Optymalizacja
Laboratorium 11

Jakub Ciszewski, Wiktor Smaga

20 czerwca 2024

1 Zadanie 1

1. Obliczamy gradienty funkcji f względem x i y.
2. Rozwiązujemy układ równań, gdzie gradienty są równe zero, aby znaleźć punkty krytyczne.
3. Dla każdego znalezionego punktu krytycznego wykonujemy operacje:
   (a) Oblicz macierz Hessego funkcji.
   (b) Podstawiamy wartości x, y do macierzy Hessego
   (c) Oblicz wartości własne macierzy Hessego (Rozwiązujemy równanie charakterystyczne)
   (d) Klasyfikuj punkt krytyczny na podstawie wartości własnych:
      - Minimum, jeśli wszystkie wartości własne są dodatnie.
      - Maksimum, jeśli wszystkie wartości własne są ujemne.
      - Punkt siodłowy, jeśli wartości własne mają różne znaki.

```
Function f1(x, y):
  Critical point: {x: 0, y: 0}, Classification: saddle point

Function f2(x, y):
  Critical point: {x: -1, y: -1}, Classification: minimum
  Critical point: {x: 0, y: 0}, Classification: saddle point
  Critical point: {x: 1, y: 1}, Classification: minimum
  Critical point: {x: -I, y: I}, Classification: complex coordinates
  Critical point: {x: I, y: -I}, Classification: complex coordinates
  Critical point: {x: sqrt(2)*(-1 - I)/2, y: sqrt(2)/2 - sqrt(2)*I/2}, Classification: complex coordinates
  Critical point: {x: sqrt(2)*(-1 + I)/2, y: sqrt(2)/2 + sqrt(2)*I/2}, Classification: complex coordinates
  Critical point: {x: sqrt(2)*(1 - I)/2, y: -sqrt(2)/2 - sqrt(2)*I/2}, Classification: complex coordinates
  Critical point: {x: sqrt(2)*(1 + I)/2, y: -sqrt(2)/2 + sqrt(2)*I/2}, Classification: complex coordinates

Function f3(x, y):
  Critical point: {x: -1, y: -1}, Classification: maximum
  Critical point: {x: 0, y: -1}, Classification: saddle point
  Critical point: {x: 0, y: 0}, Classification: saddle point
  Critical point: {x: 1, y: 0}, Classification: minimum

Function f4(x, y):
  Critical point: {x: 1, y: 1}, Classification: saddle point
```

Wykres 1: Wynik powyższej operacji

1.1 Minima i maksima globalne

Funkcja $f_1(x, y) = x^2 - 4xy + y^2$

Przy x lub y dążącym do nieskończoności, wyrażenie x^2 oraz y^2 dominuje, co sprawia, że funkcja dąży do nieskończoności. Nie posiada więc minimum globalnego ani maksimum globalnego.

Funkcja $f_2(x, y) = x^4 - 4xy + y^4$

Przy x lub y dążącym do nieskończoności, wyrażenie x^4 oraz y^4 dominuje, co sprawia, że funkcja dąży do nieskończoności. Nie posiada więc minimum globalnego ani maksimum globalnego.

Funkcja $f_3(x, y) = 2x^3 - 3x^2 - 6xy(x - y - 1)$

Przy x lub y dążącym do nieskończoności, wyrażenie $2x^3$ dominuje, co sprawia, że funkcja dąży do $\pm$ nieskończoności w zależności od znaku x. Nie posiada więc minimum globalnego ani maksimum globalnego.

Funkcja $f_4(x, y) = (x - y)^4 + x^2 - y^2 - 2x + 2y + 1$

Przy x lub y dążącym do nieskończoności, wyrażenie $(x-y)^4$ dominuje, co sprawia, że funkcja dąży do nieskończoności. Nie posiada więc minimum globalnego ani maksimum globalnego.

