Uniwersytet Technologiczno-Przyrodniczny
im. Jana i Jędrzeja Śniadeckich w Bydgoszczy

Dokumentacja projektowa
Analiza porównawcza metody Newtona i metody Broydena-Flechtera-Goldfarba-Shanno (BFGS)

Autorzy:
Bartosz Gołata
Andrzej Piątkowski

Prowadzący:
Dr inż. Teresa Chyła-Ciołczyk

21 czerwca 2012

Spis treści

1 Metody gradientowe - pojęcia podstawowe

Metody gradientowe są to metody służące do wyznaczania minimum lub maksimum funkcji celu. Metody optymalizacji gradientowej wykorzystują w swoich obliczeniach aktualne wartości funkcji celu, jej gradient lub wielkości z nim związane. Gradientem nazywamy wektor pochodnych cząstkowych funkcji celu w punkcie, który oznaczony jest następująco ∇f.
W przypadku tych metod funkcja celu musi być określona i różniczkowalna w całej przestrzeni R^n.

2 Metoda Newtona

Metoda Newtona jest metodą gradientową, dzięki której można wyznaczyć minimum funkcji celu. Stosowanie tej metody do bardziej ogólnych funkcji wymaga założenia, że w każdym punkcie istnieją wszystkie pochodne cząstkowe funkcji celu rzędu drugiego tzn. hesjanu. Hesjanem nazywamy macierz drugich pochodnych cząstkowych funkcji. Metoda ta w zastosowaniu praktycznym może okazać się bardzo czasochłonna i skomplikowana.
W naszej implementacji została wykorzystana metoda stałokrokowa, w której $h = 1$.

2.1 Warunki stopu dla metody Newtona

1. Warunek stopu określony jako: $||x^{(i)} - x^{(i-1)}||_2 < \varepsilon$,
gdzie $\varepsilon > 0$ oraz ε jest bliska 0
2. Gradient $\nabla f(x^{(i)}) = 0$
3. Hesjan $H(x)$ jest macierzą nieodwracalną

2.2 Przedstawienie metody Newtona w przykładzie

Znaleźć minimum funkcji celu daną poniższym wzorem.

1. Dane:

$f(x_1, x_2) = 3x_1^2 + 3x_2^2 - 4x_1(1 + x_2) - x_2 + 2$

$x^{(0)} = (10; 4)$

$\varepsilon = 0,01$

2. Obliczanie gradientu funkcji celu $f(x_1, x_2)$

$$\nabla f(x_1, x_2) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} f'_{x_1} \\ f'_{x_2} \end{bmatrix} = \begin{bmatrix} 6x_1 - 4x_2 - 4 \\ 6x_2 - 4x_1 - 1 \end{bmatrix}$$

3. Obliczanie hesjanu funkcji celu $f(x_1, x_2)$

$$H(x_1, x_2) = \begin{bmatrix} \frac{\partial^2 f}{\partial_{x_1} \partial_{x_1}}(x) & \frac{\partial^2 f}{\partial_{x_1} \partial_{x_2}}(x) \\ \frac{\partial^2 f}{\partial_{x_2} \partial_{x_1}}(x) & \frac{\partial^2 f}{\partial_{x_2} \partial_{x_2}}(x) \end{bmatrix} = \begin{bmatrix} f''_{x_1 x_1} & f''_{x_1 x_2} \\ f''_{x_2 x_1} & f''_{x_2 x_2} \end{bmatrix} = \begin{bmatrix} 6 & -4 \\ -4 & 6 \end{bmatrix}$$

