Laboratorium 1

Zagadnienie przeszukiwania i podstawowe podejścia do niego

Krystian Kamiński nr albumu 304013

Polecenie:

Pierwszym tematem ćwiczeń jest zagadnienie przeszukiwania i związane z tym podejścia. Państwa zadaniem będzie minimalizacja funkcji celu i porównanie wyników dla metody najszybszego spadku gradientu i metody Newtona. W raporcie oczekiwałbym, że opiszą Państwo zaobserwowane różnice w działaniu tych metod dla różnych punktów początkowych z zadanego zakresu (warto podać również wyniki czasowe). Zadane przeze mnie funkcje posiadają minima lokalne i globalne, dlatego warto, żeby udało się je Państwu znaleźć. Istnieje możliwość, że dla danego punktu startowego funkcja nie znajdzie minimum, dlatego liczę na Państwa wnioski w tym przypadku. Wizualizacje nie są przeze mnie wymagane, chociaż bardzo Państwu pomogą w interpretacji wyników np. zmiana położenia kolejnych współrzędnych punktów.

Osoby z nazwiskami od A do C

Funkcją celu będzie funkcja Bootha postaci:

$$f(x,y) = (x + 2 * y - 7)^2 + (2 * x + y - 5)^2, -5 \leq x, y \leq 5$$

Osoby z nazwiskami od D do K

Funkcją celu będzie funkcja Himmelblau postaci:

$$f(x,y) = (x * x + y - 11)^2 + (x + y * y - 7)^2, -5 \leq x, y \leq 5$$

Osoby z nazwiskami od W do Ż

Funkcją celu będzie funkcja bananowa Rosenbrocka postaci:

$$f(x,y) = (1 - x)^2 + 100 * (y - x * x)^2, -5 \leq x, y \leq 5$$

Mam nadzieję, że nikogo nie zgubiłem, ale jeśli ktoś nie wie, które powinien zrobić zadanie to proszę zapytać.

Funkcje te zazwyczaj istnieją w całej dziedzinie, dlatego też podane zakresy mają jedynie pomóc Państwu przy pisaniu programu / doborze punktów. Nie ma sensu brak współrzędnych spoza tego zakresu. Dodatkowo znamy funkcję celu, więc nie muszą Państwo losować punktów przy wyznaczaniu gradientu a jedynie użyć zadanych wzorów.

Wykorzystane biblioteki pythona:

time - na potrzeby sprawdzenia wydajności obu metod
matplotlib - w celu przedstawienia wykresów

f(x, y) = (x^2+y-11)^2+(x+y^2-7)^2

gradient funkcji wynosi:

$$\begin{bmatrix} \left(4x^3 + 4xy - 42x + 2y^2 - 14\right) \\ \left(4y^3 + 4xy - 26y + 2x^2 - 22\right) \end{bmatrix}$$

hessian funkcji wynosi:

$$\begin{pmatrix} \texttt{12*x^2+4*y-42} & \texttt{4*x+4*y} \\ \texttt{4*x+4*y} & \texttt{4*x+12*y^2-26} \end{pmatrix}$$

odwrotność hesjanu:

$$\begin{pmatrix} \dfrac{2x+6y^2-13}{24x^3-164x^2-84x+24y^3+72x^2y^2-260y^2-8xy-52y+546} & \dfrac{-x-y}{12x^3-82x^2-42x+12y^3+36x^2y^2-130y^2-4xy-26y+273} \\ \dfrac{-x-y}{12x^3-82x^2-42x+12y^3+36x^2y^2-130y^2-4xy-26y+273} & \dfrac{6x^2+2y-21}{24x^3-164x^2-84x+24y^3+72x^2y^2-260y^2-8xy-52y+546} \end{pmatrix}$$