2 Zadanie 2 - wyprowadzenie wzoru na gradient

Funkcja F jest zdefiniowana jako:

$$F(x^{(0)}, x^{(1)}, \dots, x^{(n)}) = \lambda_1 \sum_{i=0}^{n} \sum_{j=1}^{k} \frac{1}{\epsilon + \|x^{(i)} - r^{(j)}\|_2^2} + \lambda_2 \sum_{i=0}^{n-1} \|x^{(i+1)} - x^{(i)}\|_2^2$$

1. **Składnik pierwszy:**

$$\frac{\partial}{\partial x^{(i)}} \left(\sum_{j=1}^{k} \frac{1}{\epsilon + \|x^{(i)} - r^{(j)}\|_2^2} \right) = \sum_{j=1}^{k} \frac{\partial}{\partial x^{(i)}} \left(\frac{1}{\epsilon + \|x^{(i)} - r^{(j)}\|_2^2} \right)$$

Stosując regułę łańcuchową:

$$= \sum_{j=1}^{k} -\frac{1}{(\epsilon + \|x^{(i)} - r^{(j)}\|_2^2)^2} \cdot \frac{\partial}{\partial x^{(i)}} \left(\epsilon + \|x^{(i)} - r^{(j)}\|_2^2 \right)$$

Korzystając z tego, że: $\frac{\partial \|x\|^2}{\partial x} = 2x$:

$$= \sum_{j=1}^{k} -\frac{2(x^{(i)} - r^{(j)})}{(\epsilon + \|x^{(i)} - r^{(j)}\|_2^2)^2}$$

2. **Składnik drugi:**

$$\frac{\partial}{\partial x^{(i)}} \left(\|x^{(i+1)} - x^{(i)}\|_2^2 + \|x^{(i)} - x^{(i-1)}\|_2^2 \right)$$

Stosując fakt, że $\|x^{(i+1)} - x^{(i)}\|_2^2 = (x^{(i+1)} - x^{(i)})^T (x^{(i+1)} - x^{(i)})$:

$$= 2(x^{(i)} - x^{(i+1)}) + 2(x^{(i)} - x^{(i-1)})$$

3. **Łącząc gradienty:**

$$\nabla F(x^{(i)}) = \lambda_1 \sum_{j=1}^{k} -\frac{2(x^{(i)} - r^{(j)})}{(\epsilon + \|x^{(i)} - r^{(j)}\|_2^2)^2} + \lambda_2 \left(2(x^{(i)} - x^{(i+1)}) + 2(x^{(i)} - x^{(i-1)}) \right)$$

Finalny Gradient

$$\nabla F = \left[\frac{\partial F}{\partial x^{(0)}}, \frac{\partial F}{\partial x^{(1)}}, \dots, \frac{\partial F}{\partial x^{(n)}} \right]$$

Gdzie:

$$\frac{\partial F}{\partial x^{(i)}} = \lambda_1 \sum_{j=1}^{k} -\frac{2(x^{(i)} - r^{(j)})}{(\epsilon + \|x^{(i)} - r^{(j)}\|_2^2)^2} + \lambda_2 \left(2(x^{(i)} - x^{(i+1)}) + 2(x^{(i)} - x^{(i-1)}) \right)$$

3 Zadanie 3 - Algorytm największego spadku z przeszukiwaniem liniowym

Krok 1: Inicjalizacja

Algorytm generuje początkową ścieżkę złożoną z losowo wygenerowanych punktów $x^{(i)}$ na mapie

Krok 2: Obliczanie gradientu

Przy każdej iteracji algorytmu optymalizacji obliczany jest gradient funkcji celu dla każdego punktu. Znając gradient funkcji celu wiemy w którą stronę przesunąć naszą funkcję aby osiągnęła mniejszą wartość.

$$\nabla F(x) = \left[\frac{\partial F}{\partial x^{(0)}}, \frac{\partial F}{\partial x^{(1)}}, \ldots, \frac{\partial F}{\partial x^{(n)}}\right]$$

