**Wydział Elektroniki i Technik Informacyjnych**
**Politechnika Warszawska**

# Wstęp do Sztucznej Inteligencji
# Semestr 24L
# Sprawozdanie z ćwiczenia nr 1

# Metoda gradientu prostego

**Mikołaj Wewiór**

**Warszawa,**
**12 III 2024**

# 1. Opis problemu

Celem ćwiczenia było zaimplementowanie metody gradientu prostego. Jest to metoda, która poprzez obliczenie pochodnych cząstkowych danej funkcji może znaleźć jej lokalne optimum. Czy będzie to maksimum czy minimum zależy od znaku przy gradiencie. Inaczej mówiąc jest to metoda najszybszego wzrostu (lub spadku). W przypadku ćwiczenia realizowane było zadanie minimalizacji - dlatego przy $\beta$ stoi minus. Wzór na znalezienie następnych wartości współrzędnych punktu roboczego dany jest poniżej.

$$x_i[t+1] = x_i[t] - \beta \nabla q(x),$$

gdzie $q(x)$ to funkcja celu, dalej oznaczana poprzez $f(x)$, a $\beta$ jest parametrem, którego wartość ustalamy przy inicjalizacji algorytmu.

# 2. Wzory

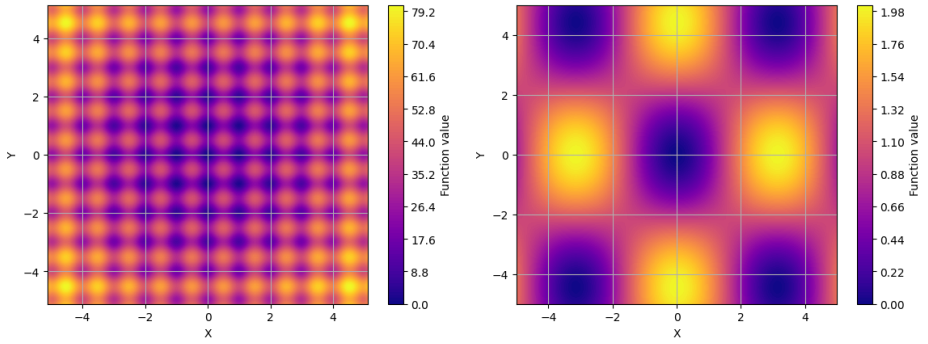
Implementacja oraz analiza zostały przeprowadzone w dwóch wymiarach ($d = 2$) dla dwóch funkcji: Rastringa oraz Griewanka. Dane są one następującymi wzorami:

Rastrigin:

$$f(x) = 10d + \sum_{i=1}^{d} [x_i^2 - 10\,\cos(2\pi x_i)],$$

Griewank:

$$f(x) = \sum_{i=1}^{d} \frac{x_i^2}{4000} - \prod_{i=1}^{d} \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1,$$




(a) Rastrigin (b) Griewank

Do obliczenia gradientów funkcji wykorzystano funkcjonalność serwisu `Wolfram Alpha`. Otrzymane gradienty dla dwóch zmiennych prezentują się w następujący sposób:

Rastrigin:

$$\frac{\partial f(x)}{\partial x_1} = 2x_1 + 20\pi\,\sin(2\pi x_1)$$

$$\frac{\partial f(x)}{\partial x_2} = 2x_2 + 20\pi\,\sin(2\pi x_2)$$

Griewank:

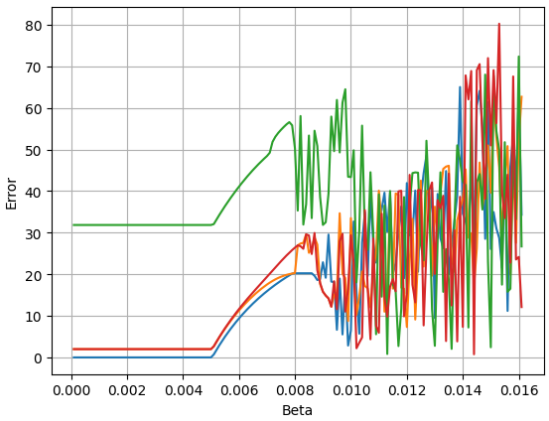
$$\frac{\partial f(x)}{\partial x_1} = \frac{x_1}{2000} + \sin(x_1)\cos\left(\frac{x_2}{\sqrt{2}}\right)$$

$$\frac{\partial f(x)}{\partial x_2} = \frac{x_2}{2000} + \frac{1}{\sqrt{2}}\sin\left(\frac{x_2}{\sqrt{2}}\right)\cos(x_1)$$

