Raport do zadania 1

Algorytm spadku wzdłuż gradientu

Kacper Siemionek

Numer indeksu: 331430

Spis treści

1. Wstępne obliczenia, wyprowadzanie wzorów

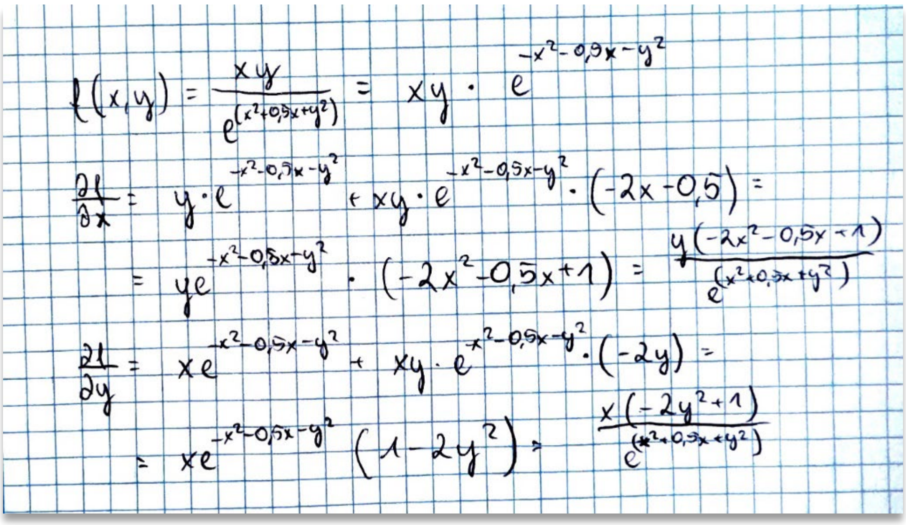
Pierwszą czynnością potrzebną do wykonania zadania jest obliczenie liczby A, czyli pierwszej niezerowej liczby po przecinku z działania:

$$\sqrt{3{,}14 * indeks\ studenta}$$

$$\sqrt{3{,}14 * 331430} \approx 1\ 020{,}14$$

$$A = 1$$

Następnie możemy przystąpić do wyprowadzenia wzorów pochodnych cząstkowych funkcji podanej w zadaniu. Zgodnie z prośbą poniżej znajdują się obliczenia wykonane ręcznie na kartce.



Rysunek 1.1 Wyprowadzenie wzorów pochodnych cząstkowych

Po obliczeniu ekstremów lokalnych bez użycia algorytmu otrzymujemy wyniki:

$$\max\left\{\frac{xy}{e^{x^2+0.5x+y^2}}\right\} = \left\{\left(-\frac{1}{8}-\frac{\sqrt{33}}{8}, -\frac{1}{\sqrt{2}}\right), \left(\frac{\sqrt{33}}{8}-\frac{1}{8}, \frac{1}{\sqrt{2}}\right)\right\} \approx$$

$$\approx \{(-0{,}8430703308, -0{,}7071067812), (0{,}5930703308, 0{,}7071067812)\}$$

$$\min\left\{\frac{xy}{e^{x^2+0.5x+y^2}}\right\} = \left\{\left(-\frac{1}{8}-\frac{\sqrt{33}}{8}, \frac{1}{\sqrt{2}}\right), \left(\frac{\sqrt{33}}{8}-\frac{1}{8}, -\frac{1}{\sqrt{2}}\right)\right\} \approx$$

$$\approx \{(-0{,}8430703308, 0{,}7071067812), (0{,}5930703308, -0{,}7071067812)\}$$

Przy użyciu algorytmu spadku wzdłuż gradientu z odpowiednio małą wartością kroku uczącego oraz dużą ilością iteracji otrzymujemy dokładnie takie same wyniki, korzystając w obu przypadkach z takiego samego przybliżenia liczb.