Gdzie:

$$\frac{\partial F}{\partial x^{(i)}} = \lambda_1 \sum_{j=1}^{k} -\frac{2(x^{(i)} - r^{(j)})}{(\epsilon + \|x^{(i)} - r^{(j)}\|_2^2)^2} + \lambda_2 \left(2(x^{(i)} - x^{(i+1)}) + 2(x^{(i)} - x^{(i-1)})\right)$$

Krok 3: Przeszukiwanie liniowe z użyciem złotego podziału

Definiujemy funkcję $\phi(\alpha) = F(x - \alpha \nabla F(x))$.
Minimalizacja jej wyniku pozwoli nam na określenie, jak daleko należy przesunąć nasz punkt (obliczyć skalar α). Używamy metody złotego podziału w celu znalezienia minimum funkcji ϕ.

Metoda złotego podziału

Wybierz początkowy przedział $[a, b]$ dla α. Iteracyjnie zawężaj przedział według metody złotego podziału:

$$\phi = \frac{1 + \sqrt{5}}{2}$$
$$c = b - \frac{b - a}{\phi}$$
$$d = a + \frac{b - a}{\phi}$$

Oblicz wartości funkcji:

$$f(c), \quad f(d)$$

Aktualizuj przedział:

- Jeśli $f(c) < f(d)$, nowy przedział to $[a, d]$.
- Jeśli $f(c) \geqslant f(d)$, nowy przedział to $[c, b]$.

Powtarzaj, aż przedział będzie wystarczająco mały.

Krok 4: Aktualizacja punktu

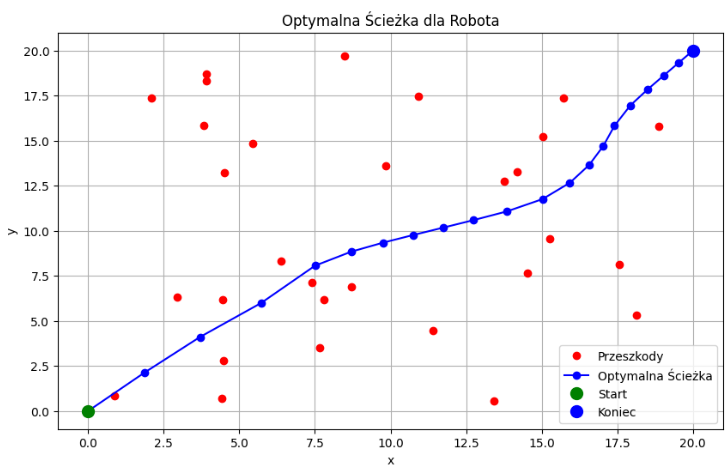
$$x_{new} = x - \alpha \nabla F(x)$$

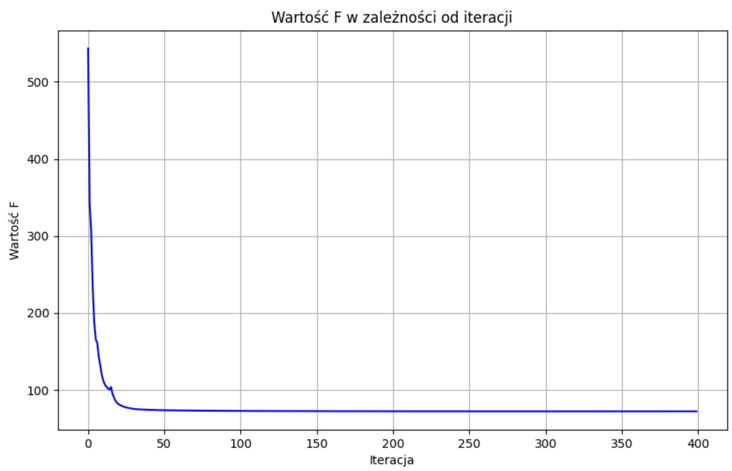
gdzie α to optymalny krok wyznaczony w poprzednim kroku.

