Name: SOLUTIONS

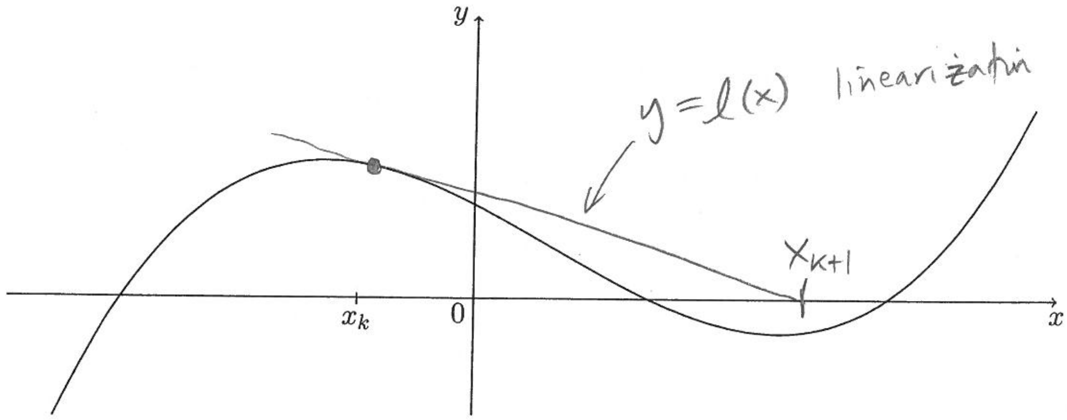
Math 310 Numerical Analysis (Bueler)

10 December 2019

Final Exam

In class. No book or electronics. 1/2 sheet of notes allowed.
120 minutes maximum. 165 points total.

1. (a) [5 points] A differentiable function $f(x)$ and an iterate x_k are shown on the axes below. Sketch, with appropriate labeling, how **Newton's method** determines the next iterate x_{k+1}.



(b) [5 points] Write **Newton's method** as a formula for determining the next iterate x_{k+1} from the previous iterate x_k:

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

(c) [5 points] Write the **secant method** as a formula for determining the next iterate x_{k+1} from the previous two iterates x_k and x_{k-1}:

$$x_{k+1} = x_k - \frac{f(x_k)}{\left(\frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}}\right)}$$

2. (a) [10 points] State **Taylor's theorem with remainder.** *Carefully state the hypotheses and the conclusion of the theorem.*

Theorem if $f \in C^{n+1}[a,b]$ and $x_0 \in (a,b)$ and $x \in [a,b]$ then there is ξ between x_0 and x so that

$$f(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)}{2}(x-x_0)^2 + \dots + \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n + \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-x_0)^{n+1}.$$

(b) [10 points] Assume that the basepoint is $a = 1$ and that the function is $f(x) = \cos(x/3)$. How close is the degree 4 Taylor polynomial $P_4(x)$ to the function $f(x)$ on the interval $[0,2]$? (*Note that you are not asked to compute $P_4(x)$; no points will be given for that.*) Answer by completing the following with a concrete upper bound. (*You may leave your concrete expression unsimplified.*)

$$|f(x) - P_4(x)| \le \frac{\max_{[0,2]} |f^{(5)}(x)|}{5!} \max_{[0,2]} |x-1|^5$$

$$\le \frac{\frac{1}{3^5}}{5!} \cdot 1^5 = \frac{1}{243 \cdot 120}$$

notes: $f(x) = \cos(x/3)$

$n = 4$

$f^{(5)}(x) = \frac{-1}{3^5} \sin\left(\frac{x}{3}\right)$

3. [15 points] Table 10.3, printed on the last page, includes the order of accuracy $O(h^2)$ for the **composite trapezoid rule**. However, we proved that if $f \in C^2[a,b]$, and if $h = (b-a)/n$ and $x_i = a + ih$ for $i = 0, 1, \ldots, n$, then there is $\xi \in [a,b]$ so that

$$\int_a^b f(x)\,dx = \frac{h}{2}\left[f(x_0) + 2f(x_1) + 2f(x_2) + \cdots + 2f(x_{n-1}) + f(x_n)\right] - \frac{1}{12}(b-a)h^2 f''(\xi).$$