2. Gradient funkcji

Gradient $\nabla f(x, y)$ to wektor utworzony z pochodnych cząstkowych.

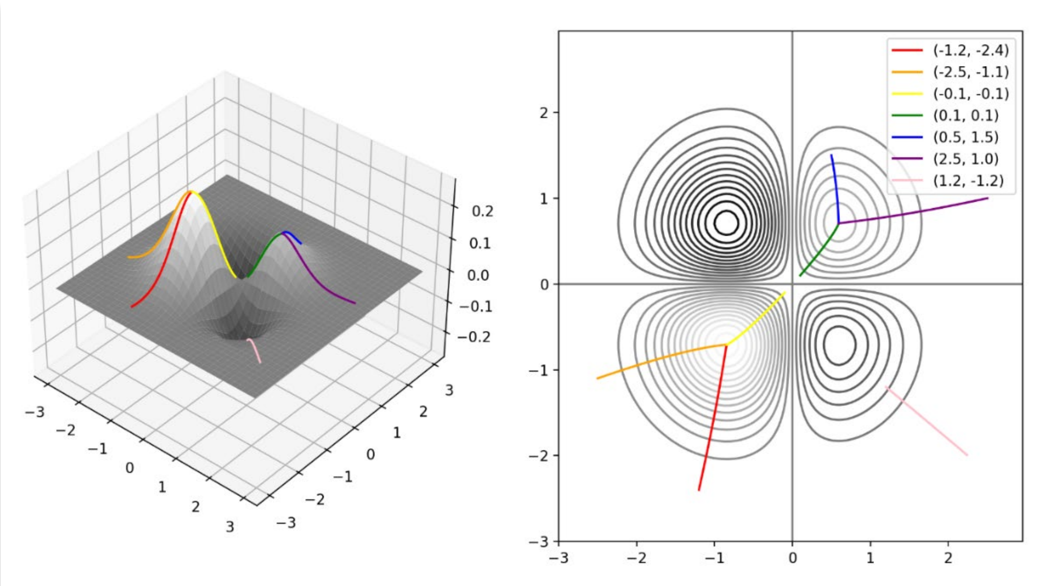
$$\nabla f(x, y) = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y})$$

Wektor ten wskazuje kierunek, w którym funkcja rośnie najszybciej. Jeśli chcemy znaleźć maksimum, powinniśmy poruszać się zgodnie z kierunkiem wektora, aby otrzymywać większe wartości funkcji. W przypadku poszukiwania minimum musimy iść w przeciwną stronę, gdzie funkcja maleje.

3. Wybór punktu startowego

W algorytmie spadku wzdłuż gradientu wybór punktu startowego ma znaczący wpływ na otrzymany wynik. Funkcja podana w zadaniu posiada kilka minimów i maksimów lokalnych, co oznacza, że wektor pochodnych cząstkowych będzie wskazywał na najbliższe z nich, zależnie od wybranego przez nas punktu. Co więcej, w konkretnych przypadkach punkt startowy, który znajduje się zbyt daleko szukanego ekstremum, może nas doprowadzić do wyniku niezgodnego z prawdą.

Podczas pierwszego eksperymentu wybrałem po 3 punkty startowe znajdujące się w pobliżu maksimów funkcji oraz jeden oddalony od nich.