iloczyn odwrotności hesjanu i gradientu

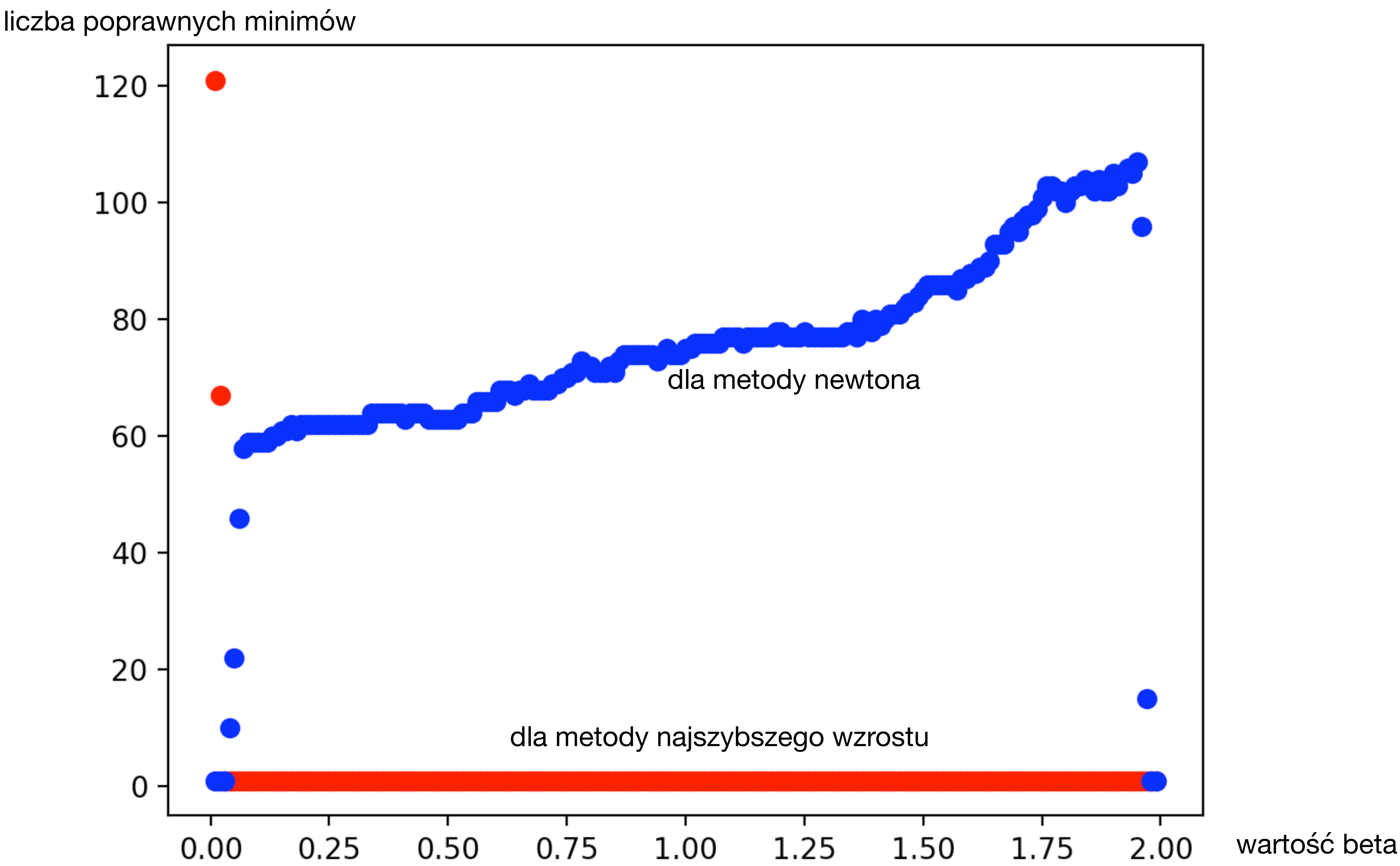
$$\begin{pmatrix} \dfrac{4x^4-28x^3-42x^2+281x+2y^4+8xy^3+12x^3y^2-128xy^2-29y^2-2x^2y+22y+91}{12x^3-82x^2-42x+12y^3+36x^2y^2-130y^2-4xy-26y+273} \\ \dfrac{2x^4-45x^2+14x+4y^4+12x^2y^3-44y^3-2xy^2-26y^2+8x^3y-80x^2y+265y+231}{12x^3-82x^2-42x+12y^3+36x^2y^2-130y^2-4xy-26y+273} \end{pmatrix}$$

W metodzie Newtona rozbiłem ten iloczyn na fragmenty, ze względu na zbyt dużą liczbę znaków w linii.
Funkcja ta posiada 4 minima lokalne równe 0, w punktach:
(x = 3,0, y = 2,0), (x = 3,58, y = -1,85), (x = -3,78, y = -3,28),

(x = -2,81, y = 3,13),
1 maksimum lokalne równe 181,62 dla x = -0,27, y = -0,92 oraz 4 punkty siodłowe równe 104,02 dla (x = -3,07, y = -0,08), 13,31 dla (x = 3,39, y = 0,07), 178,34 dla (x = -0,13, y = -1.95), 67,72 dla (x = 0,09, y = 2,88)

Obie metody były testowane dla wszystkich par liczb całkowitych (x, y) z podanego zakresu <-5, 5>.

