Sprawozdanie 10.

Poszukiwanie minimum wartości funkcji metodą największego spadku w 2D

Mirosław Kołodziej

13.05.2021

1. Wstęp teoretyczny

1.1 Kierunkowe metody poszukiwania minimum funkcji

Pochodną kierunkową funkcji celu definiujemy różniczkę zupełną funkcji celu jako iloczyn skalarny wektorów:

$$df = \frac{\partial f}{\partial x_i} dx_1 + \cdots + \frac{\partial f}{\partial x_n} dx_n = \nabla f(x) dx.$$

Zakładamy, że wektor $\vec{u}$ wyznacza kierunek prostej łączącej punkty x i x', możemy więc zapisać:

$$x(\lambda) = x' + \lambda \vec{u},$$

zaś dla bardzo małych zmian wartości λ:

$$dx = \vec{u} d\lambda$$

Na prostej łączącej x i x' funkcję celu, zależną od zmiennej λ wyraża się wzorem:

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

Wyliczając różniczkę zupełną dla tej funkcji otrzymujemy:

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

zaś pochodną kierunkową funkcji celu w punkcie x dla kierunku $\vec{u}$ wyrażamy następująco:

$$\frac{dF(\lambda)}{d\lambda} = \frac{df(x)}{d\lambda}\bigg|_{\vec{u}} = \nabla^T f(x) \vec{u}$$

Korzystając z rozwinięcia w szereg Taylora mamy:

$$F(\lambda) = F(0) + \frac{dF}{d\lambda}\lambda \left(+ \frac{d^2 F}{d\lambda^2} \frac{\lambda^2}{2} + \cdots \right) \rightarrow zaniedbujemy,$$

gdzie:

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

$$F(0) = f(x' + 0 \cdot \vec{u}) = f(x').$$

Wprowadźmy następujące założenie:

$$\lambda > 0, \qquad \frac{dF}{d\lambda} = \nabla^T f(x')\vec{u} = -|\nabla^T f(x')\vec{u}| < 0,$$

wtedy:

$$F(\lambda) = f(x') - \lambda|\nabla^T f(x')\vec{u}| < f(x').$$

Zatem, ze względu określenia gradientu funkcji w dowolnym punkcie, musimy dobrać wektor $\vec{u}$ tak, aby spełniał warunek:

$$\nabla^T f(x')\vec{u} < 0.$$

Rozpoczynając w określonym punkcie x_0, szukamy kolejnego przybliżenia x_1 w kierunku spadku wartości funkcji. Wyznaczamy tym samym ciąg kolejnych przybliżeń $x0, x1, x2$, ... poszukiwanego minimum. Procedurę iteracyjną przerywamy, gdy spełni się jeden z warunków:

- $\|x^{i+1} - x^i\| < \epsilon$
- $\nabla f(x) = \vec{0}$
- w kolejnych iteracjach rośnie wartość normy $\|x^{i+1} - x^i\|$ – oznacza to brak zbieżności.

1.2 Metoda największego spadku

Metoda największego spadku to algorytm numeryczny mający na celu znalezienie minimum zadanej funkcji celu. Korzystamy w niej z pochodnej kierunkowej funkcji w punkcie x':

$$\left.\frac{df(x')}{d\lambda}\right|_{\vec{u}} = \frac{dF(\lambda)}{d\lambda} = \nabla^T f(x')\vec{u},$$

gdzie długość wektora kierunkowego $\vec{u}$ to $\|\vec{u}\| = 1$. Korzystając z nierówności Schwartza:

$$\nabla^T f(x')\vec{u} \geq -\|\nabla^T f(x')\| \cdot \|\vec{u}\| =\geq -\|\nabla^T f(x')\| \cdot 1.$$

Gdy wybierzemy wektor kierunkowy w poniższej postaci:

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

to wskazuje on kierunek największego spadku. Pochodna kierunkowa ma wtedy największą wartość:

$$\frac{dF(\lambda)}{d\lambda} = -\nabla^T f(x')\frac{\nabla f(x')}{\|\nabla f(x')\|} = -1.$$

Sam algorytm metody największego spadku możemy przedstawić w postaci:

- wybierz x^0
- obliczaj iteracyjnie:

$$\vec{u}^i = \frac{-\nabla f\left(x^{i-1}\right)}{\|\nabla f(x^{i-1})\|}$$

$$x^i = x^{i-1} + \lambda\vec{u}^i$$

$$F(\lambda_i) = f\left(x^{i-1} + \lambda_i\vec{u}^i\right) = \min_{\lambda} f(x^{i-1} + \lambda_i\vec{u}^i)$$

