AKADEMIA GÓRNICZO-HUTNICZA
IM. STANISŁAWA STASZICA
W KRAKOWIE

Sprawozdanie - laboratorium nr 10

Poszukiwanie minimum wartości funkcji w dwóch wymiarach

metodą Newtona

Klaudia Fil, 12.05.2019 r.

1. Wstęp teoretyczny

Zadaniem optymalizacji jest poszukiwanie minimum lub maksimum funkcji, w praktyce oznacza to sprowadzenie do poszukiwania tylko tego pierwszego, czyli takiego punktu w $f: R^n \rightarrow R$, dla którego zachodzi:

$$min f(x) = f(x') \Leftrightarrow \forall f(x') < f(x) \quad , \tag{1}$$

$$x = [x_1, x_2, \dots, x_n]^T \tag{2}$$

z warunkami:

$$g_j(x) \leq 0, j = 1,2, \dots, m \quad , \tag{3}$$

$$h_j(x) = 0, j = 1,2, \dots, r \quad , \tag{4}$$

gdzie funkcje $f(x), g(x), h(x)$ są funkcjami skalarnymi, pierwsza to funkcja celu, a dwie następne określają warunki jakie musi spełnić rozwiązanie (więzy).

Gradient funkcji celu $f(x) \in C^2$ definiujemy:

$$g(x) = \nabla f(x) \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix} \quad . \tag{5}$$

Macierzą Hessego, czyli Hesjanem dla funkcji celu $f(x) \in C^2$ nazywamy:

$$H(x) = \{ \frac{\partial^2 f(x)}{\partial x_i x_j} \} = \nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x 1 x_2} & \dots & \dots \\ \frac{\partial^2 f(x)}{\partial x_2 x_1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n x_1} & \dots & \dots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix} \quad . \tag{6}$$

Minimum globalne funkcji stanowi taki x' , dla którego:

$$\forall f(x) \geq f(x') \quad , \text{dla} \quad x \in R^n \tag{7}$$

Minimum lokalne funkcji stanowi taki x', dla którego:

$$\exists \epsilon : \epsilon > 0, \epsilon \in R \rightarrow \forall f(x) > f(x') \quad \text{, dla} \quad x : \|x - x'\| \le \epsilon \tag{7}$$

Punktem siodłowym $x' = \begin{pmatrix} x^0 \\ y_0 \end{pmatrix}$ funkcji nazywamy taki punkt, który:

$$\exists \epsilon : \epsilon > 0, \epsilon \in R \rightarrow \forall f(x, y^0) \le f(x^0, y^0) \le f(x^0, y) \quad \text{dla} \quad x : \|x - x^0\| \le \epsilon \quad \text{i} \quad x : \|x - x'\| \le \epsilon \tag{8}$$

Jednym z rodzajów metod poszukiwania minimum funkcji, są metody kierunkowe, dla których definiujemy różniczkę zupełną jako iloczyn skalarny wektorów:

$$df = \frac{\partial f}{\partial x_1} dx_1 + \ldots + \frac{\partial f}{\partial x_n} dx_n = \nabla f(x) dx \quad . \tag{9}$$

Kiedy wektor *u* wyznacza kierunek prostej łączącej punkty *x* i *x'* to są one powiązane:

$$x(\lambda) = x' + \lambda u \quad , \tag{10}$$

natomiast dla bardzo małych wartości $x(\lambda) = x' + \lambda u$ możemy zapisać:

$$dx = u\, d\lambda \quad . \tag{11}$$

Na prostej łączącej ustalone punkty *x* i *x'* wartość funkcji celu zależna będzie od nowej zmiennej λ :

$$F(\lambda) = f(x' + \lambda u) = f(x) \quad , \tag{12}$$

licząc różniczkę zupełną dla funkcji celu zależnej od λ :

$$dF = df = \nabla^T f(x) u\, d\lambda \quad . \tag{13}$$

Możliwe jest wyznaczenie pochodnej kierunkowej funkcji celu w *x* dla kierunku *u:*

$$\frac{d F(\lambda)}{d\lambda} = \frac{d f(x)}{d\lambda} \Big|_u = \nabla^T f(x) u \quad . \tag{14}$$

Metodą Newtona poszukiwania minimum funkcji kwadratowej w R^n wykorzystując następująco zdefiniowaną funkcję:

$$f(x) = \frac{1}{2} x^T A x + x^T b + c \quad \text{,gdzie} \tag{15}$$

A jest pewną macierzą kwadratową oraz $a, b \in R^n$, $c \in R$.