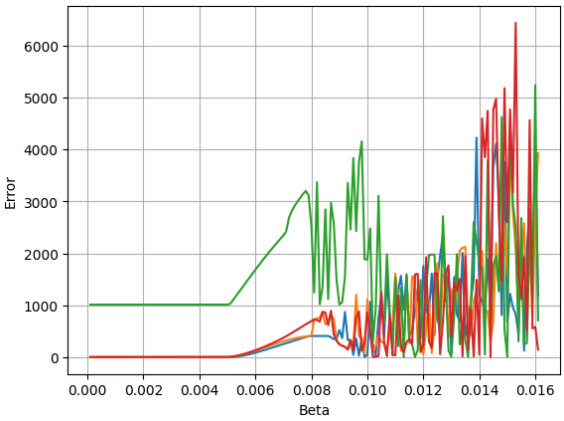
# 3. Wpływ parametru $\beta$

Na początku przebadanano wpływ parametru $\beta$, czyli tzw. parametr uczenia się. Określa on jak silnie gradient wpływa na zmianę współrzędnych punktu roboczega w kolejnej iteracji. Przebiegi oraz tabele dla analizowanych funkcji znajdują się poniżej.

## 3.1. Funkcja Rastrigina




(a) błąd absolutny (L1)



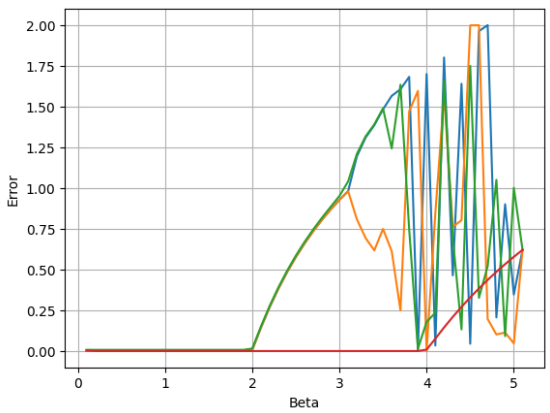

(b) błąd średniokwadratowy (MSE)

Każdy z kolorów reprezentuje wyniki uzyskane z danego punktu inicjalizacji.

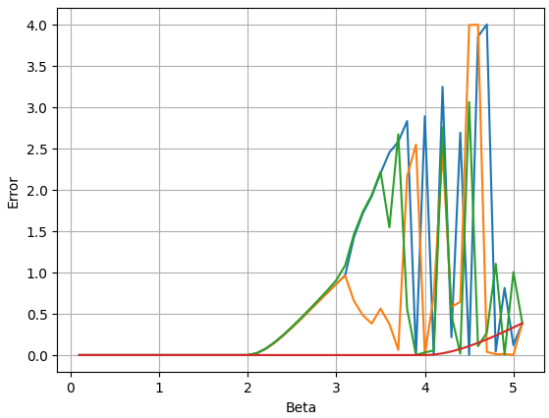
| $\beta$ | 0,0001 | 0,0049 | 0,005 | 0,0051 | 0,007 | 0,012 |
|---|---|---|---|---|---|---|
| błąd absolutny (L1) | $1{,}0 \cdot 10^{-10}$ | $1 \cdot 10^{-10}$ | $4{,}0 \cdot 10^{-8}$ | 0,7 | 15,9 | 37,6 |
| błąd średniokwadratowy (MSE) | $1{,}3 \cdot 10^{-20}$ | $6{,}0 \cdot 10^{-21}$ | 0,0 | 0,5 | 253,7 | 1414,7 |

Tab. 1: Wartości błędów dla wybranych wartości $\beta$ dla przebiegu niebieskiego

## 3.2. Funkcja Griewanka




(a) błąd absolutny (L1)




(b) błąd średniokwadratowy (MSE)

Każdy z kolorów reprezentuje wyniki uzyskane z danego punktu inicjalizacji.

