Gradient Descent algorithm in Python

Optimizing a complicated function

In this exercise, we are going to implement the Gradient Descent (GD) algorithm in Python. As the function to minimize, we consider $f(\mathbf{x}) = f(x_1, x_2) = (x_1^2 + x_2^2) \times [\cos(x_1) + \sin(x_2)]$.

1. Plot the function over $x_1 \in [-10, 10]$ and $x_2 \in [-10, 10]$.
2. Is f convex or non-convex? Is it smooth or non-smooth? Comment on the local/global extrema.
3. Write a Python function implementing f.
4. Compute the gradient of f, i.e. $\nabla f(\mathbf{x})$, and implement it in Python.
5. Write a Python function for performing the GD algorithm.
   (a) Inputs: f, ∇f, $\mathbf{x}_0$, μ (learning rate), and the number of iterations, denoted K.
   (b) Outputs: Final solution for $\mathbf{x}$, i.e. $\mathbf{x}_K$, the sequences of $f(\mathbf{x}_k)$ and $\|\nabla f(\mathbf{x}_k)\|$.
6. Apply the GD on the function considered in this exercise.
   (a) Try different $\mathbf{x}_0$ and μ, and plot the evolution of $f(\mathbf{x}_k)$ and $\|\nabla f(\mathbf{x}_k)\|$ versus the iterations. What do you suggest to alleviate the initialization problem?
   (b) Identify convergence/divergence with each choice.
   (c) Instead of using the number of iterations, can you think of other stopping criteria? Implement them, redo (a) with the best parameter settings that you have found, and describe your observations.