Name:	______________________ family name ______ given name	(please print)
SFU ID:	______________________ student number ______ SFU-email	@sfu.ca

[1] **1.** (a) Let $f(x)$ be a continuous function on an interval $[a,b]$ with $f(a)\cdot f(b) < 0$ and suppose that the bisection method is applied to $f(x)$. If $a_1 = 5$ and $b_1 = 10$, define the first iterate

$$p_1 = \frac{5+10}{2} = 7.5$$

[2] (b) Have you computed p_1 robustly? If not, provide and alternate computation for p_1 which is robust.

Not robust $p_1 = \frac{a+b}{2}$

Robust $p_1 = a + \frac{b-a}{2} \Rightarrow p_1 = 5 + \frac{10-5}{2} = 7.5$

[3] (c) How many iterations of the bisection method are needed to find the root to a tolerance of 10^{-5}?

$|p_n - p| \leq \frac{b-a}{2^n}$

$\frac{10-5}{2^n} \leq 10^{-5}$

$2^n \geq 5\cdot 10^5$

$n \geq \log_2(5\cdot 10^5)$

$n \geq 18.93$

$\therefore\ n \geq 19$

[1] (d) With what order, α, and with what constant, λ, does the bisection method converge?

$\alpha = 1$

$\lambda = 1/2$

2. Answer **briefly** but include a **justification** for your answer.

[2] (a) A "Megaflop" stands for 10^6 floating point operations. I am computing the LU factorization of a 3000 by 3000 matrix on my laptop. It takes about 13 seconds. Therefore my laptop is capable of performing roughly how many Megaflops per second?

LU factorization takes $\frac{2}{3}(3000)^3$ flops $= 1.8\times10^{10}$ flops

I can do $\frac{18{,}000 \text{ Megaflops}}{13 \text{ seconds}} = \underline{1384}$ Megaflops/second. $= 18{,}000$ Megaflops.

[2] (b) If it takes 13 seconds to find the LU factorization of a 3000 by 3000 matrix, approximately how long will it take to find the LU factorization of a 1500 by 1500 matrix?

$O(n^3) \therefore \left(\frac{1}{2}\right)^3$ time $= \frac{1}{8}$

$13 \cdot \frac{1}{8} = \underline{1.625 \text{ seconds}}$

[2] (c) The irrational number, $\pi = 3.14159265358979...$ cannot be represented exactly in a decimal (base, $\beta = 10$) floating point system. Is there an integer base floating point system in which π can be represented exactly? If yes, write down this representation. If no, explain why not.

No, because π is irrational $\therefore$ cannot be represented as a terminating decimal in any base.

[2] (d) If the floating point standard for a computer system had a base of 2 and a precision of 45, what is the maximum distance between two numbers which cannot be told apart? Assuming rounding arithmetic.

This is machine epsilon for a 45 digit precision system

$$\varepsilon = \frac{1}{2}\beta^{1-k} = \frac{1}{2}\cdot 2^{1-45} = 2^{-45} = 2.8\times10^{-14}$$

[2] (e) We solve the system $Ax = b$ with $\kappa(A) = 100$ on a computer with a machine epsilon of 10^-8. The exact solution is $x = [4, 3, 2, 1]^T$ and our computed solution is $x_c = [4.04, 3.97, 1.98, 1.01]^T$. Is the algorithm used satisfactory or not?

Note the relative error in the calculation.

$$\frac{4.04-4}{4} = \frac{0.04}{4} = 0.01 = 10^{-2}$$

We expect degradation to $\varepsilon\kappa(A) = (10^{-8})(100) = 10^{-6}$

We got much worse $\Rightarrow$ Algorithm is NOT satisfactory

[7] **3.** (a) The following are equivalent formulas for finding one of the two the roots of a quadratic:

$$s_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \quad \text{and} \quad s_2 = \frac{-2c}{b + \sqrt{b^2 - 4ac}}.$$

Calculate s_1 and s_2 for $x^2+(53.21)x-1 = 0$ using 4-digit chopping arithmetic. Compare their relative errors from the "exact" positive solution $s = 0.01878682682086$.

$$S_1 = \frac{-53.21 + \sqrt{(53.21)^2 - (4)(1)(-1)}}{(2)(1)} \qquad S_2 = \frac{-2(-1)}{53.21 + \sqrt{(53.21)^2 - (4)(1)(-1)}}$$

$$S_1 = \frac{-53.21 + \sqrt{2831 + 4.000}}{2.000} \qquad S_2 = \frac{2.000}{53.21 + \sqrt{2831 + 4}}$$

$$S_1 = \frac{-53.21 + \sqrt{2835}}{2.000} \qquad S_2 = \frac{2.000}{53.21 + \sqrt{2835}}$$

$$S_1 = \frac{-53.21 + 53.24}{2.000} \qquad S_2 = \frac{2.000}{53.21 + 53.24}$$

$$S_1 = \frac{0.03000}{2.000} \qquad S_2 = \frac{2.000}{106.4}$$

$$S_1 = 0.01500 \qquad S_2 = 0.01879$$

Relative error $(S_1) = \dfrac{|0.01500 - 0.01878682\cdots|}{|0.0187868\cdots|} = 0.201$

For S_2 $\dfrac{|0.01879 - s|}{s} = 0.0001689$

[2] (b) Provide an explanation for why one method was better than the other.

