ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Models, Algorithms and Data (MAD): Introduction to Computing

Spring semester 2018

Prof. P. Koumoutsakos, Prof. J. H. Walther
ETH Zentrum, CLT
CH-8092 Zürich

Exam

Issued: August 8, 2018

Exam directives. In order to pass the exam, the following requirements have to be met:

- Read carefully the first two pages of the exam. Write your name and Legi-ID where requested. Before handing in the exam, **PUT YOUR SIGNATURE ON PAGE 2**.
- Clear your desk (no cell phones, cameras, etc.): on your desk you should have your Legi, your pen, your notes and optionally your non-scientific calculator.
- The teaching assistants will give you the exam sheets and the necessary paper sheets. You are not allowed to use any other paper sheets. On the top-right corner of every page write your complete name and Legi-ID.
- The personal summary consists of no more than 4 pages (2 sheets). The personal summary can be handwritten or machine-typed. In case it is machine-typed, the text has to be single-spaced and the font size has to be at least 8 pts. You are not allowed to bring a copy of somebody else's summary.
- You can answer in English or in German; the answers should be handwritten and clearly readable, written in blue or black - do NOT write anything in red or green. Only one answer per question is accepted. Invalid answers should be clearly crossed out.
- If something is disturbing you during the exam, or it is preventing you from peacefully solving the exam, please report it immediately to an assistant. Later notifications will not be accepted.
- You must hand in: the exam cover, the sheets with the exam questions and your solutions. The exam cannot be accepted if the cover sheet or the question sheets are not handed back.

Family Name:	
Name:	
Legi-ID:	

Question	Maximum score	Score	TA 1	TA 2
1	4			
2	5			
3	6			
4	3			
5	7			
6	9			
7	8			
8	5			
9	10			
10	9			
11	6			
12	9			
Total	81			

With your signature you confirm that you have read the exam directives; you solved the exam without any unauthorized help and you wrote your answers following the outlined directives.

Signature:

Short Questions

Question 1: Linear Least Squares (4 points)

Let $\bar{x}$ be the linear least squares solution to the problem $A\vec{x} = \vec{b}$, and $\vec{E} = A\bar{x} - \vec{b}$ be the corresponding residual vector.

a) Describe the geometric interpretation of the least squares solution.

b) Let matrix A be equal to

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}.$$

Which of the following vectors is a possible value for $\vec{E}$, why?

$$\square \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \quad \square \begin{pmatrix} 0 \\ 1 \\ -2 \\ 1 \end{pmatrix} \quad \square \begin{pmatrix} -1 \\ -1 \\ 1 \\ 1 \end{pmatrix} \quad \square \begin{pmatrix} -1 \\ -1 \\ -1 \\ -1 \end{pmatrix}$$

Question 2: Lagrange interpolation (5 points)

State whether the following statements are true or false. Each correct answer is awarded one point, answers left blank are worth zero points. Each wrong answer is penalized with minus one point. The minimum score is zero.

a) For Lagrange interpolation of N data points, the degree of each basis function $L_i(x)$ is N.

b) The Lagrange polynomial passes through every point of the given data set.

c) The Lagrange polynomial is the polynomial of highest degree that assumes the value y_i at each point x_i of a data set $\{x_i, y_i\}, i = 1...N$.

d) If $P(x)$ is a Lagrange polynomial interpolating the function $f(x) = \frac{1}{x}$ at the points $x_i = \{0, 1, 4\}$, then $P'(4) = \frac{-1}{16}$.

e) If nodes x_k are equispaced over some interval $x \in [a, b]$, then the Lagrange interpolating function oscillates with greater amplitude near a and b than near the midpoint of the interval.

Question 3: Neural networks (6 points)

Are the following statements true or false? Each correct answer is awarded one point, answers left blank are worth zero points. Each wrong answer is penalized with minus one point. The minimum score is zero.

a) Gradient descent based methods, such as back-propagation may get stuck in local minima.

b) Increasing the number of hidden nodes in a feed-forward neural network with two layers and non-linear activations improves generalization.

c) Increasing the number of layers in a feed-forward neural network with non-linear activations makes it more prone to overfit.

d) Assume that you have two networks without activation functions (just linear hidden layers). The input and output layers of the neural networks are the same. The first one has three hidden layers, while the second only one. These two neural networks can approximate the same class of functions.

e) During training of a neural network with gradient descent, shuffling the data of each batch before performing the training step on this batch can help the network avoid local minima and converge faster to an optimum.

f) Generalization is the property of a learning system to approximate the target output values for inputs that are not included in the testing data-set.

Question 4: Newton's method (3 points)

Show that the formula for determining the square root of a

$$x_n = \frac{1}{2}\left(x_{n-1} + \frac{a}{x_{n-1}}\right)$$

is a special case of Newton's iteration. Generalize the formula for the p-th root of a.

