_k = 0, \qquad \text{dla} \;\; i = 1, \ldots, k-1$$

Dowód:

$$\begin{aligned}
\boldsymbol{v}_i^\top \boldsymbol{g}_k &= \boldsymbol{v}_i^\top (\boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{b}) \\
&= \boldsymbol{v}_i^\top \bigg(\boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{A}\Big(\sum_{j=1}^{k-1} \alpha_j \boldsymbol{v}_j \Big) - \boldsymbol{b} \bigg) \\
&= \boldsymbol{v}_i^\top (\boldsymbol{A}\boldsymbol{x}_1 - \boldsymbol{b}) + \sum_{j=1}^{k-1} \alpha_j \underbrace{\boldsymbol{v}_i^\top \boldsymbol{A}\boldsymbol{v}_j}_{=0 \text{ dla } i \neq j} \\
&= \boldsymbol{v}_i^\top (\boldsymbol{A}\boldsymbol{x}_1 - \boldsymbol{b}) + \alpha_i \boldsymbol{v}_i^\top \boldsymbol{A}\boldsymbol{v}_i
\end{aligned}$$

Ortogonalność gradientów

Gradient w danej iteracji $\boldsymbol{g}_k = \nabla f(\boldsymbol{x}_k)$ jest **ortogonalny** do wszystkich poprzednich kierunków poprawy $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{k-1}$, tzn.

$$\boldsymbol{v}_i^\top \boldsymbol{g}_k = 0, \qquad \text{dla} \;\; i = 1, \ldots, k-1$$

Dowód:

$$\begin{aligned}
\boldsymbol{v}_i^\top \boldsymbol{g}_k &= \boldsymbol{v}_i^\top (\boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{b}) \\
&= \boldsymbol{v}_i^\top \Big(\boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{A}\Big(\sum_{j=1}^{k-1} \alpha_j \boldsymbol{v}_j \Big) - \boldsymbol{b} \Big) \\
&= \boldsymbol{v}_i^\top (\boldsymbol{A}\boldsymbol{x}_1 - \boldsymbol{b}) + \sum_{j=1}^{k-1} \alpha_j \underbrace{\boldsymbol{v}_i^\top \boldsymbol{A}\boldsymbol{v}_j}_{=0 \text{ dla } i \neq j} \\
&= \boldsymbol{v}_i^\top (\boldsymbol{A}\boldsymbol{x}_1 - \boldsymbol{b}) + \alpha_i \boldsymbol{v}_i^\top \boldsymbol{A}\boldsymbol{v}_i \\
&= \boldsymbol{v}_i^\top (\boldsymbol{A}\boldsymbol{x}_1 - \boldsymbol{b}) + \frac{\boldsymbol{v}_i^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_1)}{\boldsymbol{v}_i^\top \boldsymbol{A}\boldsymbol{v}_i} \boldsymbol{v}_i^\top \boldsymbol{A}\boldsymbol{v}_i \; = \; 0
\end{aligned}$$

Jak znaleźć kierunki sprzężone?

Jak znaleźć kierunki sprzężone?

- Jako pierwszy kierunek sprzężony możemy obrać negatywny gradient:

$$\boldsymbol{v}_1 = -\nabla f(\boldsymbol{x}_1) = -\boldsymbol{g}_1$$

Jak znaleźć kierunki sprzężone?

- Jako pierwszy kierunek sprzężony możemy obrać negatywny gradient:

$$\boldsymbol{v}_1 \;=\; -\nabla f(\boldsymbol{x}_1) \;=\; -\boldsymbol{g}_1$$

- Załóżmy, że dobraliśmy już kierunki sprzężone $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ i chcemy dobrać kierunek $\boldsymbol{v}_{k+1}$. Ponownie, sensownym kandydatem jest $-\boldsymbol{g}_{k+1}$, ale niekoniecznie jest on sprzężony do $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$

Jak znaleźć kierunki sprzężone?

- Jako pierwszy kierunek sprzężony możemy obrać negatywny gradient:

$$\boldsymbol{v}_1 \;=\; -\nabla f(\boldsymbol{x}_1) \;=\; -\boldsymbol{g}_1$$

- Załóżmy, że dobraliśmy już kierunki sprzężone $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ i chcemy dobrać kierunek $\boldsymbol{v}_{k+1}$. Ponownie, sensownym kandydatem jest $-\boldsymbol{g}_{k+1}$, ale niekoniecznie jest on sprzężony do $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$
- **Pomysł:** dodajmy do gradientu kombinację liniową poprzednich kierunków aby stał się sprzężony, tzn. aby $\boldsymbol{v}_j^\top \boldsymbol{A}\boldsymbol{v}_{k+1} = 0$ dla $j \leqslant k$

$$\boldsymbol{v}_{k+1} \;=\; -\boldsymbol{g}_{k+1} + \sum_{i=1}^{k} \beta_{k,i}\boldsymbol{v}_i$$

Jak znaleźć kierunki sprzężone?

- Jako pierwszy kierunek sprzężony możemy obrać negatywny gradient:

$$\boldsymbol{v}_1 = -\nabla f(\boldsymbol{x}_1) = -\boldsymbol{g}_1$$

- Załóżmy, że dobraliśmy już kierunki sprzężone $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ i chcemy dobrać kierunek $\boldsymbol{v}_{k+1}$. Ponownie, sensownym kandydatem jest $-\boldsymbol{g}_{k+1}$, ale niekoniecznie jest on sprzężony do $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$
- **Pomysł:** dodajmy do gradientu kombinację liniową poprzednich kierunków aby stał się sprzężony, tzn. aby $\boldsymbol{v}_j^\top \boldsymbol{A} \boldsymbol{v}_{k+1} = 0$ dla $j \leqslant k$

$$\boldsymbol{v}_{k+1} = -\boldsymbol{g}_{k+1} + \sum_{i=1}^{k} \beta_{k,i} \boldsymbol{v}_i$$

- Wyznaczmy te współczynniki:

$$\boldsymbol{v}_{k+1} = -\boldsymbol{g}_{k+1} + \sum_{i=1}^{k} \beta_{k,i} \boldsymbol{v}_i$$

Jak znaleźć kierunki sprzężone?

- Jako pierwszy kierunek sprzężony możemy obrać negatywny gradient:

$$\boldsymbol{v}_1 \;=\; -\nabla f(\boldsymbol{x}_1) \;=\; -\boldsymbol{g}_1$$

- Załóżmy, że dobraliśmy już kierunki sprzężone $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ i chcemy dobrać kierunek $\boldsymbol{v}_{k+1}$. Ponownie, sensownym kandydatem jest $-\boldsymbol{g}_{k+1}$, ale niekoniecznie jest on sprzężony do $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$
- **Pomysł:** dodajmy do gradientu kombinację liniową poprzednich kierunków aby stał się sprzężony, tzn. aby $\boldsymbol{v}_j^\top \boldsymbol{A}\boldsymbol{v}_{k+1} = 0$ dla $j \leqslant k$

$$\boldsymbol{v}_{k+1} \;=\; -\boldsymbol{g}_{k+1} + \sum_{i=1}^{k} \beta_{k,i}\boldsymbol{v}_i$$

- Wyznaczmy te współczynniki:

$$\boldsymbol{A}\boldsymbol{v}_{k+1} \;=\; -\boldsymbol{A}\boldsymbol{g}_{k+1} + \sum_{i=1}^{k} \beta_{k,i}\boldsymbol{A}\boldsymbol{v}_i$$

Jak znaleźć kierunki sprzężone?

