# APL 5

Anurag Manoj ee22b003

October 2023

## 1 Introduction

The gradient descent algorithm has been adapted separately for the single variable and double variable case.
For the single variable case, the differentiation is done numerically using complex step differentiation. For this, the single variable function has to be complex differentiable. df/dx = Im(f(x+ih))/h. This gives less numerical error in differentiation, as the error is O($h^2$). The function, "descent_steps_plot_2d" plots the gradient descent, with the early iterations dark colored and the final iterations coloured lightly. The convergence rate, defined as | f(nth x) - f(last x) | is plotted against number of iterations, with log scale on y axis. This turns out to be a linear plot in all the 4 problems. This is due to strong convexity of the functions. Thus, the convergence rate is $O((1/c)^k)$, for some c>1. The vale of c is also estimated.
For the multi-variable case, a function that returns all the points which are visited in gradient descent is there, which takes the gradient also as an argument. Another function, "descent_steps_plot_3d" plots the function and gradient descent, with dark color representing early steps and light color representing later points. The convergence is also plotted similar to the single variable case.

## 2 "P1"

The below plot uses xi=-2.2. step=0.3 The algorithm converges as:

$$O((1/6.44)^k)$$

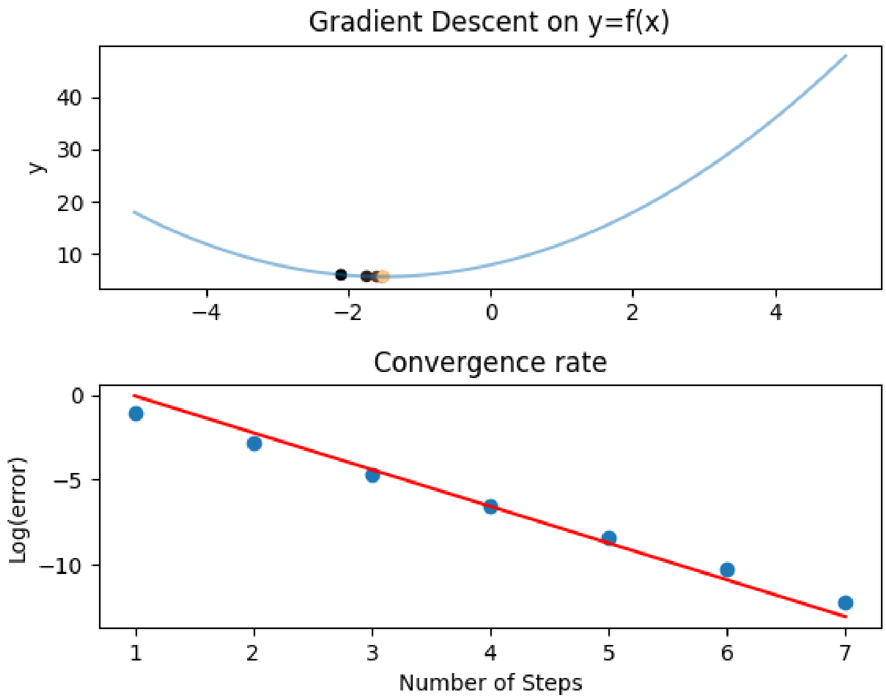


Figure 1:

# 3 "P2"

The starting point is 1,5 and step size is 0.05. 3000 iterations are done, the final value of (x,y) at which minimum is attained is (4,2). The convergence is slow as a very small step size was taken. If a larger step size is taken and a further away point is chosen, the algorithm fails to converge. This is as the slope of the quartic polynomial in x can be very large, and the algorithm can oscillate to points of larger and larger magnitudes.
The algorithm converges as:

$$O((1/1.005)^k)$$

The minimum is x,y = (4,2)

