Raport: Analiza zbieżności metody gradientu prostego

Grzegorz Topór

1 Opis implementowanego algorytmu

Zaimplementowano algorytm gradientu prostego (ang. *Gradient Descent*). Jest to iteracyjna metoda optymalizacji pierwszego rzędu, służąca do znajdowania lokalnego minimum różniczkowalnej funkcji celu. Reguła aktualizacji parametrów w kroku $t+1$ zdefiniowana jest wzorem:

$$x_{t+1} = x_t - \alpha \nabla f(x_t)$$

gdzie:

- $x_t \in \mathbb{R}^n$ – wektor parametrów w iteracji t,
- α – współczynnik uczenia (krok, ang. *learning rate*),
- $\nabla f(x_t)$ – gradient funkcji celu w punkcie x_t.

Algorytm działa iteracyjnie, aż do spełnienia jednego z warunków stopu. Zaimplementowano następujące kryteria zatrzymania:

- osiągnięcie maksymalnej liczby iteracji,
- brak poprawy wartości funkcji celu przez zadaną liczbę kroków,
- wykrycie niestabilności numerycznej (wartości `NaN`, `Inf` lub przekroczenie zakresu zmiennoprzecinkowego).

Algorithm 1 Pseudokod Metody Gradientu Prostego z zabezpieczeniem numerycznym

```
function SOLVER(x_0, α, max_iter)
    x ← x_0
    best_val ← f(x)
    for t ← 0 to max_iter do
        g ← ∇f(x)                                ▷ Obliczenie gradientu
        x_new ← x − α · g                        ▷ Krok aktualizacji
        val ← f(x_new)
        if val is NaN or val is Inf then
            Break                                ▷ Przerwanie w przypadku eksplozji gradientu
        end if
        if val < best_val then
            best_val ← val
            Zresetuj licznik patience
        end if
        x ← x_new
    end for
    return Historia wartości funkcji celu
end function
```

2 Eksperymenty numeryczne

2.1 Dane i metodyka

Badaniu poddano trzy funkcje testowe optymalizacji w przestrzeni $n = 10$:

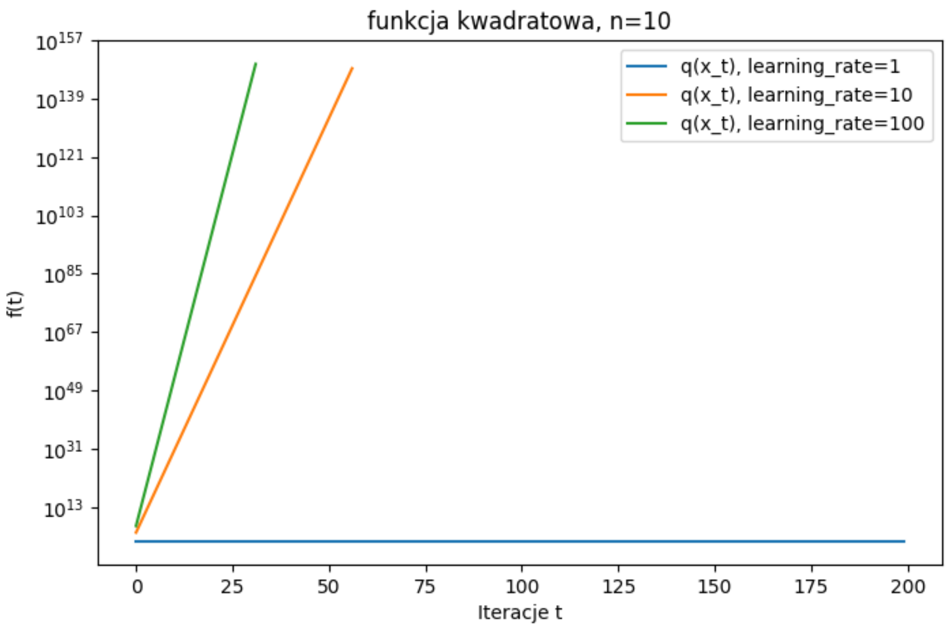
1. **Funkcja kwadratowa**
2. **Funkcja Rosenbrocka**
3. **Funkcja Ackleya**