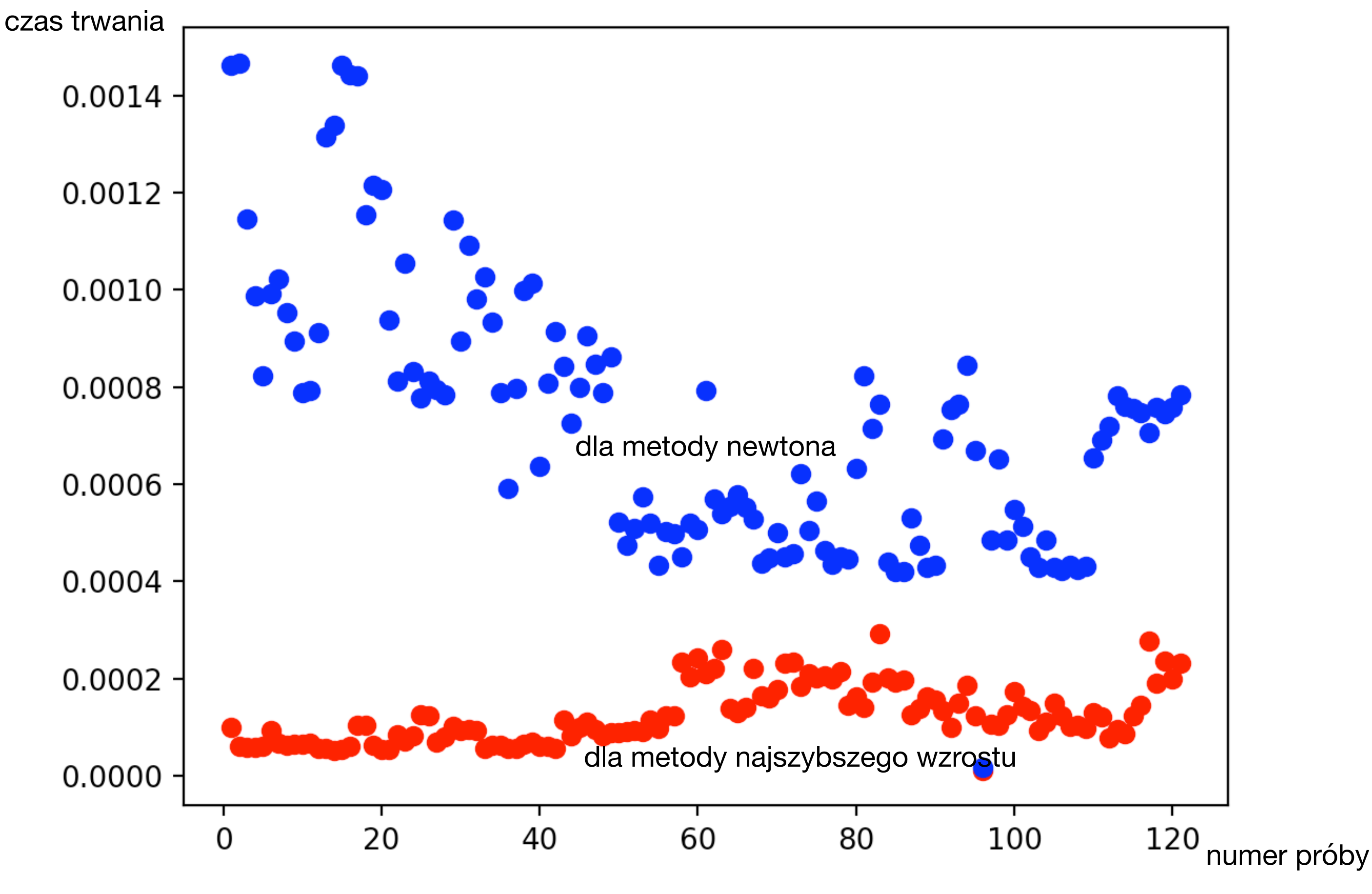
Wstępnie przetestowałem zachowanie obu metod w zależności od parametru beta.



Wykres przedstawia liczbę otrzymanych minimów dla danego parametru beta. (Czerwony wykres

reprezentuje metodę Steepest Gradient Descent, natomiast niebieski Newton Method)
Dla danego beta liczba minimów lokalnych mogła w testach wynieść od 0 do 121.
Dla pierwszej metody, optymalne beta wynosi 0,1 natomiast dla metody newtona beta = 1,95.