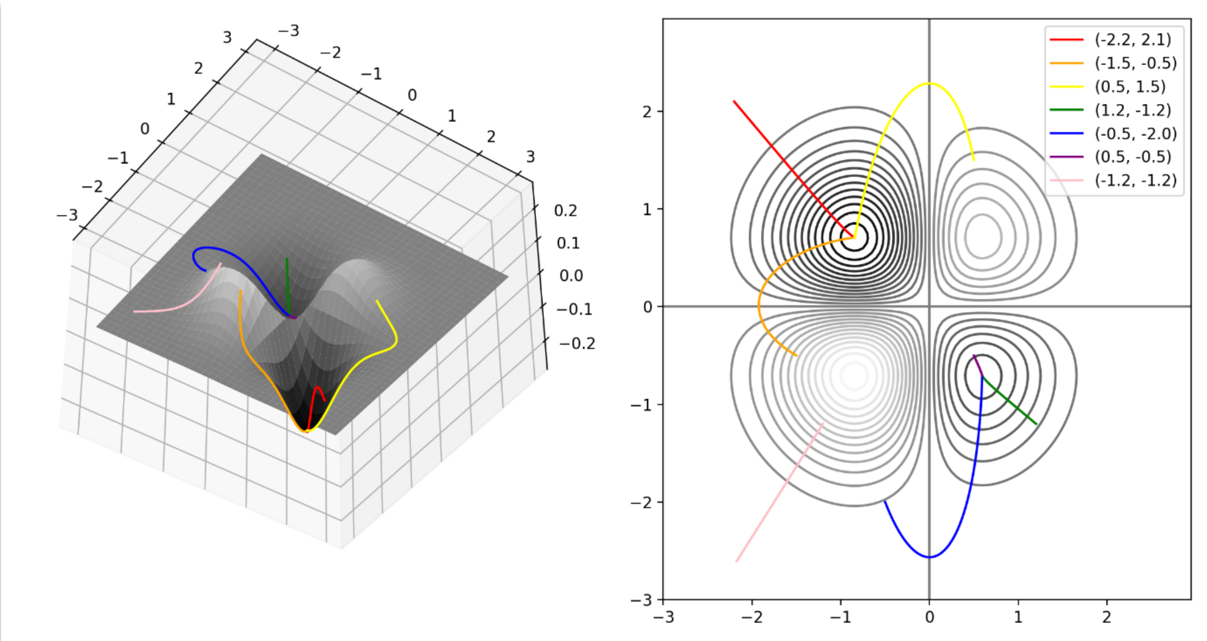
Rysunek 3.1 Wyniki dla różnych punktów startowych (maksimum)

Wyniki:

- $(-0.843,\ -0.707,\ 0.271)$ dla punktów $(-1.2,\ -2.4), (-2.5,\ -1.1), (-0.1,\ -0.1)$
- $(0.593,\ 0.707,\ 0.133)$ dla punktów $(0.1,\ 0.1), (0.5,\ 1.5), (2.5,\ 1.0)$
- $(2.239,\ -1.993,\ 0.000)$ dla punktu $(1.2,\ -1.2)$

Analiza wyników eksperymentu potwierdza początkowe założenia o znaczeniu wyboru punktu startowego. Punkty $(-1.2,\ -2.4), (-2.5,\ -1.1), (-0.1,\ -0.1)$ oraz $(0.1,\ 0.1), (0.5,\ 1.5), (2.5,\ 1.0)$ dotarły do zgodnych ze sobą wyników, które faktycznie są maksimami lokalnymi przedstawionej funkcji. Punkt $(1.2,\ -1.2)$ znajdował się blisko wzrostu funkcji, który nie prowadził do maksimum, dlatego otrzymany wynik różni się od pozostałych i nie wskazuje na żadne z maksimów.

Dobór punktów do drugiego eksperymentu wyglądał podobnie, jednak w tym przypadku szukamy minimum funkcji.