Zadanie 4 - znalezienie optymalnej ścieżki dla robota

Uruchamiamy nasz algorytm dla poniższych wartości
n = 20 - liczba punktów ścieżki
k = 30 - liczba przeszkód
x_0 = (0,0) - punkt startowy
x_n = (20, 20) - punkt końcowy
r = np.random.uniform(0, 20, (k, 2)) - wygenerowane pozycje przeszkód
$\lambda_1 = 1$ - waga pierwszej części funkcji celu
$\lambda_2 = 1$ - waga drugiej części funkcji celu
$\epsilon = 1 \times 10^{-13}$- wartość zapobiegająca dzieleniu przez zero
Liczba iteracji = 400

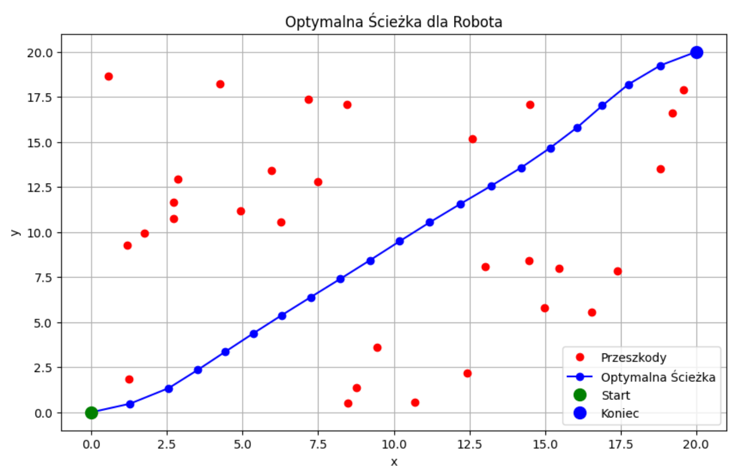