Następnie dla najbardziej optymalnych parametrów beta, przetestowałem czas trwania poszczególnych minimalizacji metod:

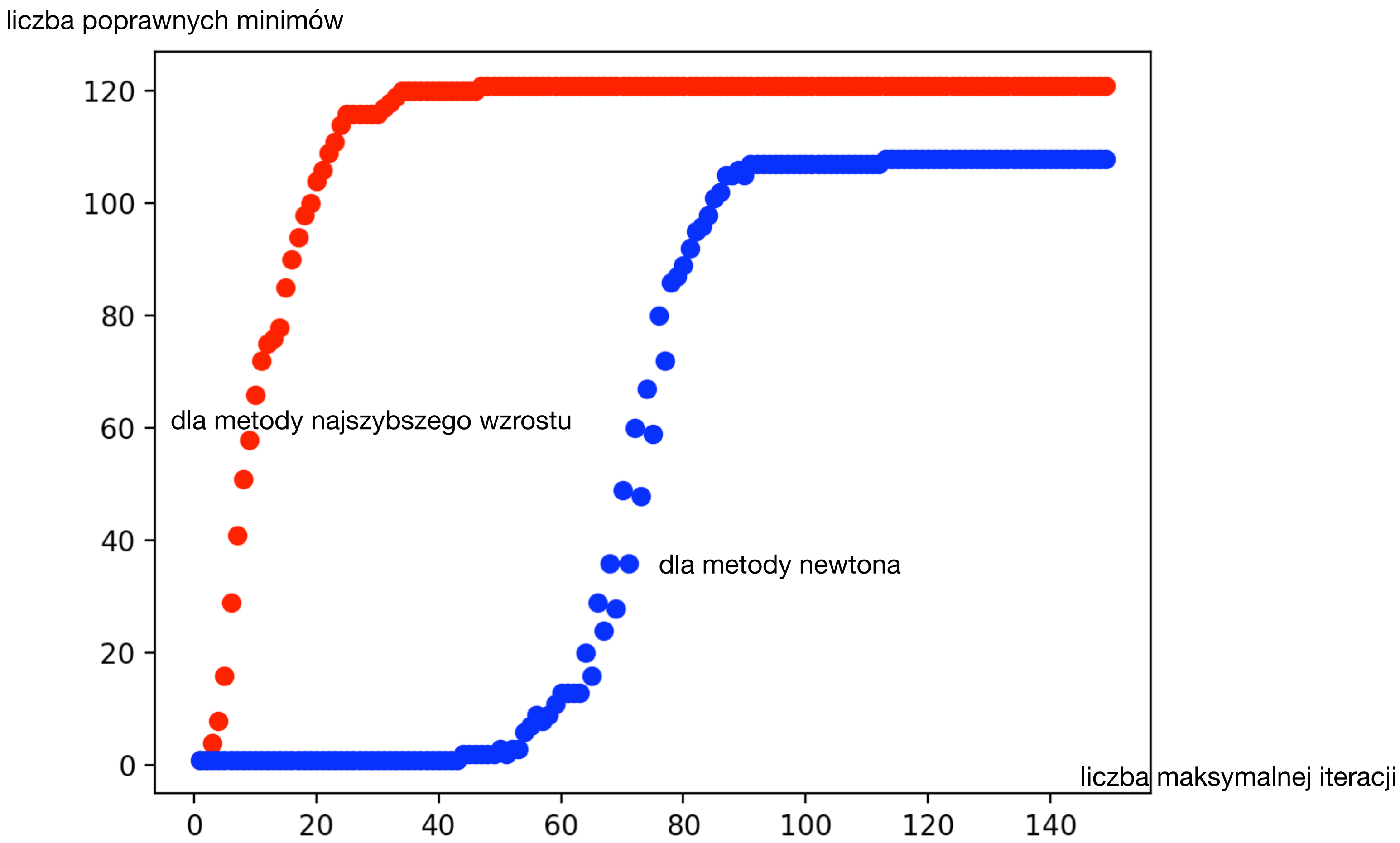


Reprezentacja kolorów metod jest identyczna jak w poprzednim przypadku.

Można zauważyć, że czas minimalizacji jest zdecydowanie lepszy w przypadku metody najszybszego wzrostu. Okazuje się jednak, że czas trwania minimalizacji w metodzie nie jest jedyną wadą. Dla pewnych wartości punktu początkowego, w tej metodzie otrzymane wyniki nie były minimalne, a okazały się punktami siodłowymi, bądź maksimum lokalnym. Taka sytuacja występuje w momencie kiedy obliczany hesjan daje wartości ujemne. Jednak przy optymalnym doborze parametru beta, aż 89% wartości jest prawidłowych, natomiast w przypadku metody najszybszego wzrostu jest to 100%.

Ostatnim testem jest zależność wydajności metody od maksymalnej liczby iteracji.

W przypadku metody najszybszego wzrostu maksymalna wydajność jest osiągana na poziomie 50 iteracji, natomiast w przypadku metody newtona liczba ta oscyluje w okolicach 90.
Oczywiście dla większej liczby iteracji czas trwania minimalizacji będzie nieoptymalnie rósł, natomiast w przypadku mniejszych wartości liczba wykrytych minimów lokalnych będzie niezadowalająco mała.


liczba poprawnych minimów
dla metody najszybszego wzrostu
dla metody newtona
liczba maksymalnej iteracji
0
20
40
60
80
100
120
140