| $\beta$ | 0,1 | 1,9 | 2,0 | 2,1 | 3,0 | 4,3 |
|---|---|---|---|---|---|---|
| błąd absolutny (L1) | $1{,}0 \cdot 10^{-10}$ | 0,0 | $2{,}3 \cdot 10^{-3}$ | $1{,}4 \cdot 10^{-1}$ | $9{,}8 \cdot 10^{-1}$ | 1,3 |
| błąd średniokwadratowy (MSE) | $8{,}3 \cdot 10^{-21}$ | $1{,}3 \cdot 10^{-21}$ | $5{,}1 \cdot 10^{-6}$ | $2{,}0 \cdot 10^{-2}$ | $8{,}6 \cdot 10^{-1}$ | 1,6 |

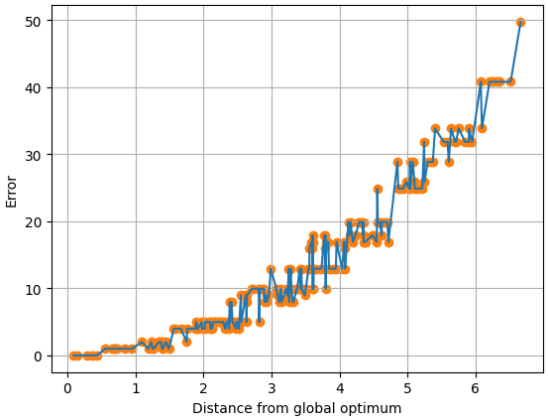
Tab. 2: Wartości błędów dla wybranych wartości $\beta$ dla przebiegu niebieskiego

Widzimy, że błąd między uzyskanym wynikiem a globalnym optimum jest do pewnego momentu stosunkowo niezmienny, a następnie znacząco wzrasta. Jest to skutek tego, że zbyt duża $\beta$ powoduje, że trajektoria punktu dążącego do optimum zaczyna zygzakować, co w niektórych przypadkach nie pozwala dojść w odpowiednio bliskie sąsiedztwo optimum globalnego. W skrajnych przypadkach sprawia to, że punkt roboczy w kolejnych iteracjach może opuścić otoczenie globalnego optimum, co skutkuje niezbieżnością algorytmu - na wykresie widać to, gdy pojawiają się bardzo duże różnice między kolejnymi wartościami funkcji celu przy niewielkim wzroście $\beta$.

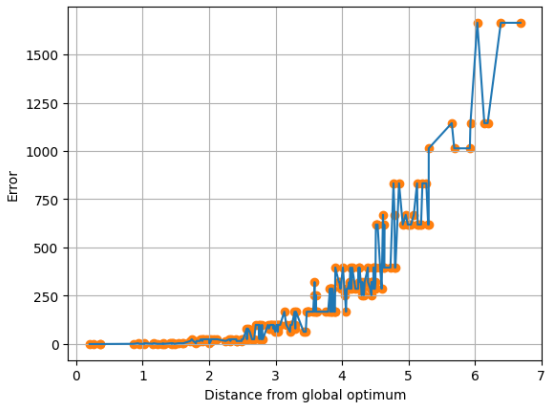
# 4. Wpływ punktów inicjalizacji

Kolejnym etapem było badanie wpływu punktu inicjalizacji na jakość rozwiązania. Przeprowadzono 100 generacji punktów. Każda z współrzędnych była losowana z rozkładem normalnym z wartością oczekiwaną 0 oraz o odchyleniu standardowym 4 z ograniczeniami do dziedziny danej funkcji. Wybrano taki rozkład ponieważ daje on nieco większe szanse na wystąpienie punktów inicjalizacji bliżej globalnego optimum, które dla dwóch badanych funkcji znajduje się w punkcie $(x_1, x_2) = (0, 0)$. Algorytm został uruchomiony z odpowiednimi wartościami parametru $\beta$, tj. 0,003 dla funkcji Rastrigina oraz 1,3 dla funkcji Griewanka. Przebiegi oraz tabele dla analizowanych funkcji znajdują się poniżej.