Suppose we want to use this rule to compute $\int_0^1 e^{-x^2}\,dx$ to an accuracy of 10^{-8}. What n is needed? *(Note you are not asked to apply the rule. You may leave your concrete expression for n unsimplified.)*

$$\left|\int_a^b f(x)\,dx - T_n\right| = \frac{1}{12}(b-a)h^2 |f''(\xi)|$$

$$\leq \frac{1}{12}\cdot 1 \cdot \frac{1}{n^2}\cdot \max_{[0,1]} \left|(4x^2-2)e^{-x^2}\right|$$

$$\leq \frac{1}{12n^2}\cdot 2 = \frac{1}{6n^2} \leq 10^{-8}$$

$$\Longleftrightarrow \quad n \geq \sqrt{\frac{10^8}{6}} = \frac{10^4}{\sqrt{6}}$$

note: $f'(x) = -2xe^{-x^2}$, $f''(x) = (4x^2-2)e^{-x^2}$

4. [10 points] Write two to four complete sentences to explain the ideas of **Clenshaw-Curtis quadrature (integration)** on the interval $[-1,1]$. *(Hint. The ideas are not that different from those of the Newton-Cotes formulas. Restate the basic ideas and then state the new ones, including a formula.)*

To approximate $\int_{-1}^1 f(x)\,dx$, Clenshaw-Curtis uses the Chebyshev points $x_j = \cos(\pi j/n)$ for $j = 0, 1, \ldots, n$. Then one computes a polynomial $p(x)$ of degree n so that $p(x_j) = f(x_j)$ for all j. Then one integrates p: $\int_{-1}^1 f(x)dx \approx \int_{-1}^1 p(x)\,dx$. Note the rule is of the form $\int_{-1}^1 f(x)dx \approx \sum_{j=0}^n c_j f(x_j)$ where $c_j = \int_{-1}^1 \prod_{\substack{k=0\\k\neq j}}^n \frac{x-x_k}{x_k-x_j}\,dx$.

5. (a) [10 points] Suppose $f(x) = e^x$. Completely set up, *but do not solve*, the **Vandermonde** linear system to find the degree 2 polynomial $p(x) = c_0 + c_1 x + c_2 x^2$ which interpolates $f(x)$ at the points $x_0 = -1, x_1 = 0, x_2 = 1$.

$$\begin{bmatrix} 1 & (-1) & (-1)^2 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} e^{-1} \\ e^0 \\ e^1 \end{bmatrix}$$

(b) [10 points] For the same $f(x)$ and interpolation points as in part **(a)**, write down **Lagrange's form** of the polynomial $p(x)$. (*Do not simplify.*)

$$p(x) = e^{-1} \frac{(x-0)(x-1)}{(-1-0)(-1-1)} + e^{0} \frac{(x+1)(x-1)}{(0+1)(0-1)}$$

$$+ e^{1} \frac{(x+1)(x-0)}{(1+1)(1-0)}$$

6. An actual proposed 8 bit version of the IEEE 754 standard for floating point representations, called *minifloat*, has this set up:

s	e_1	e_2	e_3	e_4	b_1	b_2	b_3

representing the number

$$x = (-1)^s \, (1.b_1b_2b_3)_2 \; 2^{(e_1e_2e_3e_4)_2 + 2_{10}}$$

This scheme is designed to have the (surprising) property that all representable numbers are integers. (*The "$+2_{10}$" in the exponent is* not *a misprint.*) Note the usual exception cases:

- exponent bits $(0000)_2$ define the number zero or subnormal numbers
- exponent bits $(1111)_2$ define the other exceptions: $\pm\infty$ and NaN (*... ignore the details*)

(a) [10 points] What is the **largest number** that this system can represent? (*State the number in decimal notation and show the bits. There is no need to simplify the number.*)

0	1	1	1	0	1	1	1

$= +\,(1.111)_2 \times 2^{(1110)_2+2}$

$= \left(1+\frac{1}{2}+\frac{1}{4}+\frac{1}{8}\right) 2^{14+2} = \left(2-\frac{1}{8}\right)2^{16}$

$= 2^{17} - 2^{13}.$

(b) [10 points] What is the **smallest *normal* positive number** that this system can represent? (*State the number in decimal notation and show the bits. Please simplify the number.*)

0	0	0	0	1	0	0	0

$= +\,(1.000)_2 \times 2^{(0001)_2+2}$

$= 2^3 = 8$

Extra Credit. [3 points] How do you represent 1 in the system? What is the value of "machine epsilon" in this system? Indeed, how do you define "machine epsilon"? Write complete sentences. (*Hint. Think subnormal.*)

$1 =$

0	0	0	0	0	0	0	1

(subnormal)

$= +\,(0.001)_2 \times 2^{(0001)_2+2} = \frac{1}{8}\cdot 2^3$

"machine epsilon" is not clear! a possible definition

def: $\varepsilon = \left((1.001)_2 \times 2^{(0001)_2+2}\right) - \left((1.000)_2 \times 2^{(0001)_2+2}\right) = 1$

7. Consider the ODE IVP

$$y' = t - 2y, \qquad y(0) = 2.$$

(a) [10 points] Do two steps of the Euler method with step size $h = \frac{1}{2}$. This computes approximations of $y(0.5)$ and $y(1)$.