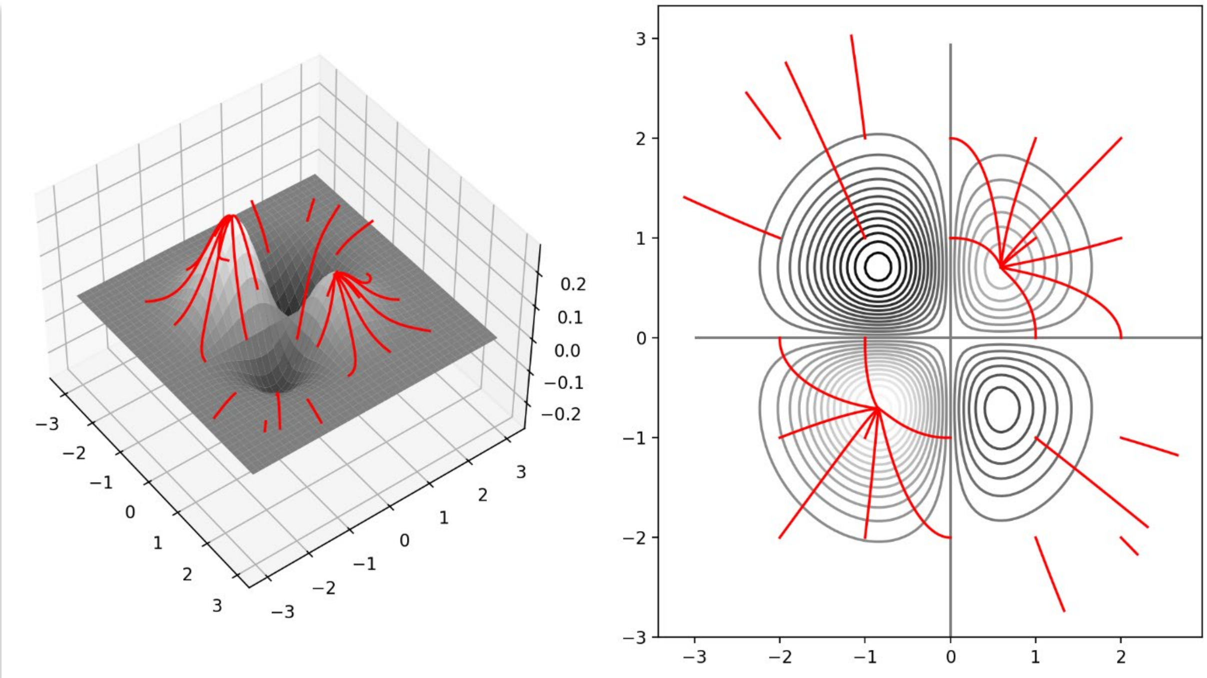
Rysunek 3.2 Wyniki dla różnych punktów startowych (minimum)

Wyniki:

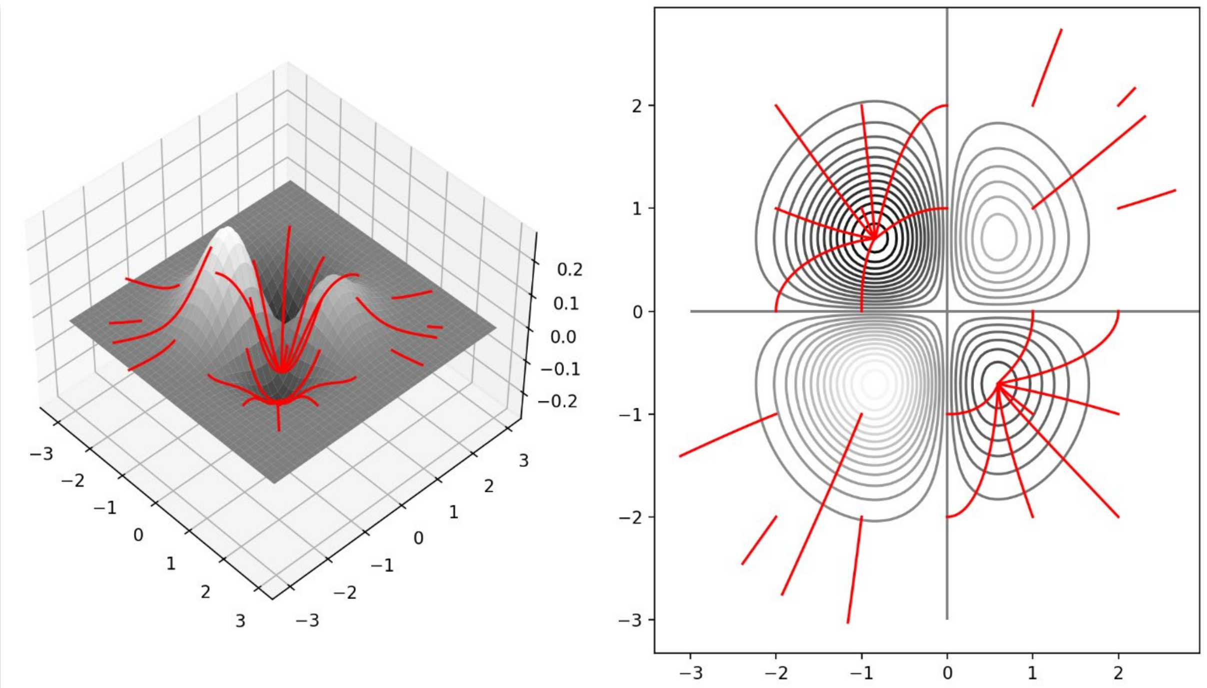
- $(-0.843,\ 0.707,\ -0.271)$ dla punktów $(-2.2,\ -2.1), (-1.5,\ -0.5), (0.5,\ 1.5)$
- $(0.593,\ -0.707,\ -0.133)$ dla punktów $(1.2,\ -1.2), (-0.5,\ -2.0), (0.5,\ -0.5)$
- $(-2.171,\ -2.605,\ 0.000)$ dla punktu $(-1.2,\ -1.2)$

