1A Let $x_n$ be a sequence of real numbers defined by $x_0 = 2$ and
\[ x_{n+1} = 1 + \sin \left( \frac{x_n}{7} \right) = g(x_n). \]

Show that
\[ |x_{n+1} - x| \leq \frac{1}{7} |x_n - x| \]
for some $x \in [1, 2]$.

**Solution:** Since the interval $[1, 2]$ containing $x_0 = 2$ is invariant under $g$ and
\[ |g'(x)| = \left| \frac{1}{7} \cos \left( \frac{x}{7} \right) \right| \leq \frac{1}{7} \]
on that interval, there is a unique fixed point $x$ with $1 \leq x \leq 2$ and
\[ |x_{n+1} - x| \leq \frac{1}{7} |x_n - x| \]
for all $n$. 
1B In floating point arithmetic with a backward stable sin function, \( x_n \) from 1A is approximated by a sequence of floating point numbers \( y_n \) satisfying

\[
y_{n+1} = \text{fl}(x_{n+1}) = \left( 1 + \sin \left( \frac{y_n(1 + \delta_n)}{7} \right) \right) (1 + \delta'_n)
\]

Ignoring terms of size \( O(\epsilon^2) \), show that

\[
|y_{n+1} - x| \leq \frac{1}{7} |y_n - x| + 3\epsilon
\]

and bound the minimum possible error \( |y_N - x| \).

**Solution:** Subtraction and the mean value theorem give

\[
|y_{n+1} - x| = \left| (1 + \sin \left( \frac{y_n(1 + \delta_n)}{7} \right))\delta'_n - \frac{1}{7} \cos(\xi)(y_n(1 + \delta_n) - x) \right|
\]

\[
\leq 2\epsilon + \frac{1}{7} |y_n|\epsilon + \frac{1}{7} |y_n - x|
\]

\[
\leq \frac{16}{7} \epsilon + \frac{1}{7} |y_n - x|
\]

since \( |y_n| \leq 2 + O(\epsilon) \). As a consequence, \( y_n - x \) decreases like \((1/7)^n\) until it reaches \( a\epsilon \) where \( a\epsilon = 16/7\epsilon + (1/7)a\epsilon \) or \( a = 8/3 \). Thus the minimum possible error is bounded by \( 8\epsilon/3 \).
2A. Let \( P(x) \) be the quadratic polynomial interpolating \( f(x) \) at \( x = 1, 1/2 \) and \( 1/3 \). Give a formula for the error \( f(x) - P(x) \) and explain why each of the three factors in the error is inevitable.

**Solution:** The error is

\[
 f(x) - P(x) = \frac{1}{3!} f'''(\xi) \omega(x)
\]

where \( \xi \) is an unknown point between 1, 1/2, 1/3 and \( x \) and \( \omega(x) = (x - 1)(x - 1/2)(x - 1/3) \). The last factor \( \omega(x) \) guarantees that the error vanishes at interpolation points, the middle factor guarantees that the error vanishes when \( f \) is a quadratic polynomial, and the first factor makes the error equal to \( f \) when \( f(x) = \omega(x) = x^3 + \cdots \) vanishes at the interpolation points, so \( P(x) = 0 \) and \( f''' = 3! \).
2B For the specific function $f(x) = 1 + 216x^3$, and the points $x = 1, 1/2$ and $1/3$ from 2A,
(a) build the divided difference table,
(b) find the Newton form of $P(x)$ from 2A,
(c) evaluate $P(0)$, and
(d) show that your error formula from 2A is satisfied at $x = 0$.

Solution: The divided difference table is

<table>
<thead>
<tr>
<th>$j$</th>
<th>$x_j$</th>
<th>$f[x_j]$</th>
<th>$f[x_j, x_{j+1}]$</th>
<th>$f[x_j, x_{j+1}, x_{j+2}]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>217</td>
<td>378</td>
<td>396</td>
</tr>
<tr>
<td>1 1/2</td>
<td>28</td>
<td>114</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 1/3</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hence the Newton form of the interpolating polynomial is

$$P(x) = 217 + 378(x - 1) + 396(x - 1)(x - 1/2)$$


$P(0) = 217 - 378 + 396/2 = 37$.

The error formula from 2A gives

$$f(x) - P(x) = \frac{1}{6} 216 (0 - 1)(0 - 1/2)(0 - 1/3) = -36,$$

while direct evaluation also gives

$$f(x) - P(x) = 1 + 2160^3 - 37 = -36$$

at $x = 0$. 

Page 4 of 10
3A Find constants $a$, $b$ and $c$ such that the numerical integration rule

$$\int_0^1 f(t) \, dt = af(-2) + bf(-1) + cf(0)$$

is exact whenever $f$ is a quadratic polynomial.