$y_0 = 2$ $\qquad t_0 = 0,\ t_1 = 0.5,\ t_1 = 1$

$\underline{k=0:}$ $\quad y_1 = y_0 + h f(t_0, y_0) = 2 + \frac{1}{2}(0 - 2\cdot 2)$

$= 0 \approx y(0.5)$

$\underline{k=1:}$ $\quad y_2 = 0 + \frac{1}{2}(0.5 - 2\cdot 0)$

$= \frac{1}{4} \approx y(1)$

(b) [10 points] The trapezoid method, an implicit rule, is

$$y_{k+1} = y_k + \frac{h}{2}\left(f(t_k, y_k) + f(t_{k+1}, y_{k+1})\right).$$

Do two steps of this method, again with step size $h = \frac{1}{2}$. This computes new approximations of $y(0.5)$ and $y(1)$. (*Hint. This requires easy algebra.*)

$\underline{k=0:}$ $\quad y_1 = 2 + \frac{1}{4}(0 - 2\cdot 2 + 0.5 - 2\cdot y_1)$

$\Leftrightarrow \quad y_1 = 2 - 1 + \frac{1}{8} - \frac{1}{2}y_1 = \frac{9}{8} - \frac{1}{2}y_1$

$\Leftrightarrow \quad \frac{3}{2}y_1 = \frac{9}{8} \quad \Leftrightarrow \quad y_1 = \frac{3}{4} \approx y(0.5)$

$\underline{k=1:}$ $\quad y_2 = \frac{3}{4} + \frac{1}{4}(0.5 - 2\cdot\frac{3}{4} + 1 - 2\cdot y_2)$

$\Leftrightarrow \quad y_2 = \frac{3}{4} + \frac{1}{4}(-1 + 1 - 2y_2)$

$\Leftrightarrow \quad y_2 = \frac{3}{4} - \frac{1}{2}y_2 \quad \Leftrightarrow \quad \frac{3}{2}y_2 = \frac{3}{4}$

$\Leftrightarrow \quad y_2 = \frac{1}{2} \approx y(1)$

[continuation of problem 7]

(c) [5 points] The exact solution to this ODE IVP is

$$y(t) = \frac{t}{2} - \frac{1}{4} + \frac{9}{4}e^{-2t}.$$

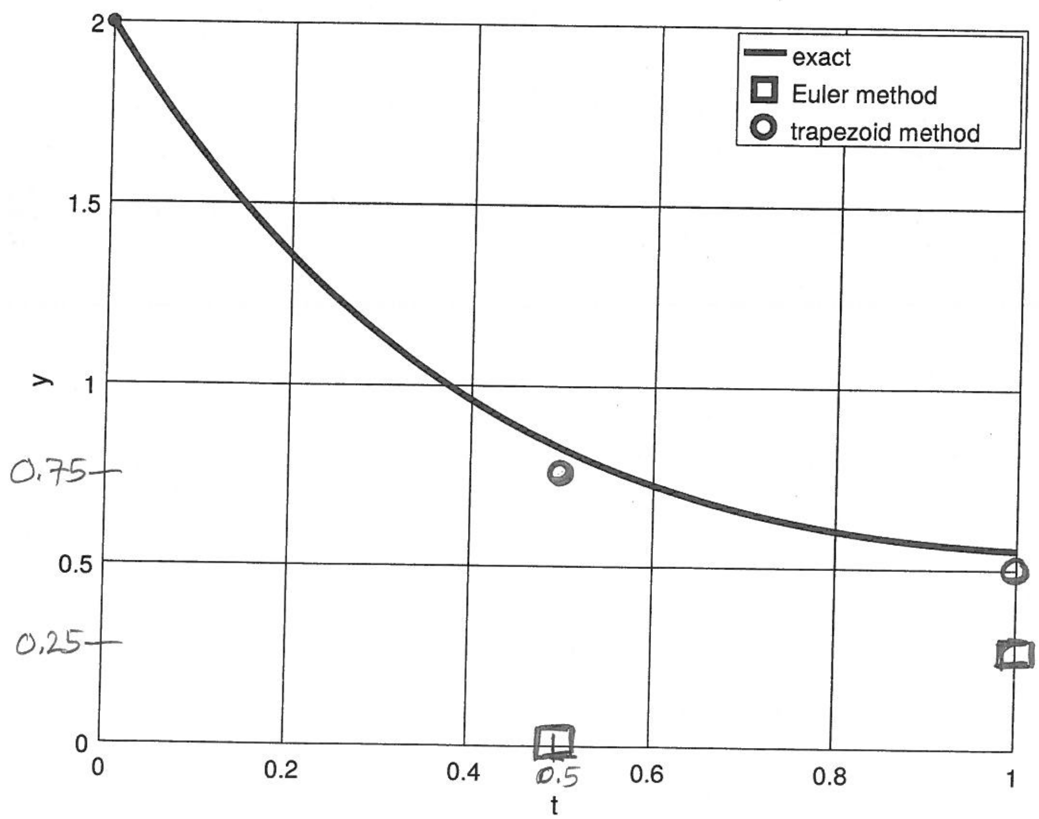
Verify this.