Wywołanie 1

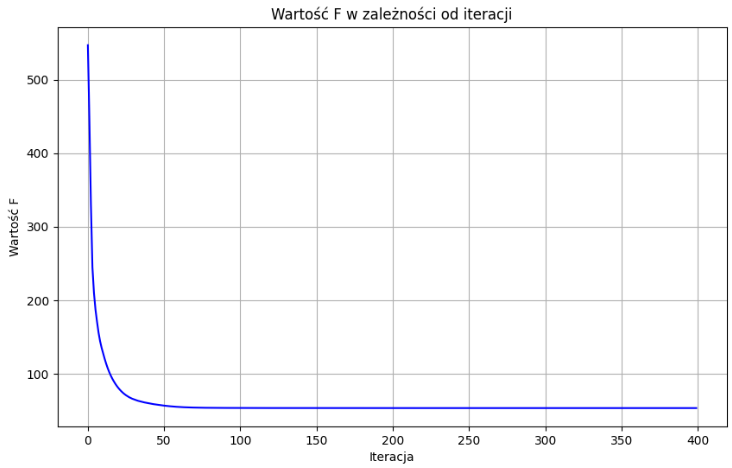




Iteracja	**Wartość funkcji F**
0	543.38
10	111.66
50	73.69
100	72.76
400	72.24

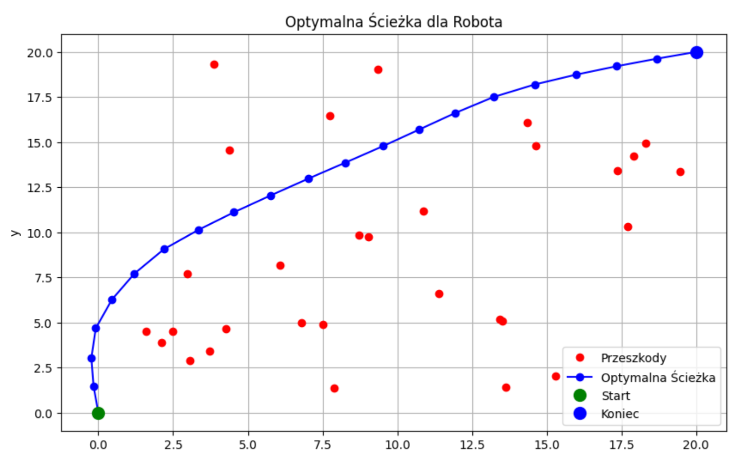
Wywołanie 2

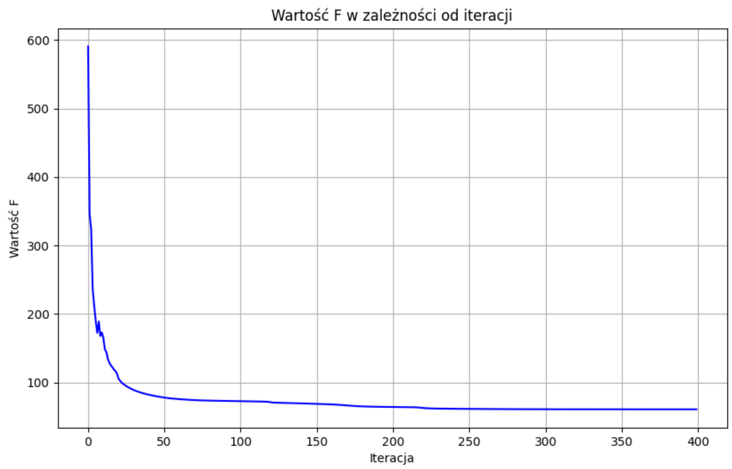




Iteracja	Wartość funkcji F
0	546.87
10	127.94
50	56.73
100	53.53
400	53.39

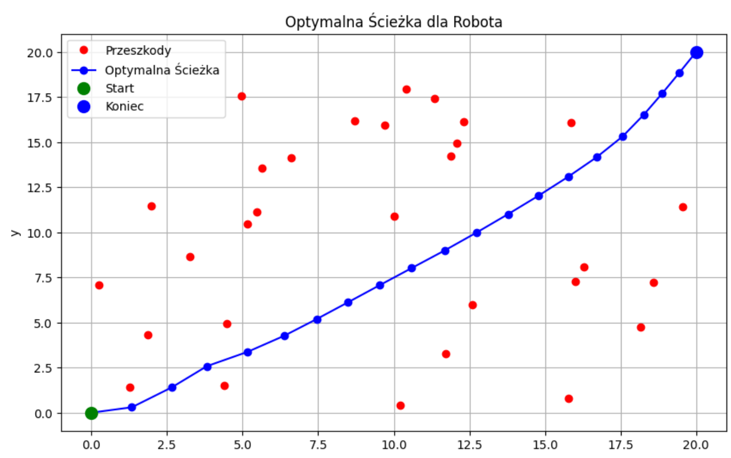
Wywołanie 3

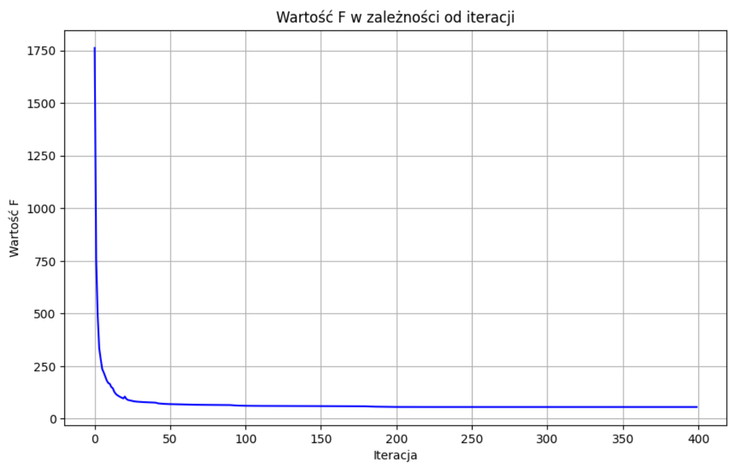




Iteracja	**Wartość funkcji F**
0	590.6506235516267
10	165.01400211803772
50	78.04074862722149
100	72.8280300065097
400	60.9391683488607