- Jako pierwszy kierunek sprzężony możemy obrać negatywny gradient:

$$\boldsymbol{v}_1 \;=\; -\nabla f(\boldsymbol{x}_1) \;=\; -\boldsymbol{g}_1$$

- Załóżmy, że dobraliśmy już kierunki sprzężone $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ i chcemy dobrać kierunek $\boldsymbol{v}_{k+1}$. Ponownie, sensownym kandydatem jest $-\boldsymbol{g}_{k+1}$, ale niekoniecznie jest on sprzężony do $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$
- **Pomysł:** dodajmy do gradientu kombinację liniową poprzednich kierunków aby stał się sprzężony, tzn. aby $\boldsymbol{v}_j^\top \boldsymbol{A}\boldsymbol{v}_{k+1} = 0$ dla $j \leqslant k$

$$\boldsymbol{v}_{k+1} \;=\; -\boldsymbol{g}_{k+1} + \sum_{i=1}^{k} \beta_{k,i}\boldsymbol{v}_i$$

- Wyznaczmy te współczynniki:

$$\boldsymbol{v}_j^\top \boldsymbol{A}\boldsymbol{v}_{k+1} \;=\; -\boldsymbol{v}_j^\top \boldsymbol{A}\boldsymbol{g}_{k+1} + \sum_{i=1}^{k} \beta_{k,i}\, \underbrace{\boldsymbol{v}_j^\top \boldsymbol{A}\boldsymbol{v}_i}_{=0 \text{ dla } j \neq i}$$

Jak znaleźć kierunki sprzężone?

- Jako pierwszy kierunek sprzężony możemy obrać negatywny gradient:

$$\boldsymbol{v}_1 \;=\; -\nabla f(\boldsymbol{x}_1) \;=\; -\boldsymbol{g}_1$$

- Załóżmy, że dobraliśmy już kierunki sprzężone $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ i chcemy dobrać kierunek $\boldsymbol{v}_{k+1}$. Ponownie, sensownym kandydatem jest $-\boldsymbol{g}_{k+1}$, ale niekoniecznie jest on sprzężony do $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$
- **Pomysł:** dodajmy do gradientu kombinację liniową poprzednich kierunków aby stał się sprzężony, tzn. aby $\boldsymbol{v}_j^\top \boldsymbol{A} \boldsymbol{v}_{k+1} = 0$ dla $j \leqslant k$

$$\boldsymbol{v}_{k+1} \;=\; -\boldsymbol{g}_{k+1} + \sum_{i=1}^{k} \beta_{k,i} \boldsymbol{v}_i$$

- Wyznaczmy te współczynniki:

$$\underbrace{\boldsymbol{v}_j^\top \boldsymbol{A} \boldsymbol{v}_{k+1}}_{=0} \;=\; -\boldsymbol{v}_j^\top \boldsymbol{A} \boldsymbol{g}_{k+1} + \beta_{k,j} \boldsymbol{v}_j^\top \boldsymbol{A} \boldsymbol{v}_j$$

Jak znaleźć kierunki sprzężone?

- Jako pierwszy kierunek sprzężony możemy obrać negatywny gradient:

$$\boldsymbol{v}_1 \;=\; -\nabla f(\boldsymbol{x}_1) \;=\; -\boldsymbol{g}_1$$

- Załóżmy, że dobraliśmy już kierunki sprzężone $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ i chcemy dobrać kierunek $\boldsymbol{v}_{k+1}$. Ponownie, sensownym kandydatem jest $-\boldsymbol{g}_{k+1}$, ale niekoniecznie jest on sprzężony do $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$
- **Pomysł:** dodajmy do gradientu kombinację liniową poprzednich kierunków aby stał się sprzężony, tzn. aby $\boldsymbol{v}_j^\top \boldsymbol{A}\boldsymbol{v}_{k+1} = 0$ dla $j \leqslant k$

$$\boldsymbol{v}_{k+1} \;=\; -\boldsymbol{g}_{k+1} + \sum_{i=1}^{k} \beta_{k,i}\boldsymbol{v}_i$$

- Wyznaczmy te współczynniki:

$$\boldsymbol{v}_j^\top \boldsymbol{A}\boldsymbol{g}_{k+1} = \beta_{k,j}\boldsymbol{v}_j^\top \boldsymbol{A}\boldsymbol{v}_j$$

Jak znaleźć kierunki sprzężone?

- Jako pierwszy kierunek sprzężony możemy obrać negatywny gradient:

$$\boldsymbol{v}_1 \;=\; -\nabla f(\boldsymbol{x}_1) \;=\; -\boldsymbol{g}_1$$

- Załóżmy, że dobraliśmy już kierunki sprzężone $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ i chcemy dobrać kierunek $\boldsymbol{v}_{k+1}$. Ponownie, sensownym kandydatem jest $-\boldsymbol{g}_{k+1}$, ale niekoniecznie jest on sprzężony do $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$
- **Pomysł:** dodajmy do gradientu kombinację liniową poprzednich kierunków aby stał się sprzężony, tzn. aby $\boldsymbol{v}_j^\top \boldsymbol{A}\boldsymbol{v}_{k+1} = 0$ dla $j \leqslant k$

$$\boldsymbol{v}_{k+1} \;=\; -\boldsymbol{g}_{k+1} + \sum_{i=1}^{k} \beta_{k,i}\boldsymbol{v}_i$$

- Wyznaczmy te współczynniki:

$$\beta_{k,j} \;=\; \frac{\boldsymbol{v}_j^\top \boldsymbol{A}\boldsymbol{g}_{k+1}}{\boldsymbol{v}_j^\top \boldsymbol{A}\boldsymbol{v}_j}$$

Jak znaleźć kierunki sprzężone?

- Jako pierwszy kierunek sprzężony możemy obrać negatywny gradient:

$$\boldsymbol{v}_1 \;=\; -\nabla f(\boldsymbol{x}_1) \;=\; -\boldsymbol{g}_1$$

- Załóżmy, że dobraliśmy już kierunki sprzężone $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ i chcemy dobrać kierunek $\boldsymbol{v}_{k+1}$. Ponownie, sensownym kandydatem jest $-\boldsymbol{g}_{k+1}$, ale niekoniecznie jest on sprzężony do $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$
- **Pomysł:** dodajmy do gradientu kombinację liniową poprzednich kierunków aby stał się sprzężony, tzn. aby $\boldsymbol{v}_j^\top \boldsymbol{A}\boldsymbol{v}_{k+1} = 0$ dla $j \leqslant k$

$$\boldsymbol{v}_{k+1} \;=\; -\boldsymbol{g}_{k+1} + \sum_{i=1}^{k} \beta_{k,i} \boldsymbol{v}_i$$

- Wyznaczmy te współczynniki:

$$\beta_{k,j} \;=\; \frac{\boldsymbol{g}_{k+1}^\top \boldsymbol{A}\boldsymbol{v}_j}{\boldsymbol{v}_j^\top \boldsymbol{A}\boldsymbol{v}_j}$$

Problem: w danej iteracji trzeba wyznaczyć k współczynników $\beta_{k,1}, \ldots, \beta_{k,k}$ i zapamiętać poprzednie kierunku poprawy $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$

Jak znaleźć kierunki sprzężone?

