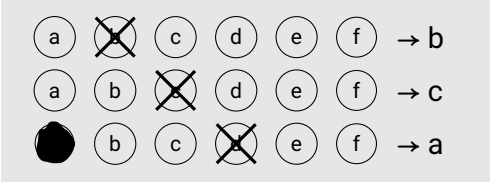
Exercises

1	2	3	4	5

1	1	1	1	1	1	1
2	2	2	2	2	2	2
3	3	3	3	3	3	3
4	4	4	4	4	4	4
5	5	5	5	5	5	5
6	6	6	6	6	6	6
7	7	7	7	7	7	7
8	8	8	8	8	8	8
9	9	9	9	9	9	9
0	0	0	0	0	0	0

Surname, First name

Modelling, Uncertainty and Data for Engineers EXAM (CEGM1000)
Exam Q1
6 Nov, 2023, 13:30-16:30



Fill in your answer(s) to the multiple-choice questions as shown above (circles = one correct answer).

Do not open the exam or turn to the back page until given permission by the instructor!
(you may write your name and student ID)

Before you start the exam, a few remarks:

- Write down your first and last name in the field on the top left corner of this page
- Fill in your student number on the top right corner of this pages. Fill in the number in the boxes on top, and mark the corresponding number. Fill the corresponding cicle as indicated above.
- You may use pen or pencil, a scientific calculator and the attached formula sheet. Any other tools and sources of information are not allowed.
- It is not allowed to release the staple.
- Scratch paper is available to use during the exam, but will not be collected or graded.
- On the following pages, some questions have a specific box for you to answer: anything written outside the boxes will not be graded. Note that we have provided a lot of space for answers.
- In case you want to erase and rewrite your answer, ask an invigilator for a white sticker to cover your incorrect answer.
- The answer space size is not an indicator of how long we expect your answers to be! Shorter is generally better.
- In case you want to correct your answer for a multiple choice question, follow the instructions above. If you mess up, put an arrow to the answer you think is correct. If you mess up again, add a comment.
- A summary of points and questions is provided on the last page, as well as examples of how to correct your multiple choice answers.

Good luck!

Exercise 1: Programming

Consider the following file structure schematic, which describes the files on your computer:

```
|-- MUDE
    |-- Project_76
        |-- Difficult_Assignment.ipynb
        |-- auxiliary_files
            |-- data.csv
            |-- figure.png
            |-- functions.py
```

In which:

```
    |-- Project_76
```

is a Git repository.

One of your group members just sent you a message on WhatsApp asking you to check their most recent commit to the file:

```
        |-- Difficult_Assignment.ipynb
```

which they just committed to GitLab.

2p **1a** Which of the following would NOT be a way of incorporating the changes in the local repository on your computer?

(a) Clone the repository
(b) Pull from remote
(c) Visit the file on gitlab.tudelft.nl
(d) Pull from remote, then review the commits
(e) Send the file over WhatsApp then save it in your local repository

You run the cells in

```
        |-- Difficult_Assignment.ipynb
```

and get the following error:

```
---------------------------------------------------------------------------
AttributeError                            Traceback (most recent call last)
Cell In[3], line 1
----> 1 print(find_squared_error(a))

File ~\code\MUDE\2023\Week_76\auxiliary_files\functions.py:17, in find_squared_error(data)
     15 def find_squared_error(data):
     16     """Evaluate squared error of data (ndarray)."""
---> 17     return (data - data.mean())**2

AttributeError: 'list' object has no attribute 'mean'
```

2p **1b** What is the most likely problem for the shown error?

- (a) The filepath to "'functions.py"' is incorrect
- (b) The function "'find_squared_error()"' is undefined
- (c) The object "'a"' is a list
- (d) The variable "'data.mean()"' is undefined

3p **1c** Describe in one sentence how you would fix this error. You may include short snippet of hand-written code, if you think it is necessary (you will not be graded on whether the syntax is 100% correct).