Wywołanie 4

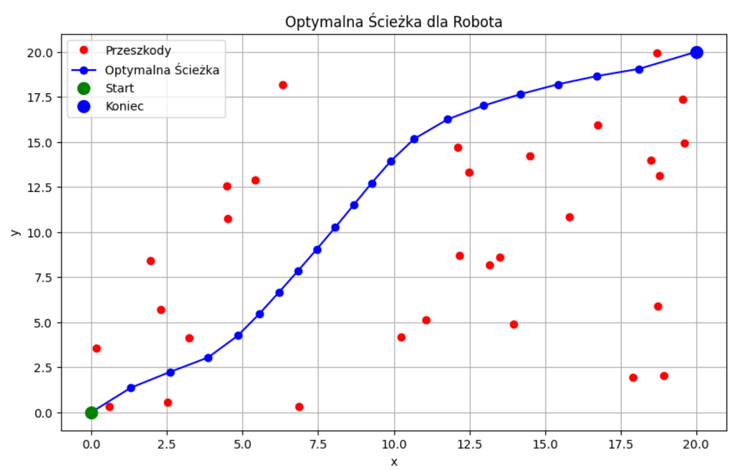


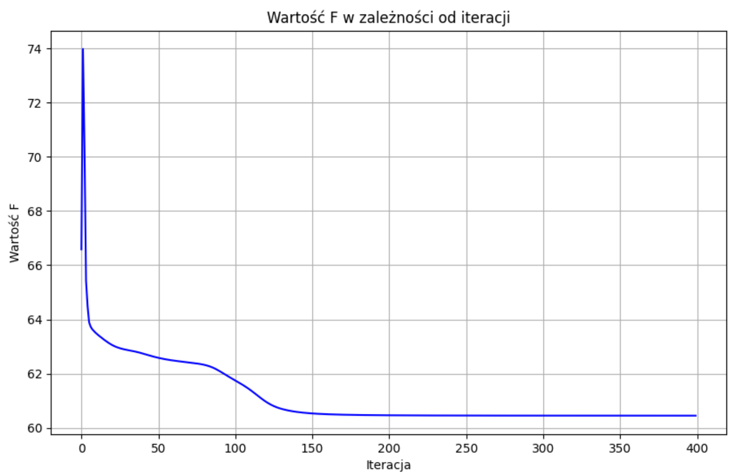


Iteracja	Wartość funkcji F
0	1762.54
10	166.02
50	69.54
100	61.73
400	55.87

Wywołanie 5

W wywołaniu piątym naszymi punktami startowymi x_i są punkty ustawione w linii prostej