- Jako pierwszy kierunek sprzężony możemy obrać negatywny gradient:

$$\boldsymbol{v}_1 \;=\; -\nabla f(\boldsymbol{x}_1) \;=\; -\boldsymbol{g}_1$$

- Załóżmy, że dobraliśmy już kierunki sprzężone $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ i chcemy dobrać kierunek $\boldsymbol{v}_{k+1}$. Ponownie, sensownym kandydatem jest $-\boldsymbol{g}_{k+1}$, ale niekoniecznie jest on sprzężony do $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$
- **Pomysł:** dodajmy do gradientu kombinację liniową poprzednich kierunków aby stał się sprzężony, tzn. aby $\boldsymbol{v}_j^\top \boldsymbol{A}\boldsymbol{v}_{k+1} = 0$ dla $j \leqslant k$

$$\boldsymbol{v}_{k+1} \;=\; -\boldsymbol{g}_{k+1} + \sum_{i=1}^{k} \beta_{k,i}\boldsymbol{v}_i$$

- Wyznaczmy te współczynniki:

$$\beta_{k,j} \;=\; \frac{\boldsymbol{g}_{k+1}^\top \boldsymbol{A}\boldsymbol{v}_j}{\boldsymbol{v}_j^\top \boldsymbol{A}\boldsymbol{v}_j}$$

Proble $\beta_{k,1}, \ldots$ $\ldots, \boldsymbol{v}_k$

Magia gradientów sprzężonych:
Zachodzi $\beta_{k,j} = 0$ dla $j = 1, \ldots, k-1$, czyli pozostaje tylko **jeden** niezerowy współczynnik $\beta_{k,k}$

Kluczowy punkt konstrukcji algorytmu

$\boxed{\text{Zachodzi } \boldsymbol{g}_{k+1}^\top \boldsymbol{A} \boldsymbol{v}_j = 0 \text{ dla } j = 1, \ldots, k-1}$

Kluczowy punkt konstrukcji algorytmu

Zachodzi $\boldsymbol{g}_{k+1}^\top \boldsymbol{A}\boldsymbol{v}_j = 0$ dla $j = 1, \ldots, k-1$

Dowód: Najpierw pokazujemy, że gradient $\boldsymbol{g}_{k+1}$ jest **ortogonalny** do wszystkich poprzednich gradientów, tzn.

$$\boldsymbol{g}_{k+1}^\top \boldsymbol{g}_j = 0, \qquad \text{dla} \;\; j = 1, \ldots, k$$

Kluczowy punkt konstrukcji algorytmu

Zachodzi $\boldsymbol{g}_{k+1}^{\top} \boldsymbol{A} \boldsymbol{v}_j = 0$ dla $j = 1, \ldots, k-1$

Dowód: Najpierw pokazujemy, że gradient $\boldsymbol{g}_{k+1}$ jest **ortogonalny** do wszystkich poprzednich gradientów, tzn.

$$\boldsymbol{g}_{k+1}^{\top} \boldsymbol{g}_j = 0, \qquad \text{dla} \;\; j = 1, \ldots, k$$

- Pokazaliśmy już, że gradient jest ortogonalny do poprzednich kierunków poprawy, tzn. dla $j \leqslant k$ mamy $\boldsymbol{g}_{k+1}^{\top} \boldsymbol{v}_j = 0$

Kluczowy punkt konstrukcji algorytmu

Zachodzi $\boldsymbol{g}_{k+1}^\top \boldsymbol{A}\boldsymbol{v}_j = 0$ dla $j = 1, \ldots, k-1$

Dowód: Najpierw pokazujemy, że gradient $\boldsymbol{g}_{k+1}$ jest **ortogonalny** do wszystkich poprzednich gradientów, tzn.

$$\boldsymbol{g}_{k+1}^\top \boldsymbol{g}_j = 0, \qquad \text{dla} \;\; j = 1, \ldots, k$$

- Pokazaliśmy już, że gradient jest ortogonalny do poprzednich kierunków poprawy, tzn. dla $j \leqslant k$ mamy $\boldsymbol{g}_{k+1}^\top \boldsymbol{v}_j = 0$
- Z zasady doboru kierunku poprawy:

$$\boldsymbol{v}_j = -\boldsymbol{g}_j + \sum_{i=1}^{j-1} \beta_{j-1,i} \boldsymbol{v}_i \quad \Longrightarrow \quad \boldsymbol{g}_j = -\boldsymbol{v}_j + \sum_{i=1}^{j-1} \beta_{j-1,i} \boldsymbol{v}_i$$

Kluczowy punkt konstrukcji algorytmu

Zachodzi $\boldsymbol{g}_{k+1}^\top \boldsymbol{A}\boldsymbol{v}_j = 0$ dla $j = 1, \ldots, k-1$

Dowód: Najpierw pokazujemy, że gradient $\boldsymbol{g}_{k+1}$ jest **ortogonalny** do wszystkich poprzednich gradientów, tzn.

$$\boldsymbol{g}_{k+1}^\top \boldsymbol{g}_j = 0, \qquad \text{dla} \;\; j = 1, \ldots, k$$

- Pokazaliśmy już, że gradient jest ortogonalny do poprzednich kierunków poprawy, tzn. dla $j \leqslant k$ mamy $\boldsymbol{g}_{k+1}^\top \boldsymbol{v}_j = 0$
- Z zasady doboru kierunku poprawy:

$$\boldsymbol{v}_j = -\boldsymbol{g}_j + \sum_{i=1}^{j-1} \beta_{j-1,i} \boldsymbol{v}_i \quad \Longrightarrow \quad \boldsymbol{g}_j = -\boldsymbol{v}_j + \sum_{i=1}^{j-1} \beta_{j-1,i} \boldsymbol{v}_i$$

- Czyli dla $j \leqslant k$:

$$\boldsymbol{g}_{k+1}^\top \boldsymbol{g}_j \;=\; -\underbrace{\boldsymbol{g}_{k+1}^\top \boldsymbol{v}_j}_{=0} + \sum_{i=1}^{j-1} \beta_{j-1,i} \underbrace{\boldsymbol{g}_{k+1}^\top \boldsymbol{v}_i}_{=0} \;=\; 0$$

Kluczowy punkt konstrukcji algorytmu

Zachodzi $\boldsymbol{g}_{k+1}^\top \boldsymbol{A}\boldsymbol{v}_j = 0$ dla $j = 1, \ldots, k-1$

Dowód: Pokazaliśmy, że gradient $\boldsymbol{g}_{k+1}$ jest **ortogonalny** do wszystkich poprzednich gradientów, tzn. $\boldsymbol{g}_{k+1}^\top \boldsymbol{g}_j = 0$ dla $j = 1, \ldots, k$

Kluczowy punkt konstrukcji algorytmu

Zachodzi $\boldsymbol{g}_{k+1}^\top \boldsymbol{A}\boldsymbol{v}_j = 0$ dla $j = 1, \ldots, k-1$

Dowód: Pokazaliśmy, że gradient $\boldsymbol{g}_{k+1}$ jest **ortogonalny** do wszystkich poprzednich gradientów, tzn. $\boldsymbol{g}_{k+1}^\top \boldsymbol{g}_j = 0$ dla $j = 1, \ldots, k$

Teraz wykorzystamy fakt, że:

$$\boldsymbol{g}_{k+1} = \boldsymbol{A}\boldsymbol{x}_{k+1} - \boldsymbol{b} = \boldsymbol{A}(\boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k) - \boldsymbol{b} = \underbrace{\boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{b}}_{=\boldsymbol{g}_k} + \boldsymbol{A}\alpha_k \boldsymbol{v}_k,$$

Kluczowy punkt konstrukcji algorytmu

Zachodzi $\boldsymbol{g}_{k+1}^\top \boldsymbol{A}\boldsymbol{v}_j = 0$ dla $j = 1, \ldots, k-1$

Dowód: Pokazaliśmy, że gradient $\boldsymbol{g}_{k+1}$ jest **ortogonalny** do wszystkich poprzednich gradientów, tzn. $\boldsymbol{g}_{k+1}^\top \boldsymbol{g}_j = 0$ dla $j = 1, \ldots, k$

Teraz wykorzystamy fakt, że:

$$\boldsymbol{g}_{k+1} = \boldsymbol{A}\boldsymbol{x}_{k+1} - \boldsymbol{b} = \boldsymbol{A}(\boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k) - \boldsymbol{b} = \underbrace{\boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{b}}_{=\boldsymbol{g}_k} + \boldsymbol{A}\alpha_k \boldsymbol{v}_k,$$

Jeśli $\alpha_k = 0$, to $\boldsymbol{g}_{k+1} = \boldsymbol{g}_k$, a ponieważ $\boldsymbol{g}_{k+1}^\top \boldsymbol{g}_k = 0$, wnioskujemy, że $\boldsymbol{g}_{k+1} = \nabla f(\boldsymbol{x}_{k+1}) = \boldsymbol{0}$, czyli **jesteśmy w optimum**.