**Solution:** Integrating the Lagrange basis functions gives

$$a = \int_0^1 \frac{(x + 1)(x - 0)}{(-2 + 1)(-2 - 0)} \, dx = \frac{5}{12},$$

$$b = \int_0^1 \frac{(x + 2)(x - 0)}{(-1 + 2)(-1 + 0)} \, dx = -\frac{16}{12},$$

and

$$c = \int_0^1 \frac{(x + 2)(x + 1)}{(0 + 2)(0 + 1)} \, dx = \frac{23}{12}.$$

Check: $a + b + c = 1, a(-2) + b(-1) + c(0) = 1/2, a(-2)^2 + b(-1)^2 + c(0)^2 = 1/3.$
Consider the multistep method

\[ u_{n+1} = u_n + h \left( cf_n + bf_{n-1} + af_{n-2} \right), \]

for \( y' = f(t, y) \), where \( t_n = nh \) and \( a, b \) and \( c \) are found from 3A. Define the local truncation error \( \tau_n \) and show the method is at least third-order accurate.

**Solution:** The local truncation error \( \tau \) is defined by

\[ h\tau_n = y_{n+1} - y_n - h \left( cy'_n + by'_{n-1} + ay'_{n-2} \right) \]

where \( y \) is the exact solution to \( y' = f(y) \). Integration gives

\[ y_{n+1} - y_n - \int_{t_n}^{t_{n+1}} y'(s) ds = 0, \]

so

\[ h\tau_n = y_{n+1} - y_n - \int_{t_n}^{t_{n+1}} y'(s) ds + \int_{t_n}^{t_{n+1}} y'(s) - P(s) ds \]

where \( P \) is the quadratic polynomial interpolating \( y' \) at \( t_n, t_{n-1} \) and \( t_{n-2} \). Hence

\[ h\tau_n = \int_{t_n}^{t_{n+1}} y'(s) - P(s) = \frac{1}{3!} \int_{t_n}^{t_{n+1}} y'''(s) (s-t_n)(s-t_{n-1})(s-t_{n-2}) ds = O(h^3) \]

and \( \tau_n = O(h^3) \). Since multistep methods of this form are stable, the method is at least third-order accurate.
4A Write the implicit numerical method
\[ u_{n+1} = u_n + \frac{h}{2} f \left( \frac{2}{3} u_n + \frac{1}{3} u_{n+1} \right) + \frac{h}{2} f \left( \frac{1}{3} u_n + \frac{2}{3} u_{n+1} \right) \]
for \( y' = f(y) \) as a two-stage Runge-Kutta method
\[
\begin{align*}
    k_1 & = f(u_n + h(a_{11}k_1 + a_{12}k_2)) \\
    k_2 & = f(u_n + h(a_{21}k_1 + a_{22}k_2)) \\
    u_{n+1} & = u_n + h(b_1k_1 + b_2k_2)
\end{align*}
\]
Define the local truncation error \( \tau \) and show the method is at least second-order accurate.

Solution: Clearly
\[
\begin{align*}
    k_1 & = f \left( \frac{2}{3} u_n + \frac{1}{3} u_{n+1} \right) = f(u_n + \frac{h}{6}(k_1 + k_2)) \\
    k_2 & = f \left( \frac{1}{3} u_n + \frac{2}{3} u_{n+1} \right) = f(u_n + \frac{h}{3}(k_1 + k_2)),
\end{align*}
\]
since
\[ u_{n+1} = u_n + \frac{h}{2}(k_1 + k_2). \]
The local truncation error is defined by
\[ h\tau_n = y_{n+1} - y_n - \frac{h}{2}(k_1 + k_2) \]
where \( k_1 \) and \( k_2 \) are defined with \( y = y_n \) in place of \( u_n \). Differentiation of \( k_i \) with respect to \( h \) gives
\[ k'_i = f'(y + h \sum_j a_{ij}k_j)(\sum_j a_{ij}k_j + h \sum_j a_{ij}k'_j) \]
and evaluation at \( h = 0 \) gives
\[ k_i(0) = f(y), \quad k'_i(0) = f'(y) \sum_j a_{ij}k_j(0) = f'(y) \sum_j a_{ij}f(y). \]
Hence
\[ h\tau_n = y + hy' + \frac{h^2}{2}y'' + O(h^3) - h \sum_i b_i(f(y) + h \sum_j a_{ij}f'(y)f(y) + O(h^2)). \]
Since \( y' = f(y) \), differentiation gives \( y'' = f'(y)y' = f'(y)f(y) \). Since one-step methods are always stable, the two order conditions
\[ \sum_i b_i = \frac{1}{2} + \frac{1}{2} = 1 \]
and
\[ \sum_i b_i \sum_j a_{ij} = \frac{11}{23} + \frac{12}{23} = \frac{1}{2} \]
guarantee at least second-order accuracy.
4B Suppose the method of 4A is applied to the test equation \( y' = \lambda y \) with a complex scalar \( \lambda \) such that the real part of \( \lambda \) is negative. Show that the numerical solution satisfies \( |u_n| \leq |u_0| \) for all \( n \geq 0 \).