- warunki zakończenia obliczeń:

$$\left\| x^{i+1} - x^{i} \right\| < \varepsilon_1$$

$$\left\| \nabla f\left(x^{i}\right) \right\| < \varepsilon_2$$

$$\left| f(x^{i}) - f(x^{i-1}) \right| < \varepsilon_3$$

2. Problem

Zadaniem naszych zajęć laboratoryjnych było wyznaczenie numerycznie minimum funkcji:

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

Korzystaliśmy przy tym z metody największego spadku. Zaczynaliśmy od przybliżenia $\overrightarrow{r_0}$, którego wartości w kolejnych iteracjach zmienialiśmy według wzoru:

$$\overrightarrow{r_{i+1}} = \vec{r_i} - h \cdot \nabla f(\vec{r})|_{\vec{r}=\overrightarrow{r_i}},$$

gdzie gradient $\nabla f(\vec{r})$ jest równy:

$$\nabla f(\vec{r}) = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right].$$

Zaczęliśmy od zaprogramowania metody największego spadku dla dwóch wymiarów. Pochodne obliczyliśmy numerycznie z poniższych wzorów:

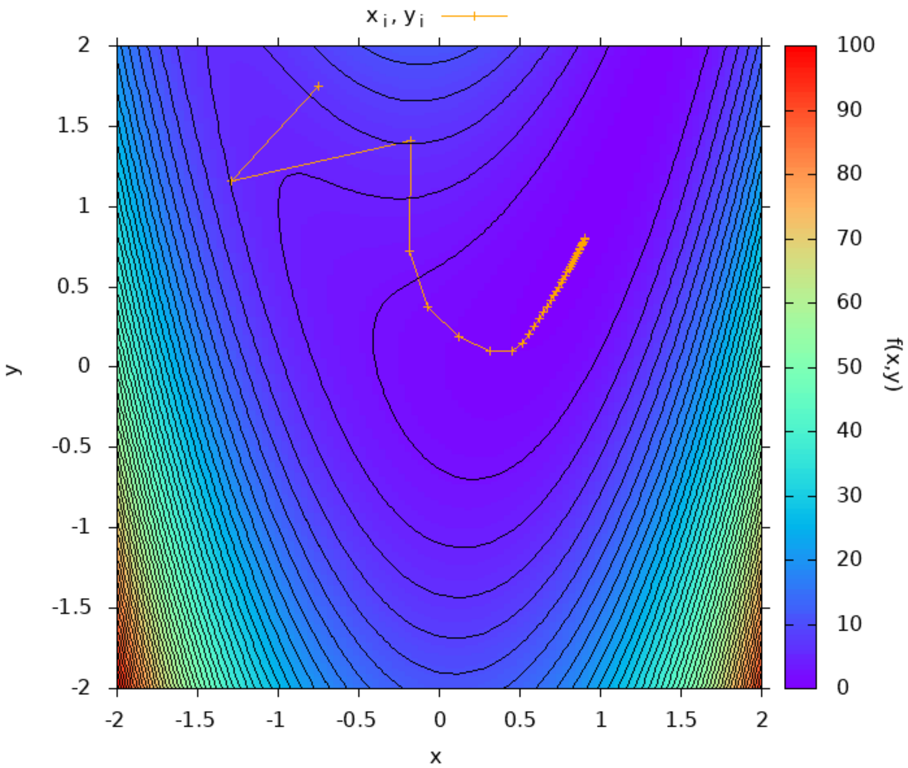
$$\frac{\partial f(\vec{r})}{\partial x} = \frac{f(\vec{r} + \Delta \cdot \overrightarrow{e_x}) - f(\vec{r} - \Delta \cdot \overrightarrow{e_x})}{2\Delta}$$

$$\frac{\partial f(\vec{r})}{\partial y} = \frac{f\left(\vec{r} + \Delta \cdot \overrightarrow{e_y}\right) - f\left(\vec{r} - \Delta \cdot \overrightarrow{e_y}\right)}{2\Delta},$$