Gdy *A* jest macierzą symetryczną to zachodzi:

$$\nabla f(x) = Ax + b \quad \text{, oraz} \quad \nabla^2 f(x) = A \Rightarrow H(x) = A \quad . \tag{16}$$

Jeśli *A* jest dodatniookreślona to rozwiązanie można łatwo znaleźć, ponieważ

$$\nabla f(x) = Ax + b = 0 \quad \text{, a więc} \tag{17}$$

$x' = -A^{-1}b$ - macierz dodatniookreślona, nieosobliwa dlatego można ją odwrócić.

W tej metodzie zakładamy:

$$x' = x^i + \delta \quad \text{, gdzie } x^i \text{ - przybliżone rozwiązanie w } i\text{-tej iteracji.} \tag{18}$$

Wykorzystując szereg Taylora możemy zapisać:

$$0 = \nabla f(x') = \nabla f(x^i + \delta) = \nabla f(x^i) + H(x^i)\delta + O(\|\delta\|^2) \quad , \tag{19}$$

wtedy w *i*-tej iteracji poprawiamy rozwiązanie $x^{i+1} = x^i + \delta$ i ostatecznie:

$$x^{i+1} = x^i - H^{-1}(x^i)\nabla f(x^i) \quad . \tag{20}$$

2. Zadanie do wykonania

2.1. Opis problemu

Celem ćwiczenia jest znalezienie miejsca zerowego funkcji:

$$f(x, y) = x^2 - 4x + 8 + y^2 - 4y + xy \quad , \tag{21}$$

jednak po zauważeniu że wykres ten jest przesunięty o stałą 8, możemy zdefiniować nową funkcje, której wyznaczymy minimum:

$$g(x, y) = f(x, y) - 8 = x^2 - 4x + y^2 - 4y + xy \quad . \tag{22}$$

Wykorzystując metodę Newtona (15) - bez stałej:

$$g(r) = \frac{1}{2} r^T A r + r^T b \quad , \tag{23}$$

gdzie $r^T = [x, y]$, *A* to macierz Hessego, $b = [b_1, b_2]$ jest wektorem wyrazów wolnych.

Znajdujemy macierz *A*, ze wzoru (6):

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \quad , \tag{24}$$

oraz bazując na wzorze (23) obliczamy $b = [-4, -4]$.

Dla porównania należy sporządzić wykres konturowy $g(x, y)$ w przedziale $x \in (-10, 10)$, $y \in (-10, 10)$ i na jego podstawie określić przybliżone minimum.

Następnie dla punktów startowych $(0,0),(10,-10),(100,100),(500,500)$ wyznaczyć minimum metodą Newtona, oraz określić ilość iteracji.

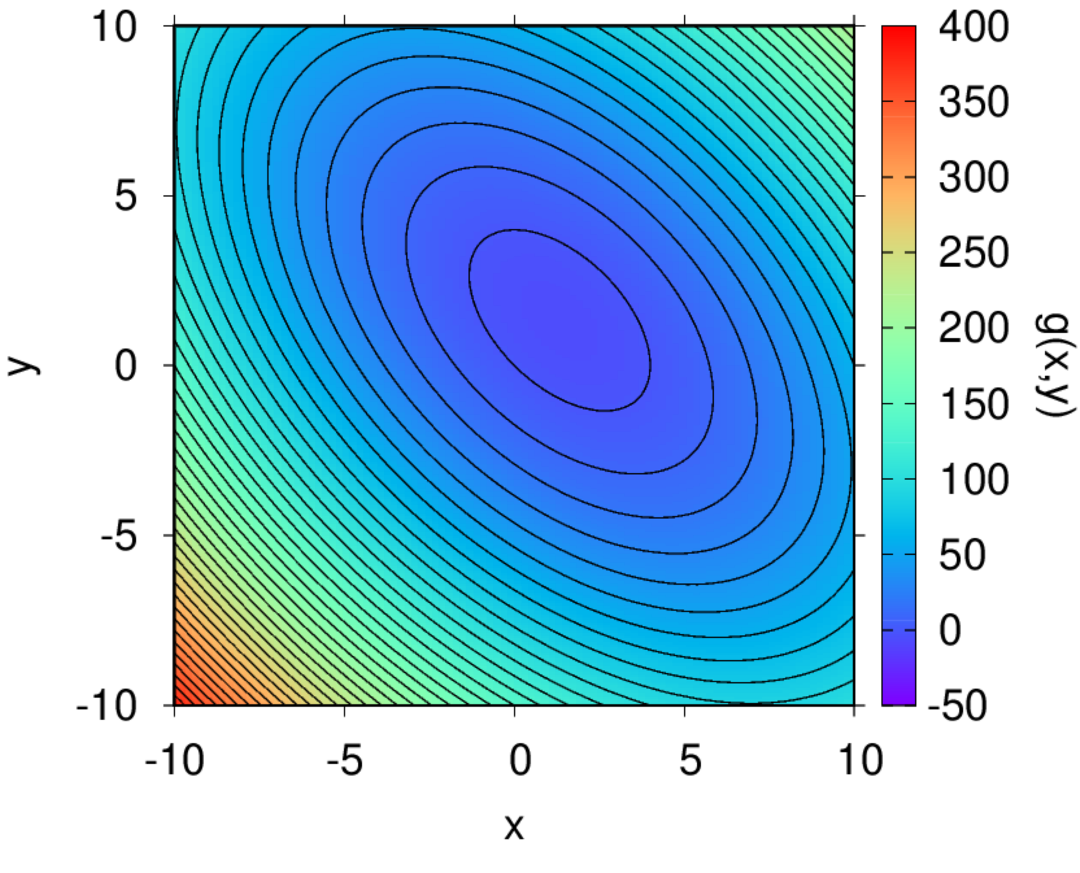
Ostatnią częścią jest wprowadzenie wagi ω do wzoru (20):