You execute a cell with python code:

```
import awesome
```

in a Jupyter notebook for which the working directoy is part of an unknown file structure. So, this Jupyter notebook is NOT part of the file structure of the previous questions.

The following error is returned:

```
import awesome
---------------------------------------------------------------------------
ModuleNotFoundError                       Traceback (most recent call last)
Cell In[3], line 1
----> 1 import awesome

ModuleNotFoundError: No module named 'awesome'
```

3p **1d** Which of the following is NOT a plausible explanation for the shown error?

(a) You have not installed the package with pip or conda yet
(b) The package is only available via conda, not pip
(c) You activated the wrong conda environment
(d) The file awesome.py does not exist in your working directory

Exercise 2: Uncertainty propagation

A single beam echo sounder is used to measure the depth in a harbor. The principle is based on transmitting a sonar pulse and measuring the 2-way travel time. The depth can then be determined by multiplying half of the travel time with the propagation speed. The propagation speed C of the water depends on the temperature T and salt concentration S as:

$$C = 1449.2 + 4.6T - 0.055T^2 + 0.0029T^3 + 1.34(S - 35)$$

All the variables are random variables, due to uncertainty in the temperature and salt concentration.

We are interested in the precision of the propagation speed, where it is known that:

$$\mu_T = 15^oC,\ \sigma_T = 2^oC$$

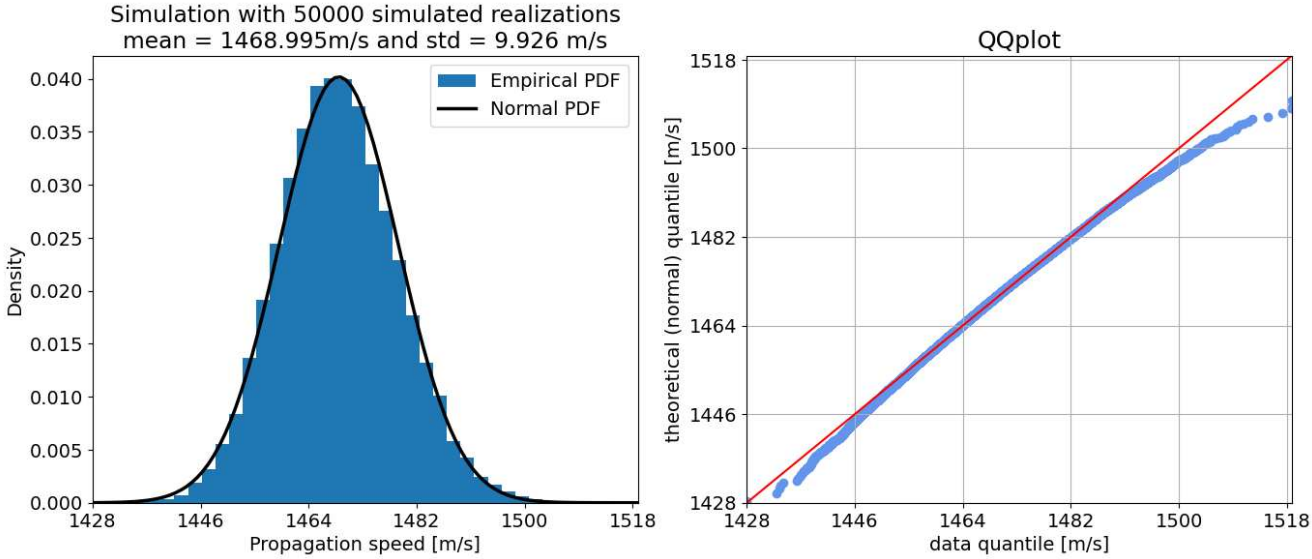
$$\mu_S = 0kg/m^3,\ \sigma_S = 0.5kg/m^3$$

$$Cov(T, S) = 0$$

7p **2a** Approximate the standard deviation of the propagation speed (give your answer to 2 decimal places). *Show how you arrived at your answer.*

4p **2b** In order to reduce σ_C, would it have more impact to reduce σ_T or σ_S? *Explain your answer.*

Assume T and S are normally distributed. We are interested in the distribution of C. We apply Monte Carlo simulations to obtain a large number of sample values for T and S and compute the corresponding sample values of the propagation speed C. Below you see the resulting histogram (left) and the QQ-plot (right), where in both cases the normal distribution is used as the model distribution.