Solution: For this equation,

\[
\begin{align*}
  u_{n+1} &= u_n + \frac{h}{2} \lambda \left( \frac{2}{3} u_n + \frac{1}{3} u_{n+1} \right) + \frac{h}{2} \lambda \left( \frac{1}{3} u_n + \frac{2}{3} u_{n+1} \right) = u_n + \frac{h}{2} \lambda u_n + \frac{h}{2} \lambda u_{n+1} \\
  &\text{or} \\
  u_{n+1} &= R(z) u_n
\end{align*}
\]

where \( z = h\lambda \) and

\[
R(z) = \frac{1 + z/2}{1 - z/2}.
\]

By induction,

\[
u_n = R(z)^n u_0.
\]

Thus \( |u_n| \leq |u_0| \) iff \( |R(z)| \leq 1 \) or \( |1 + z/2| \leq |1 - z/2| \), which means \( z \) is closer to \(-2\) than 2. This happens precisely when \( \lambda \) is in the left half plane.
**5A** Let

\[ A = \begin{bmatrix} 4 & 0 & -6 \\ 0 & 16 & -20 \\ -6 & -20 & 70 \end{bmatrix}. \]

Show that \( A \) is symmetric positive definite and find an upper triangular matrix \( R \) with positive diagonal entries \( r_{ii} > 0 \) such that \( A = R^T R \).

**Solution:** Set

\[ A = R^T R = \begin{bmatrix} r_{11} & 0 & 0 \\ r_{12} & r_{22} & 0 \\ r_{13} & r_{23} & r_{33} \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ 0 & r_{22} & r_{23} \\ 0 & 0 & r_{33} \end{bmatrix} \]

and solve successively for

\[ r_{11} = \sqrt{a_{11}} = 2, \]
\[ r_{12} = a_{12}/r_{11} = 0, \]
\[ r_{13} = a_{13}/r_{11} = -3, \]
\[ r_{22} = \sqrt{a_{22} - r_{12}^2} = 4, \]
\[ r_{23} = a_{23}/r_{22} = -5, \]
and

\[ r_{33} = \sqrt{a_{33} - r_{13}^2 - r_{23}^2} = 6. \]

Thus the invertible matrix

\[ R = \begin{bmatrix} 2 & 0 & -3 \\ 0 & 4 & -5 \\ 0 & 0 & 6 \end{bmatrix} \]

satisfies \( R^T R = A \), and therefore \( A \) is symmetric positive definite.
5B Let

\[
A = \begin{bmatrix}
  d_1 & 0 & \cdots & 0 & r_1 \\
  0 & d_2 & 0 & \cdots & r_2 \\
  \vdots & & \ddots & & \vdots \\
  0 & 0 & \cdots & d_{n-1} & r_{n-1} \\
  r_1 & r_2 & \cdots & r_{n-1} & d_n
\end{bmatrix} = D + re_n^T + e_n r^T,
\]

where \( e_n^T r = 0 \), be a symmetric arrowhead matrix. Develop an algorithm for computing the Cholesky factorization of \( A \) in \( O(n) \) scalar floating point operations. Use your algorithm to find a condition on the diagonal matrix \( D \) and the vector \( r \) which determines when the matrix \( A \) is positive definite.

**Solution:** For \( j = 1 : n - 1 \), subtract the multiple \( r_j/d_j \) of row \( j \) of \( A \) from row \( n \) to eliminate the nonzero entry \( r_j \) in row \( n \), also subtracting \( r_j^2/d_j \) from \( d_n \) in the process. Thus the matrix

\[
M = \begin{bmatrix}
  1 & 0 & \cdots & 0 & 0 \\
  0 & 1 & 0 & \cdots & 0 \\
  \vdots & & \ddots & & \vdots \\
  0 & 0 & \cdots & 1 & 0 \\
  -\frac{r_1}{d_1} & -\frac{r_2}{d_2} & \cdots & -\frac{r_{n-1}}{d_{n-1}} & 1
\end{bmatrix} = I - e_n D^{-1} r^T
\]

produces an upper triangular matrix

\[
MA = D + re_n^T - e_n r^T D^{-1} re_n^T.
\]

To preserve symmetry, we apply the same operations to the columns of \( MA \) to get the diagonal matrix

\[
MAM^T = D - e_n r^T D^{-1} re_n^T = F.
\]

Hence

\[
A = M^{-1} F M^{-T} = (I + e_n D^{-1} r^T)(D - e_n r^T D^{-1} re_n^T)(I + r D^{-1} e_n^T) = R^T R
\]

where

\[
R = F^{1/2} M^{-T} = \begin{bmatrix}
  \sqrt{d_1} & 0 & \cdots & 0 & r_1/\sqrt{d_1} \\
  0 & \sqrt{d_2} & 0 & \cdots & r_2/\sqrt{d_2} \\
  \vdots & & \ddots & & \vdots \\
  0 & 0 & \cdots & \sqrt{d_{n-1}} & r_{n-1}/\sqrt{d_{n-1}} \\
  0 & 0 & \cdots & 0 & \sqrt{d_n - \sum_j r_j^2/d_j}
\end{bmatrix}
\]

We conclude that \( A \) is positive definite if all \( d_j > 0 \) and

\[
d_n - \sum_j \frac{r_j^2}{d_j} = e_n^T D e_n - r^T D^{-1} r > 0.
\]