Kluczowy punkt konstrukcji algorytmu

Zachodzi $\boldsymbol{g}_{k+1}^\top \boldsymbol{A}\boldsymbol{v}_j = 0$ dla $j = 1, \ldots, k-1$

Dowód: Pokazaliśmy, że gradient $\boldsymbol{g}_{k+1}$ jest **ortogonalny** do wszystkich poprzednich gradientów, tzn. $\boldsymbol{g}_{k+1}^\top \boldsymbol{g}_j = 0$ dla $j = 1, \ldots, k$

Teraz wykorzystamy fakt, że:

$$\boldsymbol{g}_{k+1} = \boldsymbol{A}\boldsymbol{x}_{k+1} - \boldsymbol{b} = \boldsymbol{A}(\boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k) - \boldsymbol{b} = \underbrace{\boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{b}}_{=\boldsymbol{g}_k} + \boldsymbol{A}\alpha_k \boldsymbol{v}_k,$$

Jeśli $\alpha_k = 0$, to $\boldsymbol{g}_{k+1} = \boldsymbol{g}_k$, a ponieważ $\boldsymbol{g}_{k+1}^\top \boldsymbol{g}_k = 0$, wnioskujemy, że $\boldsymbol{g}_{k+1} = \nabla f(\boldsymbol{x}_{k+1}) = \boldsymbol{0}$, czyli **jesteśmy w optimum**.

Jeśli $\alpha_k \neq 0$, mamy:

$$\boldsymbol{A}\boldsymbol{v}_k = \frac{1}{\alpha_k}\left(\boldsymbol{g}_{k+1} - \boldsymbol{g}_k\right)$$

Kluczowy punkt konstrukcji algorytmu

Zachodzi $\boldsymbol{g}_{k+1}^\top \boldsymbol{A}\boldsymbol{v}_j = 0$ dla $j = 1, \ldots, k-1$

Dowód: Pokazaliśmy, że gradient $\boldsymbol{g}_{k+1}$ jest **ortogonalny** do wszystkich poprzednich gradientów, tzn. $\boldsymbol{g}_{k+1}^\top \boldsymbol{g}_j = 0$ dla $j = 1, \ldots, k$

Teraz wykorzystamy fakt, że:

$$\boldsymbol{g}_{k+1} = \boldsymbol{A}\boldsymbol{x}_{k+1} - \boldsymbol{b} = \boldsymbol{A}(\boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k) - \boldsymbol{b} = \underbrace{\boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{b}}_{=\boldsymbol{g}_k} + \boldsymbol{A}\alpha_k \boldsymbol{v}_k,$$

Jeśli $\alpha_k = 0$, to $\boldsymbol{g}_{k+1} = \boldsymbol{g}_k$, a ponieważ $\boldsymbol{g}_{k+1}^\top \boldsymbol{g}_k = 0$, wnioskujemy, że $\boldsymbol{g}_{k+1} = \nabla f(\boldsymbol{x}_{k+1}) = \boldsymbol{0}$, czyli **jesteśmy w optimum**.

Jeśli $\alpha_k \neq 0$, mamy:

$$\boldsymbol{A}\boldsymbol{v}_k = \frac{1}{\alpha_k} (\boldsymbol{g}_{k+1} - \boldsymbol{g}_k)$$

Stąd otrzymujemy dla $j \leqslant k-1$:

$$\boldsymbol{g}_{k+1}^\top \boldsymbol{A}\boldsymbol{v}_j \;=\; \frac{1}{\alpha_k} \boldsymbol{g}_{k+1}^\top (\boldsymbol{g}_{j+1} - \boldsymbol{g}_j) \;=\; \frac{1}{\alpha_k} \Big(\underbrace{\boldsymbol{g}_{k+1}^\top \boldsymbol{g}_{j+1}}_{=0} - \underbrace{\boldsymbol{g}_{k+1}^\top \boldsymbol{g}_j}_{=0} \Big) \;=\; 0$$

Dobór kierunków sprzężonych

Wniosek: W algorytmie gradientów sprzężonych:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k, \qquad \alpha_k = -\frac{\boldsymbol{v}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}$$

kierunki sprzężone $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ można otrzymać jako:

$$\begin{aligned} \boldsymbol{v}_1 &= -\boldsymbol{g}_1 \\ \boldsymbol{v}_{k+1} &= -\boldsymbol{g}_{k+1} + \beta_k \boldsymbol{v}_k, \qquad \beta_k = \frac{\boldsymbol{g}_{k+1}^\top \boldsymbol{A} \boldsymbol{v}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}, \end{aligned}$$

Dobór kierunków sprzężonych

Wniosek: W algorytmie gradientów sprzężonych:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k, \qquad \alpha_k = -\frac{\boldsymbol{v}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}$$

kierunki sprzężone $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ można otrzymać jako:

$$\begin{aligned} \boldsymbol{v}_1 &= -\boldsymbol{g}_1 \\ \boldsymbol{v}_{k+1} &= -\boldsymbol{g}_{k+1} + \beta_k \boldsymbol{v}_k, \qquad \beta_k = \frac{\boldsymbol{g}_{k+1}^\top \boldsymbol{A} \boldsymbol{v}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}, \end{aligned}$$

Uwaga: Zwykle stosuje się inną postać współczynników α_k:

$$\alpha_k = \frac{-\boldsymbol{v}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}$$

Dobór kierunków sprzężonych

Wniosek: W algorytmie gradientów sprzężonych:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k, \qquad \alpha_k = -\frac{\boldsymbol{v}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}$$

kierunki sprzężone $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ m

Zgodnie z $\boldsymbol{v}_k = -\boldsymbol{g}_k + \beta_{k-1} \boldsymbol{v}_{k-1}$

$$\begin{aligned} \boldsymbol{v}_1 &= -\boldsymbol{g}_1 \\ \boldsymbol{v}_{k+1} &= -\boldsymbol{g}_{k+1} + \beta_k \boldsymbol{v}_k, \qquad \beta_k = \frac{\boldsymbol{g}_{k+1}^\top \boldsymbol{A} \boldsymbol{v}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}, \end{aligned}$$

Uwaga: Zwykle stosuje się inną postać współczynników α_k:

$$\alpha_k = \frac{(\boldsymbol{g}_k - \beta_{k-1} \boldsymbol{v}_{k-1})^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}$$

Dobór kierunków sprzężonych

Wniosek: W algorytmie gradientów sprzężonych:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k, \qquad \alpha_k = -\frac{\boldsymbol{v}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}$$

kierunki sp

gradient ortogonalny do poprzednich kierunków:

$\boldsymbol{v}_j^\top \boldsymbol{g}_k = 0$ dla $j < k$

$$\boldsymbol{v}_{k+1} = -\boldsymbol{g}_{k+1} + \beta_k \boldsymbol{v}_k, \qquad \beta_k = \frac{\boldsymbol{g}_{k+1}^\top \boldsymbol{A} \boldsymbol{v}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k},$$

Uwaga: Zwykle stosuje się inną postać współczynników α_k:

$$\alpha_k = \frac{\boldsymbol{g}_k^\top \boldsymbol{g}_k - \beta_{k-1} \boldsymbol{v}_{k-1}^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}$$

