i Information sheet 1TD352

Scientific Computing for Data Analysis

The exam is divided into two parts, *part A* and *part B*. Part A is related to grade 3, and part B is to grades 4 and 5.

Part A (grade 3)

The tasks in Part A are directly linked to one of the three course objectives: *Key Concepts, Algorithms, and Analysis*. In this part, there are 2 questions related to each objective (questions 1 to 6) and maximum 2 points per question.

Questions in part A are either multiple-choice questions or require entering some texts or a numerical value in a provided box. For these types of questions, you may need to solve the problem on paper and then select or enter the correct answer. You are able to answer this part in the Inspera only. As an option, you can write down your detailed solutions on paper and hand them in for review. We can then correct the information if you have entered the wrong answer due to something that we judge to be a simple careless error.

Part B (grades 4 and 5)

In part B there are 3 questions (questions 7, 8, 9). You can either type the detailed solution directly into Inspera or hand it in on paper to the invigilators. If you hand in paper answers, make a note about it in Inspera in the solution box. If you decide to write the solution in Inspera please use the inline formula editor (icon in the menubar) for writing math letters and formulas.

Very important: questions 7, 8, and 9 can give either 0 or 10 points with no intermediate scores. This implies the necessity of writing a complete solution.

In both parts, if you hand in paper-answers please write in **English** and with neat and legible handwriting.

Grades

- **Grade 3:** At least 6 points. You must answer at least one question on each objective of part A. This corresponds to minimum 3 questions (i.e. 6 points) distributed among the three objectives. Failure to meet a course objective entirely will result in a failing grade on the exam.
- **Grade 4:** At least 18 points. You must fulfill Grade 3 + solve one more question from part A (at least 4 correct answers) + get 10 points from part B (1 complete solution).
- **Grade 5:** At least 30 points. You must fulfill Grade 3 + solve two more questions from part A (at least 5 correct answers) + get 20 points from part B (2 complete solutions).

Allowed aid

- Online Python (available as a link, the idea is that you are able to use Python as a pocket calculator, very much in the same way as on the problem solving sessions. It will be enough to be able to use numpy)
- Numpy reference manual (available as a link)
- Numpy cheat sheet (available as a link)
- The formula sheet (is available as a link)
- Online calculator (available as a link, you can also bring your own calculator if you want)

Good Luck!

1 1TD352_Concept_1

Classify methods and models.
*** In order to get 2 points, you need to provide at least 6 correct answers out of the total 7. Otherwise, you will receive 0 points (no intermediate points).**

	Deterministic model	Stochastic model	Stochastic method	Deterministic method
SSA (Gillespies algorithm)	○	○	○	○
F and R real-valued $\frac{dF}{dt} = \beta FR - \gamma F$ $\frac{dR}{dt} = \alpha R - \beta FR$	○	○	○	○
Least squares problem $\min_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2$	○	○	○	○
F and R integer-valued $R \xrightarrow{\alpha} 2R$ $R + F \xrightarrow{\beta} 2F$ $F \xrightarrow{\gamma} \emptyset$	○	○	○	○
Monte Carlo algorithm	○	○	○	○
QR-iteration	○	○	○	○
Power method	○	○	○	○

Maximum marks: 2

2 1TD352_Concept_2

Select the correct alternative.
*** In order to get 2 points, you need to provide at least 6 correct answers out of the total 7. Otherwise, you will receive 0 points (no intermediate points).**

(a) The SVD *is not* used to compute the Select alternative (low rank approximation, norm 2, norm infinity, pseudo-inverse) of a matrix.

(b) Orthogonal matrices when applied to a vector, Select alternative (either stretch or shrink it, change its Euclidean norm (norm 2), only rotate it).

(c) The Schur decomposition of a symmetric matrix is identical with its Select alternative (LU decomposition, eigen-decomposition, QR factorization, SVD).

(d) Given a uniform random number, the Select alternative (basis Monte Carlo method, inverse transform method, Gillespies algorithm) can be used to generate a random number from an arbitrary distribution f.

(e) The Brownian motion $X(t)$ is Select alternative (uniformly, exponentially,

normally) distributed with variance Select alternative (1, 0, t)

(f) A stochastic process is called a Select alternative (non-Markovian, discrete time, continuous time, Markov) process, if one can make predictions for the future of the process based solely on it's present state.

Maximum marks: 2

3 1TD352_Algorithm_1

For a quadratic least squares fitting (with ansatz $y = a_0 + a_1 x + a_2 x^2$) on four points with function values $\mathbf{y} = [-1, 2, 1, -3]^T$, the SVD of the data matrix results in factors (rounded to 2 decimal places)

$$U = \begin{bmatrix} -0.15 & 0.90 & -0.35 & -0.22 \\ -0.06 & 0.40 & 0.62 & 0.67 \\ -0.32 & 0.04 & 0.67 & -0.67 \\ -0.93 & -0.19 & -0.21 & 0.22 \end{bmatrix}, \; V = \begin{bmatrix} -0.30 & 0.68 & 0.67 \\ -0.42 & -0.72 & 0.55 \\ -0.86 & 0.11 & -0.50 \end{bmatrix}$$

and singular values $\sigma_1 = 4.90, \; \sigma_2 = 1.69, \; \sigma_3 = 1.08$.

