Course: Numerical Analysis for Machine Learning
Prof. E. Miglio - September 3rd 2024
Duration of the exam: 2.5 hours.

Exercise 1 (13 points)

Load the matrix $L = [l_{ij}]$ contained in the file `matrix_L.txt`. This matrix is the Euclidean distance matrix of a set of 50 points $\mathbf{p}_i$ where $i = 1, \ldots, 50$.

Let us call P the unknown 50×3 matrix containing the coordinates of the 50 points (the i-th row is the i-th point).

1. **(1 point)** Check that the origin of the coordinate system is placed in the centroid of the set of points.
2. **(2 points)** Use the following relation to compute the elements of $P^T P$:
$$\mathbf{p}_i^T \mathbf{p}_j = -\frac{1}{2}\left\{ l_{ij}^2 - \frac{1}{N}\sum_{j=1}^{N} l_{ij}^2 - \frac{1}{N}\sum_{i=1}^{N} l_{ij}^2 + \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} l_{ij}^2 \right\} \tag{1}$$
3. **(4 points)** Describe if and how the eigenvalue decomposition can be used to compute P starting from the matrix $P^T P$.
4. **(3 points)** Compute P using the method described in the previous point.
5. **(3 points)** Prove relation (1). (*Hint*: start by noticing that $l_{ij}^2 = (\mathbf{p}_i - \mathbf{p}_j)^T(\mathbf{p}_i - \mathbf{p}_j)$ and then sum only over i, only over j and over both i and j).

Exercise 2 (14 points)

Consider the minimization of the 2D quadratic function
$$J(\mathbf{x}) = \frac{1}{2}(x^2 + \epsilon y^2), \tag{2}$$
where $\mathbf{x} = (x, y)$ and $\epsilon > 0$.

1. **(1 point)** How does ϵ impact on the numerical solution of the problem ?
2. **(1 point)** Let us consider the Gradient Descent (GD) algorithm with variable learning rate $\tau_k > 0$. Write the recursive relation that describes the method.
3. **(2 points)** In the convex case, if J is of class C^2, in order to have convergence, we must have
$$0 < \tau_k < \frac{2}{\sup_{\mathbf{x}} \|HJ(\mathbf{x})\|}, \tag{3}$$
where $HJ(\mathbf{x})$ is the Hessian of J. Compute the maximum value of τ_k to have convergence.
4. **(2 points)** Implement the GD method with constant step size. Assume $\eta = 4$ and $\mathbf{x}_0 = (0.9, 0.3)$. Display the iterations on the contour figure for different values of the step size. Comment on the results.
5. **(2 points)** Consider the *exact line search method* for the dynamic choice of the step size. In this method at each time step we choose
$$\tau_k = \underset{s}{\operatorname{argmin}}\, J(\mathbf{x}^{(k)} - s\nabla J(\mathbf{x}^{(k)})). \tag{4}$$
Compute the explicit expression of τ_k given by the *exact line search method* for the function (2).
6. **(3 points)** Implement the GD method with variable step size using the expression derived at the previous point and apply the method to the minimization of function (2). Display the iterations on the contour graphs. Comment on the results.
7. **(3 points)** Proof expression (3).

Exercise 3 (6 points)
Consider the expression

$$y = f(x_1, x_2, x_3, x_4) = (x_1 + x_2) * x_3 - x_4. \tag{5}$$

1. **(1 point)** Draw the computational graph and write the Wengert list corresponding to equation (5). How many intermediate variables do you need ?
2. **(1 point)** Compute the values of the *derivatives on the edges*; report these values on the computational graph.
3. **(1 point)** Assume $x_1 = 1, x_2 = -2, x_3 = -1$ and $x_4 = 5$. Write the Wengert list for these input values.
4. **(1 point)** By using the *derivatives on the edges* compute the values of $\frac{\partial y}{\partial x_i}$ for $i = 2$ and $i = 4$. For the same cases write the Wengert list for the derivatives.
5. **(2 points)** Write the reverse list and use it to compute $\frac{\partial y}{\partial x_i}$ for $i = 1, 2, 3, 4$.