Question 5: Numerical Integration (7 points)

You are asked to approximate the integral $\int_a^b f(x)$ by:

a) Deriving the one-point quadrature rule, starting with the expression $\int_a^b f(x)\,dx = c_1 f(x_1)$.

Given the formula obtained, you are requested to provide answers regarding:

b) A different name that is commonly given to the previously derived formula.

c) The appropriateness of the obtained formula for exact computation of $\int_0^{\pi/2} \cos x dx$.

d) The expected maximum order of polynomial that could be accurately integrated if a Simpson's integration rule was used instead.

Question 6: Monte Carlo Integration (9 points)

Your task is to compute the electrostatic energy of an electric charge being moved in the electric field of a uniformly charged solid sphere:

$$E_{el} = -\int_{r_1}^{r_2} \frac{1}{r^2} dr,$$

where r is the radius from the center of the sphere, and r_1 and r_2 are constants and $r_2 > r_1$.

a) From a computational perspective does it make sense to use monte carlo integration? Briefly explain.

b) We assume that we are drawing the radius from a random uniform distribution $r \sim uniform(r_1, r_2)$. Write down the expectation value of $\frac{-1}{r^2}$ over the distribution, you do not need to compute it.

c) Manipulate the expectation value as such that you reobtain the integral E_{el} you are trying to compute. Then write how you can approximate the integral if you only have a finite number of M samples.

You now try to compute the electrostatic energy through measurements. You have a device that measures the electric field at random distances. However, your measurements are not uniformly distributed but according to the following probability density function, due to the device:

$$P(r) \sim C \cdot r^2$$

Where C is a constant.

d) Calculate C such that equation is a valid probability density function. Explicitly show that the conditions are met.

Numerical Problems

Question 7: Lagrange Interpolation (8 points)

You sample the function $f(x) = \frac{-2}{3}x^3 + \frac{5}{3}x^2 + 4x$ at the points $\{x_i\} = \{0, 1, 3\}$ and wish to compute the Lagrange Interpolating polynomial for this dataset.

a) Compute all the Lagrange basis functions for the above data.

b) Construct the Lagrange interpolation of the above data.

c) Interpolate the above data at $x = 2$ using your previous results and compute the interpolation error. How would your interpolation error change if you sampled a fourth point?

Question 8: Gram-Schmidt orthonormalization (5 points)

Assume you are given a set of functions $\varphi = \{x, \sin(x)\}$ on the interval $[-\pi, \pi]$.

a) Write down the appropriate condition for orthonormality between two functions ϕ_i and ϕ_j and orthonormalise the basis if necessary. *Hint: use integration by parts* $\int_a^b u(x)v'(x)dx = [u(x)v(x)]_a^b - \int_a^b u'(x)v(x)dx$ *and* $\int_{-\pi}^{\pi} \sin^2(x)dx = \pi$.

b) What is the main reason to use orthonormal basis functions for interpolations?

Question 9: Backpropagation (10 points)

In this exercise, we are going to perform steps of the back-propagation method in a trivial neural network with a $\tanh$ activation function. The input is two dimensional $x = (x_1, x_2)^T$ while the output is one dimensional $o \in \mathbb{R}$. The neural network architecture is

$$o = \tanh(z), \quad z = w_1x_1 + w_2x_2 + b,$$

where $w_1, w_2, b \in \mathbb{R}$ are the weight parameters.

a) Prove that the derivative of the tanh activation function $f(x) = \tanh(x)$ takes the form

$$f'(x) = 1 - f^2(x),$$

using the fact that

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}.$$

b) Assume that you are given only one data-point $(x_1, x_2, \tilde{o}) = (1, 5, 1.5)$, where $\tilde{o}$ is the target value. The weights are initialized according to $(w_1, w_2, b) = (0.2, 0.07, 0)$. Compute the forward pass. Report values for z and o. Use the approximation $\tanh(0.55) \approx 0.5$.

c) The loss function is given as $\mathcal{L}(w_1, w_2, b) = (o - \tilde{o})^2$. Compute $\frac{\partial \mathcal{L}}{\partial w} = \frac{\partial \mathcal{L}}{\partial z}\frac{\partial z}{\partial w}$ for $w \in \{w_1, w_2, b\}$. After deriving the three analytic expressions, replace values for the sample $(x_1, x_2, \tilde{o}) = (1, 5, 1.5)$.

d) Use the error gradients computed in the previous question to perform one step of the gradient-descent algorithm with a step-size $\eta = 0.01$. The formula of gradient descent is

$$w^{i+1} = w^i - \eta \left.\frac{\partial \mathcal{L}}{\partial w}\right|_{w_1^i, w_2^i, b^i},$$

where i is the iteration number.

e) Did the loss decrease after the gradient-descent step? If yes, is this always the case? If not, what may be the cause of the increase? Use the approximation $\tanh(0.955) \approx 0.75$.