4. Obliczamy gradient funkcji celu w punkcie $x^{(0)}$

$$\nabla f(x^{(0)}) = \begin{bmatrix} 40 \\ -17 \end{bmatrix}$$

5. Obliczamy odwrotność hesjanu funkcji celu

Wzór ogólny: $H^{-1}(x^{(i)}) = \frac{1}{det(H(x^{(i)}))} \cdot D^T$

$d_{11} = (-1)^{1+1} \cdot 6 = 6$

$d_{12} = (-1)^{1+2} \cdot (-4) = 4$

$d_{21} = (-1)^{2+1} \cdot (-4) = 4$

$d_{22} = (-1)^{2+2} \cdot 6 = 6$

$det(H(x^{(0)})) = 6 \cdot 6 - 4 \cdot 4 = 20$

$$D = \begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix} = \begin{bmatrix} 6 & 4 \\ 4 & 6 \end{bmatrix}$$

$$H^{-1}(x^{(0)}) = \frac{1}{20} \cdot \begin{bmatrix} 6 & 4 \\ 4 & 6 \end{bmatrix}^T = \begin{bmatrix} 0,3 & 0,2 \\ 0,2 & 0,3 \end{bmatrix}$$

6. Obliczamy wektor kierunkowy

Wzór ogólny: $d^{(i)} = -H^{-1}(x^{(i)}) \nabla f(x^{(i)})$

$$d^{(0)} = \begin{bmatrix} 0,3 & 0,2 \\ 0,2 & 0,3 \end{bmatrix} \cdot \begin{bmatrix} 40 \\ -17 \end{bmatrix} = \begin{bmatrix} -8,6 \\ -2,9 \end{bmatrix}$$

7. Obliczamy kolejny punkt x

Wzór ogólny: $x^{(i+1)} = x^{(i)} + h^{(i)} \cdot d^{(0)}$, gdzie $h^{(i)} = 1$

$$x^{(1)} = \begin{bmatrix} 10 \\ 4 \end{bmatrix} + \begin{bmatrix} -8,6 \\ -2,9 \end{bmatrix} = \begin{bmatrix} 1,4 \\ 1,1 \end{bmatrix}$$

8. Sprawdzanie warunków

Warunek nr 1: $||x^{(i+1)} - x^{(i)}||_2 < \epsilon$

$||x^{(1)} - x^{(0)}||_2 < 0,01$

$\sqrt{(1,4-10)^2 + (1,1-4)^2} < 0,01$

$9,075 < 0,01$ (Nie jest to prawdą, więc sprawdzany jest kolejny warunek)

Warunek nr 2: $\nabla f(x^{(i)}) = 0$

$\nabla f(x^{(1)}) = 0$

$\nabla f(1,4;1,1) = 0$ (Podstawiamy dane wartości do wyrażenia $\nabla f(x_1, x_2)$)

$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0$ (Warunek został spełniony, algorytm kończy działanie)

Warunek 3 nie został sprawdzony, ponieważ wcześniejszy warunek został spełniony. Z powyższego przykładu wynika, że minimum funkcji celu zostało wyliczone w jednym kroku. I wynosi ono $x = (1,4;1,1)$.

3 Metoda Broydena-Fletchera-Goldfarba-Shanno

Metoda Broydena-Fletchera-Goldfarba-Shanno (BFGS) jest jedną z metod quasi-Newtonowskich (metody zmiennej metryki). Agorytm bazuje na metodzie Newtona. Zaletą w porównaniu do metody Newtona jest to, że nie mamy obowiązku obliczania hesjanu funkcji celu, co ułatwia obliczenia. Zamiast tego obliczana jest macierz $V^{(i)}$. Na początku obliczeń macierz $V^{(i)}$ ma postać macierzy jednostkowej. W kolejnych krokach macierz $V^{(i)}$ obliczana jest według wzoru podanego poniżej.
$V^{(i+1)} = V^{(i)} + (1 + C^{(i)})A^{(i)} + D^{(i)}$

3.1 Warunki stopu dla metody BFGS

W metodzie Broydena-Flechtera-Goldfarba-Shanno wykorzystuje się standardowy warunek stopu:
$||x^{(i)} - x^{(i-1)}||_2 < \varepsilon$, gdzie $\varepsilon > 0$ oraz ε jest bliska 0

3.2 Przedstawienie metody BFGS w przykładzie

Znaleźć minimum funkcji celu daną poniższym wzorem:

1. Dane:

$f_{(x_1,x_2)} = 4x_1^2 + 2x_1x_2 + 4x_2^2 + 3x_1$

$x^{(-1)} = (2; 2)$

$x^{(0)} = (1,2; 1,3)$

$\varepsilon = 0,01$

$h^{(i)} = 1$

$V^{(0)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

2. Obliczamy gradient funkcji celu

Wzór ogólny: $\nabla f(x1,x2) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix}$

$\nabla f(x1,x2) = \begin{bmatrix} 8x1 + 2x2 + 3 \\ 2x1 + 8x2 \end{bmatrix}$

3. Obliczamy $a^{(i)}$

Wzór ogólny: $a^{(i)} = x^{(i)} - x^{(i-1)}$

$a^{(0)} = x^{(0)} - x^{(-1)} = \begin{bmatrix} 1,2 \\ 1,3 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} -0,8 \\ -0,7 \end{bmatrix}$

4. Obliczamy $s^{(i)}$

Wzór ogólny: $s^{(i)} = \nabla f(x^{(i)}) - \nabla f(x^{(i-1)})$

$s^{(0)} = \nabla f(x^{(0)}) - \nabla f(x^{(-1)}) = \begin{bmatrix} 15,2 \\ 12,8 \end{bmatrix} - \begin{bmatrix} 23 \\ 20 \end{bmatrix} = \begin{bmatrix} -7,8 \\ -7,2 \end{bmatrix}$

5. Obliczamy $A^{(i)}$

Wzór ogólny: $A^{(i)} = \frac{a^{(i)}(a^{(i)})^T}{(a^{(i)})^T s^{(i)}}$

$A^{(0)} = \frac{a^{(0)}(a^{(0)})^T}{(a^{(0)})^T s^{(0)}} = \dfrac{\begin{bmatrix} -0,8 \\ -0,7 \end{bmatrix} \cdot \begin{bmatrix} -0,8 & -0,7 \end{bmatrix}}{\begin{bmatrix} -0,8 & -0,7 \end{bmatrix} \cdot \begin{bmatrix} -7,8 \\ -7,2 \end{bmatrix}} = \dfrac{\begin{bmatrix} 0,64 & 0,56 \\ 0,56 & 0,49 \end{bmatrix}}{11{,}28} = \begin{bmatrix} 0,0567 & 0,0496 \\ 0,0496 & 0,0434 \end{bmatrix}$

6. Obliczamy $C^{(i)}$

Wzór ogólny: $C^{(i)} = \frac{(s^{(i)})^T V^{(i)} s^{(i)}}{(a^{(i)})^T s^{(i)}}$

$C^{(0)} = \frac{(s^{(0)})^T V^{(0)} s^{(0)}}{(a^{(0)})^T s^{(0)}} = \dfrac{\begin{bmatrix} -7,8 & -7,2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} -7,8 \\ -7,2 \end{bmatrix}}{\begin{bmatrix} -0,8 & -0,7 \end{bmatrix} \cdot \begin{bmatrix} -7,8 \\ -7,2 \end{bmatrix}} = \frac{112{,}68}{11{,}28} = 9,98936$

7. Obliczamy $D^{(i)}$

Wzór ogólny: $D^{(i)} = -\frac{a^{(i)}(s^{(i)})^T V^{(i)} + V^{(i)} s^{(i)} (a^{(i)})^T}{(a^{(i)})^T s^{(i)}}$

$D^{(0)} = -\frac{a^{(0)}(s^{(0)})^T V^{(0)} + V^{(0)} s^{(0)} (a^{(0)})^T}{(a^{(0)})^T s^{(0)}} =$

$$-\frac{\begin{bmatrix} -0,8 \\ -0,7 \end{bmatrix} \cdot \begin{bmatrix} -7,8 & -7,2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} -7,8 \\ -7,2 \end{bmatrix} \cdot \begin{bmatrix} -0,8 & -0,7 \end{bmatrix}}{\begin{bmatrix} -0,8 & -0,7 \end{bmatrix} \cdot \begin{bmatrix} -7,8 \\ -7,2 \end{bmatrix}} =$$

$$\frac{\begin{bmatrix} 6,24 & 5,76 \\ 5,46 & 5,04 \end{bmatrix} + \begin{bmatrix} 6,24 & 5,46 \\ 5,76 & 5,04 \end{bmatrix}}{11{,}28} = \begin{bmatrix} -1,1064 & -0,9947 \\ -0,9947 & -0,8936 \end{bmatrix}$$