6p **2c** Based on the figures, do you conclude that the normal distribution can be used as the model distribution for the propagation speed? *Explain your answer.*

6p **2d** The standard deviation of the simulated sample values of C (from question *c*) is equal to 9.93 m/s. Give one reason why it is not the same as the answer you (should have) obtained in question *a*?

Exercise 3: Observation Theory

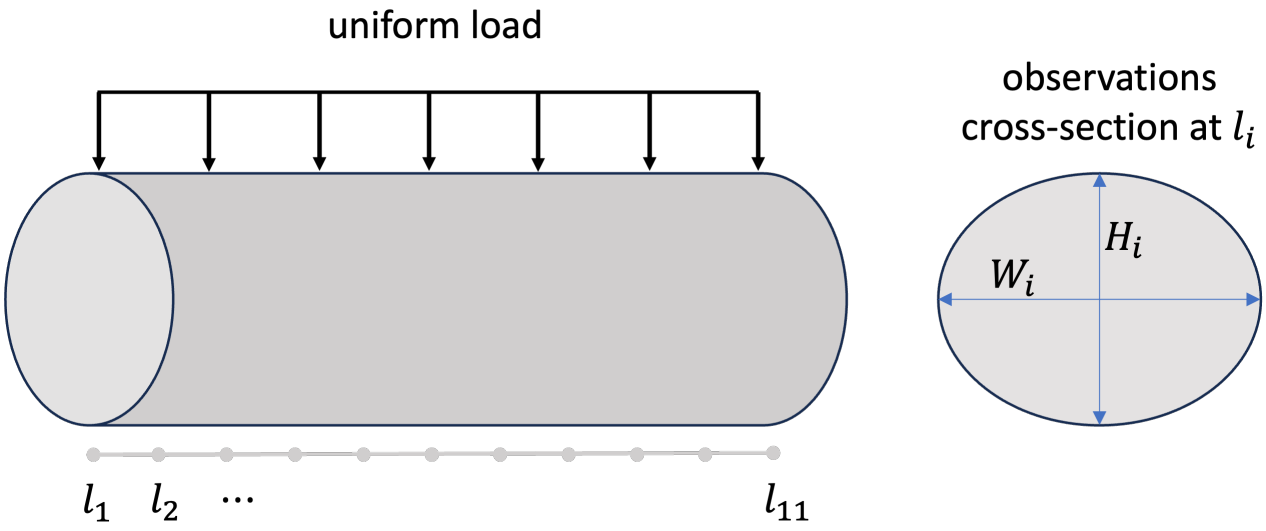
Will the tunnel deform?

An engineer wants to test how a perfectly circular tunnel segment (see Figure) will deform when a uniform load is applied from above for a period of time.

The null hypothesis is that the segment will not deform such that cross-section remains circular over the full length of the segment.

The alternative hypothesis is that the segment deforms uniformly across the full length, such that each cross-section will have the same ellipse-shape.

After applying the load, the width W_i $[mm]$ and height H_i $[mm]$ of the cross-sections are measured at $l_i = 0, 1, \ldots, 10$ meters, such that we have 22 observations in total.