Dobór kierunków sprzężonych

Wniosek: W algorytmie gradientów sprzężonych:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k, \qquad \alpha_k = -\frac{\boldsymbol{v}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}$$

kierunki sprzężone $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ można otrzymać jako:

$$\begin{aligned} \boldsymbol{v}_1 &= -\boldsymbol{g}_1 \\ \boldsymbol{v}_{k+1} &= -\boldsymbol{g}_{k+1} + \beta_k \boldsymbol{v}_k, \qquad \beta_k = \frac{\boldsymbol{g}_{k+1}^\top \boldsymbol{A} \boldsymbol{v}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}, \end{aligned}$$

Uwaga: Zwykle stosuje się inną postać współczynników α_k:

$$\alpha_k = \frac{\boldsymbol{g}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}$$

Dobór kierunków sprzężonych

Wniosek: W algorytmie gradientów sprzężonych:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k, \qquad \alpha_k = -\frac{\boldsymbol{v}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}$$

kierunki sprzężone $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ można otrzymać jako:

$$\begin{aligned} \boldsymbol{v}_1 &= -\boldsymbol{g}_1 \\ \boldsymbol{v}_{k+1} &= -\boldsymbol{g}_{k+1} + \beta_k \boldsymbol{v}_k, \qquad \beta_k = \frac{\boldsymbol{g}_{k+1}^\top \boldsymbol{A} \boldsymbol{v}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}, \end{aligned}$$

Uwaga: Zwykle stosuje się inną postać współczynników α_k:

$$\alpha_k = \frac{\boldsymbol{g}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}$$

Podobnie, dla współczynnika β_k:

$$\beta_k = \frac{\boldsymbol{g}_{k+1}^\top \boldsymbol{A} \boldsymbol{v}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}$$

Dobór kierunków sprzężonych

Wniosek: W algorytmie gradientów sprzężonych:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k, \qquad \alpha_k = -\frac{\boldsymbol{v}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A}\boldsymbol{v}_k}$$

kierunki sprzężone $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ można otrzymać jako:

$$\begin{aligned}
\boldsymbol{v}_1 &= -\boldsymbol{g}_1 \\
\boldsymbol{v}_{k+1} &= -\boldsymbol{g}_{k+1} + \beta_k \boldsymbol{v}_k, \qquad \beta_k = \frac{\boldsymbol{g}_{k+1}^\top \boldsymbol{A}\boldsymbol{v}_k}{\boldsymbol{v}_k^\top \boldsymbol{A}\boldsymbol{v}_k},
\end{aligned}$$

Uwaga: Zwykle stosuje się inną postać współczynników α_k:

$$\alpha_k = \frac{\boldsymbol{g}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A}\boldsymbol{v}_k}$$

Podobnie, dla współczynnika β_k:

$$\beta_k = \frac{\boldsymbol{g}_{k+1}^\top (\alpha_k \boldsymbol{A}\boldsymbol{v}_k)}{\boldsymbol{v}_k^\top (\alpha_k \boldsymbol{A}\boldsymbol{v}_k)}$$

Dobór kierunków sprzężonych

Wniosek: W algorytmie gradientów sprzężonych:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k, \qquad \alpha_k = -\frac{\boldsymbol{v}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}$$

kierunki sprzężone $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ można otrzymać jako:

W dowodzie na poprzednim slajdzie pokazaliśmy, że:

$$\alpha_k \boldsymbol{A} \boldsymbol{v}_k = \boldsymbol{g}_{k+1} - \boldsymbol{g}_k$$

$$\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k$$

Uwaga: Zwykle stosuje się inną postać współczynników α_k:

$$\alpha_k = \frac{\boldsymbol{g}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}$$

Podobnie, dla współczynnika β_k:

$$\beta_k = \frac{\boldsymbol{g}_{k+1}^\top (\alpha_k \boldsymbol{A} \boldsymbol{v}_k)}{\boldsymbol{v}_k^\top (\alpha_k \boldsymbol{A} \boldsymbol{v}_k)}$$

Dobór kierunków sprzężonych

Wniosek: W algorytmie gradientów sprzężonych:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k, \qquad \alpha_k = -\frac{\boldsymbol{v}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}$$

kierunki sprzężone $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ można otrzymać jako:

$$\begin{aligned} \boldsymbol{v}_1 &= -\boldsymbol{g}_1 \\ \boldsymbol{v}_{k+1} &= -\boldsymbol{g}_{k+1} + \beta_k \boldsymbol{v}_k, \qquad \beta_k = \frac{\boldsymbol{g}_{k+1}^\top \boldsymbol{A} \boldsymbol{v}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}, \end{aligned}$$

Uwaga: Zwykle stosuje się inną postać współczynników α_k:

$$\alpha_k = \frac{\boldsymbol{g}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}$$

Podobnie, dla współczynnika β_k:

$$\beta_k = \frac{\boldsymbol{g}_{k+1}^\top (\boldsymbol{g}_{k+1} - \boldsymbol{g}_k)}{\boldsymbol{v}_k^\top (\boldsymbol{g}_{k+1} - \boldsymbol{g}_k)}$$

Dobór kierunków sprzężonych

Wniosek: W algorytmie gradientów sprzężonych:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k, \qquad \alpha_k = -\frac{\boldsymbol{v}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A}\boldsymbol{v}_k}$$

kierunki sprzężone $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ można otrzymać jako:

gradient ortogonalny do poprzednich gradientów:

$$\boldsymbol{g}_{k+1}^\top \boldsymbol{g}_j = 0 \text{ dla } j \leqslant k$$

$\boldsymbol{v}_k^\top \boldsymbol{A}\boldsymbol{v}_k$

Uwaga: Zwykle stosuje się inną postać współczynników α_k:

$$\alpha_k = \frac{\boldsymbol{g}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A}\boldsymbol{v}_k}$$

Podobnie, dla współczynnika β_k:

$$\beta_k = \frac{\boldsymbol{g}_{k+1}^\top \boldsymbol{g}_{k+1} - \boldsymbol{g}_{k+1}^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{g}_{k+1} - \boldsymbol{v}_k^\top \boldsymbol{g}_k}$$

Dobór kierunków sprzężonych

Wniosek: W algorytmie gradientów sprzężonych:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k, \qquad \alpha_k = -\frac{\boldsymbol{v}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}$$

kierunki sprzężone $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ można otrzymać jako:

$$\boldsymbol{v}_1 = -\boldsymbol{g}_1$$

$$\boldsymbol{g}_{k+1}^\top \boldsymbol{A} \boldsymbol{v}_k$$

gradient ortogonalny do poprzednich kierunków:

$$\boldsymbol{v}_j^\top \boldsymbol{g}_{k+1} = 0 \text{ dla } j \leqslant k$$

Uwaga: Zwy

$$\alpha_k = \frac{\boldsymbol{g}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}$$

Podobnie, dla współczynnika β_k:

$$\beta_k = \frac{\boldsymbol{g}_{k+1}^\top \boldsymbol{g}_{k+1}}{\boldsymbol{v}_k^\top \boldsymbol{g}_{k+1} - \boldsymbol{v}_k^\top \boldsymbol{g}_k}$$

Dobór kierunków sprzężonych

Wniosek: W algorytmie gradientów sprzężonych:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k, \qquad \alpha_k = -\frac{\boldsymbol{v}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}$$

kierunki sprzężone $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ można otrzymać jako:

$$\boldsymbol{v}_1 = -\boldsymbol{g}_1$$

$$\boldsymbol{v}_{k+1} = -\boldsymbol{g}_{k+1} + \ldots \quad \boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k$$

Pokazaliśmy przy przekształceniu α_k, że

$$-\boldsymbol{v}_k^\top \boldsymbol{g}_k = \boldsymbol{g}_k^\top \boldsymbol{g}_k$$

Uwaga: Zwykle stosuje się inną postać współczynników α_k:

$$\alpha_k = \frac{\boldsymbol{g}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}$$

