Sprawozdanie - laboratorium nr 10
Poszukiwanie minimum wartości funkcji metodą największego spadku w 2D

Marek Kiełtyka

16 maja 2019

1 Wstęp

1.1 Minimalizacja funkcji

Inaczej nazywana optymalizacją, jej zadaniem jest poszukiwanie odpowiednio minimum lub maksimum funkcji wielu zmiennych. Chodzi o znalezienie punktu spełniającego warunek

$$\min f(\vec{x}) = f(\vec{x^*}) \Leftrightarrow \bigwedge_{\vec{x}\in R^n} f(\vec{x^*}) < f(\vec{x}), \quad \text{gdzie } \vec{x} = [x_1, x_2, \dots, x_n]^T. \tag{1}$$

1.2 Gradient funkcji

Dla funkcji celu $f(\vec{x}) \in C^2$, tj. poszukującej minimum badanej funkcji wejściowej definiuje się funkcję wektorową będącą gradientem funkcji

$$g(\vec{x}) = \nabla f(\vec{x}) = \left[\frac{\partial f(\vec{x})}{\partial x_1}, \frac{\partial f(\vec{x})}{\partial x_2}, \dots, \frac{\partial f(\vec{x})}{\partial x_n},\right]^T. \tag{2}$$

Należy pamiętać, iż gradient skierowany jest zawsze w stronę narastających wartości.

1.3 Pochodna kierunkowa funkcji celu

Różniczkę zupełną funkcji celu definiuje się jako iloczyn skalarny wektorów

$$df = \frac{\partial f}{\partial x_1}dx_1 + \dots + \frac{\partial f}{\partial x_n}dx_n = \nabla f(\vec{x})dx. \tag{3}$$

Punkty $\vec{x}$ i $\vec{x'}$ nazywane są powiązanymi ze sobą, jeżeli wektor $\vec{u}$ wyznacza kierunek prostej je łączącej, stąd

$$\vec{x}(\lambda) = \vec{x'} + \lambda\vec{u}. \tag{4}$$

Dla bardzo małych zmian wartości λ można uogólnić wzór (4).

$$d\vec{x} = \vec{u}d\lambda \tag{5}$$

Na prostej łączacej ustalone punkty $\vec{x}$ oraz $\vec{x'}$ wartość funkcji celu zależna jest od zmiennej λ.

$$F(\lambda) = f(\vec{x'} + \lambda\vec{u}) = f(\vec{x}) \tag{6}$$

Mając na uwadze powyższe powiązania, oblicza się różniczkę zupełną dla funkcji celu zależnej od λ

$$dF = df = \nabla^T f(\vec{x})\vec{u}d\lambda. \tag{7}$$

Finalnie, wyrażenie na **pochodną kierunkową funkcji** celu w punkcie $\vec{x}$ dla kierunku $\vec{u}$ jest postaci

$$\frac{dF(\lambda)}{d\lambda} = \left.\frac{df(\vec{x})}{d\lambda}\right|_{\vec{u}} = \nabla^T f(\vec{x})\vec{u}, \tag{8}$$

jednak korzystając z niej należy ją wyznaczać w każdej iteracji.

1.4 Znajdowanie minimum funkcji przy pomocy pochodnej kierunkowej

Przybliżanie należy rozpocząć z punktu $\vec{x_0}$ przechodząc przez kolejne punkty $\vec{x_1}, \vec{x_2}, \ldots, \vec{x_n}$ w kierunku spadku wartości funkcji. Pozwala to wyznaczyć ciąg przybliżeń poszukiwanego minimum. Należy przerwać algorytm iteracyjny w momencie, gdy zostanie spełniony jeden z warunków:

1. Norma różnicy wektorów z sąsiednich kroków jest mniejsza od zadanego progu: $||\vec{x}^{i+1} - \vec{x}^{i}|| < \epsilon$
2. $\nabla f(\vec{x}) = 0$
3. W kolejnych iteracjach wartość normy $||\vec{x}^{i+1} - \vec{x}^{i}||$ wzrasta, co oznacza brak zbieżności.

1.5 Metoda największego spadku

Wykorzystuje się

- wzór (8) w punkcie $\vec{x'}$ dla wektora kierunkowego $\vec{u}$ o długości równej $||\vec{u}|| = 1$

$$\left.\frac{\mathrm{d}f(\vec{x'})}{\mathrm{d}\lambda}\right|_{\vec{u}} = \frac{\mathrm{d}F(0)}{\mathrm{d}\lambda} = \nabla^T f(\vec{x'})\vec{u}, \tag{9}$$

- nierówność Schwartza

$$\nabla^T f(\vec{x'})\vec{u} \geq -||\nabla^T f(\vec{x'})|| \cdot ||\vec{u}|| = -||\nabla^T f(\vec{x'})|| = \min. \tag{10}$$

Należy wybrać wektor kierunkowy o postaci

$$\vec{u} = \frac{-\nabla f(\vec{x'})}{||\nabla f(\vec{x'})||} \tag{11}$$

aby wskazywał kierunek największego spadku, a pochodna kierunkowa mogła osiągnąć najmniejszą wartość.

$$\frac{\mathrm{d}F(0)}{\mathrm{d}\lambda} = -\nabla^T f(\vec{x'})\frac{\nabla f(\vec{x'})}{||\nabla f(\vec{x'})||} = \min \tag{12}$$