8. Obliczamy $V^{(i)}$

Wzór ogólny: $V^{(i+1)} = V^{(i)} + (1 + C^{(i)}) \cdot A^{(i)} + D^{(i)}$

$V^{(1)} = V^{(0)} + (1 + C^{(0)}) \cdot A^{(0)} + D^{(0)}$

$$V^{(1)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + (1 + 9,98936) \cdot \begin{bmatrix} 0,0567 & 0,0496 \\ 0,0496 & 0,0434 \end{bmatrix} + \begin{bmatrix} -1,1064 & -0,9947 \\ -0,9947 & -0,8936 \end{bmatrix} =$$

$$\begin{bmatrix} 0,5167 & -0,4496 \\ -0,4496 & 0,5833 \end{bmatrix}$$

9. Obliczamy $d^{(i)}$

Wzór ogólny: $d^{(i)} = -V^{(i+1)} \cdot \nabla f(x^{(i)})$

$d^{(0)} = -V^{(1)} \cdot \nabla f(x^{(0)})$

$$d^{(0)} = -\begin{bmatrix} 0,5167 & -0,4496 \\ -0,4496 & 0,5833 \end{bmatrix} \cdot \begin{bmatrix} 15,2 \\ 12,8 \end{bmatrix} = \begin{bmatrix} -2,0991 \\ -0,6325 \end{bmatrix}$$

10. Obliczamy $x^{(i)}$

Wzór ogólny: $x^{(i+1)} = x^{(i)} + h^{(i)} \cdot d^{(i)}$

$x^{(1)} = x^{(0)} + h^{(i)} \cdot d^{(0)}$

$$x^{(1)} = \begin{bmatrix} 1,2 \\ 1,3 \end{bmatrix} + 1 \cdot \begin{bmatrix} -2,0991 \\ -0,6325 \end{bmatrix} = \begin{bmatrix} 1,2 \\ 1,3 \end{bmatrix} + \begin{bmatrix} -2,0991 \\ -0,6325 \end{bmatrix} = \begin{bmatrix} -0,8991 \\ 0,6675 \end{bmatrix}$$