$$y' = \frac{1}{2} - \frac{9}{2}e^{-2t}$$

$$t - 2y = t - \left(t - \frac{1}{2} + \frac{9}{2}e^{-2t}\right) = \frac{1}{2} - \frac{9}{2}e^{-2t}$$

$$y(0) = 0 - \frac{1}{4} + \frac{9}{4}\cdot 1 = \frac{8}{4} = 2 \checkmark$$

(d) [5 points] The graph below already shows the exact solution. Using the symbols shown in the legend, add your results from parts **(a)** and **(b)** to this graph.



8. [15 points] The midpoint method for the ODE IVP $y' = f(t,y)$, $y(t_0) = y_0$ is the pair of formulas

$$y_{k+1/2} = y_k + \frac{h}{2} f(t_k, y_k)$$
$$y_{k+1} = y_k + h f(t_{k+1/2}, y_{k+1/2})$$

Write a MATLAB algorithm which does n steps of the midpoint method to approximately solve the ODE IVP on the interval $t_0 \le t \le t_f$.

Note that the time interval is described by a pair of numbers `tspan = [t0 tf]` *and that* $h = (t_f - t_0)/n$. *You may assume that the ODE is scalar, so* y_k *in the above formulas is a single number, not a column vector. Also, the returned variables should be suitable for plotting with* `plot(tt,yy)`. *Finally, do not worry about adding comments.*

```
function [tt,yy] = midpoint(f,tspan,y0,n)

h=(tspan(2)-tspan(1))/n;

tt = tspan(1) : h : tspan(2);

yy = zeros (size (tt));

for k=1:n
    ytmp = yy(k) + (h/2)* f (tt(k), yy(k));
    yy(k+1) = yy(k) + h* f(tt(k)+h/2, ytmp);
end
```

10. [10 points] Solve the following system of linear equations by Gauss elimination with partial pivoting and back substitution. Show your steps, that is, indicate your row operations.

$$x_1 + x_2 = 3$$
$$4x_1 - 3x_2 = -2.$$

GEPP

$1 < 4$ so $R_1 \Leftrightarrow R_2$:

$$4x_1 - 3x_2 = -2$$
$$x_1 + x_2 = 3$$

$R_2 \Leftarrow R_2 - \frac{1}{4}R_1$:

$$4x_1 - 3x_2 = -2$$
$$\frac{7}{4}x_2 = \frac{7}{2}$$

BS

$$x_2 = \frac{7/2}{7/4} = 2$$

$$x_1 = \frac{-2 + 3 \cdot 2}{4} = 1$$

11. [10 points] Suppose you have n linear equations in n unknowns,

$$A\mathbf{x} = \mathbf{b}.$$

The algorithm used to solve such linear systems has two stages:

(1) Gauss elimination with partial pivoting, and
(2) back substitution.

Approximately how many floating-point operations occur in these stages? Which stage is more costly? Answer quantitatively in a complete sentence or two. (*You do not need to count operations exactly, or prove your claims either.*)

Stage (1) requires $\frac{2}{3}n^3 + O(n^2) = O(n^3)$ operations. Stage (2) requires only $O(n^2)$. Thus (1) is more costly and the cost of (2) can be ignored when n is large.

TABLE 10.3
Quadrature formulas and their errors.

Method	Approximation to $\int_a^b f(x)\,dx$	Error
Trapezoid rule	$\frac{b-a}{2}[f(a)+f(b)]$	$-\frac{1}{12}(b-a)^3 f''(\eta),\ \eta \in [a,b]$
Simpson's rule	$\frac{b-a}{6}\left[f(a)+4f\left(\frac{a+b}{2}\right)+f(b)\right]$	$\frac{1}{2880}(b-a)^5 f^{(4)}(\xi),\ \xi \in [a,b]$
Composite trapezoid rule	$\frac{h}{2}[f_0+2f_1+\ldots+2f_{n-1}+f_n]$	$O(h^2)$
Composite Simpson's rule	$\frac{h}{6}[f_0+4f_{1/2}+2f_1+\ldots+2f_{n-1}+4f_{n-1/2}+f_n]$	$O(h^4)$

[BLANK SPACE FOR SCRATCH WORK]