## 4.1. Funkcja Rastrigina




(a) błąd absolutny (L1)



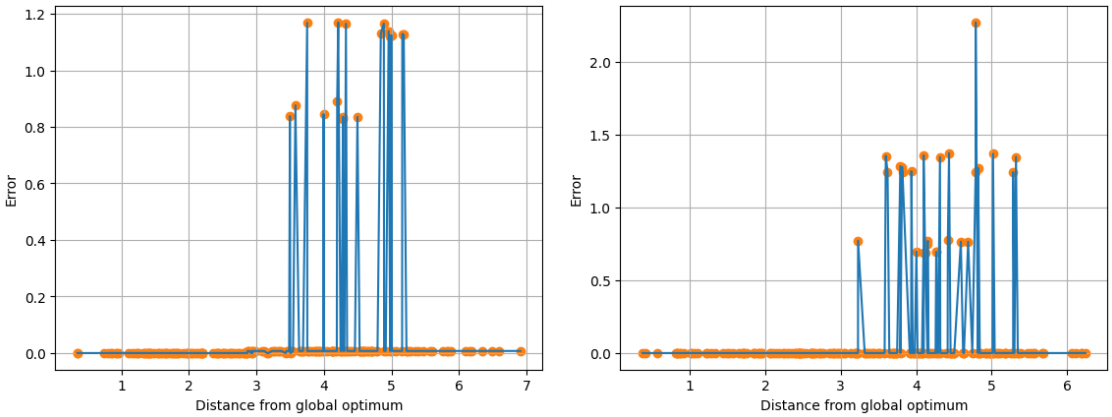

(b) błąd średniokwadratowy (MSE)

| Odległość od optimum globalnego | 0,4 | 0,6 | 1,3 | 3,6 | 4,3 | 5,6 |
|---|---|---|---|---|---|---|
| błąd absolutny (L1) | 0,0 | 1,0 | 2,0 | 12,9 | 16,9 | 33,8 |
| błąd średniokwadratowy (MSE) | $1{,}2 \cdot 10^{-23}$ | 1,0 | 3,9 | 167,3 | 286,1 | 1144,4 |

Tab. 3: Wartości błędów dla wybranych wartości odległości od optimum

Z wykresów bardzo dobrze widać, że wraz z wzrastającą odległością od optimum globalnego, punkty końcowe wpadają w optima lokalne funkcji, co jest spodziewanym skutkiem. Powyższy wykres przekłada się na krajobraz funkcji celu. Widać również, że dla ograniczonych wycinków dziedziny, różne punkty przyjmują te same wartości funkcji celu, co przedstawia obszary przyciągania danych lokalnych optimum.

## 4.2. Funkcja Griewanka




(a) błąd absolutny (L1)

(b) błąd średniokwadratowy (MSE)

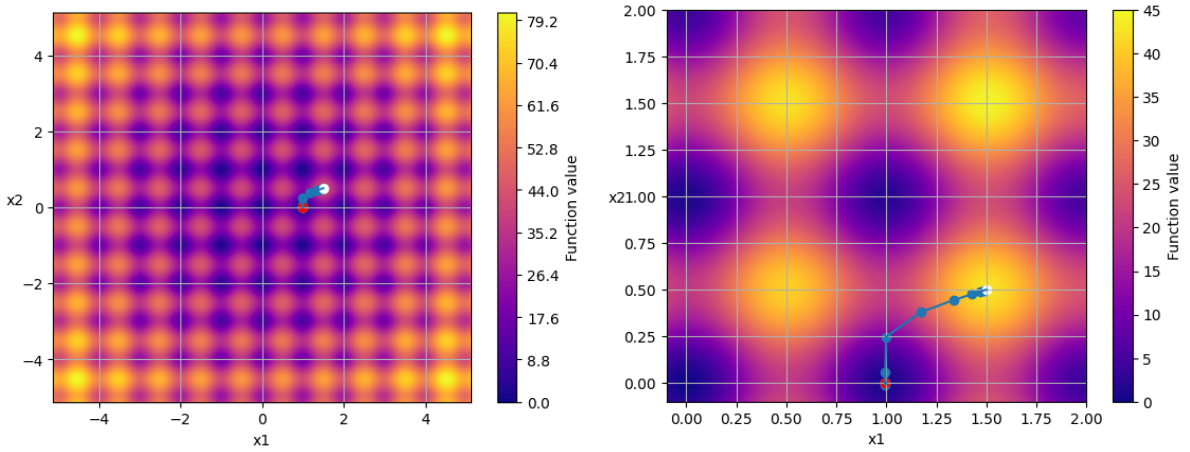
| Odległość od optimum globalnego | 0,5 | 2,3 | 3,1 | 3,7 | 4,4 | 6,4 |
|---|---|---|---|---|---|---|
| błąd absolutny (L1) | 0,0 | 0,0 | $7{,}4 \cdot 10^{-3}$ | 1,0 | 1,1 | $7{,}4 \cdot 10^{-3}$ |
| błąd średniokwadratowy (MSE) | $3{,}2 \cdot 10^{-24}$ | $5{,}2 \cdot 10^{-25}$ | $5{,}5 \cdot 10^{-5}$ | 1,0 | 1,2 | $5{,}5 \cdot 10^{-5}$ |