What is the computed coefficient a_2 and what is the residual of the solution (round to 2 decimal places)?
(You can use python as a calculator)

Select one alternative:

- $a_2 = -1.74, \; residual = 0.23$
- $a_2 = 12.00, \; residual = 0.23$
- $a_2 = 12.00, \; residual = 2.90$
- $a_2 = -1.74, \; residual = 2.90$

Maximum marks: 2

4 1TD352_Algorithm_2

The task is to approximate the value of integral

$$\int_0^\infty (x^2 - 3x)\, e^{-0.5x}\, dx$$

using the Monte Carlo method on six random points

$$0.050, \; 0.588, \; 0.694, \; 0.817, \; 1.387, \; 6.211$$

which are exponentially distributed according to probability density function (pdf) $f(x) = 0.5\, e^{-0.5x}$. What is the approximate value? (rounded to 3 decimal places)

Select one alternative:

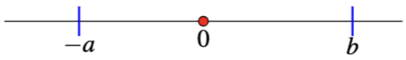
- 4.252
- -0.624
- 2.126
- 3.951
- -1.247

Maximum marks: 2

5 1TD352_Analysis_1

*** This question consists of 2 items (a) and (b). You must answer both items correctly to get 2 points. Otherwise you'll receive 0 points (we do not allocate 1 point for 1 correct answer).**

Assume that a and b are two positive integers. Consider the following interval on the real axis:



One starts at 0 and randomly walks either left or right with steplength 1 until hitting one of the boundaries $-a$ or b. The aims are to use Monte-Carlo method to estimate the number of steps needed to hit one of the boundaries, and analyze the **error** at a certain probability.

For two specific values for a and b, we provided 600 realizations using a fair coin to choose between left or right directions, and obtained a vector solution of size 600 as below (the first three and the last two numbers are given):

$$result = [31, 64, 27, \ldots, 35, 46]$$

The sample mean of this vector is $mean(result) \doteq 35.0$ and the sample standard deviation is $std(result) \doteq 38.0$.

According to the information above, answer the following questions (solve and enter the final solution in the designated box):

(a) With 95% probability, the error of this estimation satisfies $|e| \leq$ (round to one decimal place; for example 4.3)

(b) To decrease the error by a factor of 3, the number of realizations must be increased from 600 to (enter an integer number)

Maximum marks: 2

6 1TD352_Analysis_2

*** This question consists of two items (a) and (b). To get 2 points, you must select the correct answers for all the questions. Otherwise you'll receive 0 points (no intermediate points).**

(a) The QR-iteration algorithm, used to compute the eigenvalues of a matrix A, converges to the following quasi-triangular Schur form:

$$\tilde{T} = \begin{bmatrix} 4.0 & 1.0 & -1.3 & 2.0 \\ 0.0 & 2.0 & 1.0 & 3.1 \\ 0.0 & -0.5 & 1.0 & 2.5 \\ 0.0 & 0.0 & 0.0 & 0.5 \end{bmatrix}$$

What is the list of eigenvalues of A?
Notation: in items below, letter j stands for the imaginary number (like in Python).

Select one alternative:

- $\lambda_1 = 4.0,\ \lambda_2 = 3 + 1j,\ \lambda_3 = 3 - 1j,\ \lambda_4 = 0.5$
- $\lambda_1 = 4.0,\ \lambda_2 = 1.5 + 0.5j,\ \lambda_3 = 1.5 - 0.5j,\ \lambda_4 = 0.5$
- $\lambda_1 = 4.0,\ \lambda_2 = 2,\ \lambda_3 = -0.5,\ \lambda_4 = 0.5$
- $\lambda_1 = 4.0,\ \lambda_2 = 2,\ \lambda_3 = 1,\ \lambda_4 = 0.5$

(b) The SVD of a matrix A results in factor

$$\Sigma = \begin{bmatrix} 3.0 & 0 & 0 & 0 \\ 0 & 1.5 & 0 & 0 \\ 0 & 0 & 0.2 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Also assume that A_2 is the closet rank 2 matrix to A. Answer the following questions:
b-1) What is $rank(A) =$?

2	6	3	4
○	○	○	○

b-2) What is the condition number of A in norm 2?

15	3.0	$\frac{1}{15}$	0.2	3.36
○	○	○	○	○

b-3) What is $\|A - A_2\|_2 =$?