Method S_2 is better because method S_1 has <u>cancellation</u> error.

[2] **4.** (a) Identify all fixed points for the iteration $x_n = g(x_{n-1})$ with $g(x) = \frac{2}{3}x + \frac{1}{x}$.

$x = \frac{2}{3}x + \frac{1}{x}$ $\quad x^2 = 3$ $\quad$ There are two fixed points

$\frac{1}{3}x = \frac{1}{x}$ $\quad x = \pm\sqrt{3}$ $\quad P_1 = \sqrt{3}$, $P_2 = -\sqrt{3}$

[4] (b) Calculate three iterations of the fixed-point method starting from $x_0 = 2$. Which fixed point is the method converging to?

$g(2) = \left(\frac{2}{3}\right)(2) + \frac{1}{2} = \frac{4}{3} + \frac{1}{2} = \frac{11}{6}$

$g(11/6) = \frac{2}{3}(11/6) + 1/2 = 1.7\overline{6}$

$g(1.7\overline{6}) = 1.744\overline{165464}$

It is converging to $\sqrt{3}$

[3] (c) For fixed point interation starting with any $x_0 > 1$, explain whether or not the fixed point iteration will converge.

$g(x) = \frac{2}{3}x + \frac{1}{x}$ $\quad$ For $x_0 > 1$ $\frac{1}{x^2}$ is less than 1,

$g'(x) = \frac{2}{3} - \frac{1}{x^2}$ $\quad$ positive & decreasing.

Therefore $\max |g'(x)| < 1$ for $x_0 > 1$

$\therefore$ For any $x_0 > 1$, the fixed point iteration will converge.

[4] (d) State the Newton method iteration for calculating this same fixed point. Calculate three iterations of the Newton method, again starting from $x_0 = 2$.

The two fixed points are $\pm\sqrt{3}$.

Therefore an equivalent root finding problem is

$f(x) = x^2 - 3 \Rightarrow f'(x) = 2x \Rightarrow P_{n+1} = P_n - \frac{P_n^2 - 3}{2P_n}$

$P_{n+1} = P_n - \frac{1}{2}P_n + \frac{3}{2P_n} = \frac{1}{2}P_n + \frac{3}{2P_n}$ $\quad P_0 = 2$

$P_1 = \frac{1}{2}\cdot 2 + \frac{3}{2\cdot 2} = 1.75$ $\quad P_2 = \frac{1.75}{2} + \frac{3}{(2)(1.75)} = 1.732142$ $\quad P_3 = 1.73205080\ldots$

[1] (e) At what rate does the Newton method iteration converge for this function?

Quadratically, since $\sqrt{3}$ is a simple zero.

[5] **5.** (a) Use Gaussian Elimination with **partial pivoting** to solve the linear system:

$$2x_1 - 3x_2 + 2x_3 = 5$$
$$-4x_1 + 2x_2 - 6x_3 = 14$$
$$2x_1 + 2x_2 + 4x_3 = 8$$

$$\left[\begin{array}{ccc|c} 2 & -3 & 2 & 5 \\ -4 & 2 & -6 & 14 \\ 2 & 2 & 4 & 8 \end{array}\right] \quad \begin{array}{c}\text{Pivot}\\ E_1 \leftrightarrow E_2\end{array} \quad \left[\begin{array}{ccc|c} -4 & 2 & -6 & 14 \\ 2 & -3 & 2 & 5 \\ 2 & 2 & 4 & 8 \end{array}\right]$$

$E_2 + \frac{1}{2}E_1 \to E_2$

$E_3 + \frac{1}{2}E_1 \to E_3$

$$\sim \left[\begin{array}{ccc|c} -4 & 2 & -6 & 14 \\ 0 & -2 & -1 & 12 \\ 0 & 3 & +1 & 15 \end{array}\right] \quad \begin{array}{c}\text{Pivot}\\ E_2 \leftrightarrow E_3\end{array} \quad \left[\begin{array}{ccc|c} -4 & 2 & -6 & 14 \\ 0 & 3 & 1 & 15 \\ 0 & -2 & -1 & 12 \end{array}\right]$$

$E_3 + 2/3\, E_2 \to E_3$

$$\sim \left[\begin{array}{ccc|c} -4 & 2 & -6 & 14 \\ 0 & 3 & 1 & 15 \\ 0 & 0 & -1/3 & 22 \end{array}\right]$$

$\therefore\ x_3 = -66$

$3x_2 - 66 = 15$

$3x_2 = 15 + 66 = 81$

$x_2 = 27$

$-4x_1 + 2(27) - 6(-66) = 14 \quad x_1 = \dfrac{14 - 396 - 54}{-4}$

$x_1 = 109$

$$\therefore\ X = \begin{bmatrix} -116 \\ 27 \\ 109 \end{bmatrix}$$

[2] (b) Consider factoring the coefficient matrix, A, for the above linear system into LU such that $PA = LU$. Find P and U.

$$U = \begin{bmatrix} -4 & 2 & -6 \\ 0 & 3 & 1 \\ 0 & 0 & -1/3 \end{bmatrix}$$

Pivots were $E_1 \leftrightarrow E_2$ $\{1, 2, 3\} \to \{2, 1, 3\}$

and $E_2 \leftrightarrow E_3$ $\to \{2, 3, 1\}$

$$\therefore\ P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

$\frac{2^3}{3!} = \frac{8}{6} = \frac{4}{3}$

[3] **6.** (a) Consider the function $f(x) = e^{2x}$. Find $P_3(x)$, the third degree Taylor polynomial which approximates the function $f(x)$ expanded about $x_0 = 0$.

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots \quad \therefore e^{2x} = 1 + 2x + \frac{(2x)^2}{2!} + \frac{(2x)^3}{3!} + \cdots$$

$$\therefore P_3(x) = 1 + 2x + 2x^2 + \frac{4}{3}x^3$$

[1] (b) Use $P_3(x)$, the third degree Taylor polynomial, to approximate $f(0.43)$ of the function $f(x)$.

$$f(0.43) \approx P_3(0.43) = 1 + 2(0.43) + 2(0.43)^2 + \frac{4}{3}(0.43)^3 = 2.335809\overline{3}$$

[3] (c) What is the maximum error incurred in using $P_3(x)$ to approximate $f(x)$ on the interval $[0, 5]$?

Taylor remainder theorem.

$$\text{Error} \leq \max_{0 \leq x \leq 5} \frac{|f^4(x)|}{4!} (x)^4$$

$f^4(x) = 2^4 e^{2x}$ ← an increasing function, maximum at $x = 5$

x^4 maximum @ $x = 5$ on $[0,5]$

$$\text{Error} \leq \frac{2^4 e^{10}}{4!} 5^4$$ Yikes!!

[3] (d) Find $P_L(x)$, the third degree Lagrange polynomial which approximates the function $f(x)$ given the sampling points $(0, 1)$, $(0.25, 1.64872)$, $(0.5, 2.71828)$, $(0.75, 4.48169)$.

This was too many computations so any reasonable attempt that showed you knew what to do was given the mark

$$L_0 = \frac{(x-0.25)(x-0.5)(x-0.75)}{(-0.25)(-0.5)(-0.75)} \qquad L_1 = \frac{x(x-0.5)(x-0.75)}{(0.25)(-0.25)(-0.5)}$$

$$L_2 = \frac{(x)(x-0.25)(x-0.75)}{(0.5)(0.25)(-0.25)} \qquad L_3 = \frac{(x)(x-0.25)(x-0.5)}{(0.75)(0.5)(0.25)}$$

$(0.25)(0.5)(0.75) = 0.09375$

$(0.25)^2(0.5) = 0.03125$

$$P_L(x) = L_0(x) + 1.64872 L_2(x) + 2.71828 L_2(x) + 4.48169 L_3(x)$$

[1] (e) Use $P_L(x)$, the third degree Lagrange polynomial, to approximate $f(0.43)$ of the function $f(x)$.

$$L_0(0.43) = \frac{(0.43-0.25)(0.43-0.5)(0.43-0.75)}{-0.09375}$$

$$L_1 = \frac{(0.43)(0.43-0.5)(0.43-0.75)}{0.03125}$$

$$L_2 = \frac{(0.43)(0.43-0.25)(0.43-0.75)}{-0.03125}$$

$$L_3 = \frac{(0.43)(0.43-0.25)(0.43-0.5)}{0.09375}$$

[2] (f) One the interval $[0, 5]$, which polynomial approximation ($P_3(x)$ or $P_L(x)$) of $fx) = e^{2x}$ is likely to give better accuracy. Why?

f) $P_L(x)$ because the Taylor series is about a point.

OR

Also: note that $P_L(x)$ should only be used between 0 ; 0.75

$L_0(0.43) = -0.043008$

$L_1(0.43) = 0.308224$

$L_2(0.43) =$

$L_3(0.43) = -0.057792$

$\therefore P_L(x) =$