WSTĘP DO SZTUCZNEJ INTELIGENCJI

Ćwiczenie 1 – Gradient Prosty

MICHAŁ MIZIA 331407

SPIS TREŚCI

Wstęp

Treść zadania:

Znana jest funkcja celu:

1. F(x) = 7x + 4sin(x)
2. G(x) = $\frac{x*y}{e^{x^2+y^2}}$

Zaimplementować metodę gradientu prostego opisaną na wykładzie. Użyć zaimplementowany algorytm do wyznaczenia ekstremów funkcji.

Zbadać wpływ następujących parametrów na proces optymalizacji:

- długość kroku uczącego

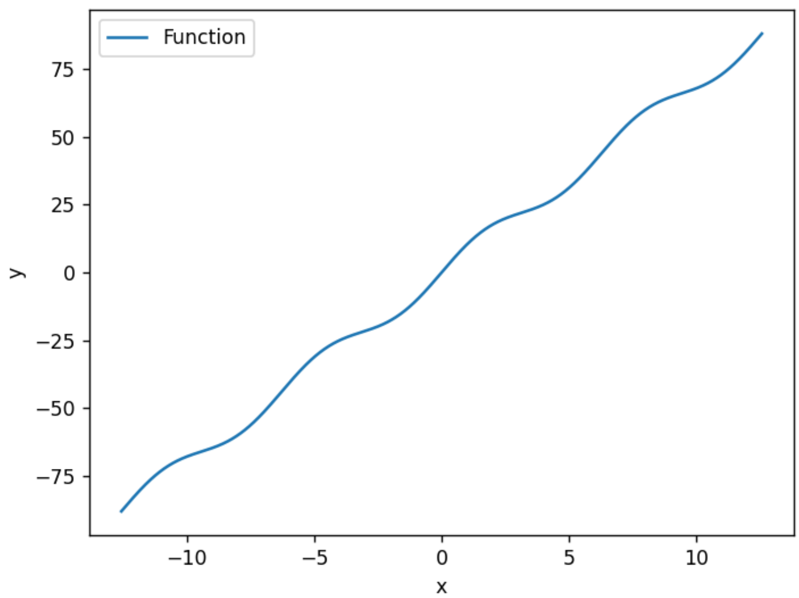
- limit maksymalnej liczby kroków algorytmu

- rozmieszczenie punktu startowego.

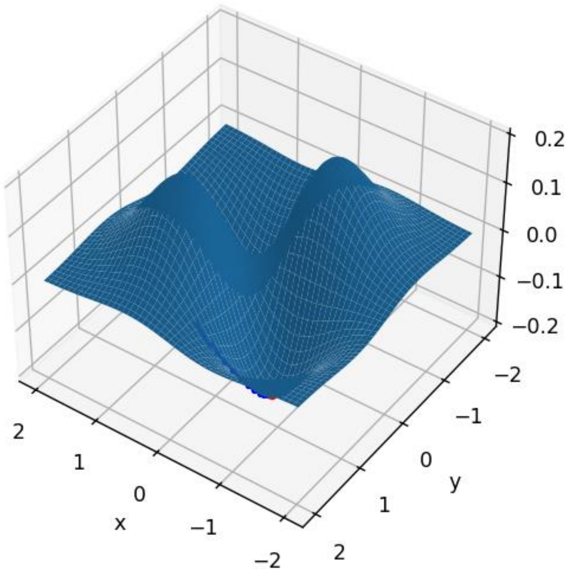
Zinterpretować wyniki w kontekście kształtu badanej funkcji.

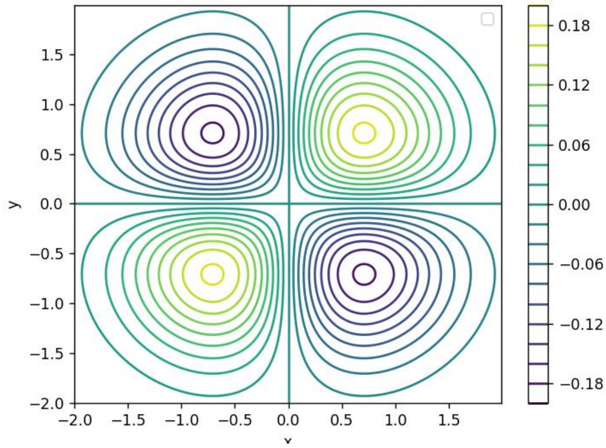
Działanie funkcji zostało przedstawione w pliku README.md oraz w konkretnych plikach py. W sprawozdaniu skupię się na badaniu wpływu parametrów na działanie gradientu.