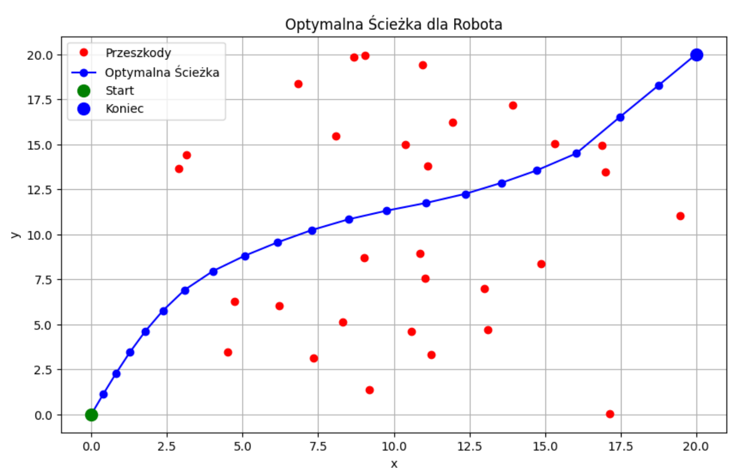


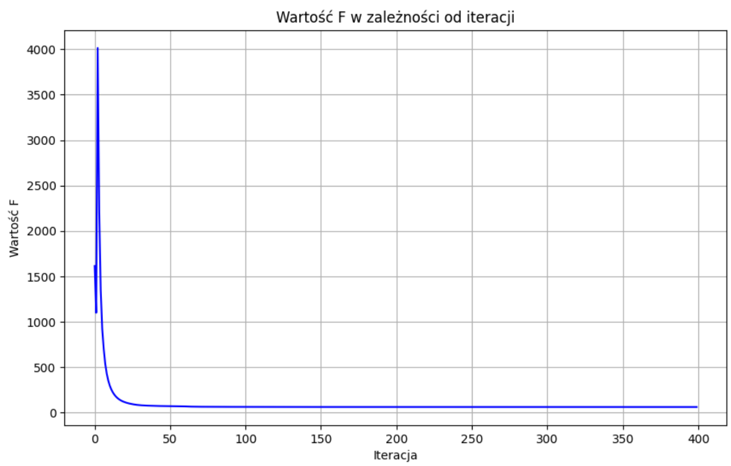
Iteracja	**Wartość funkcji** F
0	66.58
10	63.46
50	62.58
100	61.74
400	60.45

4 Testowanie dla stałej wartości skalara

Testowanie dla skalara $\alpha = 0.05$

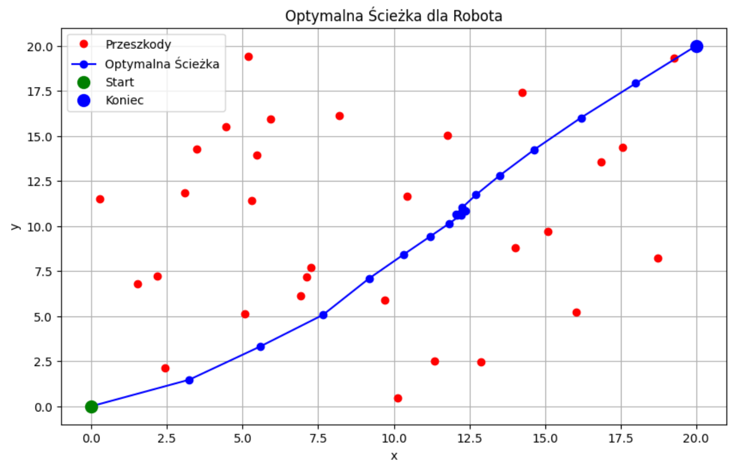
Zamiast złotego podziału użyliśmy wartości stałej skalara

