Solution 1 Note that
\[ \lim_{x \to +\infty} f(x) = +\infty \quad \text{and} \quad \lim_{x \to -\infty} f(x) = -\infty. \]

Hence for any value of \( k \), there exist numbers \( a \) and \( b \) (which depend on \( k \)) such that \( a < b \) and \( f(a) < 0 < f(b) \). As \( f \) is continuous, by the Intermediate Value Theorem, there must exist \( c \in (a, b) \) such that \( f(c) = 0 \).

Since \( f'(x) = 3x^2 + 2 > 0 \) for all \( x \), the fundamental theorem of calculus guarantees that
\[ f(x_2) - f(x_1) = \int_{x_1}^{x_2} f'(x)dx > 0 \]
whenever \( x_2 > x_1 \), so any zero of \( f \) must be unique.

Altogether, \( f \) has a unique zero, that is, \( f \) crosses the \( x \)-axis exactly once.

Solution 2 Let
\[ f(x) = x - (\ln(1 + x))^x = x - \exp(x \ln[\ln(1 + x)]) \]
on \( x \in [\pi, 2\pi] \).

On the interval \([\pi, 2\pi]\), the function \( \ln(1 + x) \) is continuous and positive. Thus the function \( \ln[\ln(1 + x)] \) is also continuous there. Thus \( f \) is continuous as it is the composition of continuous functions.

One can compute (e.g. with a calculator) that
\[ f(\pi) \approx 0.1254 > 0, \]
and
\[ f(2\pi) \approx -68.1327 < 0. \]

Then the Intermediate Value Theorem immediately implies that a number \( x \) exists in \([\pi, 2\pi]\) with
\[ 0 = f(x) = x - (\ln(1 + x))^x. \]

Solution 3 We compute the following derivatives:
\[ f'(x) = e^x(\cos x - \sin x), \quad f''(x) = -2e^x \sin x, \quad \text{and} \quad f^{(3)}(x) = -2e^x(\sin x + \cos x). \]

Thus the second Taylor polynomial centered at \( x_0 = 0 \) is
\[ P_2(x) = f(0) + f'(0)x + \frac{1}{2}f''(0)x^2 = 1 + x. \]

a. Taylor’s remainder formula tells us that
\[ |f(0.5) - P_2(0.5)| = |R_2(0.5)| = \left| \frac{f^{(3)}(c)}{3!} (0.5)^3 \right| = \left| \frac{e^c(\sin c + \cos c)}{3} (0.5)^3 \right|. \]
for some $c$ between 0 and 0.5. One example of an upper bound is achieved using the crude bounds $|e^c| < e < 3$ for $c \in [0, 0.5]$, $|\sin c| \leq 1$, and $|\cos c| \leq 1$. Thus $|\sin c + \cos c| < 2$, so

$$|R_2(0.5)| = \frac{|e^c||\sin c + \cos c|}{3}(0.5)^3 \leq \frac{6}{3}(0.5)^3 = \frac{1}{4} = 0.25.$$  

Tighter upper bounds can be found by examining the behavior of $f^{(3)}(c)$ more closely (for example, one could compute the maximum and minimum of $f^{(3)}(c)$ for $c \in [0, 0.5]$).

Because $f(0.5) \approx 1.4469$ and $P_2(x) = 1.5$, the actual error is about 0.0531.

b. Using the bounds $|e^c| < e < 3$ for $c \in [0, 1]$, $|\sin c| \leq 1$, $|\cos c| \leq 1$, and $|x| \leq 1$ for $x \in [0, 1]$, we find that

$$|f(x) - P_2(x)| = \left|\frac{e^c(\sin c + \cos c)}{3}x^3\right| \leq \frac{3(1 + 1)}{3}1^3 = 2.$$  

c. We find

$$\int_0^1 P_2(x) \, dx = \int_0^1 (1 + x) \, dx = 1.5.$$  

d. We use again the bounds $|e^c| < e < 3$ for $c \in [0, 1]$, $|\sin c| \leq 1$, $|\cos c| \leq 1$. For $x \in [0, 1]$, $|x|^3 = x^3$. The error is bounded above by

$$\left|\int_0^1 f(x) \, dx - \int_0^1 P_2(x) \, dx\right| = \left|\int_0^1 R_2(x) \, dx\right| \leq \int_0^1 |R_2(x)| \, dx \leq \int_0^1 2x^3 \, dx = 0.5.$$  

One can compute that

$$\int_0^1 f(x) \, dx \approx 1.378,$$

so that the actual error is approximately $1.5 - 1.378 = 0.122$.