11. Sprawdzanie warunków

Warunek nr 1: $||x^{(i+1)} - x^{(i)}||_2 < \epsilon$

$||x^{(1)} - x^{(0)}||_2 < 0,01$

$\sqrt{(-0,8991 - 1,2)^2 + (0,6675 - 1,3)^2} < 0,01$

$2,1923 < 0,01$

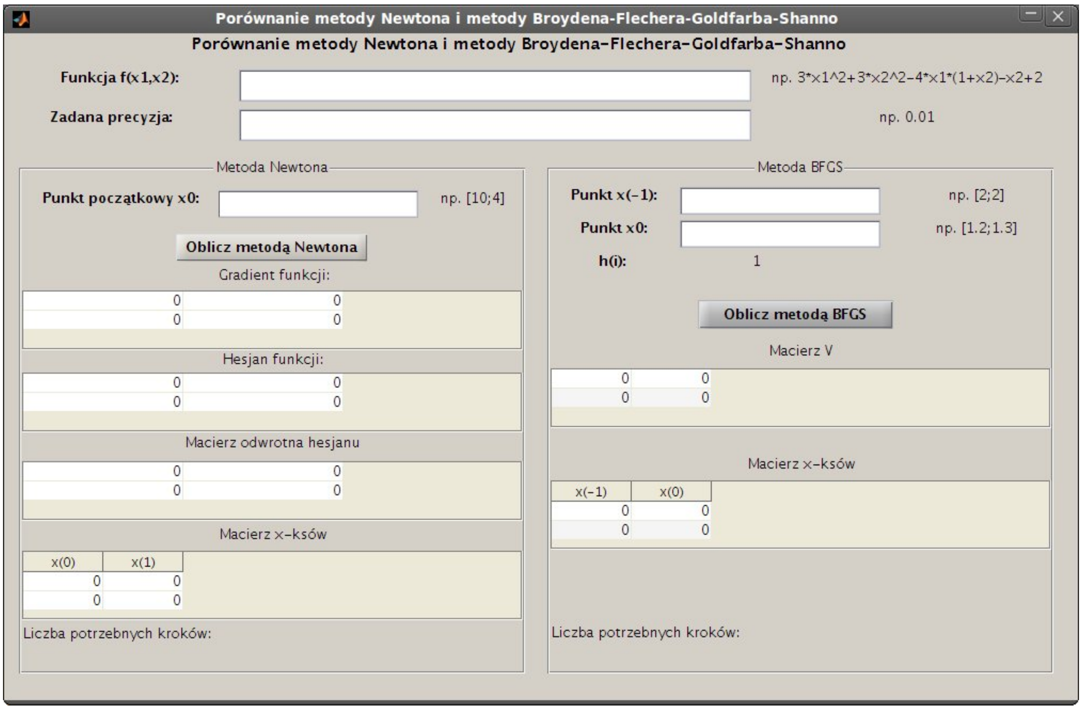
Widzimy, że warunek nie został spełniony dlatego wykonujemy cały algorytm zaczynając od punkt numer 3.
W tym przypadku potrzeba było aż 5 kroków aby znaleźć minimum funkcji celu.
Wynikiem jest punkt $x^{(4)} = \begin{bmatrix} -0,4001 \\ -0,1000 \end{bmatrix}$

4 Program komputerowy

4.1 Opis programu

Aplikacja została stworzona w środowisku matematycznym Matlab. Interfejs programu jest bardzo prosty i przejrzysty. W jego górnej części mamy miejsce na wpisanie funkcji celu oraz określenie precyzji warunku stopu. Reszta programu została podzielona na dwie części. Z lewej strony Metoda Newtona, z prawej strony Metoda Broydena-Fletchera-Goldfarba-Shanno. Określamy punkty startowe dla wybranej przez nas metody i klikamy przycisk Oblicz. Na samym dole otrzymujemy informacje w ilu korkach dana metoda znalazła minimum funkcji celu. Przy każdym polu edycji mamy przykładowe dane jakie możemy wprowadzić. Wprowadzane dane powinny zachować wskazny format.

4.2 Interfejs aplikacji



Rysunek 1: Interfejs aplikacji

4.3 Testowanie aplikacji

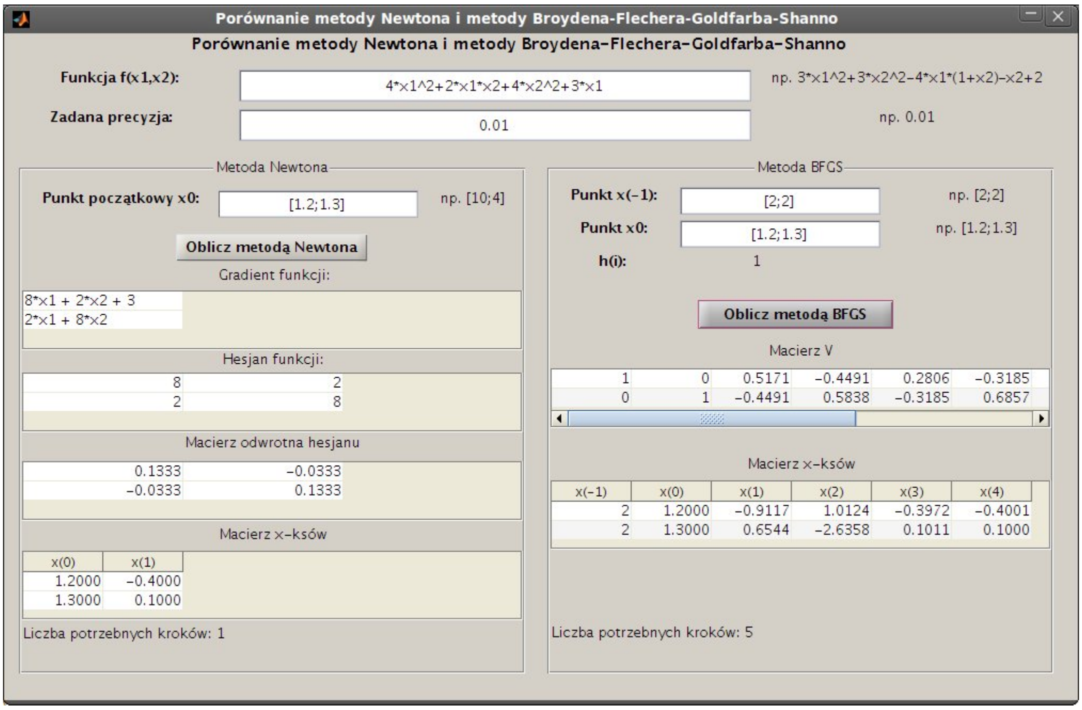
Ażeby określić czy algorytm zaimplementowany w programie został poprawnie napisany program został przetestowany. Do testów zostały użyte poniżej dane wejściowe:

Funkcja celu: $f_{(x)} = 4x1^2 + 2x1x2 + 4x2^2 + 3x1 \leftarrow minimum$

Precyzja warunku stopu: $e = 0.01$

Punkt $x^{(0)} = [1.2; 1.3]$

Punkt $x^{(-1)} = [2; 2]$



Rysunek 2: Praca aplikacji

Na rysunku powyżej widzimy, że minimum zostało znalezione przez obie zaimplementowane metody. Metoda Newtona wykonała to w 1 kroku, natomiast metoda Broydena-Fletchera-Goldfarba-Shanno potrzebowała aż 5 kroków.
Wyniki, które otrzymaliśmy z obydwu metod są identyczne.

Metoda Newtona:

$$x^{(1)} = \begin{bmatrix} -0,4000 \\ -0,1000 \end{bmatrix}$$

Metoda Broydena-Fletchera-Goldfarba-Shanno:

$$x^{(1)} = \begin{bmatrix} -0,4001 \\ -0,1000 \end{bmatrix}$$

5 Wnioski końcowe

Metoda Newtona jest metodą dość skomplikowaną, ponieważ jesteśmy zmuszeni do wyliczenia hesjanu funkcji celu czyli pochodnych cząstkowych drugiego rzędu. W metodzie Broydena-Fletchera-Goldfarba-Shanno nie musimy obliczać hesjanu, natomiast mamy do obliczenia gradient lecz ten występuje w obu metodach. W drugiej prezentowanej metodzie zrezygnowano z hesjanu na rzecz macierzy V, która jest przybliżonym hesjanem lecz te przybliżenia nie wpływają znacząco na otrzymywane wyniki.

W czasie testowania aplikacji zauważono, że dla niektórych postaci funkcji metoda Newtona sprawdza się o wiele lepiej niż metoda BFGS. Pierwsza metoda potrzebuje tylko i wyłącznie jednego korku do znalezienia minimum, gdzie druga potrzebuje ok. 2-5 kroków więcej. Zastosowanie metody BFGS to takiego rodzaju funkcji jest stosunkowo niekorzyste. Jednakże metoda Newtona ustępuje drugiej metodzie jeżeli będziemy rozpatrywali dośc zaawansowane funkcje celu, ponieważ w metodzie Newtona w tym momencie hesjan będzie bardzo trudny do obliczenia i zajmie to bardzo dużo czasu.

Tak więc jeżeli zależy nam na łatwość obliczeń lepsza jest metoda Broydena-Fletchera-Goldfarba-Shanno. Jeżeli natomiast szukamy minimum w grupie funkcji gdzie metoda Newtona znajduje rozwiązanie w jednym kroku, to na pewno jest to wybór bardziej optymalny.