gdzie $\overrightarrow{e_x}$ i $\overrightarrow{e_y}$ są wersorami układu kartezjańskiego. Przyjęliśmy $\Delta = 10^{-4}$. Następnie szukaliśmy przybliżonego położenia minimum funkcji startując od punktu $\overrightarrow{r_0} = [-0.75, 1.75]$. Przyjęliśmy $h = 0.1$, jako warunek stopu $\|\overrightarrow{r_{i+1}} - \vec{r_i}\|_2 < \varepsilon$, a jako maksymalną liczbę iteracji 1000. Obliczenia wykonaliśmy dla $\varepsilon = 10^{-2}$ i $\varepsilon = 10^{-3}$, a wyniki zapisaliśmy do pliku Dla każdej wartości ε sporządziliśmy rysunek, na którym widoczny był kontur funkcji oraz kolejne przybliżenia minimum połączone linią. Skorzystaliśmy przy tym z Gnuplota.

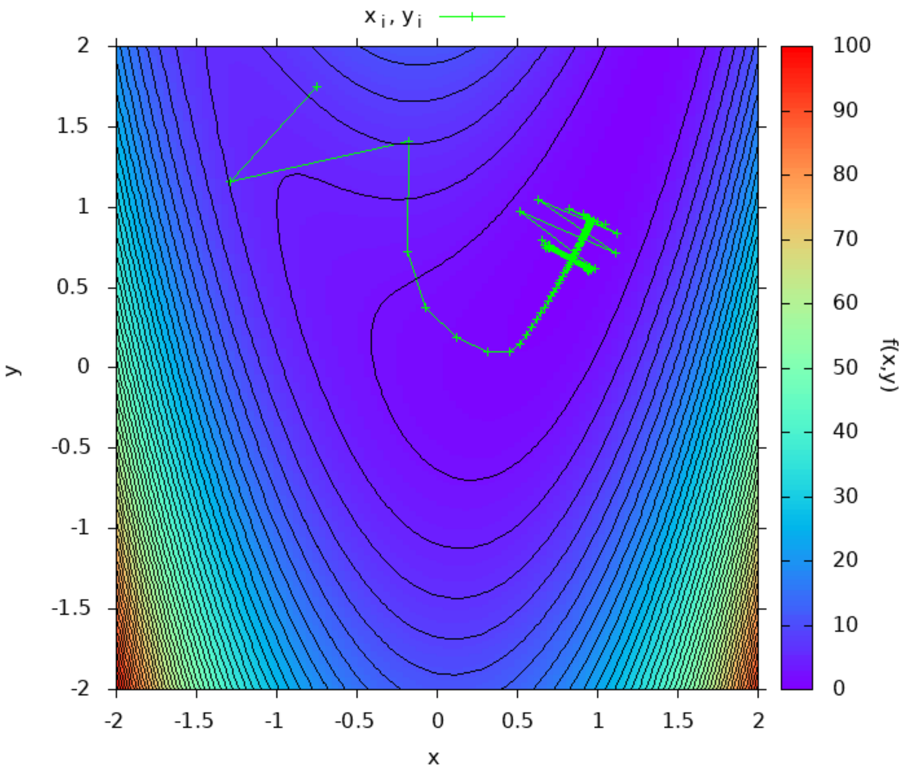
3. Wyniki

3.1 Kontur funkcji z kolejnymi przybliżeniami minimum dla $\varepsilon = 10^{-2}$



Ostateczne przybliżenie minimum wynosi $(0.904394, 0.801381)$.

3.2 Kontur funkcji z kolejnymi przybliżeniami minimum dla $\varepsilon = 10^{-3}$



Ostateczne przybliżenie minimum wynosi $(0.949205, 0.621386)$.

4. Wnioski

Znaleźliśmy minimum wartości funkcji metodą największego spadku w 2D. Algorytm podał niedokładne wartości przybliżeń. Dla przypadku z większą ilością iteracji, wartości te zaczynają oscylować. Możemy więc uznać, że w pierwszym przypadku warunek stopu został obrany zbyt restrykcyjnie – wartości są zbyt dalekie od minimum, w drugim zaś jest on nie do końca właściwy, ze względu na oscylacje. Uzyskane przybliżenie jest dalekie od dokładnego ze względu na kształt funkcji. Kontur wartości funkcji celu jest wydłużony – ma kształt elipsy. Z tego powodu na końcu drugiego rozwiązania możemy zaobserwować zygzak. Mogliśmy otrzymać niedokładne rozwiązania również z powodu braku regulacji wartości h. Zmienianie tej wartości pozwoliłoby na zwiększenie dokładności metody.