- 0.2
- 0
- 3.0
- 1.5

Maximum marks: 2

7 1TD352_HigherGrade_1

*** This question gives either 0 or 10 points (no intermediate points). Minor errors are acceptable.**

After a person takes medicine, the amount of drug left in the person's body decreases over time. When testing a new drug, a pharmaceutical company develops a mathematical model to quantify this relationship. To find such a model, suppose a dose of 1000 mg of a certain drug is absorbed by a person's bloodstream. Blood samples are taken every five hours, and the amount of drug remaining in the body is calculated. Possible data from an experiment are shown in the table below:

Drug Absorption Data

Hours since drug was administered	0	5	10	15	20	25	30
Amount of drug in body (mg)	1000	550	316	180	85	56	31

What is an **appropriate model (ansatz)** for this data and why? **Fit** your suggested model to data and finally use the model to **estimate** the amount of drug remaining in the body after 40 hours.

Use a **stable algorithm** to find your solution.

No code is required but use Python as a calculator and write down all steps of your solution. Round numbers to 2 decimal places.

Fill in your answer here or on a paper. If you hand in paper-answer, make a note about it here.

Format

Words: 0

Maximum marks: 10

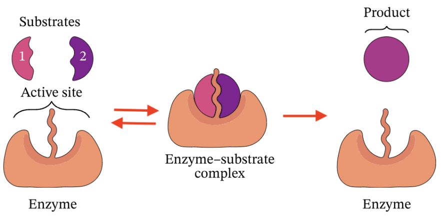
8 1TD352_HigherGrade_2

*** This question gives either 0 or 10 points (no intermediate points). Minor errors are acceptable.**

Consider the Michaelis-Menten model

$$S + E \underset{c_2}{\overset{c_1}{\rightleftarrows}} C \xrightarrow{c_3} P + E$$

which is the standard model for enzyme catalysts. See also the figure below:



This phenomenon involves an enzyme E binding to a substrate S to form a complex (enzyme-substrate) C that releases a product P and regenerating the original form of the enzyme. Here, c_1 (forward rate), c_2 (reverse rate), and c_3 (catalytic rate) denote the constant rates of the reactions. Our intention is to solve this model using the Gillespie's algorithm (SSA).

Write down **propensity functions and state-change vectors** for this model. Assume that for a certain enzyme the reactions rates are $c_1 = 0.002\ \mathrm{mol}^{-1}\mathrm{sec}^{-1}$, $c_2 = 0.1\ \mathrm{sec}^{-1}$ and $c_3 = 0.75\ \mathrm{sec}^{-1}$ where sec is the unit of time and mol is the unit for number of proteins. Furthermore, assume that at time $t = 0.1\ \mathrm{sec}$ the number of proteins have been computed as $E(t) = 300\ \mathrm{mol}$, $S(t) = 200\ \mathrm{mol}$, $C(t) = 100\ \mathrm{mol}$, and $P(t) = 50\ \mathrm{mol}$. The task is to compute the number of proteins in the next time level $t + \tau$. To this aim, we have generated two uniform random numbers $u_1 = 0.64$ and $u_2 = 0.83$ from the $U(0,1)$ distribution. The number u_1 must be used to determine the steplength τ, and u_2 to determine the specific reaction that will occur. Given these conditions, proceed to compute the **next time level** (i.e. $t + \tau$) and the **number of proteins** at this new time.

No code is required but use Python as a calculator and **write down all steps and details** of your solution.

* The image downloaded from http://nagwa.com

Fill in your answer here or on a paper. If you hand in paper-answer, make a note about it here.

Format

Words: 0

Maximum marks: 10

9 1TD352_HigherGrade_3

*** This question gives either 0 or 10 points (no intermediate points). Minor errors are acceptable.**

The **Weibull distribution** models a broad range of random variables, largely in the nature of a time to failure or time between events. The distribution is named after Swedish mathematician Waloddi Weibull, who described it in detail in 1939. The cdf (cumulative density function) of Weibull distribution is defined by two parameters $\lambda > 0$ (scale parameter) and $k > 0$ (shape parameter), and is given by

$$F(x) = 1 - e^{-(x/\lambda)^k}, \quad x \geq 0$$

Suppose someone tried to model this distribution by, first, sampling numbers u from the Uniform distribution $U(0,1)$ and, then, applying the Inverse Transform Method (ITM) to obtain Weibull samples x. The results were approximated to two decimals and look as follows:

u	0.50	0.61	0.82	0.53	0.20	0.45
x	0.98	1.80	5.86	1.15	0.10	0.73

Find the parameter values λ and k that best fit these samples, with an accuracy of two decimals (2 numbers after the dot).

No code is required but use Python as a calculator and **write down all steps and details** of your solution.

Fill in your answer here or on a paper. If you hand in paper-answer, make a note about it here.

Format

Words: 0

Maximum marks: 10