2 Zadanie do wykonania

2.1 Opis problemu

Celem laboratorium było znalezienie minimum zadanej funkcji (13) przy pomocy opisywanej metody

$$f(\vec{r}) = f(x, y) = \frac{5}{2}(x^2 - y)^2 + (1 - x)^2, \tag{13}$$

począwszy od przybliżenia $\vec{r_0}$ postaci

$$\vec{r_{i+1}} = \vec{r_i} - h \cdot \nabla f(\vec{r})|_{\vec{r}=\vec{r_i}} \qquad \text{gdzie } \nabla f(\vec{r}) = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right], \tag{14}$$

które było korygowane w kolejnych iteracjach. Składowe gradientu ze wzoru (14) obliczano numerycznie korzystając z faktu, iż

$$\frac{\partial f(x_i, y_i)}{\partial x} \approx \frac{f(x_i + \Delta x, y_i) - f(x_i - \Delta x, y_i)}{2\Delta x} \tag{15}$$

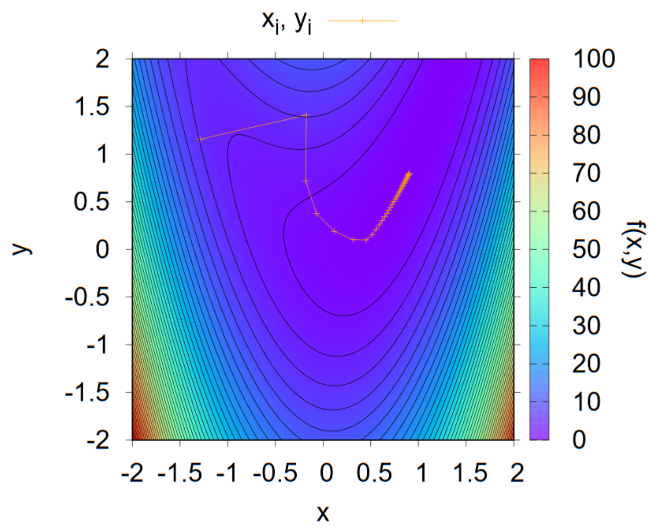
$$\frac{\partial f(x_i, y_i)}{\partial y} \approx \frac{f(x_i, y_i + \Delta y) - f(x_i, y_i - \Delta y)}{2\Delta y} \tag{16}$$

Przyjęto następujące założenia:

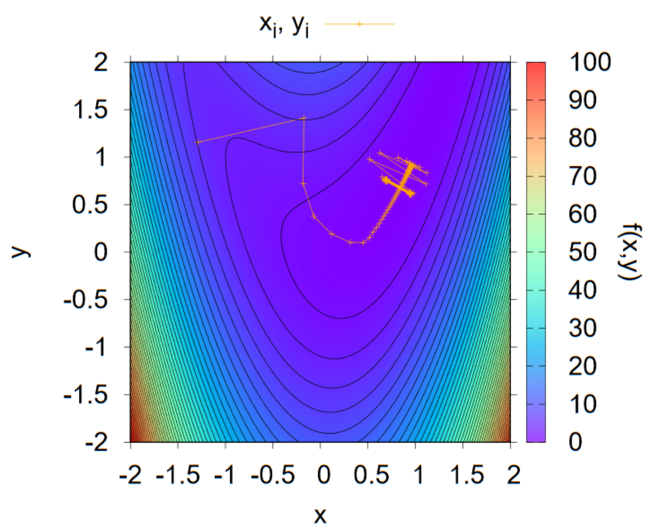
- punkt początkowy $\vec{r_0} = [-0.75, 1.75]$
- stała $h = 0.01$
- warunki stopu
  - $||\vec{r_{i+1}} - \vec{r_i}||_2 < \epsilon$ dla $\epsilon_1 = 10^{-2}$, $\epsilon_2 = 10^{-3}$
  - maksymalna liczba iteracji $maxIter = 1000$
- $\Delta x = \Delta y = \Delta = 10^{-4}$
- $x, y \in [-2, 2]$ z krokiem w obu kierunkach równym 0.02

2.2 Wyniki

Korzystając z programu napisanego w języku C++ i spreparowanego wcześniej skryptu programu *Gnuplot* sporządzono wykresy dla obu przypadków. Rozważano jedną funkcję, zatem dokładne jej minimum było jedno i wyniosło $x_{min} = 1, y_{min} = 1$.



Wykres 1: Położenia kolejnych przybliżeń minimum funkcji $f(x, y)$ w poszczególnych iteracjach dla $\epsilon = 10^{-2}$. W tle: kontury oraz mapa wartości funkcji $f(x, y)$. Program wykonał 37 iteracji.



Wykres 2: Położenia kolejnych przybliżeń minimum funkcji $f(x, y)$ w poszczególnych iteracjach dla $\epsilon = 10^{-3}$. W tle: kontury oraz mapa wartości funkcji $f(x, y)$. Program wykonał 1000 iteracji.

3 Wnioski

Wykresy dowodzą, iż ciężko jest wybrać trafne kryterium zbieżności. Dla pierwszego przypadku nie udało się osiągnąć minimum, choć przebieg (x_i, y_i) zdawał się być na dobrej drodze ku temu. Zmniejszając rząd zbieżności o jeden, zastosowanie znalazł drugi warunek stopu z uwagi na niekończące się oscylacje wokół poszukiwanego minimum. Potwiedziło to również teoretyczny kształt przebiegu (x_i, y_i) dla funkcji z wydłużonym konturem.

Jedynym parametrem, który można by zmodyfikować dla lepszej efektywności obliczeń, jest h. Z racji jego stałości podczas laboratorium skoki nie były regulowane, co znalazło szczególne odbicie w punktach funkcji o niewielkim gradiencie, a w konsekwencji nie osiągnięto zbieżności.

Mimo tych niedoskonałości, metoda największego spadku prowadzi do znalezienia minimum funkcji wielu zmiennych w sposób szybki i prosty.