Exercise class 9

Learning Objectives:

- Finding extreme points
- Gradient method

Demo 1: Finding minima and maxima of functions

Find the minima and/or maxima of the following functions.

a) $f(x_1, x_2) = x_1^3 + x_2^3 - 3x_1x_2$

b) $f(x_1, x_2, x_3) = x_1^2(x_1 - 3) + (x_2 - 1)^2 + (x_3 - 1)^2$

Hint. Use the Hessian to verify necessary and sufficient conditions.

Demo 2: Linear regression using gradient method

Linear regression is a key prediction technique in machine learning and statistics. It consists of obtaining the linear function

$$y = a^\top x + b$$

that best fit some m data points $(x_i, y_i)_{i=1,\dots,m}$ available for an input x with n features (that is, $x \in \mathbb{R}^n$) and an output y. Then, given a new observation $m+1$, we can predict y_{m+1} to be

$$\hat{y}_{m+1} = a^\top(x_{m+1}) + b$$

In these applications, the measurement of fitness of the predictor is given by the *accumulated (or sum of) squared error* $f : \mathbb{R}^{n+1} \mapsto \mathbb{R}$ for the predictions obtained for a given (a, b) for x_i and the observed y_i, for $i = 1, \dots, m$. Notice that it simply amounts to the difference between the prediction $\hat{y}$ and the actual observation y, squared to compensate for positive and negative deviations. That is

$$f(a, b) = \sum_{i=1}^{m} \left[\left(\sum_{j=1}^{n} a_j x_{ij} + b_i \right) - y_i \right]^2 = \sum_{i=1}^{m} e_i^2 = e^\top e = ||e||_2^2.$$

Finding the best fitting (a, b) can be achieved by employing optimisation to find (a, b) that minimise the accumulate squared error, a method that is commonly referred to as the *least squared error* (LSE) estimation.

Given the data below (with $m = 7$ and $n = 1$), estimate the parameters a and b of estimate $y = ax + b$ using the LSE estimation. To find the optimal parameters, minimise the squared error function f using the gradient method, with starting point $(a, b) = (0, 0)$ and step size $\lambda = 0.01$. Use a tolerance $|\nabla f(a_k, b_k)| \leq 0.01$.

x_i	0	1	2	3	4	5	6
y_i	1	3	1.5	4	6.5	5	8

Problem 1: Finding minima and maxima of functions

Find the minima and/or maxima of the following functions.

a) $f(x_1, x_2) = x_1^3(x_1 - 4) + (x_2 - 5)^2$

b) $f(x_1, x_2, x_3) = (1 - x_2)(1 - x_3) + x_1^2 - 1$

Hint. Use the Hessian.

Problem 2: Extreme points

Determine the nature of the extreme points of the following function:

$$f(\mathbf{x}) = 2x_1^2 + x_2^2 + x_3^2 + 6(x_1 + x_2 + x_3) + 2x_1x_2x_3$$

Examine the points $(1,\ -4.2,\ 1.2)$, $(1,\ 1.2,\ -4.2)$, and $(-2.82,\ 1.65,\ 1.65)$.

Problem 3: Stationary and extreme points

Verify that the function

$$f(x_1, x_2, x_3) = 2x_1x_2x_3 - 4x_1x_3 - 2x_2x_3 + x_1^2 + x_2^2 + x_3^2 - 2x_1 - 4x_2 + 4x_3$$

has the stationary points (0, 3, 1), (0, 1, -1), (1, 2, 0), (2, 1, 1), and (2, 3, -1). Use the sufficiency condition to identify the extreme points.

Problem 4: The Gradient method

Calculate by hand the first two steps (x_1 and x_2) of the gradient method for the minimization of the function f. Initial value is $x_0 = (0, 0)$. Compute optimal step sizes at each iteration.

$$f(x_1, x_2) = (1 - x_1)^2 + (1 - x_2 - x_1)^2$$

Hint. The optimal step size can be obtained from first-order optimality conditions, namely $\min_{\alpha \in \mathbb{R}} f(x_{k+1}) = \min_{\alpha \in \mathbb{R}} f(x_k - \alpha \nabla f(x_k))$.

Problem 5: Analytical LSE estimation*

Linear regression, as presented in Demo 2, can be alternatively performed by finding a point $\alpha = (a, b) \in \mathbb{R}^{n+1}$ that satisfies the optimality conditions of the accumulated squared error function $f(\alpha) = e^\top e$, where e is defined as in Demo 2.

Formulate the minimisation problem for the LSE estimation in a general manner and provide its optimality conditions.

Hint. You might need the following differentiation rules:

1. $\nabla(a^\top x) = a$
2. $\nabla(x^\top Ax) = A^\top x + Ax$

Home Exercise 9: Gradient method with line search

Perform one iteration of the gradient method to solve

$$\max . f(x_1, x_2) = 2x_1x_2 + 2x_2 - x_1^2 - 2x_2^2$$

from the initial point $x_0 = (0.5, 0.5)$. Use the bisection method to find the optimal step size with interval [0,2] and tolerance $\varepsilon = 0.01$. Is the new point obtained optimal (considering the tolerance of $\varepsilon = 0.01$)?

Hint. Do it by hand and notice it is a maximisation.