Instructions

- You may rip off the last two pages as soon as you sit down.
- There are 43 marks available. It will be marked out of 40.
- No aides.
- Turn off all electronic media and store them under your desk.
- You may ask only one question during the examination: “May I go to the washroom?”
- Asking any other question will result in a deduction of 5 marks from the exam grade.
- If you think a question is ambiguous, write down your assumptions and continue.
- Do not leave during first hour or after there are only 15 minutes left.
- Do not stand up until all exams have been picked up.
- There are questions on both sides of the pages.
- If a question only asks for an answer, you do not have to show your work to get full marks; however, if your answer is wrong and no rough work is presented to show your steps, no part marks will be awarded.
- Answer the questions in the spaces provided. If you require additional space to answer a question, please use another page that is more blank, but refer the marker to that page.

1. (4 points) What are the seven "tools" we have been using and will continue to use throughout this course? -1 for each incorrect answer or missing tool.

2. (3 points) Add the following two numbers shown in the double-precision floating-point representation:

```
4047300000000000     0 10000000100 01110011...0
c003000000000000     1 10000000000 00110000...0
```

   Each row has the same number twice, only the first is in the hexadecimal representation, and the second is in the binary representation. You may give your answer in either hexadecimal or in binary, as you wish. If all subsequent digits or bits are zero, just write `0···`.

3. (3 points) The error term for the centered three-point derivative is $-\frac{1}{6}f^{(3)}(\xi)h^2$. If $h = 0.1$, what is, in absolute value, the maximum possible error when approximating the derivative of $\sin(2x)$ for some value of $0 \leq x \leq 2\pi$?

   What is, in absolute value, the maximum possible error when approximating the derivative of $\tan(x)$ on that same interval?

4. (2 points) Given that $0 < x_1 < x_2 < \cdots < x_n$ and if $w_1 + w_2 + \cdots + w_n = 1$ is a convex combination, is it possible to find one value of x such that e^{-x} equals

$$w_1 e^{-x_1} + w_2 e^{-x_2} + \cdots + w_n e^{-x_n}?$$

If yes, what are the restrictions on x, and if no, give a counter example.

5. (4 points) Show that the error of the approximation of the first derivative

$$f^{(1)}(x) \approx \frac{f(x+h) - f(x-h)}{2h}$$

is equal to $-\frac{1}{6}f^{(3)}(\xi)h^2$ where $x - h \leq \xi \leq x + h$. You must show and explain each step in your derivation and calculations. You should use 2^{nd}-order Taylor series for $f(x+h)$ and $f(x-h)$. Show where you use the intermediate-value theorem. Be sure to explain how we know that the resulting ξ lies on the interval $x - h \leq \xi \leq x + h$. We will start you with the following, where we know that for $x \leq \xi_+ \leq x + h$:

$$f(x+h) = f(x) + f^{(1)}(x)h + \frac{1}{2}f^{(2)}(x)h^2 + \frac{1}{6}f^{(3)}(\xi_+)h^3$$

6. (2 points) Write down the system of linear equations that need to be solved to find the least-squares best-fitting linear polynomial passing through the points $(0, 1)$, $(1, 3)$ and $(4, 5)$. You don't have to solve this system of linear equations; you only need to write the augmented matrix of numbers that must be solved to find the coefficients of the linear polynomial.

7. (4 points) Use Gaussian elimination with partial pivoting to reduce the following augmented matrix to row-echelon form.

$$\left(\begin{array}{rrr|r} -5.0 & 3.0 & 2.0 & -13.0 \\ 4.0 & -4.0 & -8.8 & 30.4 \\ -0.5 & 4.3 & -1.8 & 8.7 \end{array}\right)$$

You do not have to find the solution to this system of linear equations (that is, you do not have to apply backward substitution); however, if you wish to check your answer, the solution is $\begin{pmatrix} 2 \\ 1 \\ -3 \end{pmatrix}$.

8. (4 points) You are performing forensic engineering on a system where a sequence of readings (including times and voltages) are $(7234, 7.5)$, $(7236, 6.2)$, $(7238, 5.2)$ and $(7240, 4.3)$. The last reading is the last recorded reading before the system failed. You would like to estimate when the voltage was five volts (5.0V). How would you approach this problem? What steps would you take to minimize any numeric error in your estimation?

   Hint: There are at least two very different solutions that will get full marks.

9. (4 points) Assume that we want to apply Simpson's rule for approximating the integral from a to b by breaking the interval into n equally-sized sub-intervals where $x_k = a + hk$ and $h = \frac{b-a}{n}$, evaluating each interval at the end-points and at the mid-point. To do the error analysis, we must sum all the errors to get

$$\sum_{k=1}^{n} -\frac{1}{90} f^{(4)}(\xi_k) h^5$$

   where $x_{k-1} \le \xi_k \le x_k$. Show how this formula can be simplified to

$$-\frac{(b-a)}{90} f^{(4)}(\xi) h^4$$

   and describe how we know that $a \le \xi \le b$. Recall that we can multiply by $1 = \frac{n}{n}$.

10. (2 points) You know that the function $f(x) = e^{-x}\sin(14x)+e^{-2x}\cos(12x)$ has the values $f(0.89) = -0.09593832566$, $f(0.90) = -0.01845240662$ and $f(0.91) = 0.5730607040$. Given this information, how would you proceed to find a better approximation of the root with one application of one algorithm? Justify why you chose the method you did. Do not actually try to apply the method you are recommending.

11. (2 points) Apply one step of the secant method given the two approximations of the root $(1.7, 0.6)$ and $(1.9, 0.2)$. Suppose that the value of the function at this next point is 0.1. What is the next approximation of the root using the secant method?

12. (2 points) Suppose you have applied the techniques in class to find a least-squares best-fitting quadratic through the last 11 points, each measuring electric current, and you get that the least-squares best-fitting quadratic is $1.50 + 0.36s + 0.18s^2$, as described in class. If these readings are being taken once per one hundred seconds, what is the best approximation of the charge passing through that point in the most recent time step?

    Remember, in class, we shifted and scaled so that the most recent reading is at 0, the previous reading is at -1, etc.

13. (3 points) Apply one step of Jacobi's method to find a better approximation of a solution to the system of linear equations

$$\begin{pmatrix} 5 & 1 & -1 & 0 \\ 1 & 10 & 2 & 1 \\ -1 & 2 & 20 & -2 \\ 0 & 1 & -2 & 10 \end{pmatrix} \mathbf{u} = \begin{pmatrix} 2 \\ 4 \\ -2 \\ 3 \end{pmatrix} \quad \text{with } \mathbf{u}_0 = \begin{pmatrix} 0.4 \\ 0.4 \\ -0.1 \\ 0.3 \end{pmatrix}.$$

Recall $A\mathbf{u} = \mathbf{b}$ is equivalent to $(A_{\text{diag}} + A_{\text{off}})\mathbf{u} = \mathbf{b}$ and a diagonal matrix with non-zero entries on the diagonal is easily invertible.

14. (4 points) The non-linear system of equations

$$x + xy + 2xz = -2, \quad xy - 3y + 4yz = 2, \quad xz - yz - z = 1$$

has a solution close to $\mathbf{u}_0 = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} = \begin{pmatrix} 5 \\ 0 \\ 0 \end{pmatrix}$. Write down the system of linear equations as an augmented matrix of numbers that must be solved to find the solution $\Delta\mathbf{u}_0$ and explain how you will use the solution $\Delta\mathbf{u}_0$ to get the next approximation $\mathbf{u}_1$.

Floating-point representations: $\pm$`EENMMM` represents $\pm$`N.MMM` $\times 10^{\texttt{EE}-49}$ and the 64 bits `seeeeeeeeeeebbbbbb`$\cdots$`b` represents

$$(-1)^{\texttt{s}}\mathbf{1}.\texttt{bbbbbb}\cdots\texttt{b} \times 2^{\texttt{eeeeeeeeeee}-01111111111}$$

where `0b01111111111` $= 1023 =$ `0x3ff`. Recall 1 is `+491000` or `0x3ff0000000000000`.

0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111

Given n real or complex numbers or vectors $x_1, \ldots, x_n$ and n real or complex numbers $w_1, \ldots, w_n$, then $\sum_{k=1}^{n} w_k x_k$ is:

1. a linear combination of the x-values if there are no restrictions on the weights,
2. a weighted average if $\sum_{k=1}^{n} w_k = 1$, and
3. a convex combination if the weights form a weighted average and each $w_k \geq 0$.

Fixed-point theorem: To approximate a solution to $x = f(x)$, choose x_0 and let $x_k \leftarrow f(x_{k-1})$.
Gaussian elimination with partial pivoting: This is the Gaussian elimination algorithm but always swapping appropriate rows so that the largest entry in absolute value is in the pivot position (the row that will be used to eliminate entries in that column in subsequent rows).
***n^{th}*-order Taylor series:** If h is small, expanding around x yields:

$$f(x+h) = \left(\sum_{k=0}^{n} \frac{1}{k!} f^{(k)}(x)\, h^k \right) + \frac{1}{(n+1)!} f^{(n+1)}(\xi)\, h^{n+1}$$

where $x \leq \xi \leq x + h$. Otherwise, if x is close to x_0, expanding around x_0 yields:

$$f(x) = \left(\sum_{k=0}^{n} \frac{1}{k!} f^{(k)}(x_0)\,(x - x_0)^k \right) + \frac{1}{(n+1)!} f^{(n+1)}(\xi)\,(x - x_0)^{n+1}$$

where $x_0 \leq \xi \leq x$.

The examples of binary search and interpolation search are not required for this course: they are provided as examples of different bracketing algorithms.

```
double horner( double       const a[],
               unsigned int const degree,
               double       const x ) {
    // The coefficient of x^k is a[k]
    double result{ a[degree] };

    for ( std::size_t k{degree - 1}; k < degree; --k ) {
        result = result*x + a[k];
    }

    return result;
}
```

Noise: Averaging noisy values with zero bias mitigates the effect, while differentiating noisy values magnifies the effect. Use interpolating polynomials if the data is accurate and precise, but use least squares best-fitting polynomials if the data is accurate but not precise (that is, the data has significant noise). If the data is not accurate, we cannot recover the underlying signal.

Evaluating interpolating polynomials: For interpolating between t_k and t_{k-1} where t_k is the time of the most recent data point, shift and scale to $\ldots, -2.5, -1.5, -0.5$ and 0.5 to ensure that $-0.5 < \delta < 0.5$ to evaluate the polynomial at the point $\frac{t_{k-1}+t_k}{2} + \delta h$ where h is the time step between readings. Note, you do not have to know these formulas explicitly; rather, you must understand the idea behind deriving these. For example, why to we shift and scale so that our choice of δ is such that $|\delta| < 0.5$.

Derivatives:

Centered three-point:

$$f^{(1)}(x) = \frac{f(x+h) - f(x-h)}{2h} - \frac{1}{6} f^{(3)}(\xi)\, h^2$$

Backward two-point:

$$y^{(1)}(t) = \frac{y(t) - y(t-h)}{h} + \frac{1}{2} y^{(2)}(\tau)\, h$$

Backward three-point:

$$y^{(1)}(t) = \frac{3y(t) - 4y(t-h) + y(t-2h)}{2h} + \frac{1}{3} y^{(3)}(t)\, h^2 + \mathrm{O}\left(h^3\right)$$

Second derivatives:

Centered three-point:

$$f^{(2)}(x) = \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} - \frac{1}{12} f^{(4)}(\xi)\, h^2$$

Backward three-point:

$$y^{(2)}(t) = \frac{y(t) - 2y(t-h) + y(t-2h)}{h^2} + y^{(3)}(\tau)\, h$$

Backward four-point:

$$y^{(2)}(t) = \frac{2y(t) - 5y(t-h) + 4y(t-2h) - y(t-3h)}{h^2} + \frac{11}{12} y^{(4)}(t)\, h^2 + \mathrm{O}\left(h^3\right)$$

Integrals:
Two-point (trapezoidal rule):

$$\int_{x_{k-1}}^{x_k} f(x)\,\mathrm{d}x = \left(\frac{1}{2}f(x_{k-1}) + \frac{1}{2}f(x_k)\right)h - \frac{1}{12}f^{(2)}(\xi)\,h^3$$

Centered four-point:

$$\int_{x_{k-1}}^{x_k} f(x)\,\mathrm{d}x = \left(-\frac{1}{24}f(x_{k-2}) + \frac{13}{24}f(x_{k-1}) + \frac{13}{24}f(x_k) - \frac{1}{24}f(x_{k+1})\right)h - \frac{11}{720}f^{(4)}(t_k)\,h^5 + \mathrm{O}\left(h^6\right)$$

Simpson's rule:

$$\int_{x_{k-1}}^{x_{k+1}} f(x)\,\mathrm{d}x = \left(\frac{1}{6}f(x_{k-1}) + \frac{4}{6}f(x_k) + \frac{1}{6}f(x_{k+1})\right)(2h) - \frac{1}{90}f^{(4)}(\xi)\,h^5$$

Backward three-point (half Simpson's rule):

$$\int_{t_{k-1}}^{t_k} y(t)\,\mathrm{d}x = \left(\frac{5}{12}y(t_k) + \frac{8}{12}y(t_{k-1}) - \frac{1}{12}y(t_{k-2})\right)h - \frac{1}{24}y^{(3)}(t_k)\,h^4 + \mathrm{O}\left(h^5\right)$$

Backward four-point:

$$\int_{t_{k-1}}^{t_k} y(t)\,\mathrm{d}x = \left(\frac{9}{24}y(t_k) + \frac{19}{24}y(t_{k-1}) - \frac{5}{24}y(t_{k-2}) + \frac{1}{24}y(t_{k-3})\right)h + \frac{19}{720}y^{(4)}(t_k)\,h^5 + \mathrm{O}\left(h^6\right)$$

As Simpson's rule spans two time intervals, it is less useful, but it is interesting with its comparison with the trapezoidal rule applied twice versus one application of Simpson's rule. It also corresponds with the 4th-order Runge Kutta method. Any integral formula can be applied repeatedly on the interval $[a, b]$ by dividing the interval into n equally-spaced sub-intervals of width $h = \frac{b-a}{n}$ and then setting $x_k = a + kh$ or $t_k = a + kh$.

Least squares: In general, if we want to find the best approximation of an n-dimensional vector $\mathbf{y}$ by a linear combination of m vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$ (where $m < n$), we create the matrix $V = (\mathbf{v}_1 \cdots \mathbf{v}_m)$ and solve $V^\top V \boldsymbol{\alpha} = V^\top \mathbf{y}$. More specific to this course, having shifted and scaled the n most recent t-values onto $0, -1, -2, \ldots, -n+1$, with y values $\mathbf{y} = (y_k, y_{k-1}, y_{0-2}, \ldots, y_{k-n+1})$, we solve $V^\top V \boldsymbol{\alpha} = V^\top \mathbf{y}$ for the coefficients of the least-squares best-fitting polynomial, generally of degree one (linear or $\alpha_1 t + \alpha_0$) or two (quadratic or $\alpha_2 t^2 + \alpha_1 t + \alpha_0$). We can find the $2 \times n$ or $3 \times n$ matrix to calculate $\boldsymbol{\alpha} = \left(V^\top V\right)^{-1} V^\mathrm{T} \mathbf{y}$.

Value being estimated	Linear estimation
$y(t_k)$	α_0
$y(t_k + h)$	$\alpha_0 + \alpha_1$
$y^{(1)}(t_k)$	α_1 / h
$\int_{t_k - h}^{t_k} y(\tau) \mathrm{d}\tau$	$(\alpha_0 - \alpha_1/2)h$
$\int_{t_k}^{t_k + h} y(\tau) \mathrm{d}\tau$	$(\alpha_0 + \alpha_1/2)h$

Value being estimated	Quadratic estimation
$y(t_k)$	α_0
$y(t_k + h)$	$\alpha_0 + \alpha_1 + \alpha_2$
$y^{(1)}(t_k)$	α_1 / h
$y^{(2)}(t_k)$	$2\alpha_2 / h^2$
$\int_{t_k - h}^{t_k} y(\tau) \mathrm{d}\tau$	$(\alpha_0 - \alpha_1/2 + \alpha_2/3)h$
$\int_{t_k}^{t_k + h} y(\tau) \mathrm{d}\tau$	$(\alpha_0 + \alpha_1/2 + \alpha_2/3)h$

Root finding:

- Bisection: Let $m_k \leftarrow \frac{a_k + b_k}{2}$ and update that endpoint that has the value of the function have the same sign as $f(m_k)$.
- Newton's method: $x_{k+1} \leftarrow x_k - \frac{f(x_k)}{f^{(1)}(x_k)}$.
- Secant method: $x_{k+1} \leftarrow x_k - \frac{f(x_k)}{\frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}}} = x_k - \frac{f(x_k)(x_k - x_{k-1})}{f(x_k) - f(x_{k-1})}$.
- Inverse quadratic interpolation: Find the constant coefficient of the polynomial interpolating (y_{k-2}, x_{k-2}), (y_{k-1}, x_{k-1}) and (y_k, x_k).
- Newton's method in n dimensions: Given the approximation $\mathbf{u}_k$ to $\mathbf{f}(\mathbf{u}) = \mathbf{0}$, solve $J(\mathbf{f})(\mathbf{u}_k)\Delta\mathbf{u}_k = -\mathbf{f}(\mathbf{u}_k)$ and then let $\mathbf{u}_{k+1} \leftarrow \mathbf{u}_k + \Delta\mathbf{u}_k$.