Podobnie, dla współczynnika β_k:

$$\beta_k = \frac{\boldsymbol{g}_{k+1}^\top \boldsymbol{g}_{k+1}}{-\boldsymbol{v}_k^\top \boldsymbol{g}_k}$$

Dobór kierunków sprzężonych

Wniosek: W algorytmie gradientów sprzężonych:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k, \qquad \alpha_k = -\frac{\boldsymbol{v}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}$$

kierunki sprzężone $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ można otrzymać jako:

$$\begin{aligned} \boldsymbol{v}_1 &= -\boldsymbol{g}_1 \\ \boldsymbol{v}_{k+1} &= -\boldsymbol{g}_{k+1} + \beta_k \boldsymbol{v}_k, \qquad \beta_k = \frac{\boldsymbol{g}_{k+1}^\top \boldsymbol{A} \boldsymbol{v}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}, \end{aligned}$$

Uwaga: Zwykle stosuje się inną postać współczynników α_k:

$$\alpha_k = \frac{\boldsymbol{g}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}$$

Podobnie, dla współczynnika β_k:

$$\beta_k = \frac{\boldsymbol{g}_{k+1}^\top \boldsymbol{g}_{k+1}}{\boldsymbol{g}_k^\top \boldsymbol{g}_k}$$

Dobór kierunków sprzężonych

Wniosek: W algorytmie gradientów sprzężonych:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k, \qquad \alpha_k = -\frac{\boldsymbol{v}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}$$

kierunki sprzężone $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ można otrzymać jako:

$$\begin{aligned} \boldsymbol{v}_1 &= -\boldsymbol{g}_1 \\ \boldsymbol{v}_{k+1} &= -\boldsymbol{g}_{k+1} + \beta_k \boldsymbol{v}_k, \qquad \beta_k = \frac{\boldsymbol{g}_{k+1}^\top \boldsymbol{A} \boldsymbol{v}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}, \end{aligned}$$

Uwaga: Zwykle stosuje się inną postać współczynników α_k:

$$\alpha_k = \frac{\boldsymbol{g}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}$$

Podobnie, dla współczynnika β_k:

$$\beta_k = \frac{\|\boldsymbol{g}_{k+1}\|^2}{\|\boldsymbol{g}_k\|^2}$$

Kierunki sprzężone są kierunkami poprawy

Używając definicji kierunku sprzężonego:

$$\boldsymbol{v}_{k+1} = -\boldsymbol{g}_{k+1} + \beta_k \boldsymbol{v}_k$$

mamy:

$$\boldsymbol{g}_{k+1}^\top \boldsymbol{v}_{k+1} = -\boldsymbol{g}_{k+1}^\top \boldsymbol{g}_{k+1} + \beta_k \boldsymbol{g}_{k+1}^\top \boldsymbol{v}_k$$

Kierunki sprzężone są kierunkami poprawy

Używając definicji kierunku sprzężonego:

$$\boldsymbol{v}_{k+1} = -\boldsymbol{g}_{k+1} + \beta_k \boldsymbol{v}_k$$

mamy:

$$\boldsymbol{g}_{k+1}^\top \boldsymbol{v}_{k+1} = -\boldsymbol{g}_{k+1}^\top \boldsymbol{g}_{k+1} + \beta_k \boldsymbol{g}_{k+1}^\top \boldsymbol{v}_k$$

gradient ortogonalny do poprzednich kierunków:

$$\boldsymbol{g}_{k+1}^\top \boldsymbol{v}_j = 0 \text{ dla } j \leqslant k$$

Kierunki sprzężone są kierunkami poprawy

Używając definicji kierunku sprzężonego:

$$\boldsymbol{v}_{k+1} = -\boldsymbol{g}_{k+1} + \beta_k \boldsymbol{v}_k$$

mamy:

$$\boldsymbol{g}_{k+1}^\top \boldsymbol{v}_{k+1} = -\boldsymbol{g}_{k+1}^\top \boldsymbol{g}_{k+1}$$

Kierunki sprzężone są kierunkami poprawy

Używając definicji kierunku sprzężonego:

$$\boldsymbol{v}_{k+1} = -\boldsymbol{g}_{k+1} + \beta_k \boldsymbol{v}_k$$

mamy:

$$\boldsymbol{g}_{k+1}^\top \boldsymbol{v}_{k+1} = -\|\boldsymbol{g}_{k+1}\|^2 < 0,$$

więc v_{k+1} jest **kierunkiem poprawy**

Metoda gradientów sprzeżonych dla funkcji kwadratowej

Wybierz punkt startowy $\boldsymbol{x}_1$

Wyznacz gradient $\boldsymbol{g}_1 = \nabla f(\boldsymbol{x}_1)$ oraz początkowy kierunek $\boldsymbol{v}_1 = -\boldsymbol{g}_1$

Powtarzaj dla $k = 1, 2, \ldots$:

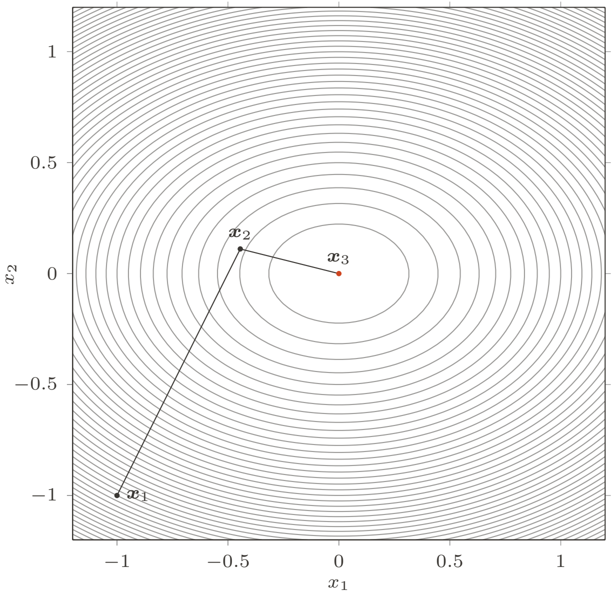
1. Wyznacz optymalną długość kroku $\alpha_k = \frac{\boldsymbol{g}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}$
2. Uaktualnij rozwiązanie: $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k$
3. Wyznacz gradient $\boldsymbol{g}_{k+1} = \nabla f(\boldsymbol{x}_{k+1})$
4. Jeśli $\boldsymbol{g}_{k+1} = \boldsymbol{0}$, zatrzymaj algorytm, w przeciwnym przypadku wyznacz nowy kierunek poprawy:

$$\boldsymbol{v}_{k+1} \;=\; -\boldsymbol{g}_{k+1} + \beta_k \boldsymbol{v}_k, \qquad \beta_k = \frac{\|\boldsymbol{g}_{k+1}\|^2}{\|\boldsymbol{g}_k\|^2},$$

Algorytm zatrzymuje się po co najwyżej n krokach zwracając rozwiązanie optymalne $\boldsymbol{x}^*$

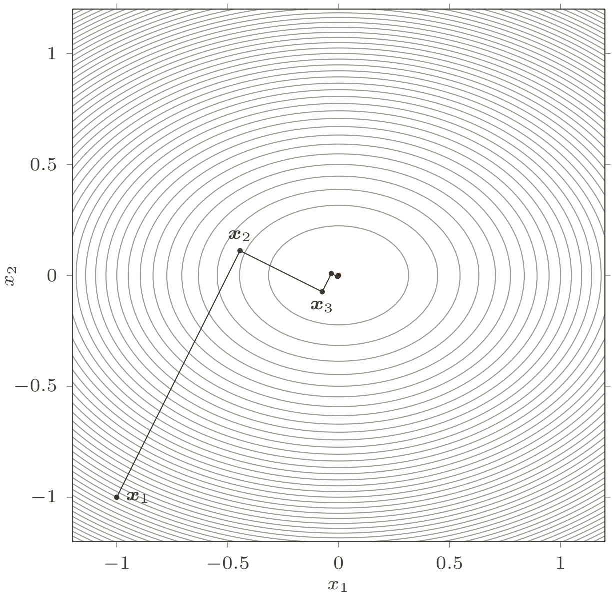
Metoda gradientów sprzężonych – przykład

$$f(x_1, x_2) = x_1^2 + 2x_2^2$$



Dla porównania: algorytm najszybszego spadku

$$f(x_1, x_2) = x_1^2 + 2x_2^2$$



Metoda gradientów sprzeżonych dla dowolnej funkcji

Metodę można łatwo uogólnić na minimalizację dowolnych, niekoniecznie kwadratowych, funkcji, wprowadzającą następujące zmiany:

Metoda gradientów sprzeżonych dla dowolnej funkcji

Metodę można łatwo uogólnić na minimalizację dowolnych, niekoniecznie kwadratowych, funkcji, wprowadzającą następujące zmiany:

- Dla funkcji kwadratowej optymalną długość kroku $\alpha_k = \frac{\boldsymbol{g}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}$ uzyskaliśmy poprzez:

  $$\alpha_k = \operatorname*{argmin}_{\alpha} f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k)$$

  Dla funkcji nie-kwadratowej trzeba rozwiązać powyższy problem jednowymiarowej optymalizacji numerycznie

Metoda gradientów sprzeżonych dla dowolnej funkcji

Metodę można łatwo uogólnić na minimalizację dowolnych, niekoniecznie kwadratowych, funkcji, wprowadzającą następujące zmiany:

- Dla funkcji kwadratowej optymalną długość kroku $\alpha_k = \frac{\boldsymbol{g}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}$ uzyskaliśmy poprzez:

  $$\alpha_k = \operatorname*{argmin}_{\alpha} f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k)$$

  Dla funkcji nie-kwadratowej trzeba rozwiązać powyższy problem jednowymiarowej optymalizacji numerycznie
- Ponieważ nie mamy już gwarancji, że poszczególne gradienty $\boldsymbol{g}_k$ są ortogonalne, można zastąpić:

  $$\beta_k = \frac{\|\boldsymbol{g}_{k+1}\|^2}{\|\boldsymbol{g}_k\|^2} = \frac{\boldsymbol{g}_{k+1}^\top \boldsymbol{g}_{k+1}}{\|\boldsymbol{g}_k\|^2} \quad \Rightarrow \quad \beta_k = \frac{\boldsymbol{g}_{k+1}^\top (\boldsymbol{g}_{k+1} - \boldsymbol{g}_k)}{\|\boldsymbol{g}_k\|^2}$$

  W zależności, którego wzoru używamy, mamy dwie wersje algorytmu: **Fletchera-Reevesa** oraz **Polaka-Ribière'a**.

Metoda gradientów sprzeżonych dla dowolnej funkcji

Metodę można łatwo uogólnić na minimalizację dowolnych, niekoniecznie kwadratowych, funkcji, wprowadzającą następujące zmiany:

- Dla funkcji kwadratowej optymalną długość kroku $\alpha_k = \frac{\boldsymbol{g}_k^\top \boldsymbol{g}_k}{\boldsymbol{v}_k^\top \boldsymbol{A} \boldsymbol{v}_k}$ uzyskaliśmy poprzez:

  $$\alpha_k = \operatorname*{argmin}_{\alpha} f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k)$$

  Dla funkcji nie-kwadratowej trzeba rozwiązać powyższy problem jednowymiarowej optymalizacji numerycznie
- Ponieważ nie mamy już gwarancji, że poszczególne gradienty $\boldsymbol{g}_k$ są ortogonalne, można zastąpić:

  $$\beta_k = \frac{\|\boldsymbol{g}_{k+1}\|^2}{\|\boldsymbol{g}_k\|^2} = \frac{\boldsymbol{g}_{k+1}^\top \boldsymbol{g}_{k+1}}{\|\boldsymbol{g}_k\|^2} \quad \Rightarrow \quad \beta_k = \frac{\boldsymbol{g}_{k+1}^\top (\boldsymbol{g}_{k+1} - \boldsymbol{g}_k)}{\|\boldsymbol{g}_k\|^2}$$

  W zależności, którego wzoru używamy, mamy dwie wersje algorytmu: **Fletchera-Reevesa** oraz **Polaka-Ribière'a**.
- Nie mamy gwarancji, że algorytm skończy się po n iteracjach, więc jako warunek stopu przyjmujemy $\|\nabla f(\boldsymbol{x}_k)\| < \epsilon$ dla ϵ bliskiego zeru.

Metoda gradientów sprzeżonych dla dowolnej funkcji (Fletchera-Reevesa i Polaka-Ribière'a)

Wybierz punkt startowy $\boldsymbol{x}_1$

Wyznacz gradient $\boldsymbol{g}_1 = \nabla f(\boldsymbol{x}_1)$ oraz początkowy kierunek $\boldsymbol{v}_1 = -\boldsymbol{g}_1$

Powtarzaj dla $k = 1, 2, \ldots$:

1. Wyznacz optymalną długość kroku $\alpha_k = \operatorname{argmin}_\alpha f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k)$
2. Uaktualnij rozwiązanie: $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k$
3. Wyznacz gradient $\boldsymbol{g}_{k+1} = \nabla f(\boldsymbol{x}_{k+1})$
4. Jeśli $\|\boldsymbol{g}_{k+1}\| < \epsilon$, zatrzymaj algorytm, w przeciwnym przypadku wyznacz nowy kierunek poprawy:

$$\boldsymbol{v}_{k+1} = -\boldsymbol{g}_{k+1} + \beta_k \boldsymbol{v}_k, \qquad \text{gdzie:}$$

$$\text{(F-R)} \quad \beta_k = \frac{\|\boldsymbol{g}_{k+1}\|^2}{\|\boldsymbol{g}_k\|^2} \qquad \text{lub} \qquad \text{(P-R)} \quad \beta_k = \frac{\boldsymbol{g}_{k+1}^\top(\boldsymbol{g}_{k+1} - \boldsymbol{g}_k)}{\|\boldsymbol{g}_k\|^2}$$

Metoda Polaka-Ribière'a preferowana w praktyce

Kierunki sprzężone są kierunkami poprawy

Ponieważ:

$$\alpha_k \;=\; \operatorname*{argmin}_{\alpha} f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k),$$

Mamy:

$$\frac{\mathrm{d} f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k)}{\mathrm{d}\alpha}\Big|_{\alpha=\alpha_k} \;=\; 0$$

Kierunki sprzężone są kierunkami poprawy

Ponieważ:

$$\alpha_k = \operatorname*{argmin}_{\alpha} f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k),$$

Mamy:

$$\frac{\mathrm{d} f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k)}{\mathrm{d}\alpha}\Big|_{\alpha=\alpha_k} = 0$$

Z reguły łańcuchowej:

$$\frac{\mathrm{d} f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k)}{\mathrm{d}\alpha} = \nabla f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k)^\top \boldsymbol{v}_k,$$

a tym samym:

$$\nabla f(\boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k)^\top \boldsymbol{v}_k = 0$$

Kierunki sprzężone są kierunkami poprawy

Ponieważ:

$$\alpha_k = \operatorname*{argmin}_{\alpha} f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k),$$

Mamy:

$$\frac{\mathrm{d} f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k)}{\mathrm{d}\alpha}\Big|_{\alpha=\alpha_k} = 0$$

Z reguły łańcuchowej:

$$\frac{\mathrm{d} f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k)}{\mathrm{d}\alpha} = \nabla f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k)^\top \boldsymbol{v}_k,$$

a tym samym:

$$\nabla f(\boldsymbol{x}_{k+1})^\top \boldsymbol{v}_k = 0$$

Kierunki sprzężone są kierunkami poprawy

Ponieważ:

$$\alpha_k = \operatorname*{argmin}_{\alpha} f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k),$$

