Friedrich-Alexander-Universität Erlangen-Nürnberg
CML: Control, Machine Learning, and Numerics
Assignment 1
Due Date: 6:00pm, May 18, 2023. Assignment submission address: ycsong.math@gmail.com

1. Detect and classify (with a short justification) the stationary point of $f : \mathbf{R}^2 \to \mathbf{R}$ with $f(x_1, x_2) = x_1^2 - x_2^2$.

2. Consider the gradient descent method.
Show with the help of an example that if the step length $\alpha$ is badly chosen, the gradient descent might diverge. [**Hint:** The iteration for gradient descent is given by $x_{k+1} = x_k - \alpha \nabla f(x_k)$. Consider the 1D case with a quadratic function for instance. ]

3. In this exercise we investigate that Newton is really a *local* method and does not converge from arbitrary starting points $x_0 \in \mathbf{R}$. Let $f : \mathbf{R} \to \mathbf{R}$ with $f(x) = \sqrt{x^2 + 1}$.

   (1) Investigate for which starting points $x_0$ Newton does converge and for which starting points Newton will diverge.

   (2) What are possible procedures to globalize Newton's method and to increase the convergence radius?

4. Consider the following optimization problem

$$\min_{x_1, x_2} f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (x_1 - 1)^2$$

It is easy to see that the exact solution to the above problem is $(1, 1)^T$.

Implementing the following methods using Matlab (or other software you prefer) to solve the above optimization problem:

- Gradient descent method with constant step sizes: 0.1, 0.01, 0.001, 0.0001.
- Gradient descent method with backtracking line search.
- Classic Newton method (i.e. step size is 1).
- Newton method with backtracking line search.

Parameters setting-up: initial value $x^0 = [0, 0]^T$, maximum iteration: 50000, stoping criterion: norm of gradient $< 10^{-8}$. (You can verify your result by comparing the output with the exact solution)

Please report all the results of the four methods listed above, and discuss briefly what you can conclude from the results. (For backtracking line search, please choose the parameters carefully such that the algorithm converges as fast as possible)