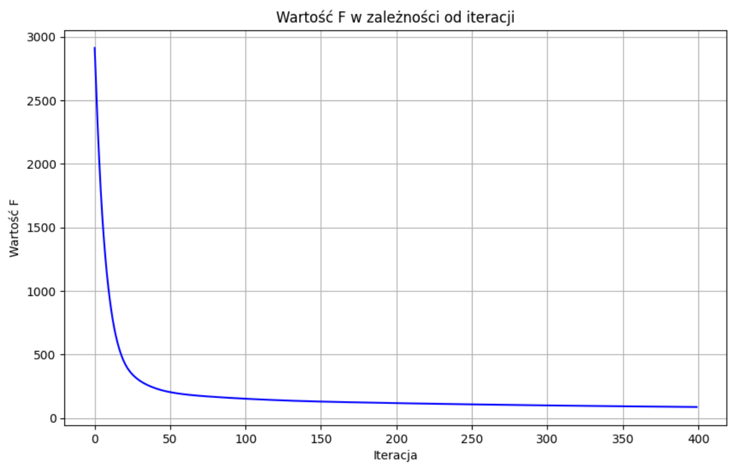




Iteracja	Wartość funkcji F
0	1615.00
10	298.50
50	72.21
100	64.26
400	63.14

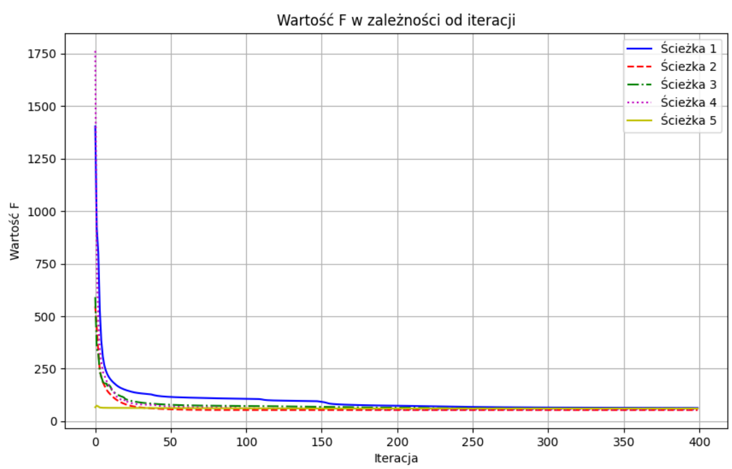
Testowanie dla skalara $\alpha = 0.01$



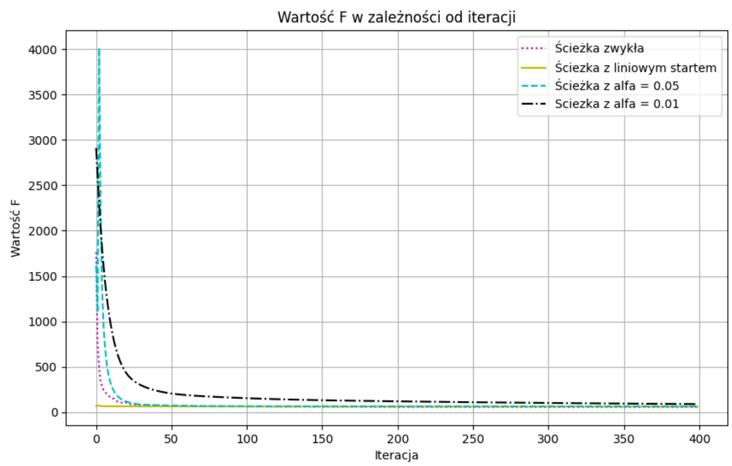


Iteracja	Wartość funkcji F
0	2911.95
10	935.21
50	204.73
100	152.77
400	87.89

5 Podsumowanie



Wykres 2: Porównanie wartości funkcji F pierwszych 5 wywołań dla skalara obliczanego przy pomocy złotego podziału



Wykres 3: Porównanie wartości funkcji F wywołań z losowymi punktami startowymi, linią prostą jako punkty startowe oraz skalara = 0.05 i 0.01

Jak widać skalara równy 0.01 radzi sobie dla iteracji od 10 do 50 gorzej od inny obranych sposobów

6 Wnioski

- Gradient funkcji celu pozwala nam na określenie kierunku, w którym funkcja rośnie najbardziej, przez co wiemy, w którym kierunku powinniśmy zmienić położenie punktu, aby zmniejszyć wartość funkcji celu.
- Przeszukiwanie liniowe jest techniką używaną do znalezienia optymalnej wartości skalarnej (kroku α) wzdłuż danego kierunku. Dzięki przeszukiwaniu liniowemu możemy precyzyjnie dostosować krok, aby zmniejszyć funkcję celu wzdłuż kierunku przeciwnym do gradientu.
- Metoda złotego podziału (golden section search) jest efektywną metodą przeszukiwania liniowego, która minimalizuje liczbę ewaluacji funkcji celu potrzebnych do znalezienia optymalnej wartości α. Wykorzystuje ona proporcje złotego podziału do zawężania przedziału, w którym znajduje się minimum, co przyspiesza proces optymalizacji.
- Wyznaczanie skalara przy pomocy metody złotego podziału pozwala nam na szybsza optymalizację dla małej ilości iteracji w porównaniu do stałego skalara
- Punkty startowe ułożone w linię początkowo dają lepszy rezultat. Różnica następnie zmniejsza się przy większej ilości iteracji