Mamy:

$$\left.\frac{\mathrm{d} f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k)}{\mathrm{d}\alpha}\right|_{\alpha=\alpha_k} = 0$$

Z reguły łańcuchowej:

$$\frac{\mathrm{d} f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k)}{\mathrm{d}\alpha} = \nabla f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k)^\top \boldsymbol{v}_k,$$

a tym samym:

$$\boldsymbol{g}_{k+1}^\top \boldsymbol{v}_k = 0$$

Kierunki sprzężone są kierunkami poprawy

Ponieważ:

$$\alpha_k = \operatorname*{argmin}_{\alpha} f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k),$$

Mamy:

$$\left.\frac{\mathrm{d} f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k)}{\mathrm{d}\alpha}\right|_{\alpha=\alpha_k} = 0$$

Z reguły łańcuchowej:

$$\frac{\mathrm{d} f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k)}{\mathrm{d}\alpha} = \nabla f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k)^\top \boldsymbol{v}_k,$$

a tym samym:

$$\boldsymbol{g}_{k+1}^\top \boldsymbol{v}_k = 0$$

$$\boldsymbol{v}_{k+1} = -\boldsymbol{g}_{k+1} + \beta_k \boldsymbol{v}_k$$

Kierunki sprzężone są kierunkami poprawy

Ponieważ:

$$\alpha_k \ = \ \operatorname*{argmin}_{\alpha} f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k),$$

Mamy:

$$\left.\frac{\mathrm{d} f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k)}{\mathrm{d}\alpha}\right|_{\alpha=\alpha_k} \ = \ 0$$

Z reguły łańcuchowej:

$$\frac{\mathrm{d} f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k)}{\mathrm{d}\alpha} \ = \ \nabla f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k)^\top \boldsymbol{v}_k,$$

a tym samym:

$$\boldsymbol{g}_{k+1}^\top \boldsymbol{v}_k \ = \ 0$$

$$\boldsymbol{g}_{k+1}^\top \boldsymbol{v}_{k+1} = -\boldsymbol{g}_{k+1}^\top \boldsymbol{g}_{k+1} + \beta_k \underbrace{\boldsymbol{g}_{k+1}^\top \boldsymbol{v}_k}_{=0}$$

Kierunki sprzężone są kierunkami poprawy

Ponieważ:

$$\alpha_k = \operatorname*{argmin}_{\alpha} f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k),$$

Mamy:

$$\left.\frac{\mathrm{d}f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k)}{\mathrm{d}\alpha}\right|_{\alpha=\alpha_k} = 0$$

Z reguły łańcuchowej:

$$\frac{\mathrm{d}f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k)}{\mathrm{d}\alpha} = \nabla f(\boldsymbol{x}_k + \alpha \boldsymbol{v}_k)^\top \boldsymbol{v}_k,$$

a tym samym:

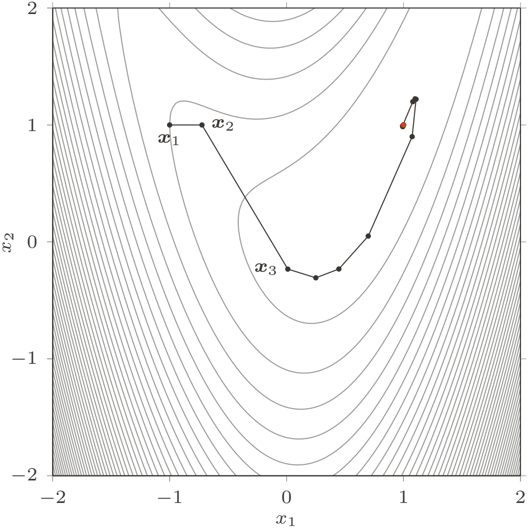
$$\boldsymbol{g}_{k+1}^\top \boldsymbol{v}_k = 0$$

$$\boldsymbol{g}_{k+1}^\top \boldsymbol{v}_{k+1} = -\|\boldsymbol{g}_{k+1}\|^2 < 0,$$

więc $\boldsymbol{v}_{k+1}$ jest **kierunkiem poprawy**

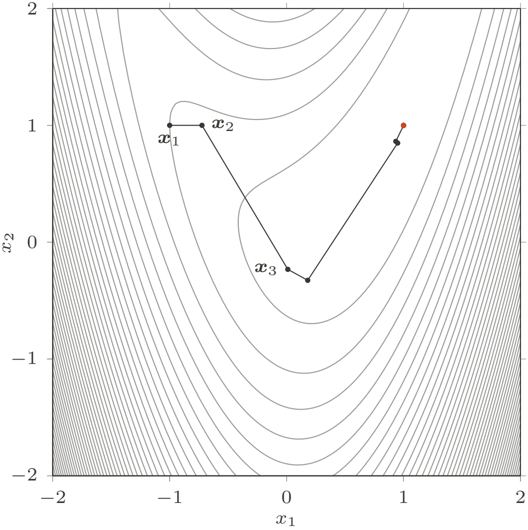
Metoda Fletchera-Reevesa – przykład

$$f(x_1, x_2) = 2.5(x_1^2 - x_2)^2 + (1 - x_1)^2$$



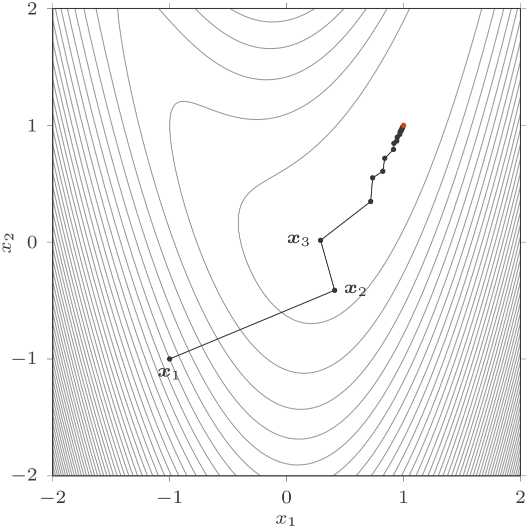
Metoda Polaka-Ribière'a – przykład

$$f(x_1, x_2) = 2.5(x_1^2 - x_2)^2 + (1 - x_1)^2$$



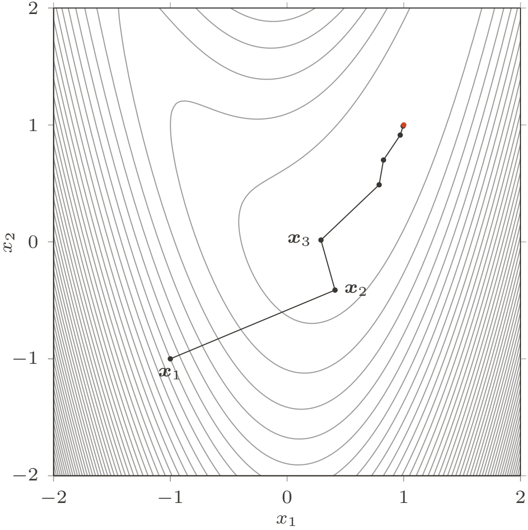
Metoda Fletchera-Reevesa – przykład

$$f(x_1, x_2) = 2.5(x_1^2 - x_2)^2 + (1 - x_1)^2$$



Metoda Polaka-Ribière'a – przykład

$$f(x_1, x_2) = 2.5(x_1^2 - x_2)^2 + (1 - x_1)^2$$



Dla porównania: algorytm najszybszego spadku

$$f(x_1, x_2) = 2.5(x_1^2 - x_2)^2 + (1 - x_1)^2$$