Funkcja F okazuje się być zwykłą funkcją rosnącą połączona z oscylacją sinusa



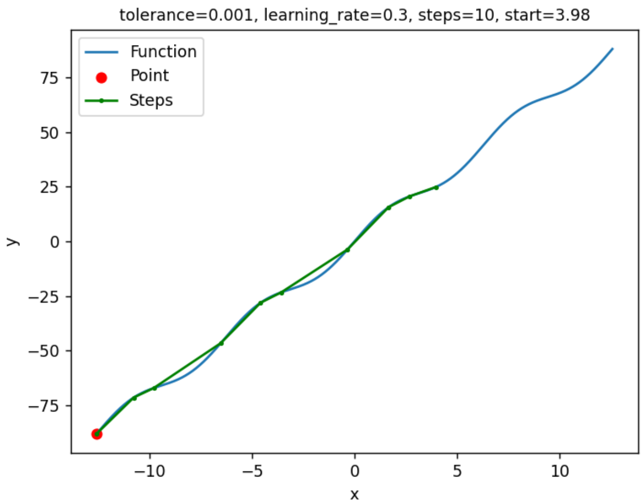
Z kolei funkcja G ma bardziej skomplikowany kształt zawierający 2 lokalne maksima oraz minima. Poniżej przedstawiono wizualizację 3d oraz wykres topograficzny.

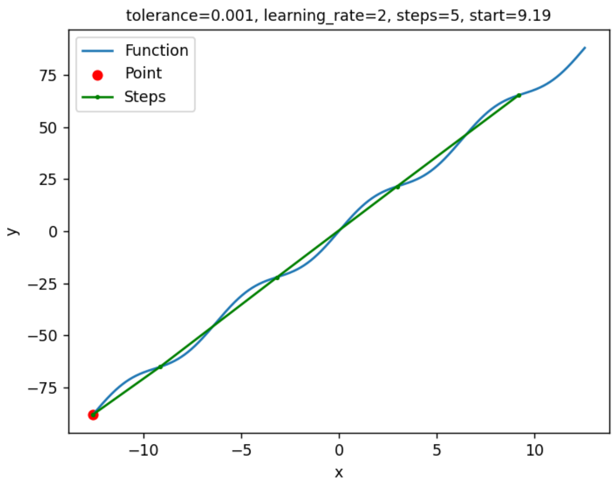




Długość kroku uczącego

Dla funkcji F długość kroku zmienia jedynie to w ilu krokach dojdzie ona do minimum ponieważ stale rosnąca, dla funkcji o innych parametrach zachowanie to byłoby inne ale w tym przypadku z każdego punktu startowego dojdziemy do minimum.