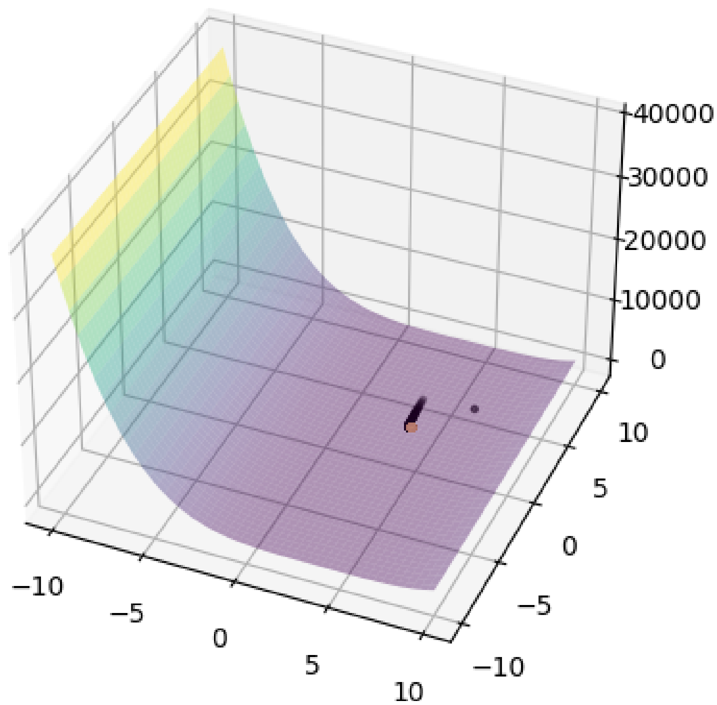


Figure 2:

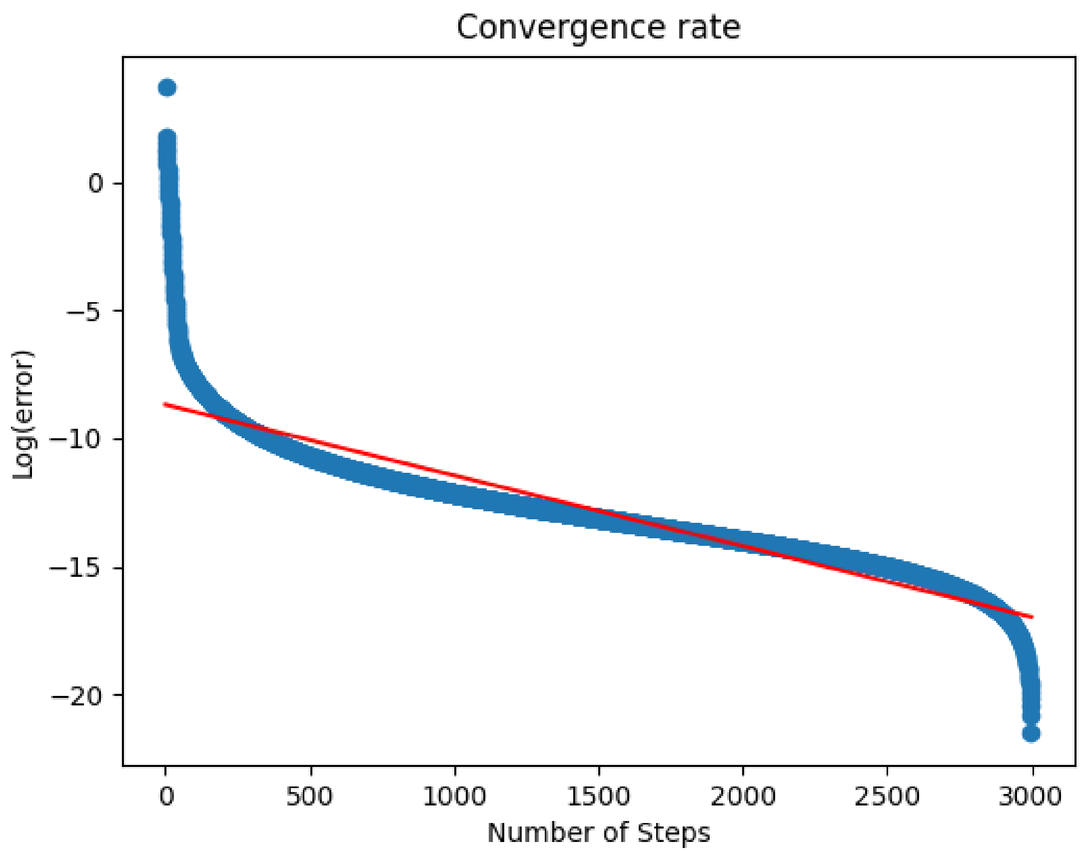


Figure 3:

## 4 "P3"

The starting point is 0.1,-0.1 and step size is 0.05, 175 iterations are done. Convergence is very slow at this starting point as the function is not concave here. The algorithm converges as:

$$O((1/1.05)^k)$$

x,y = -1.517,-1.53 at the minimum point chosen. If a point like (3,3) is chosen, the minima found is outside the given limits.

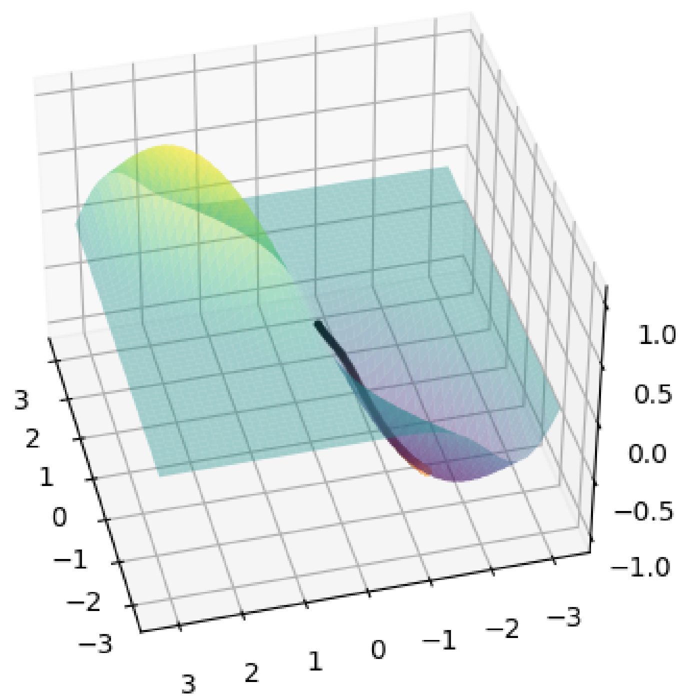


Figure 4:

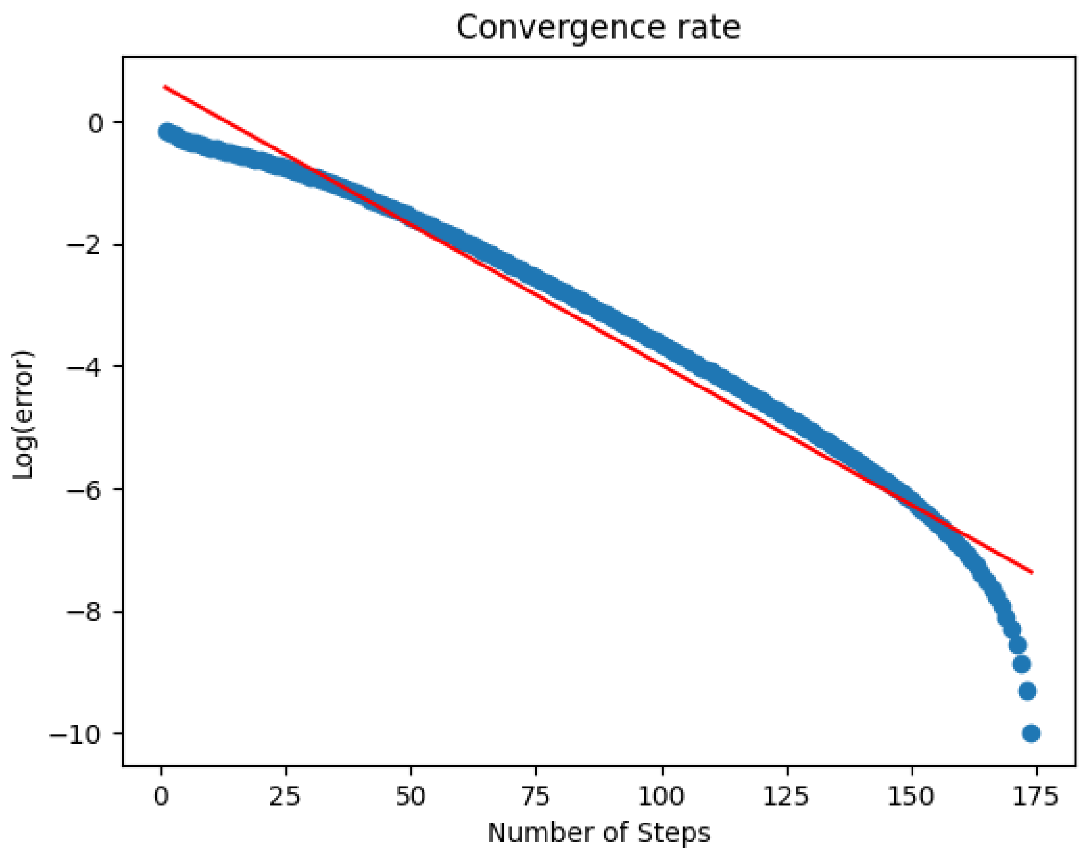


Figure 5:

## 5 "P4"

The below plot is generated with xi=4.3, step=0.3. The algorithm converges as:

$$O((1/2.68)^k)$$

X=-1.5 is the minimum

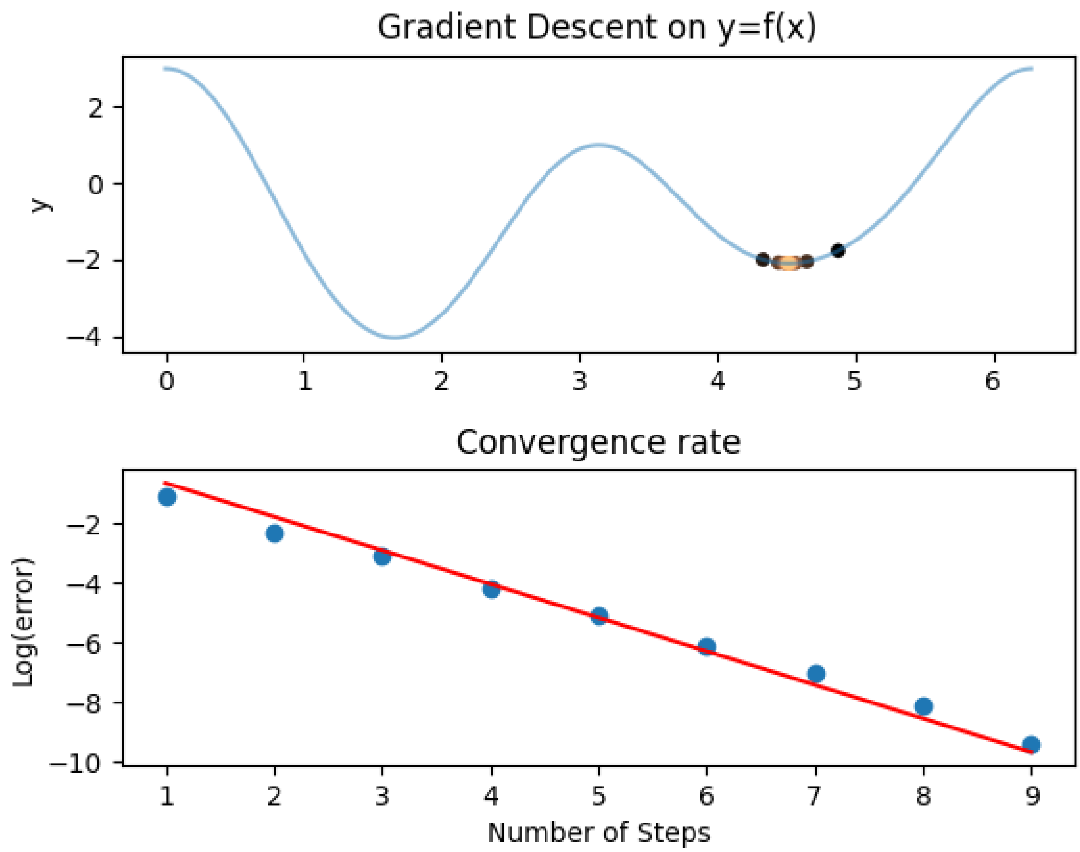


Figure 6:

The below plot is generated with xi=1.2, step=0.125. If a larger step of 0.5 is chosen, the algorithm does not converge. The algorithm converges as:

$$O((1/7.43)^k)$$

X=4.519 is the minimum

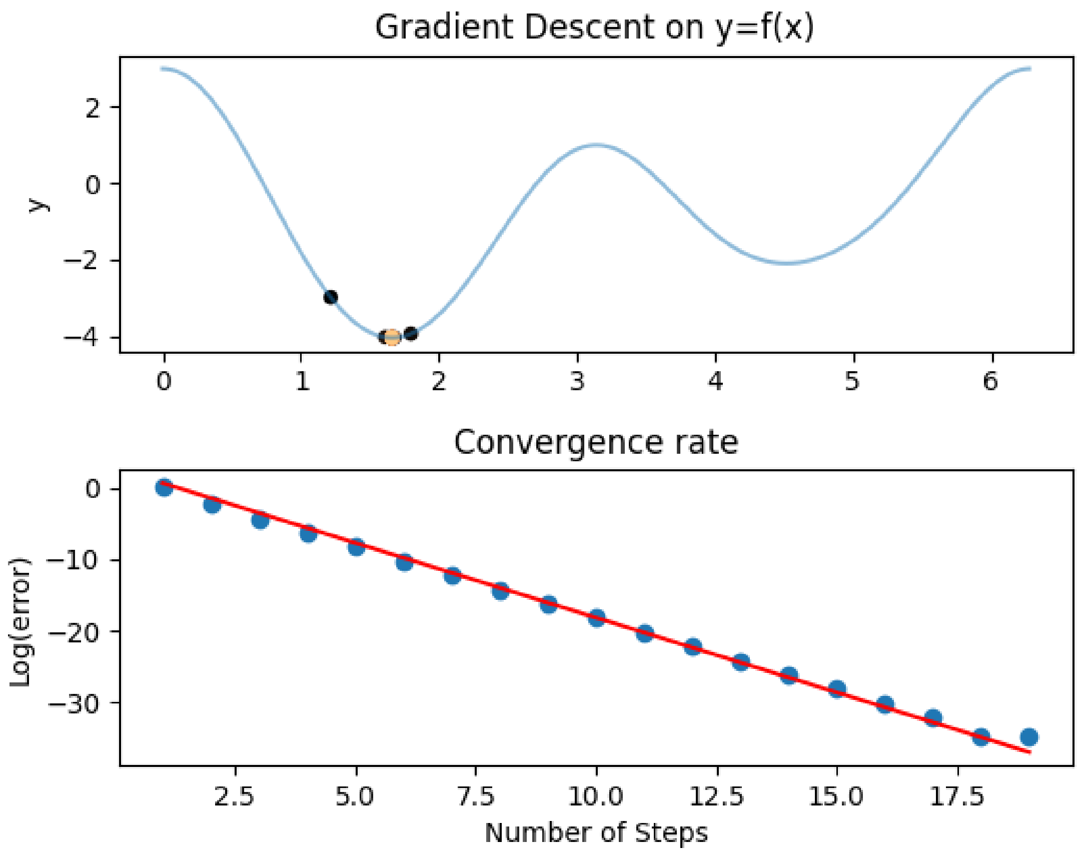
Gradient Descent on y=f(x)
y
2
0
−2
−4
0
1
2
3
4
5
6
Convergence rate
Log(error)
0
−10
−20
−30
2.5
5.0
7.5
10.0
12.5
15.0
17.5
Number of Steps

Figure 7: