**Course: Numerical Analysis for Machine Learning**

Prof. E. Miglio - February 6th 2025

Duration of the exam: 2.5 hours.

---

**IMPORTANT:** During the exam you are allowed to use your notes, books and resources on the web but the use of ChatGPT (or other LLM) is strictly forbidden. The use of such tools will result in the invalidation of the exam.

---

**Exercise 1**

The Fashion-MNIST dataset is a dataset of 60000 $28 \times 28$ grayscale images of 10 fashion categories, along with a test set of 10000 images. You can load the dataset using the following commands

```
import tensorflow as tf
(x_train, y_train), (x_test, y_test) = tf.keras.datasets.fashion_mnist.load_data()
```

*Hint: This dataset is almost a drop-in replacement for MNIST dataset. Be careful on which axis there are the samples and on which the features.*

1. Create a reduced train dataset with only T-shirt (label 0) and sneakers (label 7).
2. Plot the first 20 training images of the reduced dataset.
3. Implement a function to compute the rSVD of rank $k$.
4. Compute the PCA of the reduced dataset using the rSVD. Propose a good choice of $k$ by analyzing the plots of the the singular values, the cumulate fraction of singular values and the fraction of the "explained variance".
5. Visualize the first 20 principal axes. Comment what you are seeing.
6. Create a scatterplot for the first two principal components `PC1, PC2` of all the training images, grouped by label. Consider only the first 500 pictures.
7. Define a classifier, by proposing a linear separator $y = mx + q$, choose the values of $m$ and $q$ by inspecting the plot. *Hint: the classifier predicts one class if* $m$`PC1` $+ q >$ `PC2` *and the other if* $m$`PC1` $+ q \leq$ `PC2`.
8. Test the classifier on the test dataset. Compute the accuracy and print the confusion matrix.

**Exercise 2**

Consider the McCormick function defined as:

$$f(x, y) = \sin(x + y) + (x - y)^2 - 1.5x + 2.5y + 1, \tag{1}$$

where $-1.5 \leq x \leq 4$ and $-3 \leq y \leq 4$.

1. Draw the graph of the McCormick function and find the global minimum.
2. Implement the Gradient Descent (GD) method with fixed learning rate. Apply the GD method to find the minimum of (1) using a learning rate $\eta = 0.01$ and starting from the point $(-1, -1)$. Use the norm of the gradient as stopping criterium. Try with different learning rates.
3. Implement the Newton method and apply it to find the minimum starting from the point $(-1, -1)$. Apply the method with different starting point. When the Newton method can fail ?
4. Compare the results obtained with the two methods. Plot on the same graph the function and the iterates of the two methods.

**Exercise 3**
Consider the function:

$$f(x, y) = (x + y)e^{xy}$$

We want to compute the gradient $\nabla f(x, y) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)$ for $x = 1$ and $y = 2$ using

1. the forward mode of AD via Wengert list;
2. the reverse mode of AD via Wengert list.

Draw the computational graph of the function along with the derivatives on the edges.