**Instructions**

- You may rip off the last two pages as soon as you sit down.
- There are 33 marks available. It will be marked out of 30.
- 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. (3 points) What are the seven tools we have been using and will continue to use throughout this course? $-0.5$ for each incorrect answer or missing tool.

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

```
b5830000000000       1 01101011000 00110...0
b598000000000000     1 01101011001 10000...0
```

Each row has the same twice, only the first is in the hexadecimal representation, and the second is in the binary representation. Your answer must be in the same representation, but you may give your answer in either hexadecimal or in binary, as you wish.

3. (2 points) From Taylor series, given that $\sin(1.1) = \sin(1) + 0.1\cos(1) - 0.005\sin(\xi)$ where $1 \le \xi \le 1.1$, what is the maximum possible value of the absolute error of $\sin(1) + 0.1\cos(1)$ as an approximation to $\sin(1.1)$ and what is the maximum possible value of the relative error of this approximation?

   Do not try to calculate $|\sin(1.1) - (\sin(1) + 0.1\cos(1))|$, but rather maximize the error term of the Taylor series.

4. (1 point) Given that you know that the weights $w_1$ through $w_5$ form a convex combination, what restrictions are there as to the possible value of the convex combination of the values of $\sin(0.3), \sin(0.4), \sin(0.5), \sin(0.6)$ and $\sin(0.7)$?

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 $x - h \le \xi \le x + h$. You must show and explain each step in your derivation and calculations. You should use $2^{\text{nd}}$-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 $\xi$ lies on the interval $x - h \le \xi \le x + h$. We will start you with the following, where we know that for $x \le \xi_+ \le x + h$:

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

6. (1 point) Write down the system of linear equations you would have to solve to find the best approximation of the vector $\mathbf{b} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ as a linear combination of the two vectors $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ and $\mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$. If your system of linear equations contains matrix-matrix or matrix-vector products, you don't have to multiply them out. Just be sure to define the entries of any matrix and how you would use that matrix to create a system of linear equations.

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

$$\left(\begin{array}{rrr|r} 2.8 & -5.2 & -7.0 & -7.8 \\ 2.1 & 1.1 & 3.0 & 8.9 \\ -7.0 & 3.0 & 5.0 & 2.0 \end{array}\right)$$

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 $\begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix}$.

8. (4 points) You have a system that is sampling the speed of a vehicle 10 times per second, and the speedometer is reasonably accurate. You'd like to estimate how far the vehicle travelled in the last 0.1 seconds, but you only track the last three speeds $s_0, s_{-1}$ and $s_{-2}$, where $s_0$ is the most recent reading. Normally, you would use the half-Simpson's rule.

   Suppose, however, that when taking the most recent reading, $s_0$, there was noise in the signal, so that that most recent reading and the last two readings are $s_0 = 2.7$ m/s, $s_{-1} = 153.2$ m/s and $s_{-2} = 2.3$ m/s. Clearly, the reading $s_{-1}$ is incorrect, but you'd still like an estimate as to how far the vehicle travelled in the last 0.1 seconds. This is all the data you have: what would you use as your best estimate as to how far the vehicle moved in the last 0.1 seconds using $s_0$ and possibly $s_{-2}$? The vehicle can be assumed to be not rapidly changing its acceleration in the last second. Justify your answer.

   **Hint**: The numbers are set up so that the most reasonable approximations require negligible arithmetic and can be deduced geometrically without finding explicit interpolating polynomials.

9. (3 points) Assume that we want to apply the trapezoidal formula for approximating the integral from $a$ to $b$ by breaking the interval into $n$ equally-sized sub-intervals where $x_k = a + hk$ and $h = \frac{b-a}{n}$. To do the error analysis, we must sum all the errors to get

$$\sum_{k=1}^{n} -\frac{1}{12} f^{(2)}(\xi_k) h^3$$

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

$$-\frac{(b-a)}{12} f^{(2)}(\xi) h^2$$

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

10. (2 points) Apply one step of Newton's method to find a better approximation of the root of $x^2 - x - 1$ when the initial approximation is $x_0 = 2$.

11. (2 points) Apply one step of the bisection method to find a better approximation of the root of $x^3 - 0.1$ given that you know the root is on the interval $[0, 1]$. Which end-point will you update?

12. (1 point) Suppose you have applied the techniques in class to find a least-squares best-fitting quadratic through the last 10 points, and you get that the least-squares best-fitting quadratic is $1.50 + 0.36t + 0.18t^2$, as described in class. What is the best approximation of the value of the system being sampled one step into the future?

13. (3 points) For polynomial interpolation, for each of these steps we take in shifting and scaling, we do so with a particular purpose. The purposes may be one or more of the following:

    1. Reduce the number of floating-point operations.
    2. Reduce the condition number when we find the interpolating polynomial.
    3. Reduce the likelihood of subtractive cancellation occurring while evaluating the polynomial at $\delta$.

    For each of the following, write one and only one number next to it. For some, there are multiple answers, but you only need to pick one, and both answers will be accepted.

    Shift to zero. ______________
    Scale to integer or half-integer values. ______________
    Use Horner's rule to evaluate the polynomial. ______________

    You do not have to use each number once.

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

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

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

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 $x_k \leftarrow f\left(x_{k-1}\right)$.

**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:

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

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

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

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 $\delta$ 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.
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 |
|---|---|
| $y(t_k)$ | $\alpha_0$ |
| $y(t_k + h)$ | $\alpha_0 + \alpha_1$ |
| $y^{(1)}(t_k)$ | $\alpha_1/h$ |
| $\int_{t_k-h}^{t_k} y(t)\mathrm{d}t$ | $(\alpha_0 - \alpha_1/2)h$ |
| $\int_{t_k}^{t_k+h} y(t)\mathrm{d}t$ | $(\alpha_0 + \alpha_1/2)h$ |

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

**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: $x_{k+1} \leftarrow x_k - \frac{f(x_k)}{\frac{f(x_k)-f(x_{k-1})}{x_k-x_{k-1}}} = x_k - \frac{f(x_k)(x_k-x_{k-1})}{f(x_k)-f(x_{k-1})}$.