W powyższym eksperymencie również potwierdza się założenie. Punkty znajdujące się w pobliżu minimów bez problemu dotarły do celu, jednak ostatni punkt wraz z kolejnymi krokami oddalał się od poprawnego wyniku.

Dodatkowo przeprowadziłem dwa eksperymenty, startując z punktów otrzymanych po obliczeniu iloczynu kartezjańskiego $\{-2,-1,0,1,2\} \times \{-2,-1,0,1,2\}$, aby pokazać więcej przykładów działania algorytmu zgodnego z wcześniejszymi założeniami.



Rysunek 3.3 Wyniki dla siatki punktów startowych (maksimum)

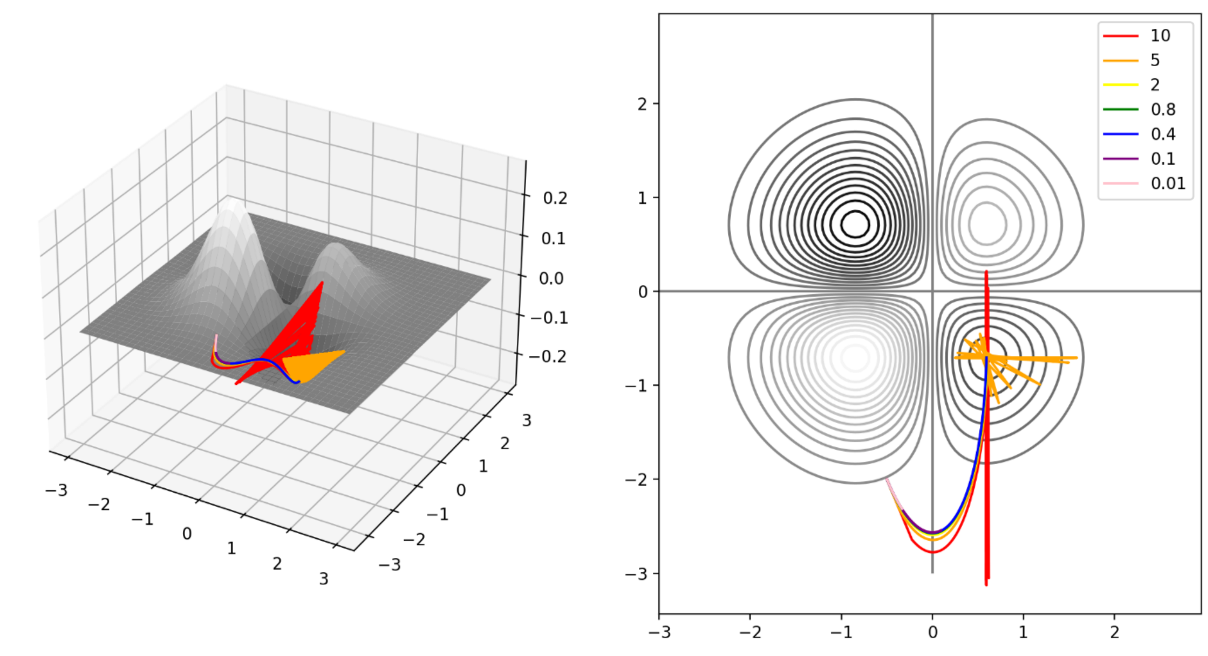


Rysunek 3.4 Wyniki dla siatki punktów startowych (minimum)

4. Wartość kroku uczącego

Kluczowym elementem działania algorytmu spadku wzdłuż gradientu jest dobór odpowiedniej wartości kroku uczącego. Jeśli wybierzemy zbyt dużą wartość, kierunek wektora gradientu może zmieniać się w nieoczekiwany sposób, co doprowadzi do błędnych „przeskoków" oraz potencjalnego pominięcia optymalnego rozwiązania. Z drugiej strony za mały krok uczący powoduje nieadekwatne działanie algorytmu, ponieważ potrzebna będzie większa ilość iteracji ze względu na minimalne zmiany w odległościach.

W celu sprawdzenia wyżej wymienionych zależności wykonałem dwa eksperymenty w poszukiwaniu minimum oraz maksimum różnych wartości kroku uczącego, zaczynając z tego samego punktu. W obu przypadkach ilość iteracji algorytmu wynosiła 1000.