Jako dziedzinę przyjęto zbiór $[-10, 10]^{10} \subset \mathbb{R}^{10}$. Punkt startowy x_0 został wylosowany z rozkładu jednostajnego wewnątrz dziedziny. Celem eksperymentu było zbadanie wpływu dużych wartości parametru kroku na zbieżność. Przetestowano parametr:

$$\alpha \in \{1, 10, 100\}$$

3 Wyniki

3.1 Funkcja Kwadratowa

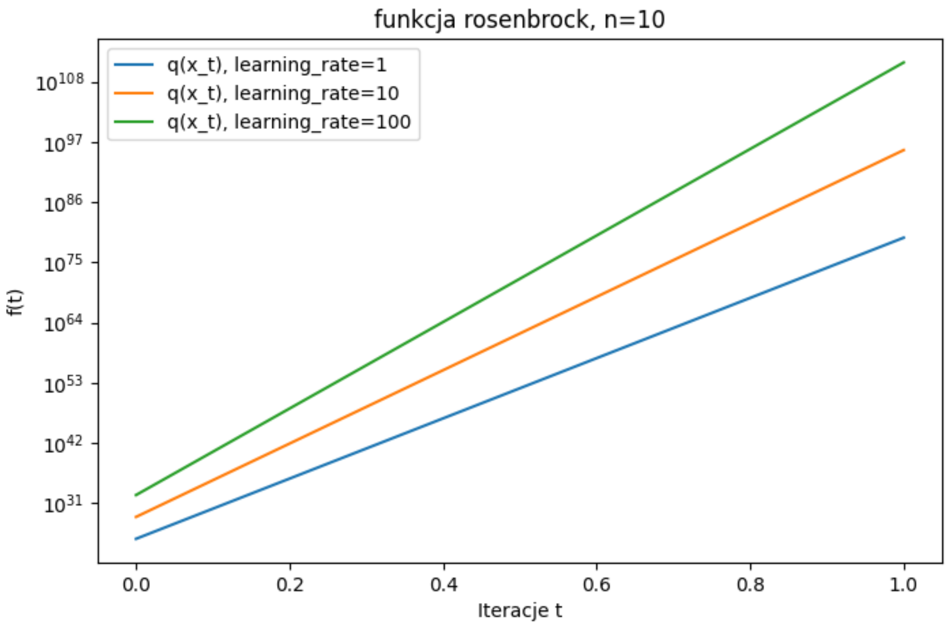


Rysunek 1: Dywergencja funkcji kwadratowej

Dla wartości $\alpha = 10$ i $\alpha = 100$ algorytm uległ dywergencji. Ze względu na brak ograniczeń funkcji kwadratowej, wartości błędu rosną wykładniczo (liniowy wzrost na skali logarytmicznej).

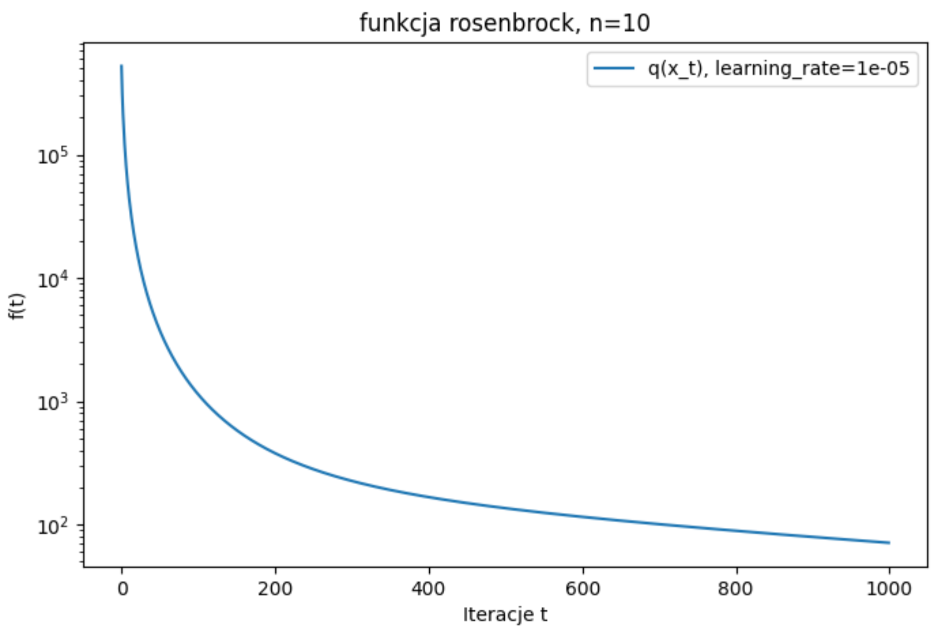
- Dla $\alpha = 100$: Natychmiastowe osiągnięcie progu 10^{150} (ok. 30 iteracji).
- Dla $\alpha = 10$: Prawie natychmiastowe osiągnięcie progu 10^{150} (ok. 50 iteracji).
- Dla $\alpha = 1$: Krok sprawia, że algorytm nie jest zbieżny – oscyluje dookoła jednej wartości.