4p **3a** What is the functional model for the null-hypothesis? What is the unknown parameter?

$$\mathbb{E}\left(\begin{bmatrix} W_1 \\ H_1 \\ W_2 \\ H_2 \\ \vdots \\ W_{11} \\ H_{11} \end{bmatrix}\right) =$$

4p **3b** What is the functional model for the alternative hypothesis?

$$\mathbb{E}\left(\begin{bmatrix} W_1 \\ H_1 \\ W_2 \\ H_2 \\ \vdots \\ W_{11} \\ H_{11} \end{bmatrix}\right) =$$

8p **3c** After applying best linear unbiased estimation, we obtain $\hat{\epsilon}^T \Sigma_Y^{-1} \hat{\epsilon} = 31.5$ with the null-hypothesis, and $\hat{\epsilon}_a^T \Sigma_Y^{-1} \hat{\epsilon}_a = 29.6$ with the alternative hypothesis. Apply an appropriate test to decide between the null and alternative hypothesis, use a false alarm rate of 0.025. Show all your steps and explain what your decision will be based on the test outcome.

Height change over time

In another experiment, the goal is to investigate at which rate the height of the tunnel segment is changing over time once the load is applied. The rate is assumed to be constant.

Therefore the *change of height* H_i [mm] with respect to the *known* initial height is measured at $t = 1, 2, \ldots, 4$ $months$ after the load is applied. The functional model is thus given by:

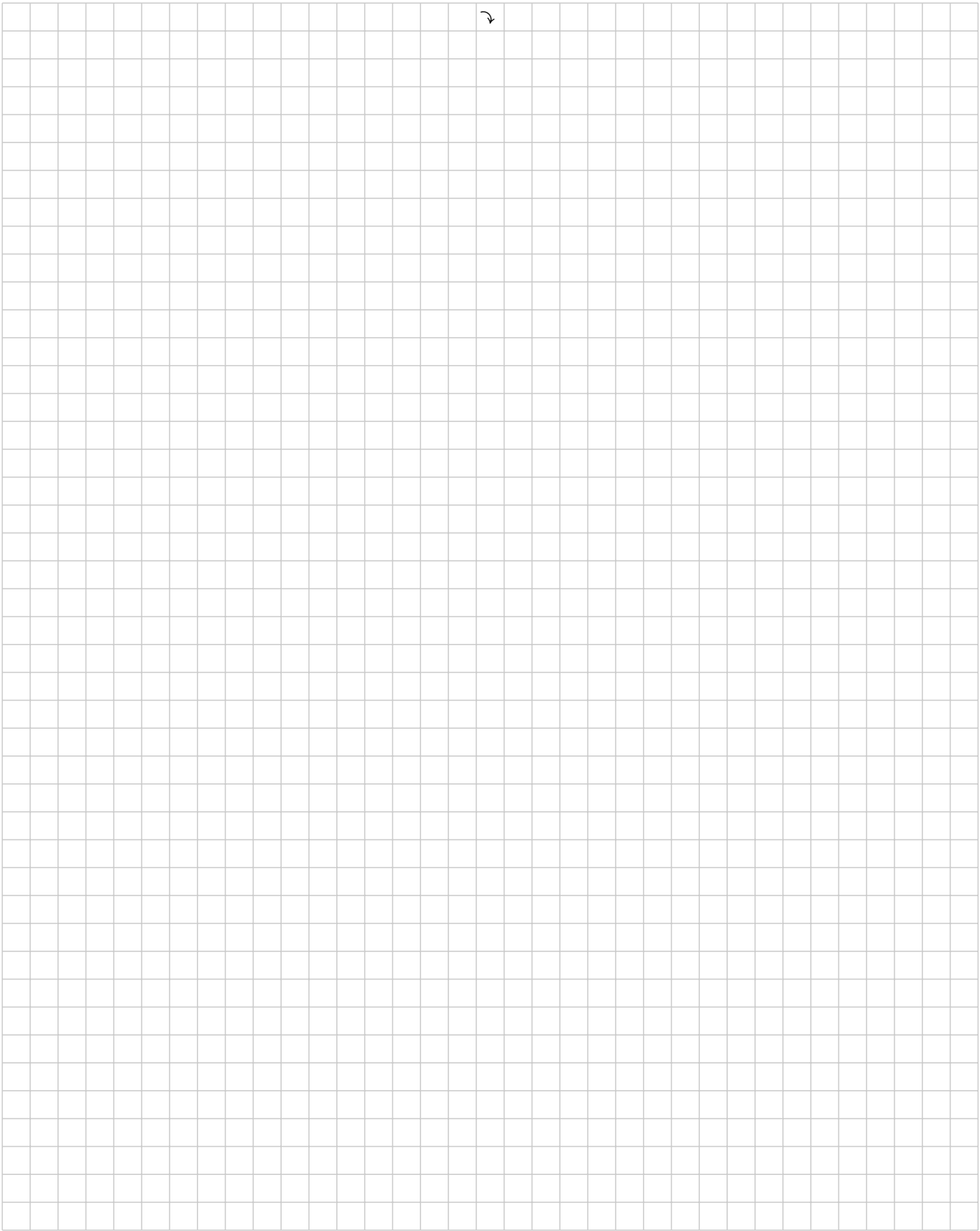
$$\mathbb{E}\left(\begin{bmatrix} H_1 \\ H_2 \\ H_3 \\ H4 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} v$$