Question 10: Newton's method convergence (9 points)

Consider 4th order polynomials, i.e., $f(x) = a + bx + cx^2 + dx^3 + ex^4$, with $e = 1$. Newton's method is used to find the roots of the functions $f(x)$.

a) What is the order of convergence of Newton's method in the vicinity of simple roots?

b) Assume that α is a root of $f(x)$. Under what **condition** is the convergence order in the vicinity of the root $x = \alpha$ linear?

c) Assume that $-\alpha$ is also a root of $f(x)$. Under what **conditions** is the convergence order in the vicinity of the root $x = -\alpha$ cubic?

d) How many roots does $f(x)$ have? Express the polynomial $f(x)$ in its root factored form $\prod_i (x - \beta_i)$ where β_i's are the roots. Substitute the roots you know.

e) Use the conditions you found in the previous sub-questions and express the remaining root(s) in terms of α.

f) What is/are the convergence rate of Newton's method in the vicinity of the remaining root(s).

Question 11: Romberg Integration (6 points)

You are asked to compute the integral of the simplest form of the so called rational function equation $f(x) = \int_1^2 \frac{1}{x} dx$. You have forgotten that the formula can be directly integrated, its result being the $ln(2) = 0.69314718$ and you try to approximate the integral numerically. You are asked to:

a) Compute its numerical approximation of the integral using the trapezoidal rule on a **single** interval.

b) Proceed to a more accurate computation resorting to a Romberg integration scheme of order 1 and subsequently of order 2, i.e. R_1^1, R_2^1. In the process you build up a table in a form of a tree diagram to keep track of the Romberg computations.

Pseudocode

Question 12: Pseudocode (9 points)

a) The following pseudo-code is supposed to perform numerical integration based on the Simpson's rule of a function $f(x)$ on the interval $[a, b]$ using N subintervals. Find the errors.

Algorithm 1 Simpson's quadrature with bugs

Input:

a, b, {integration interval}
N, {number of subdivisions}

Output:

$\mathcal{I}$, {estimate of the integral}

Steps:

$h = (b - a)/N$
$\mathcal{I}_0 = f(a) + f(b)$
$\mathcal{I}_1 = 0.0$
$\mathcal{I}_2 = 0.0$
for $i \leftarrow 1, \ldots, N$ **do**
  $\tilde{x} = a + hi$
  if i even **then**
    $\mathcal{I}_2 = f(\tilde{x})$
  else
    $\mathcal{I}_1 = f(\tilde{x})$
  end if
  $\mathcal{I} = \mathcal{I} + h(\mathcal{I}_0 + 2\mathcal{I}_2 + 4\mathcal{I}_1)/3$
end for
return $\mathcal{I}$

---

b) The following pseudo-code is supposed to perform least squares regression to compute the coefficients $x \in \mathbb{R}^m$ of the problem $Ax = b$. Find the errors.

Algorithm 2 Least squares regression with bugs

Input:

A, {matrix $A \in \mathbb{R}^{n \times m}$}
b, {vector $b \in \mathbb{R}^n$}

Output:

x, {$x \in \mathbb{R}^m$}

Steps:

Compute the transpose of A, i.e. A^T
Compute $C = AA^T$
Compute C^{-1}
Compute $d = Ab$
Compute $x = C^{-1}d$.
return x

c) Assume you are given a **closed** surface in $\mathbb{R}^3$ described by the relationship $g(x, y, z) = 0$. The interior of this surface is given by $\{(x, y, z)|g(x, y, z) < 0\}$. You assume that your curve lies in the rectangular box $\{(x, y, z)\,|\,x_- \leq x \leq x_+,\ y_- \leq y \leq y_+,\ z_- \leq z \leq z_+\}$. The following pseudo-code is used to approximate the interior volume defined by the surface based on Monte Carlo. Can you spot the errors?

Algorithm 3 Monte Carlo with bugs

Input:

$x_-, x_+, y_-, y_+, z_-, z_+$, {boundaries of the rectangular box}

Output:

V, {approximation of the interior volume}

Steps:

Compute $|\Omega| = (x_+ - x_-)(y_+ - y_-)$
$K = 0$ # counter
for $i \leftarrow 1, \ldots, N$ **do**
  Sample $x_i \sim uniform(-1, 1)$
  Sample $y_i \sim uniform(-1, 1)$
  Sample $z_i \sim uniform(-1, 1)$
  Compute $\tilde{g} = g(x_i, y_i, z_i)$.
  if $\tilde{g} \geq 0$ **then**
    $K = K + 1$
  end if
end for
$V = K|\Omega|$
return V

Good luck!