3.2 Funkcja Rosenbrocka

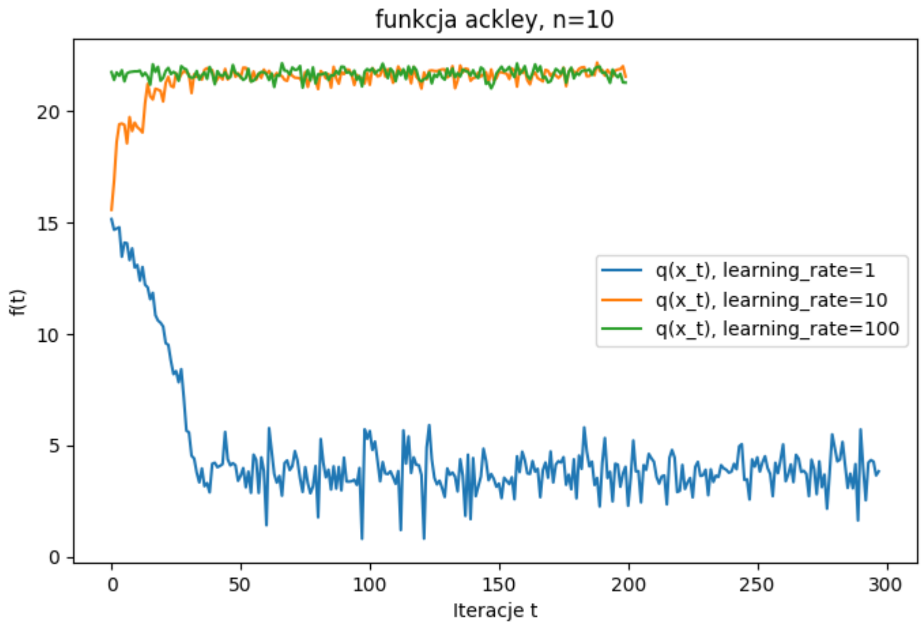


Rysunek 2: Natychmiastowa eksplozja numeryczna dla f. Rosenbrocka.

Funkcja Rosenbrocka w $n = 10$ charakteryzuje się bardzo stromymi zboczami (czwarta potęga zmiennych). Zastosowanie $\alpha \geq 1$ prowadzi do natychmiastowej niestabilności numerycznej. Wykres urywa się po 1. iteracji, ponieważ już w pierwszym kroku wartości funkcji przekraczają zakres precyzji zmiennoprzecinkowej (osiągając rzędy wielkości $10^{30} - 10^{108}$), co wymusiło awaryjne zatrzymanie solvera. Aby przetestować zbieżność algorytmu, α powinna być znacząco mniejsza od 1, np: $\alpha = 0.00001$



3.3 Funkcja Ackleya



Rysunek 3: Stagnacja funkcji Ackleya na plateau.

W przeciwieństwie do funkcji wielomianowych, funkcja Ackleya jest ograniczona (asymptotycznie dąży do stałej wartości przy dużych x).

- Duży krok ($\alpha = 10, 100$) wyrzucił punkt w obszar płaski (plateau), gdzie wartość funkcji wynosi ok. 20-22.
- Algorytm utknął w lokalnym minimum lub na płaskim obszarze z powodu zaniku gradientu ($\nabla f(x) \approx 0$).
- Dla $\alpha = 1$ widoczna jest próba optymalizacji, jednak parametr jest nadal zbyt duży, by osiągnąć zbieżność do minimum globalnego ($f(x) = 0$).

4 Wnioski

1. Badane wartości współczynnika uczenia $\alpha \in \{1, 10, 100\}$ okazały się **zbyt duże** dla problemu optymalizacji w przestrzeni 10-wymiarowej.
2. Metoda gradientu prostego jest wysoce wrażliwa na dobór parametru kroku. Przekroczenie krytycznej wartości prowadzi do niestabilności numerycznej (funkcja Rosenbrocka) lub utknięcia w obszarach zerowego gradientu (funkcja Ackleya).
3. Aby skutecznie zbadać zbieżność, należałoby rozpatrywać parametry α mniejsze od 1.