$$x^{i+1}=x^i-\omega H^{-1}(x^i)\nabla f(x^i) \quad , \tag{25}$$

oraz wyznaczenia dla minimum dla stałego punktu $(10,10)$, przyjmując wagi $\omega=0.1,0.4,0.7$.

2.2. Wyniki

Dzięki programowi w języku C, oraz wykorzystując skrypt Gnuplota wygenerowano:



Rysunek 1.: Wykres konturowy na tle mapy wartości funkcji *g(x, y)*. Czerwonym krzyżykiem zaznaczono dokładne minimum funkcji: $x_{min}=\frac{4}{3}, y_{min}=\frac{4}{3}$.

Ponieważ funkcja jest kwadratowa podjęto próbę wyznaczenia gotowego rozwiązania znajdując macierz odwrotną do *A*, gdzie wyszło $\vec{r}_{min}=\begin{pmatrix}1.33333\\1.33333\end{pmatrix}$ co zgadza się z odczytanymi z rys. 1 danymi.

(x_0, y_0)	Liczba iteracji	x_{min}	y_{min}
(0, 0)	2	1.33333	1.33333
(10, -10)	2	1.33333	1.33333
(100, 100)	2	1.33333	1.33333
(500, 500)	2	1.33333	1.33333

Tabela 1. Wyniki poszukiwania minimum funkcji (x_{min}, y_{min}) dla różnych punktów startowych (x_0, y_0). W drugiej kolumnie podano liczbę iteracji wykonanych przed uzyskaniem zbieżności.

Ilość iteracji w tabeli 1. dla każdego punktu startowego wynosi tyle samo i jest stosunkowo mała, wyniki są zgodne z wykresem konturowym, algorytm dla wagi $\omega = 1$ jest efektywny.

(ω)	Liczba iteracji	x_{min}	y_{min}
0.1	135	1.33334	1.33334
0.4	32	1.33333	1.33333
0.7	15	1.33333	1.33333

Tabela 2. Wyniki poszukiwania minimum funkcji (x_{min}, y_{min}) przy różnych wartościach wagi ω dla punktu startowego $(x_0, y_0)=(10,10)$. W drugiej kolumnie podano liczbę iteracji wykonanych przed uzyskaniem zbieżności.

Kiedy jednak zmienimy wagę zgodnie z wzorem (25) część, którą odejmujemy jest mała przez co ilość iteracji jest największa dla najmniejszej wagi i analogicznie odwrotnie proporcjonalna podczas zwiększania ω.

3. Wnioski

Metoda Newtona w naszym przypadku dla wagi równej 1 jest bardzo efektywna i niezależna od wartości punktu startowego, co jest ogromnym plusem, wystarczyło zaledwie 2 iteracje. Kiedy jednak waga maleje zwiększa się ilość iteracji, a co za tym idzie ukazują się minusy metody:

- konieczność wyznaczania hesjanu w każdym punkcie, problem z liczeniem pochodnych, jeśli nie analitycznie to numerycznie, co może skutkować błędami przybliżenia,
- nie zawsze jest zbieżna nawet w pobliżu minimum, co skutkuje potrzebą nałożeniem kolejnych warunków, spadkiem efektywności algorytmu.