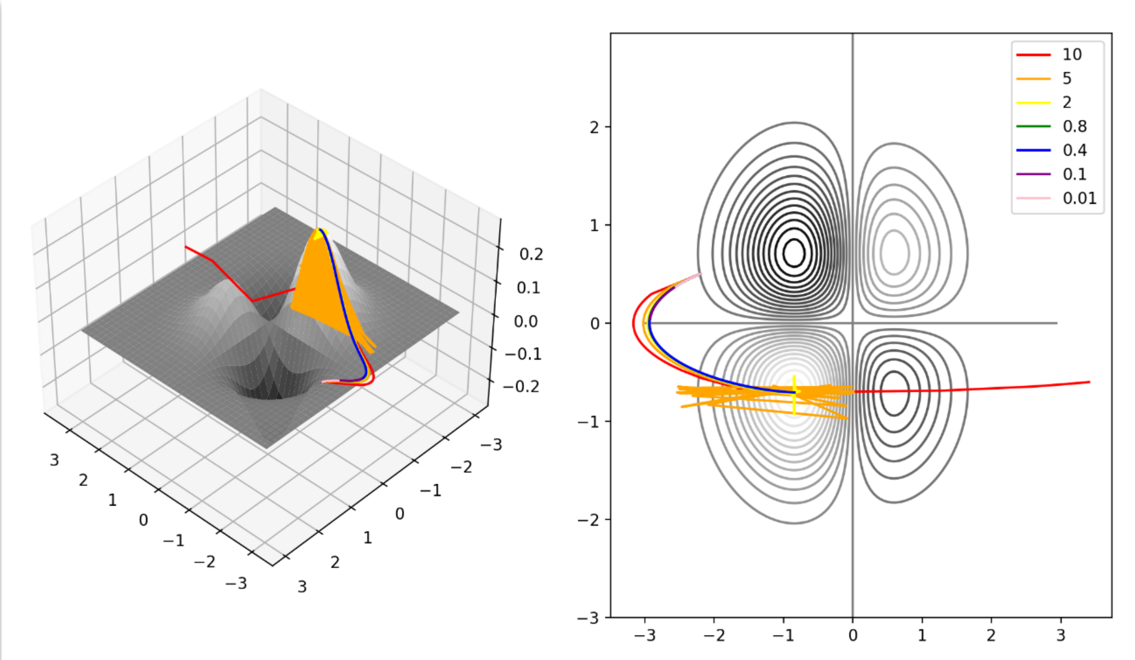


Rysunek 4.1 Wyniki dla różnych wartości kroku uczącego (minimum)

Wyniki:

- $(0.593,\ -2.908,\ 0.000)$ dla wartości 10
- $(0.997,\ -0.707,\ -0.096)$ dla wartości 5
- $(0.593,\ -0.707,\ -0.133)$ dla wartości $2,\ 0.8,\ 0.4$
- $(0.093,\ -2.544,\ 0.000)$ dla wartości 0.1
- $(-0.334,\ -2.314,\ 0.004)$ dla wartości 0.01

Jedynie wartości kroku uczącego z przedziału $[0.4;\ 2]$ pozwoliły na uzyskanie poprawnych wyników. Wyższe wartości spowodowały, że algorytm zaczął „przeskakiwać” w nieprzewidywalny sposób wzdłuż jednej prawidłowej współrzędnej. Z kolei niższe wartości nie zdążyły doprowadzić algorytmu do minimum w ustalonej liczbie iteracji. Gdyby ta liczba nie była ograniczona, wartości 0.01 oraz 0.1 prawdopodobnie kiedyś również przekazałyby nam prawidłowy wynik, jednak byłoby to znacznie mniej efektywne rozwiązanie.



Rysunek 4.2 Wyniki dla różnych wartości kroku uczącego (maksimum)

Wyniki:

- $(3.406,\ -0.601,\ 0.000)$ dla wartości 10
- $(-2.181,\ -0.707,\ 0.024)$ dla wartości 5
- $(-0.843,\ -0.542,\ 0.255)$ dla wartości 2
- $(-0.843,\ -0.707,\ 0.271)$ dla wartości $0.8,\ 0.4$
- $(-2.922,\ 0.042,\ 0.000)$ dla wartości 0.1
- $(-2.551,\ 0.373,\ 0.004)$ dla wartości 0.01

W drugim eksperymencie prawidłowe wyniki udało się uzyskać przy wartościach kroku uczącego z zakresu $[0.4;\ 0.8]$. Podobnie jak wcześniej, większe wartości powodowały, że algorytm omijał prawidłowe rozwiązanie. Natomiast te niższe pomimo właściwego kierunku były ograniczone przez ilość iteracji.

Na podstawie obu eksperymentów możemy jasno stwierdzić, że dobór odpowiedniej wartości kroku uczącego jest istotny, jeśli chcemy uzyskać optymalne rozwiązanie.

5. Zwiększenie precyzji algorytmu

Obecne działanie algorytmu pozwala na bezproblemowe znalezienie szukanych ekstremów lokalnych z dużą dokładnością. Aby otrzymać optymalne rozwiązanie, możemy:

- Zwiększyć liczbę iteracji oraz użyć odpowiednio małej wartości kroku uczącego, aby bardziej zbliżyć się do ekstremum.
- Uwzględnić różnice pomiędzy kolejnymi przejściami jako warunek zatrzymania, co spowoduje, że dotrzemy do wyniku z określonym przez nas przybliżeniem.
- Zaimplementować dynamiczną zmianę wartości kroku uczącego, czyli zmniejszać ją wraz ze zbliżaniem się do ekstremum, aby kolejne kroki były dokładniejsze.

Stosując się do powyższych przykładów, możemy dojść do rozwiązania, które jest zgodne z wynikiem uzyskanym metodami analitycznymi, bez konieczności korzystania z algorytmu.