DS401 - Optimization for Data Science

Assignment-2
*Assigned Date: 16/02/2023*
11:59pm, Due Date:24/02/2023

Instructions

- Work on the assignments on your own. You are free to discuss among your selves,but don't copy. If we find the assignments of a group of (two or more) students very similar, the group will get zero points towards this assignment. Plagiarism will be cheked with tools. Please use Python for writing code. You can submit the code as a Jupyter notebook.

- **Plagiarism will be checked with tools and will be penalized heavily**

1 Programming
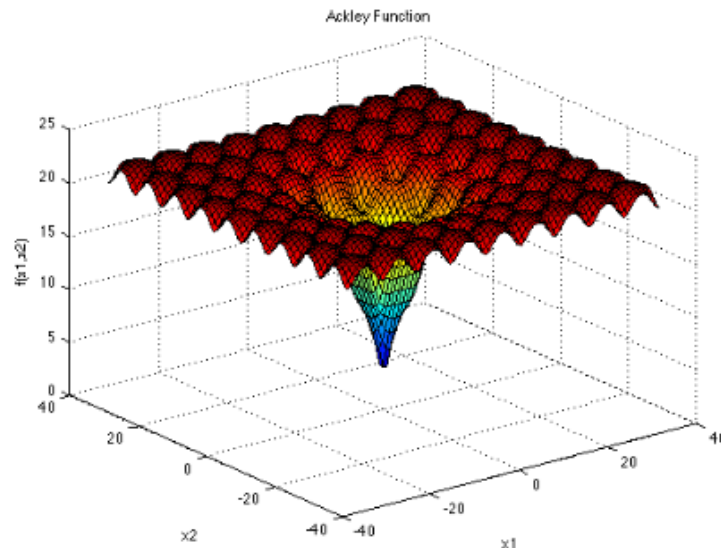
$$f(\mathbf{x}) = -a \exp\left(-b\sqrt{\frac{1}{d}\sum_{i=1}^{d} x_i^2}\right) - \exp\left(\frac{1}{d}\sum_{i=1}^{d}\cos(cx_i)\right) + a + \exp(1)$$

Description:

Dimensions: d

The Ackley function is widely used for testing optimization algorithms. In its two-dimensional form, as shown in the plot above, it is characterized by a nearly flat outer region, and a large hole at the centre. The function poses a risk for optimization algorithms, particularly hillclimbing algorithms, to be trapped in one of its many local minima.

Recommended variable values are: a = 20, b = 0.2 and c = 2π.
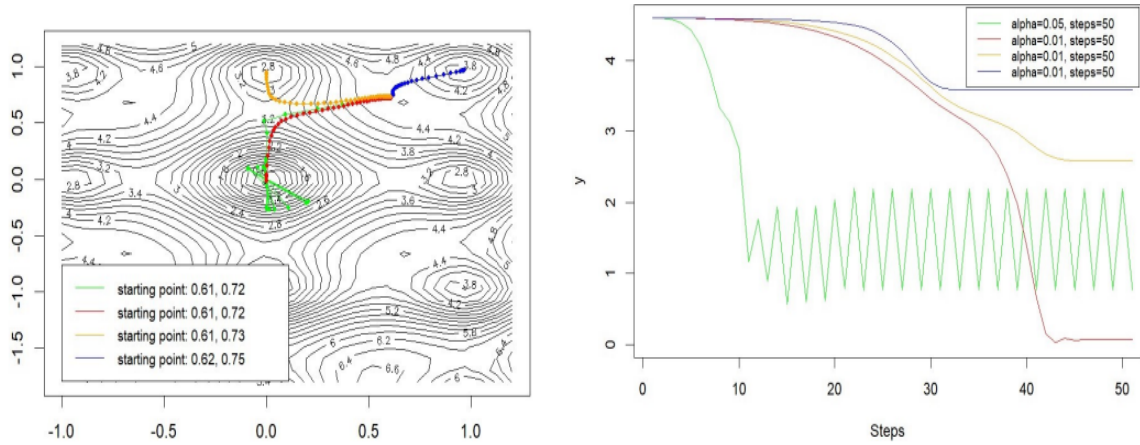
Figure 1: Figure-1

Figure 2: Figure-2

1. (20 Marks) Implement Gradient Descent (GD) Algorithm (Compute gradients using theory or Numerical Gradient) to minimize the function given below. You can use certain error tolerance/difference between x_{k+1} to x_k to stop the gradient descent

   - Experiment GD with fixed alphas and different initialization. (You are expected to replicate the results given in figure-2)

   - Implement GD with backtracking and experiment with different C_1 and ρ values.

   - Compare both.

2. (10 Marks) Run your GD algorithm implemented in previous question with and without backtracking for well and ill conditioned functions discussed in the class. Visualize contour plots and updates of independent variables. Answer the questions written in backtracking notes.

   - Try to initialize x_0 on the vector along v_{max} and v_{min} and check the convergence with different alpha values.

   - With certain value of alpha α GD with fixed α works better than GD with backtracking. Explain ?