ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Computational Methods for Engineering Applications I

Spring semester 2015

Prof. P. Koumoutsakos, J. Šukys
ETH Zentrum, CLT F 12, E 11, C 11
CH-8092 Zürich

Exam

Issued: August 14, 2015, 9:00
Hand in: August 14, 2015, 12:00

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

- Clear your desk (no cell phones, cameras, etc.): on your desk you should have your Legi, your pen and your notes. We provide you with the necessary paper and the exam sheets.
- 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**.
- 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.
- The teaching assistants will give you the necessary paper sheets. You are not allowed to use any other paper sheets.
- 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. Whenever you write a C++-compatible pseudo code, include also the associated comments.
- For questions from 11 to 19, always use a new page for answering each new question (not for sub-questions!). On the top-right corner of every page write your complete name and Legi-ID. Unless otherwise noted in the question, you should hand-in your answers on paper!
- 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	8			
2	6			
3	3			
4	12			
5	6			
6	8			
7	6			
8	12			
9	4			
10	15			
11	18			
12	16			
13	12			
14	18			
15	14			
16	18			
17	14			
18	30			
19	20			
Total	240			

This exam sums up to 240 points. This is more than the score you are supposed to achieve; your ultimate goal is to reach 180 points.

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:

Theory

Question 1: Lagrange vs Cubic (8 points)

Compare pros and cons between interpolation using Lagrange and Cubic Splines. Write two features that Lagrange is better than Cubic Splines and vice versa.

Question 2: Least Squares on a circle (6 points)

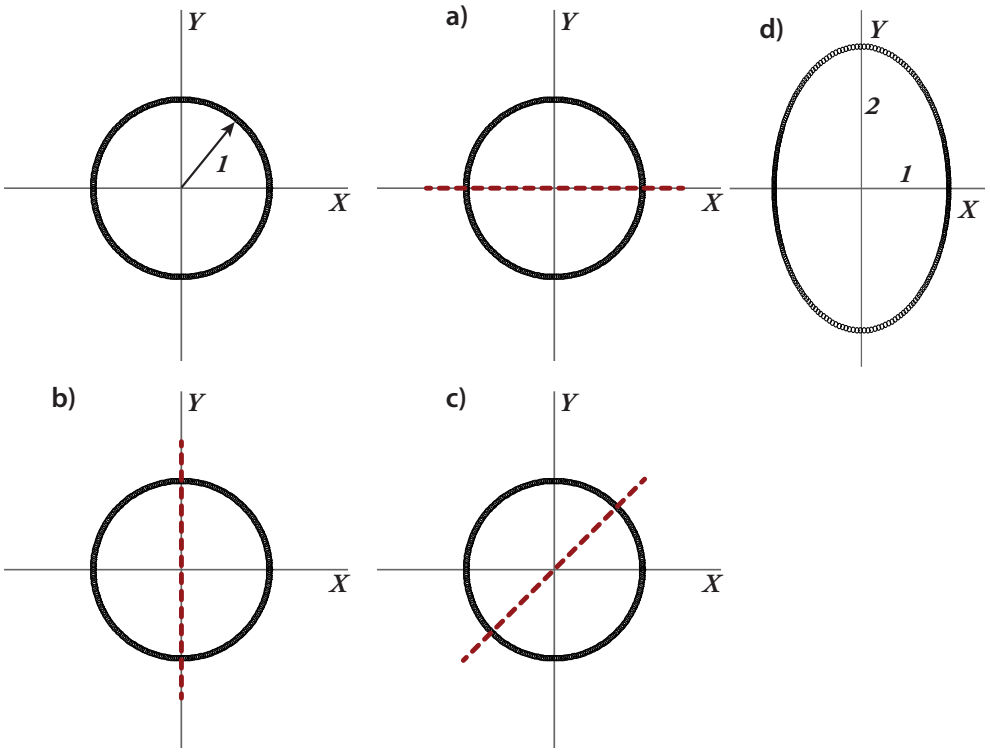


Figure 1: Data points (small black dots) are located on a circle (top left panel). In a) - c), dashed line depicts possible LSQ fittings $y = ax$ of this data. In d) - data on an ellipse.

The data points in a very big dataset ($N \gg 1$) are uniformly (and densely) located on a unit circle with its center in origin (see Fig. 1).

a) What is the correct straight line obtained with LSQ method for $y = ax$ on this data? Choose one variant and explain your answer:

- Horizontal line through origin, Fig. 1 a)
- Vertical line through origin, Fig. 1 b)
- Diagonal line through origin, Fig. 1 c)
- Any line through origin is correct
- Other variant, please elaborate.

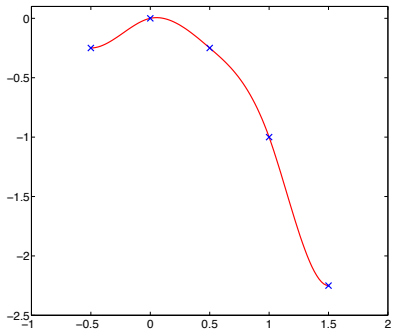
Hint: $a^2 + (b-a)^2 \geq \frac{1}{2}b^2$.

b) How will the answer change if the data is located on a vertically stretched ellipse (see Fig. 1, d)?

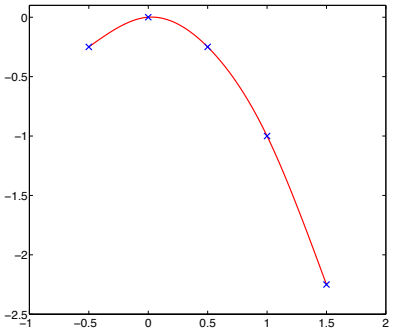
Question 3: Splines with different boundary conditions (3 points)

On the figure below you see three different cubic spline fits for the same data points. The only difference between them is boundary conditions which are the same for the right and left ends in every case. Write down which picture was generated with which boundary conditions:

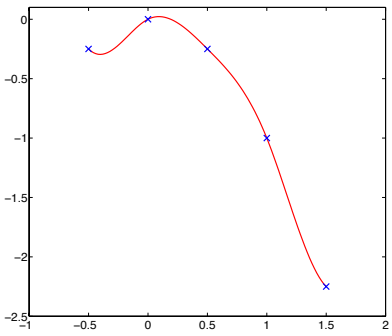
1) natural,
2) clamped,
3) periodic.



(a) Fit a.



(b) Fit b.



(c) Fit c.

Question 4: B-splines (12 points)

a) Give the condition that needs to be true in order to have "clamped" B-splines.

b) What is the maximum level of continuity that quadratic B-splines are guaranteed to have?
(a) C^0
(b) C^1
(c) C^2
(d) C^3

c) Write the knot vector in the case of 6 clamped B-splines constructed with polynomials of 5^{th} degree.

d) For a spline constructed with 6 B-splines of 3^{rd} order polynomials, up to what degree does the spline need to have continuous derivatives? Up to what degree do the B-splines need to have continuous derivatives?

Question 5: Cubic splines compared to cubic B-splines (6 points)

a) Assume N data points are given as $\{x_i, y_i\}$, $i = 1, \ldots, N$. We choose N cubic B-splines $(d = 3)$ with a knot vector which includes all N values x_i. Since we have N unknowns and N data points we can solve the system of equations for B-splines with $S_{d,t}(x_i) = y_i, i = 1, \ldots, N$. What do you expect as a result in comparison with the cubic spline interpolation? Explain your answer.

Question 6: Orthonormal Functions (8 points)

Assume that you are given a set of basis functions $\{\psi_1(x), \psi_2(x)\}$ on the interval $[-1, 1]$ which are orthonormal with respect to the inner product $< \cdot, \cdot >$. We approximated some function $f(x)$ on the interval $[-1, 1]$ using $\psi_1(x)$ and $\psi_2(x)$ as $f(x) \approx \alpha_1\psi_1(x) + \alpha_2\psi_2(x)$.

Answer the following questions:

1) You wish to add a new basis function $\psi_3(x)$. Write a formula to make this function orthonormal with respect to the two others.
2) You wish now to approximate f by the set $\{\psi_1(x), \psi_2(x), \psi_3(x)\}$ as

$$f(x) \approx \tilde{\alpha}_1\psi_1(x) + \tilde{\alpha}_2\psi_2(x) + \tilde{\alpha}_3\psi_3(x).$$

   What is the relation between α_1 and $\tilde{\alpha}_1$, α_2 and $\tilde{\alpha}_2$?

Question 7: Radial Basis Functions (6 points)

Choose an appropriate answer:

1) An advantage of using Radial Basis Functions (RBF) is that one doesn't need to recompute previous coefficients when adding a new basis function. (yes / no)
2) Centers of RBF are always located at the data points. (yes / no)
3) RBF approximation problem can always be represented as a linear system of equations. (yes / no)

Question 8: Overfitting vs underfitting (12 points)

When trying to find the "correct complexity" of a model one often faces problems of overfitting to the data vs underfitting.

a) Briefly explain the difference between these two problems and why the problems happen.

b) On the two pictures below (Figure 3) draw how underfitting could look like on the left plot and how overfitting could look like on the right plot for the given data set.

c) In case of overfitting do you expect a small or a big prediction error for a new data point?

d) Are there methods to avoid overfitting? Name one.

Question 9: Newton method (4 points)

Newton method will always converge to a solution for $f(x) = 0$ on the interval $a \leq x \leq b$ if certain conditions are met. Which of the following is *not* one of these conditions?
(a) f is continuous on the interval $a \leq x \leq b$
(b) $f(a)$ and $f(b)$ have opposite signs
(c) $f''(x)$ does not change sign on the interval $a \leq x \leq b$
(d) $f'(x) = 0$ on the interval $a \leq x \leq b$

Question 10: Trapezoidal rule (15 points)

a) Mark the following statements as true or false.

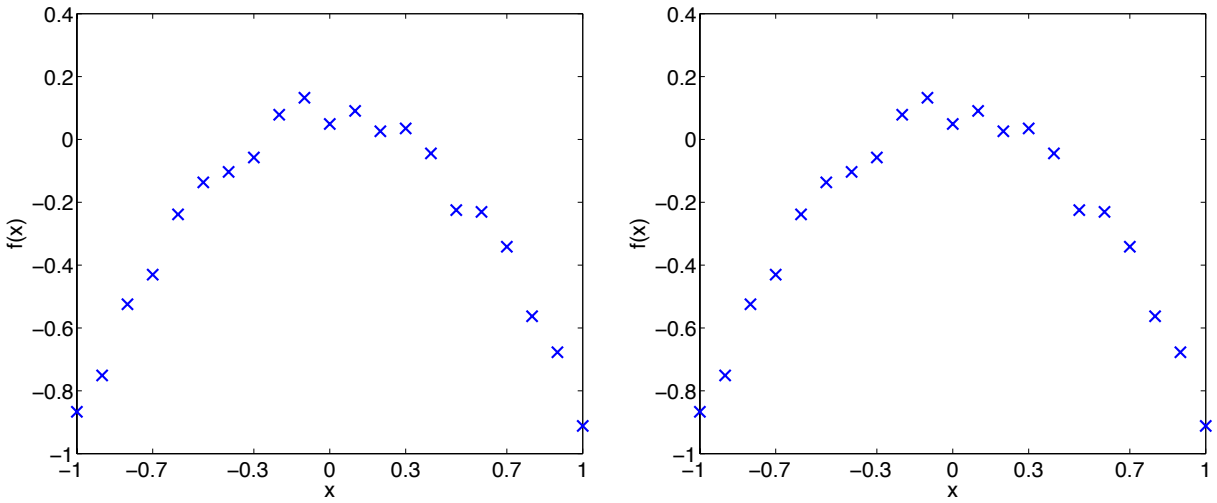


Figure 3: Overfitting vs underfitting.

1. In the trapezoidal rule there are points that are used in two different subintervals. Therefore we must evaluate the function we want to integrate twice at every point.
2. The formula for the trapezoidal rule in an interval $[a, b]$ is given by the equation:

$$I = \int_a^b f(x)dx \approx \frac{h}{2}\left[f(a) + f(b) + \sum_{j=1}^{n-1} f_j\right] \tag{1}$$

3. When h is halved in the trapezoidal rule, half of the function values used with step length h/2 are the same as those used for step length h.
4. The trapezoidal rule with two sub-intervals is exact for integrating at most second order polynomials.
5. The rectangle and the trapezoidal rule have the same order of accuracy.

b) You are interested in buying a very modern apartment in Zurich. The owner has given you a plan of the apartment (see figure). Estimate the area of the entire apartment (*including* areas below furniture and kitchen appliances) in m^2 *using the Trapezoidal rule*. Explicitly show all the intermediate steps you have performed, including the choice of quadrature points. Is your estimate exact?

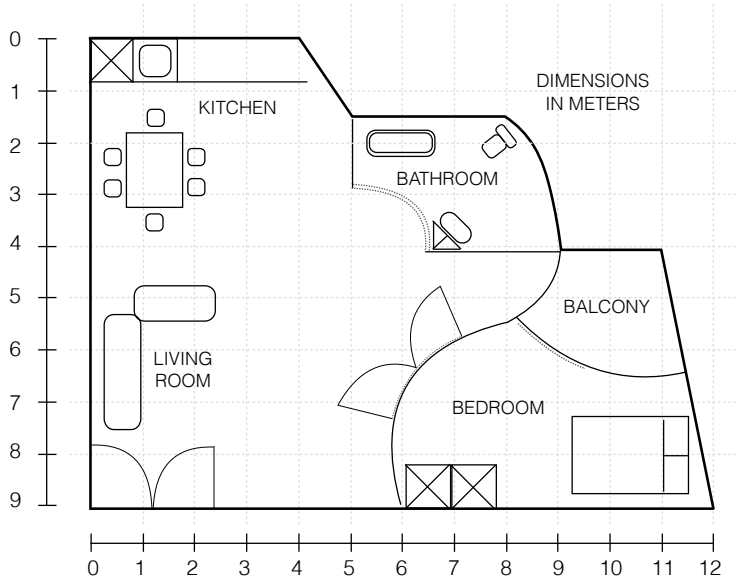


Figure 4: Figure for Q2: Top plan of the apartment of interest.

Numerical Problems

Question 11: Least Squares method (18 points)

You are preparing software for automatic processing of specific scientific data. The data you receive from the experimentalists is a set of points (x_i, y_i) for $i = 1, 2, \ldots, N$ with $N \geq 3$. This data is often perfectly linear, namely $y_i = \alpha x_i + \beta$. But unfortunately errors are inevitable in experiments, and sometimes several points (called "*outliers*") appear to deviate from the straight line: if j-th point is an outlier, then $y_j = \alpha x_j + \beta + e_j$. We will assume that parameter α is known, and you are planning to use Linear Least Squares (LSQ) to identify parameter β of the linear dependence, but the outliers may introduce errors in the fit. Therefore you want to try and minimize the bad influence of the outliers.

a) What method (algorithm) studied in the course can be used to automatically identify the outliers? Explain how.

For the following questions assume there is only one outlier (x_j, y_j) in the dataset.

b) Given α, β, dataset (x_i, y_i), with $y_i = \alpha x_i + \beta$ for $i = 1, 2, \ldots, N$ if $i \neq j$, and an outlier $y_j = \alpha x_j + \beta + e_j$, what is the coefficient β^{LSQ} of the linear fit computed with the Linear Least Squares method? In this special case, the answer should be simplified to depend only on the true value β, the outlier error e_j and the total number of points N.

c) You have an idea to change the LSQ method to the Least Absolute Residuals (LAR). Similarly to LSQ, LAR method can be written as a minimization problem (note that explicit solution for LAR cannot be easily derived analytically, unlike LSQ),

$$\beta^{LAR} = \arg\min_{\hat{\beta}} \sum_{i=1}^{N} |(\alpha x_i + \hat{\beta}) - y_i|. \tag{2}$$

With the same assumptions as in the previous question, what is the coefficient β^{LAR} of the linear fit computed with the Linear Absolute Residuals method? Again, the answer should

be simplified to depend only on the true value β, the outlier error e_j and the total number of points N.
Hint: $\frac{d|c|}{dc} = \frac{c}{|c|}$; for simplicity, you can assume that $e_j > 0$ and that $\beta \leq \beta^{LAR} \leq \beta + e_j$.

d) Compare the exact β with estimated β^{LSQ} and β^{LAR}. Which method is better to deal with outliers?

Question 12: Boundary conditions for cubic splines (16 points)

When constructing a cubic spline from a given set of data points (x_i, y_i), the system of equations that must be solved can be written as:

$$A_i f''_{i-1} + B_i f''_i + C_i f''_{i+1} = D_i \tag{3}$$

where

$$A_i = \frac{\Delta_{i-1}}{6} \tag{4}$$

$$B_i = \left(\frac{\Delta_{i-1} + \Delta_i}{3}\right) \tag{5}$$

$$C_i = \frac{\Delta_i}{6} \tag{6}$$

$$D_i = \frac{y_{i+1} - y_i}{\Delta_i} - \frac{y_i - y_{i-1}}{\Delta_{i-1}} \tag{7}$$

and

$$\Delta_i = x_{i+1} - x_i. \tag{8}$$

You decide that for your boundary conditions, you'll assume constant curvature in the two intervals closest to the end points (i.e., $f''_1 = f''_2$, and $f''_N = f''_{N-1}$). Write down the matrix system that you'd need to solve for constructing the desired spline. Make sure to explicitly write out at least the first three and the last three rows of the system, in terms of $A_1, B_1, \cdots$ and any other variables you deem necessary.

Question 13: Interpolation and Extrapolation (12 points)

You are the number one trader in Silverman Sachs, and your boss often consults you for opinions. Today, your boss comes to you with a set of revenue data from a new startup company. He is interested in predicting the performance of this company in the future.

Figure 5 shows the complete data set $D = \{(t_i, R_i)|i = 1, ..., 9\}$.

a) As a first trial, your boss wants you to use Lagrange Interpolation for the prediction. Write down the expression to evaluate the revenue $\tilde{R}$ given time $\tilde{t}$ using Lagrange Interpolation. Your answer shall be written in terms of t_i and R_i for $i = 1, \ldots, 9$. (Tips: use summation or product notation for simpler form of answer.)

b) Your boss wants to estimate the revenue of this company on March, 2013 using the Lagrange Interpolation. He asks you to comment on the performance of this estimation. Do you expect the estimation to be accurate? Why?

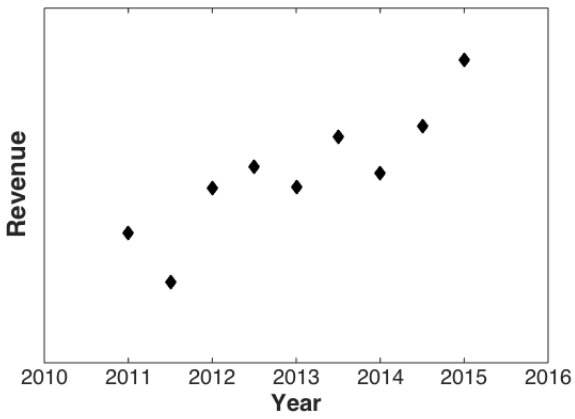


Figure 5: Revenue data for a new startup company.

c) To show off your knowledge base, you introduce two more interpolation methods to your boss:

1. Cubic Splines
2. A set of radial basis functions F_i for $i = 1, \ldots, 9$ with the following form,

$$F_i = \begin{cases} f(|t - t_i|), & \text{if } |t - t_i| < 2 \,\text{years} \\ 0, & \text{otherwise,} \end{cases}$$

where t_i are from the data set D, and $f(\,\cdot\,)$ is some arbitrary function.

Do you expect the estimation for March 2013 using these two methods to be better or worse than the estimation using Lagrange Interpolation? Why?

d) Your boss wants you to predict the revenue of the company in 2018. What would you expect for the performance of the three methods mentioned above? (i.e., Can you foresee what will be the prediction value from each method? If yes, what is the value and why? If no, why not?)

Question 14: Negative square root of 34 (18 points)

Estimate $-\sqrt{34}$ up to accuracy of 2 decimal places by performing two iterations of the Newton's method. To estimate the error of your approximation, assume that exact solution is *not* known. Show all of the intermediate steps involved. You can carry out all computations in fractions, no need to compute final decimal representations of the results.

Question 15: Newton Cotes formulas (14 points)

a) Use the Newton-Cotes formulas for $n = 1$ to compute the coefficients

$$C_k^n = \frac{1}{b-a} \int_a^b l_k^n(x) dx, \quad k = 0, \ldots, n, \tag{9}$$

where $l_k^n(x)$ are Lagrange polynomials in interval $[a, b]$ of degree n:

$$l_k^n(x) = \frac{(x - x_0)(x - x_1) \ldots (x - x_{k-1})(x - x_{k+1}) \ldots (x - x_n)}{(x_k - x_0)(x_k - x_1) \ldots (x_k - x_{k-1})(x_k - x_{k+1}) \ldots (x_k - x_n)}, \tag{10}$$

and where x_i are equidistant points in $[a, b]$. For $n = 1$ that is: $x_0 = a,\ x_1 = b$.

b) Using the computed coefficients C_k^n from (9), derive the resulting numerical integration rule using the Newton-Cotes formula

$$I \approx (b-a)\sum_{k=0}^{n} C_k^n f(x_k). \tag{11}$$

c) How is this integration rule called? Which order polynomials are integrated exactly using this rule? Which order accurate is this rule (just state the order, no need to prove)?

Question 16: Romberg integration with Simpson's rule (18)

The Simpson's integration rule reads as follows,

$$I = \int_a^b f(x)dx \approx I_S = \frac{b-a}{6}\left(f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)\right), \tag{12}$$

and is 5-th order accurate in a given interval $[a, b]$, i.e.

$$I_S - I = \mathcal{O}(h^5).$$

a) Compute the error of the composite Simpson's integration on an interval $[0, 1]$ divided into N subintervals.

b) Similarly to a classical Romberg method, we define I_0^n as an approximation of I with the Simpson's rule using n intervals. Moreover, I_k^n is a higher-order approximation obtained by Richardson extrapolation using I_{k-1}^{2n} and I_{k-1}^n. Using the error expression obtained in previous subquestion, derive a formula for I_k^n which defines the Romberg integration with underlying Simpson's rule, for any arbitrary n and k.

Question 17: Two-point Gauss Quadrature (14 points)

We claim, that the Legendre polynomials $P_n(x)$ of order n can be obtained from the Bonnet's recursion:

$$\begin{aligned} P_0(x) &= 1, \\ P_1(x) &= x, \\ nP_n(x) &= (2n-1)xP_{n-1}(x) - (n-1)P_{n-2}(x), \qquad n > 1. \end{aligned} \tag{13}$$

a) Using Bonnet's recursion, compute the Legendre polynomial $P_2(x)$ of degree 2.

b) What is the main property of Legendre polynomials used to construct Gauss quadrature? Verify, that the obtained polynomial $P_2(x)$ satisfies this property.

c) Compute the locations x_1, x_2 for the two-point Gauss quadrature using $P_2(x)$.

d) Determine the weights w_1, w_2 for this specific case of two-point Gauss quadrature.
Hint: use symmetry of the weights.

Pseudo Codes

Question 18: Adaptive Quadrature using Trapezoidal rule (30 points)

In the class, you learned that an adaptive scheme can be used to improve efficiency of a numerical scheme. The basic idea is that you only refine subintervals which have error higher than your demanded threshold. This can avoid wasting computational power in subintervals that already reach sufficient accuracy. One important issue in this approach is to estimate the error in a subinterval based on a chosen numerical scheme. In practice, calculation of the exact error is not feasible. In this question, you will investigate on optimizing the error estimation method for the trapezoidal rule.

a) For a given subinterval $\{x_i, x_{i+1}\}$, the numerical integration of $I_i = \int_{x_i}^{x_{i+1}} f(x)\,\mathrm{d}x$ using the trapezoidal rule, denoted as I_{T_i}, can be written as

$$I_{T_i} = I_i + \frac{1}{12} f''(x_{i+1/2}) h^3 + O(h^5),$$

where $h = x_{i+1} - x_i$ and $x_{i+1/2}$ is the midpoint between x_i and x_{i+1}, i.e., $x_{i+1/2} = (x_i + x_{i+1})/2$. Assuming the full integral $I = \int_a^b f(x)\,\mathrm{d}x$ is subdivided into N equally spaced subintervals with length h, show that the total error for approximating I using trapezoidal rule on each subinterval is in the order of h^2, i.e.,

$$I_T = \sum_{i=0}^{N-1} I_{T_i} = I + O(h^2).$$

b) Following Richardson's idea for error estimation, let the target quantity $G = I$, and the approximation method $G(h) = I_T(h)$, the trapezoidal rule as a function of subinterval length h. Since the trapezoidal rule is a higher order approximation scheme, the expression for error estimation shown in the notes, $\epsilon(h/2) \approx G(h/2) - G(h)$, is an overestimation of the actual error. Please improve the error estimation based on the expression $I_T = \sum_{i=0}^{N-1} I_{T_i} = I + O(h^2)$.

c) Write a C++ compatible pseudo-code for a subroutine called *AdaptiveTrapez* that estimates a 1D integral $I = \int_a^b f(x)\,\mathrm{d}x$ by applying adaptive Trapezoidal rule for given tolerance ε. The input of *AdaptiveTrapez* includes the boundaries a and b and the tolerance ε. The output of *AdaptiveTrapez* is a single scalar estimate of the integral I. In *AdaptiveTrapez*, you may call a function evaluation subroutine *F(x)* that takes in one scalar input x and return a scalar that is the function value evaluated at x. You may write any necessary additional subroutines and use them inside *AdaptiveTrapez*.

Question 19: Monte Carlo Sampling (20 points)

a) Write a C++-compatible pseudo code which calculates the integral of the multidimensional function

$$f(x_1, \ldots, x_5) = \prod_{i=1}^{5} \cos\left(x_i + \frac{\pi}{4} i\right) \tag{14}$$

in the box $x_i \in [-1.5, 1)$ using a simple Monte Carlo integration scheme. Assume that you have a function `random()`, which returns a uniformly distributed random number in interval $[0, 1)$.

b) Write a C++-compatible pseudo code that performs error estimation of the Monte Carlo sampling. In the pseudo code, you can access any variables that you computed in the previous sub-problem.

c) How does the error of the MC approximation scale with respect to the number of sample points M? Assume you perfomed Monte Carlo integration for $M = 100'000$ samples. If you want to reduce the error by a factor of 4, how many samples M do you have to take?

d) Is the integral estimated by the Monte Carlo sampling always within the error bars? Explain why.

Good luck!