Tab. 4: Wartości błędów dla wybranych wartości odległości od optimum

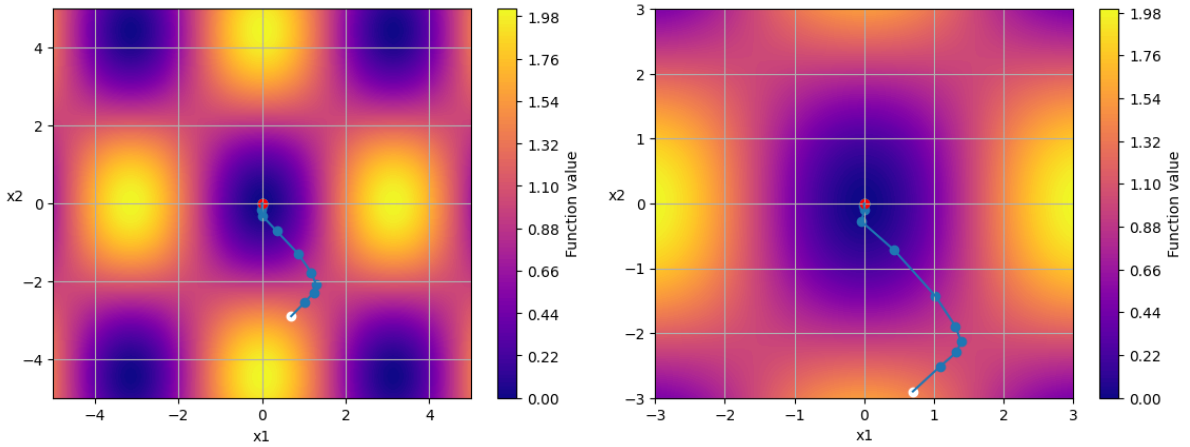
Na powyższych wykresach widać trzy dominujące obszary wartości błędów. Dla punktów zainicjowanych w odległości mniejszej niż 3 od punktu centralnego widzimy, że algorytm dochodzi do globalnego optimum. Kolejnym obszarem jest zakres w odległości pomiędzy około 2,75 do 3,5 oraz powyżej około 5,5. Są to takie pozycje, gdzie algorytm kończy działanie w lokalnych optimach o nieco gorszej wartości funkcji celu. Charakterystyczne są pojedyncze wartości między 3,5 a 5, gdzie funkcja celu osiąga względnie wysokie wyniki. Są to takie punkty, które zostały zainicjalizowane na granicy dziedziny zmiennej $x_1$, co w połączeniu z naprawą rozwiązań poprzez zawijanie, generowało punkty terminalne w lokalnym optimum powstałym przez ograniczenia.

## 5. Wizualizacja

Przykładowe trajektorie. Biały punkt to punkt inicjalizacji, a czerwony to punkt terminalny.




Rys. 6: Rastrigin




Rys. 7: Griewank

## 6. Podsumowanie

Metoda gradientu prostego nie daje gwarancji znalezienia optimum globalnego. Uzależniona jest od punktu inicjalizacji, który implikuje, do jakiego optimum lokalnego dojdzie dana trajektoria. Jednak nawet jeżeli punk inicjalizacji będzie w polu przyciągania optimum globalnego, ale parametr $\beta$ będzie zbyt duży, to najprawdopodobniej nie osiągniemy tego ekstremum. Im większa wartość parametru uczenia się, tym większe kroki wykonuje algorytm, dlatego aby na pewno osiągnąć jakiekolwiek optimum w odpowiedniej liczbie iteracji warto przyjąć betę na tyle dużą, aby trajektoria nie zaczęła zygzakować i przestrzeliwać otoczenia optimum, ale również nie zbyt małą, ponieważ zajmie to algorytmowi zbyt wiele iteracji, aby osiągnął ekstremum.