with v the unknown rate at which the height is changing in $mm/month$.

The first 2 measurements have a standard deviation of σ, the last 2 measurements have a standard deviation of 0.5σ due to a change in instrument. All measurements are independent.

It is required to obtain a 96% confidence level of $\hat{v} \pm 0.2mm/month$.

8p **3d** What should σ be to realize this? *Round your answer to two decimal places.*



Exercise 4: Numerical modelling

Given the differential equation:

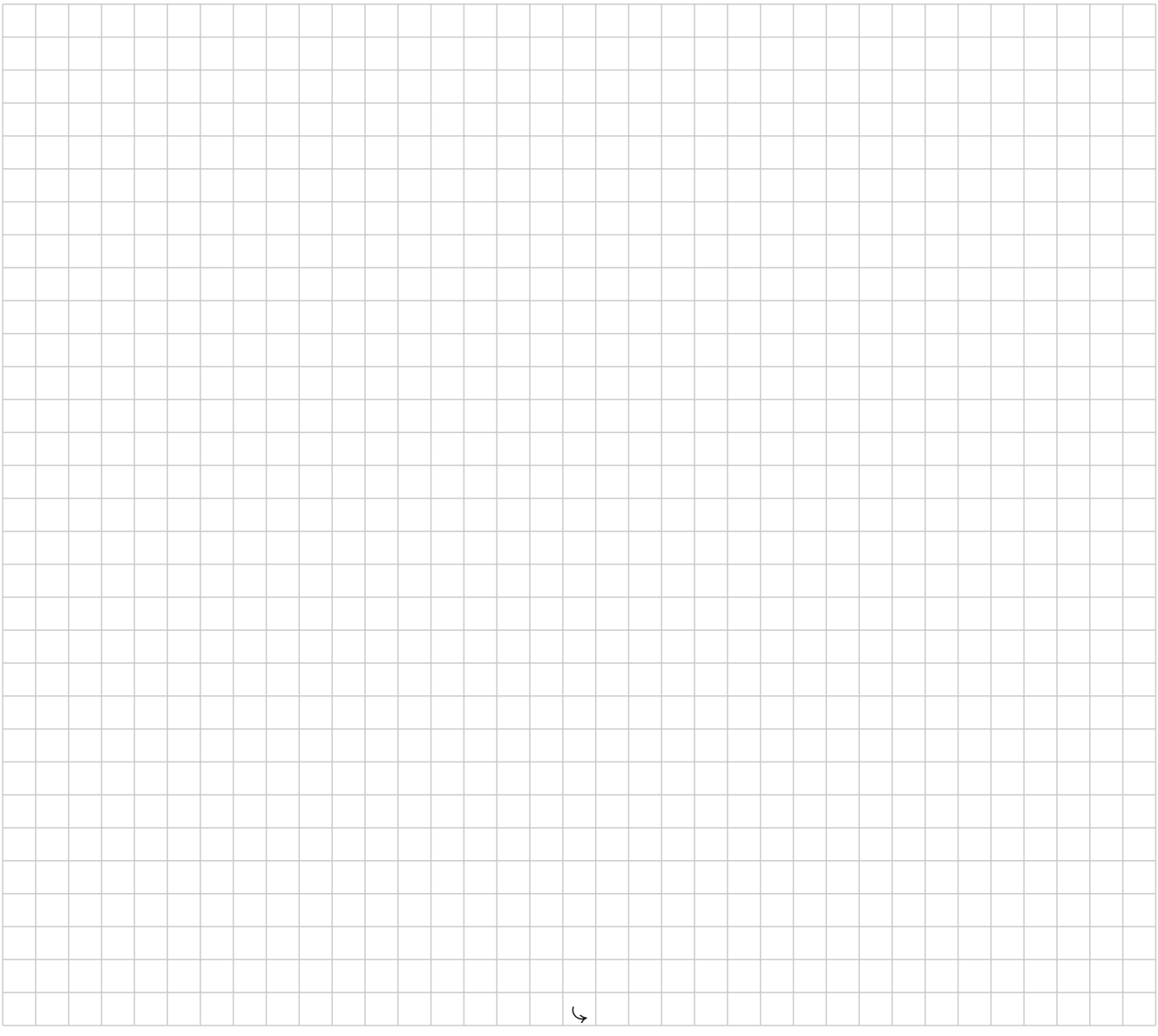
$$\frac{df\,(x)}{dx} = g\,(f\,(x))$$

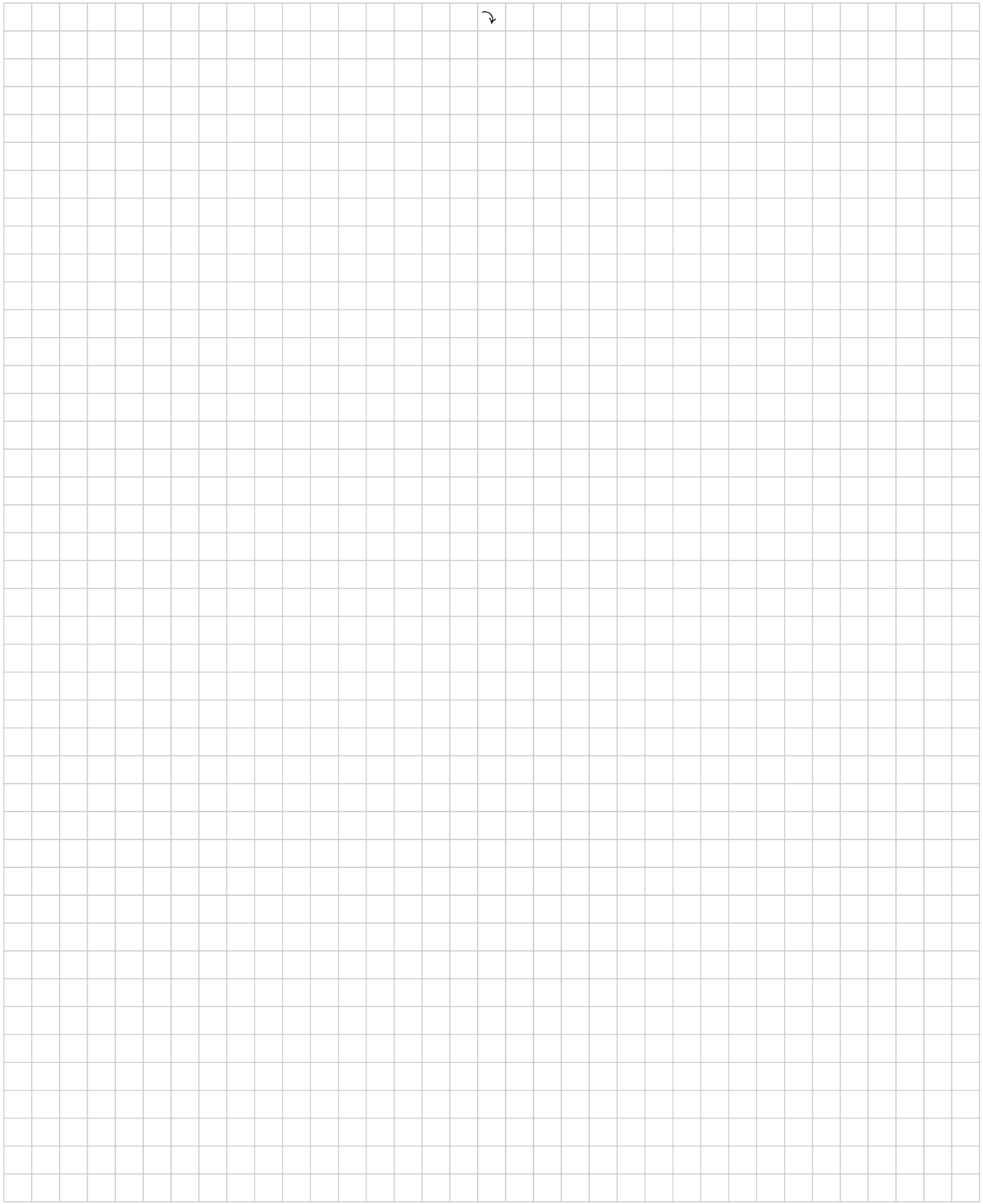
with:

$$g\,(f\,(x)) = -f\,(x) \cdot \frac{\cos\,(\pi f\,(x))}{3}$$

In this assignment you'll apply numerical methods to solve a linearised version of this differential equation.

4p **4a** Find the taylor series expansion of $g\,(f\,(x))$ as a function of $f\,(x)$ about the point $f\,(x) = 4$. Give the taylor series expansion up to and including both the first and second order. Calculate all derivatives and simplify your expression.





4p **4b** Discretise the differential equation with the taylor series up to and including the first order and apply the Forward Euler.

7p **4c** Assess the stability of using Forward Euler for discretisation. Is your solution conditionally or unconditionally stable? Include the stability assessment and the criterion of the stability for this case.

7p **4d** Which other methods for numerical derivatives could you apply (name at least 3 more methods), and what would be the effect on stability and accuracy? Explain without making any calculations.

Exercise 5: Probability and reliability

X and Y are two (unit-less) quantities that have been obtained from field measurements in order to investigate certain properties. The cumulative distribution function of X is given by:

$$F_X(x) = \begin{cases} 0, & x < 0 \\ \frac{x}{2}, & 0 \le x \le ~~1~~ 2 \\ 1, & x > 2 \end{cases}$$

2p **5a** What is $P[X \le 0.5]$?

2p **5b** The engineer wants to design for the value of X which is exceeded with a probability of 0.05. What is the design value?

The distribution $F_Y(y)$ is unknown. However, the following statistics could be calculated from the observations in the field:

	Random Variable Y
mean	0.88
std	1.28
min	0.0
25%	0.42
50%	0.78
75%	1.22
max	2.31

1p **5c** Which of the following distributions would be the best fit to the variable Y based on the previous statistics:

(a) Gaussian
(b) Uniform
(c) Lognormal

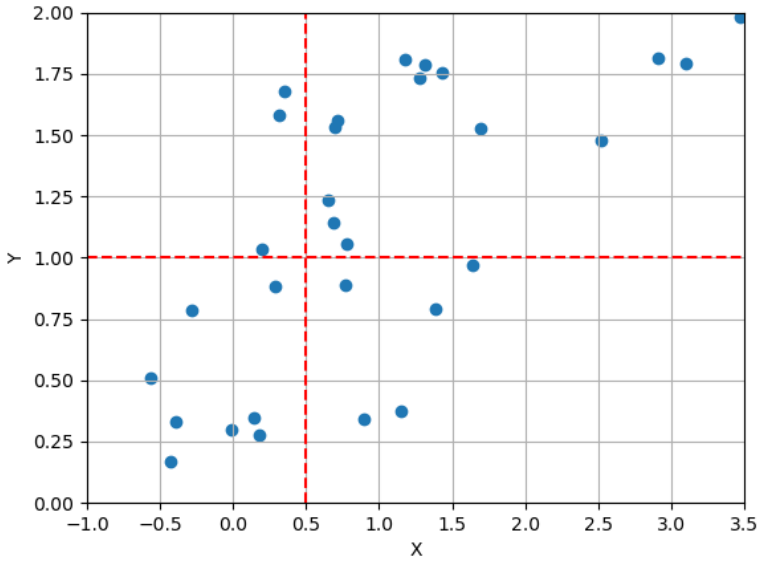
3p **5d** Justify your previous choice.

4p **5e** The engineer decides to fit by moments a Gumbel distribution to the observations. Compute the distribution parameters using the given formulas. Round the results to one decimal figure. (Hint: you may want to refer to the formula sheet)

If you have not computed the parameters for the Gumbel distribution in the previous question, use $\alpha = 0.3$ and $\beta = 1$ in the subsequent questions.

2p **5f** The engineer wants now to consider both X and Y in the design. Assume that X and Y are independent. What is $P[X \leq 0.5, Y \leq 1]$? Round your answer to two decimal places.

You finally managed to get some paired observations of X and Y and plot them in the figure below



3p **5g** Using the figure, what is $P[X > 0.5 \text{ OR } Y > 1]$? The number of observations is 30.

Finally, the engineer decides to model the multivariate uncertainty of X and Y using a multivariate normal distribution, where X **and** Y **are not independent**. The engineer defines it in python as:

```
joint_distr = scipy.stats.multivariate_normal(mu, sigma)
```

4p **5h** What would be suitable values of mu and sigma in the code above? You don't need to compute each value but assess whether they are suitable.

(a) mu = 0.88, sigma = 1.64

(b) mu = 1, sigma = 1.28

(c) mu = [1, 0.88], sigma = [[0.33, 0],[0, ~~0.~~1.64]]

(d) mu = [1, 0.88], sigma = [[0.33, 0.55],[0.55, 1.64]]

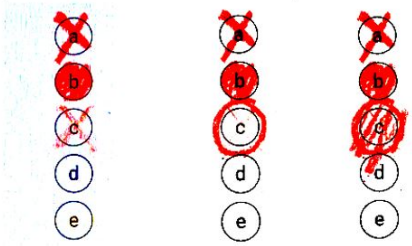
(e) mu = [1, 0.88], sigma = [[0.33, 0.55],[0, 1.64]]

The table below gives an overview of the questions to help you plan your time during the exam.

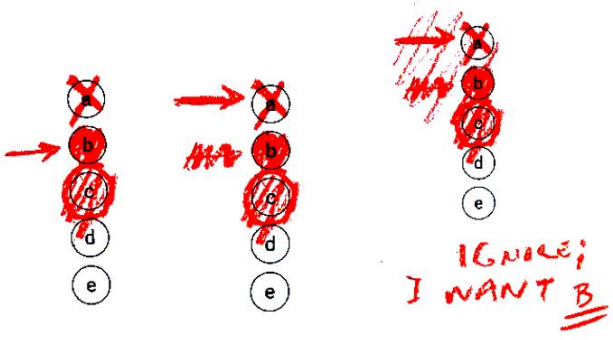
No.	Topic	Number of Sub-questions	Total Points
1	Coding	4	10
2	Propagation of Uncertainty	4	23
3	Observation Theory	4	24
4	Numerical Modelling	4	22
5	Probability	8	21
		Total:	**100**

In case you want to correct your answer for a multiple choice question put an ARROW in front of your final answer. If you also make a mistake with your arrow, write a clear message on the page. Here are a few examples:

Examples of UNCLEAR multiple choice response:



Examples of CLEAR multiple choice response:



Answer: B Answer: A Answer: B