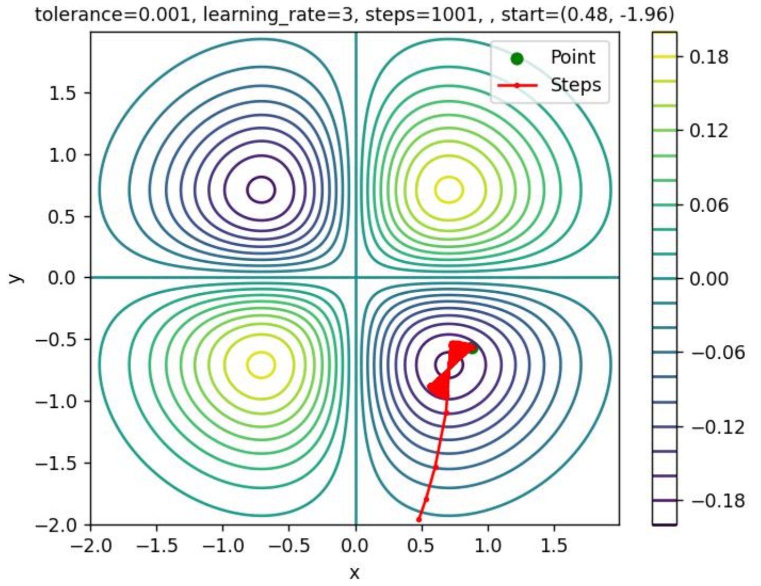




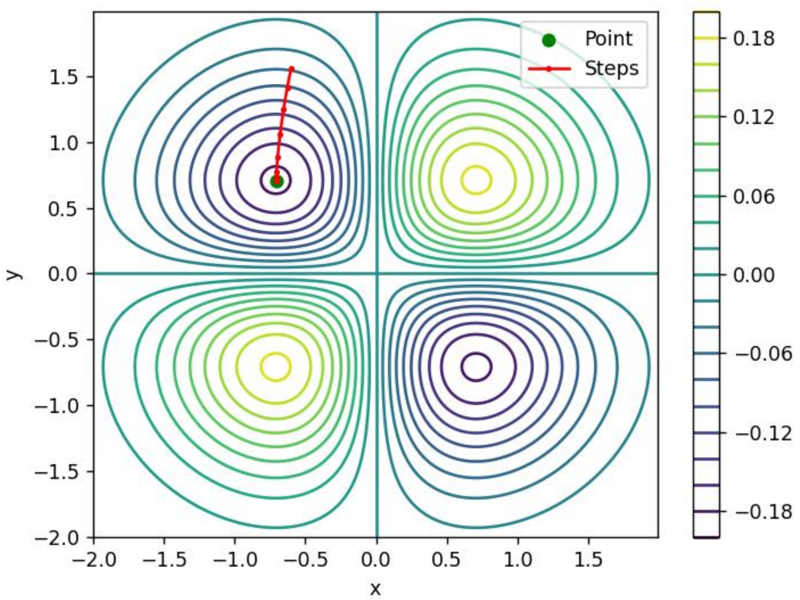
$f(x) = 7x + 4\sin(x)$

Learning Rate	Start Point	End Point	Steps
0.01	-1.68	[-12.56637061]	202
0.51	-1.68	[-12.56637061]	5
1.01	-1.68	[-12.56637061]	3
1.51	-1.68	[-12.56637061]	3
2.01	-1.68	[-12.56637061]	2
2.51	-1.68	[-12.56637061]	2
3.01	-1.68	[-12.56637061]	2

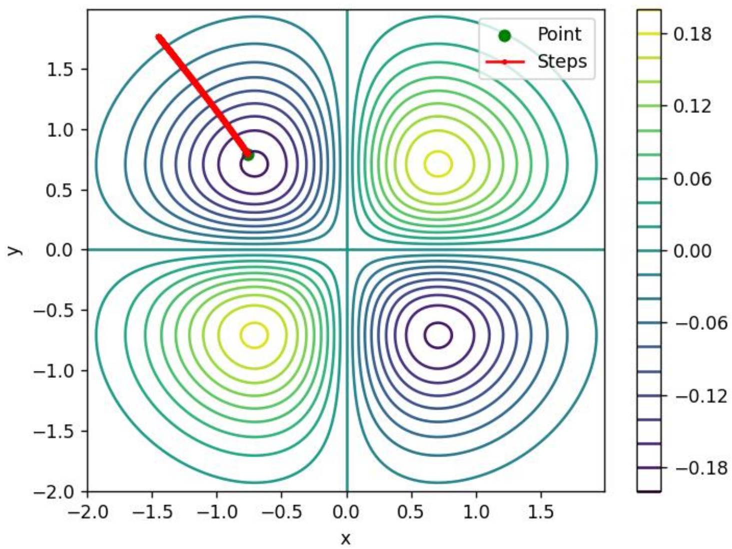
Sytuacja jest ciekawsza dla funkcji G, w przypadku bardzo dużych wartości kroku gradient zapętla się w okolicy minimum co widać na wykresie gdzie odbyło się 1001 kroków czyli maksymalna liczba dla algorytmu podawana jako parametr funkcji.



Dla wartości 1, algorytm dociera do najbliższego minimum szybko i zatrzymuje się tam.



Dla wartości jeszcze mniejszych jak 0,01 algorytm zbliża się jeszcze wolniej czasem nie mieszcząc się w limicie kroków.

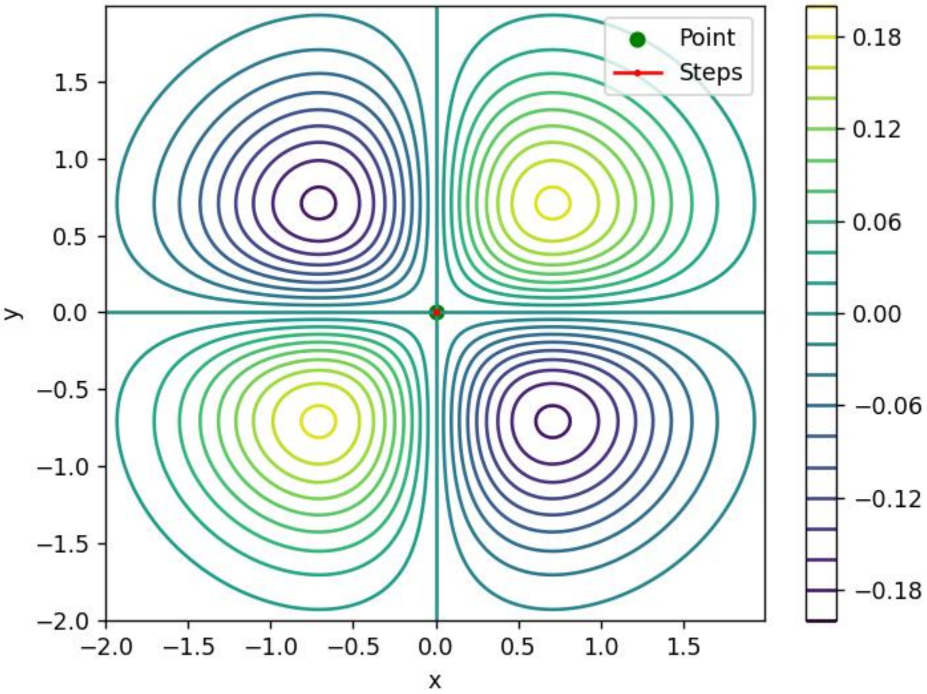


Można zatem wnioskować że istnieje optymalna wartość parametru uczącego dla każdej funkcji.

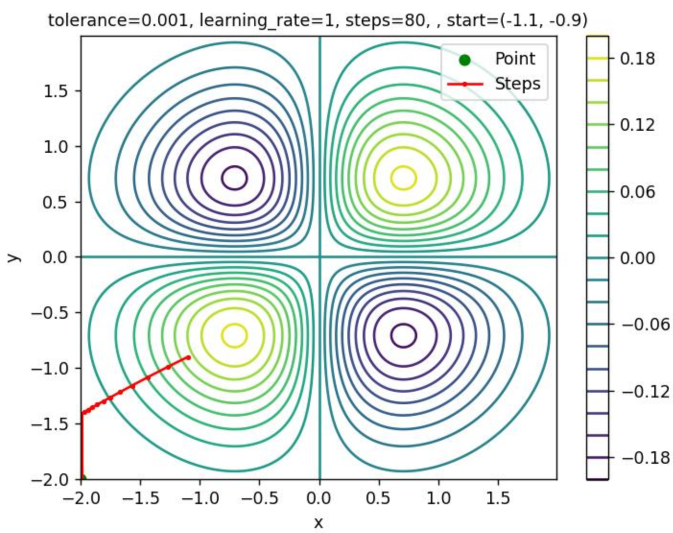
Learning Rate	Start Point	End Point	Steps
0.01	(-0.55, -0.5)	[-0.70533661 0.70487548]	1001
0.51	(-0.55, -0.5)	[-0.70631741 0.70627031]	21
1.01	(-0.55, -0.5)	[-0.70655107 0.70655108]	10
1.51	(-0.55, -0.5)	[-0.70640365 0.70634931]	7
2.01	(-0.55, -0.5)	[-0.70719235 0.70829982]	11
2.51	(-0.55, -0.5)	[-0.70621623 0.70809835]	31
3.01	(-0.55, -0.5)	[-0.88346458 0.88323839]	1001

Rozmieszczenie punktu startowego

Algorytm gradientu prostego jest wadliwy w przypadku punktów z zerowym gradientem czyli tak zwanych „saddle point”. Wtedy algorytm zakończy swoje działanie bez ruchu np. jak dla punktu (0,0)



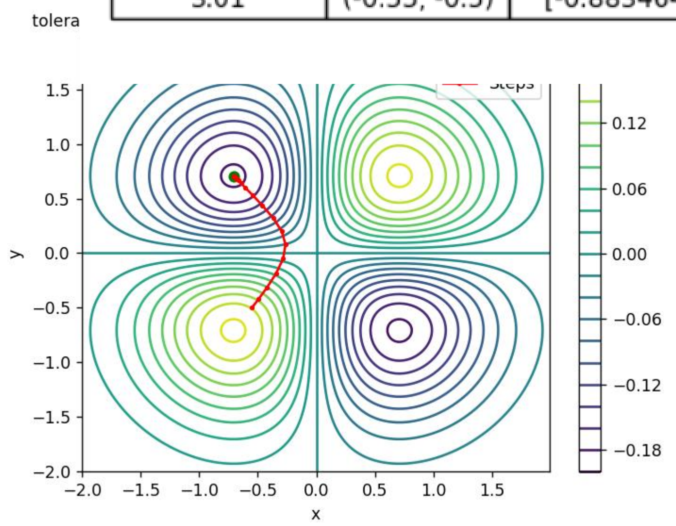
Dla większości punktów algorytm stopniowo zejdzie do minimum jednak gdy punkt startowy pojawi się np. w 3 ćwiartce układu na lewo i w dół od lokalnego maksimum, algorytm dojdzie do końca dziedziny i tam zatrzyma się ze względu na implementację zatrzymania punktu wewnątrz dziedziny (clamping).

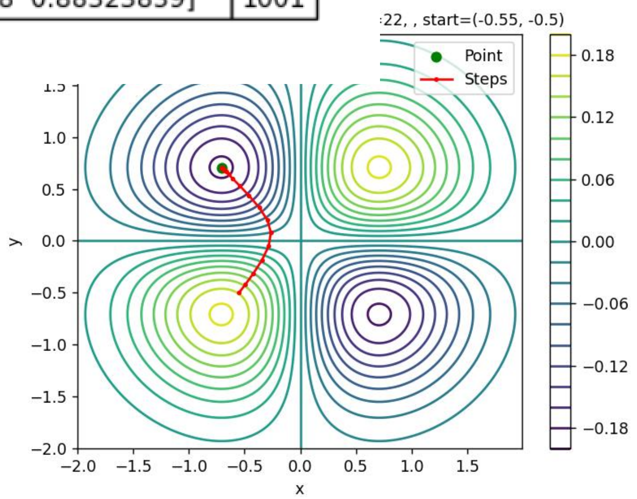


Learning Rate	Start Point	End Point	Steps
0.01	(-0.55, -0.5)	[-0.70533661 0.70487548]	1001
0.51	(-0.55, -0.5)	[-0.70709964 0.70709921]	31
1.01	(-0.55, -0.5)	[-0.70709738 0.70709738]	13
1.51	(-0.55, -0.5)	[-0.70709808 0.7070974]	9
2.01	(-0.55, -0.5)	[-0.70710629 0.7070999]	18
2.51	(-0.55, -0.5)	[-0.70709832 0.70711618]	59
3.01	(-0.55, -0.5)	[-0.88346458 0.88323839]	1001

Tolerancja algorytmu

Zmniejszenie tolerancji algorytmu będzie miało bezpośredni wpływ na zwiększenie dokładności jednak będzie się to odbywało kosztem większej ilości iteracji. Przy starcie z punktu (0.55, 0,5) zwiększenie tolerancji z 0.0001 na 0.001 przy zmniejszyło ilość kroków z 27 na 22.





Tak wygląda tabela zależności tolerancji w tym punkcie z ilością kroków, długoś kroku uczącego to 0.5

Z tego co udało mi się zaobserwować, tolerancja nie miała wpływu na optymalny krok uczący, dalej oscylację pojawiały się w okolicy wartości 3. Wykres kroku uczącego przy tolerancji zmniejszonej do tol = 0.00001

Tolerance	Start Point	End Point	Steps
1e-05	(-0.55, -0.5)	[-0.70710056 0.70710017]	32
0.0001	(-0.55, -0.5)	[-0.70704512 0.70704129]	27
0.001	(-0.55, -0.5)	[-0.70649544 0.70645741]	22
0.01	(-0.55, -0.5)	[-0.70101298 0.70063134]	17