**Solution 4**

a. 3224

b. -3224

c. 1.32421875

d. 1.324218750000000222044604925031308084726336181640625

**Solution 5**

a. In nested form, we have

$$f(x) = (((1.01e^x - 4.62)e^x - 3.11)e^x + 12.2)e^x - 1.99.$$  

b. -6.79
c. \(-7.07\)

d. For part a and part b, respectively, the absolute errors are

\[-7.61 - (-6.71)| = 0.82 \text{ and } | -7.61 - (7.07)| = 0.54.\]

The relative errors are, respectively,

\[
\left| \frac{0.82}{-7.61} \right| = 0.108 \text{ and } \left| \frac{0.54}{-7.61} \right| = 0.071.
\]

Solution 6 Using the Taylor series of the exponential and the binomial theorem,

\[
e^{a+b} = \sum_{n=0}^{\infty} \frac{(a+b)^n}{n!}
= \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}
= \lim_{N \to \infty} \sum_{n=0}^{N} \sum_{k=0}^{n} \frac{a^k}{k!} \frac{b^{n-k}}{(n-k)!}
= \lim_{N \to \infty} \sum_{n=0}^{N} \sum_{k=0}^{N} \frac{a^k}{k!} \frac{b^{n-k}}{(n-k)!}
= \sum_{k=0}^{\infty} \frac{a^k}{k!} \sum_{n=0}^{\infty} \frac{b^{n-k}}{n!}
= \sum_{k=0}^{\infty} \frac{a^k}{k!} \sum_{n=0}^{\infty} \frac{b^{n-k}}{n!}
= e^a e^b.
\]

Solution 7 We define

\[
a_n = \frac{x^n}{n!},
\]

and look at the ratio

\[
b_n = \frac{a_{n+1}}{a_n} = \frac{x^{n+1}}{(n+1)!} = \frac{x}{n + 1}.
\]

We consider two cases:

If \(x\) is an integer, then \(0 < b_n < 1\) when \(n > x - 1\), \(b_n = 1\) when \(n = x - 1\), and \(b_n > 1\) when \(n < x - 1\), which indicates that \(a_{x-1}\) and \(a_x\) are the largest terms.

If \(x\) is not an integer, then \(0 < b_n < 1\) when \(n \geq \lfloor x \rfloor\), and \(b_n > 1\) when \(n < \lfloor x \rfloor\), which implies that \(a_{\lfloor x \rfloor}\) is the largest term.

Finally, the largest term can be estimated by Stirling’s approximation. This gives the estimate

\[
\frac{x^n}{n!} \sim \frac{x^n}{\sqrt{2\pi n} \left( \frac{x}{e} \right)^n} = \frac{1}{\sqrt{2\pi n}} \left( \frac{x e}{n} \right)^n,
\]
with \( n = \lfloor x \rfloor \).

**Solution 8** Sample code follows:

```matlab
function stirling_error(n)

    n = n(:);
    % Compute the error
    abs_err_low = abs(factorial(n) - sqrt(2*pi*n).*n/exp(1).*n);
    abs_err_upp = abs(factorial(n) - exp(1)*sqrt(n).*n/exp(1).*n);
    rel_err_low = abs_err_low ./ factorial(n);
    rel_err_upp = abs_err_upp ./ factorial(n);

    % Some fancy formatting
    err_tab = table(n, rel_err_low, rel_err_upp, 'VariableNames', ...
    {'n', 'RelErrorLower', 'RelErrorUpper'});
    disp(err_tab)
end
```

The output is as follows:

```matlab
>> stirling_error(1:20)
```

<table>
<thead>
<tr>
<th>n</th>
<th>RelErrorLower</th>
<th>RelErrorUpper</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0778629911042111</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0.0404978242555089</td>
<td>0.0405201900457777</td>
</tr>
<tr>
<td>3</td>
<td>0.0272984014423566</td>
<td>0.0548341398014194</td>
</tr>
<tr>
<td>4</td>
<td>0.0205760361294467</td>
<td>0.062124125181097</td>
</tr>
<tr>
<td>5</td>
<td>0.0165069336876474</td>
<td>0.0665368126705292</td>
</tr>
<tr>
<td>6</td>
<td>0.0137802991080777</td>
<td>0.0694936775966392</td>
</tr>
<tr>
<td>7</td>
<td>0.0118262238864176</td>
<td>0.0716127501453056</td>
</tr>
<tr>
<td>8</td>
<td>0.0103572556384755</td>
<td>0.073205754752165</td>
</tr>
<tr>
<td>9</td>
<td>0.00921276223008172</td>
<td>0.0744468861046304</td>
</tr>
<tr>
<td>10</td>
<td>0.00829596044393941</td>
<td>0.07544110003887313</td>
</tr>
<tr>
<td>11</td>
<td>0.00754506747591373</td>
<td>0.076253969203548</td>
</tr>
<tr>
<td>12</td>
<td>0.006918794217380843</td>
<td>0.0769345510981655</td>
</tr>
<tr>
<td>13</td>
<td>0.00638850038967187</td>
<td>0.0775096216994112</td>
</tr>
<tr>
<td>14</td>
<td>0.00593369578918022</td>
<td>0.078002828867424</td>
</tr>
<tr>
<td>15</td>
<td>0.0053933454520014</td>
<td>0.078430489028539</td>
</tr>
<tr>
<td>16</td>
<td>0.00519411958723724</td>
<td>0.0788048530922656</td>
</tr>
<tr>
<td>17</td>
<td>0.0048894037082221</td>
<td>0.0791352984339833</td>
</tr>
<tr>
<td>18</td>
<td>0.00461845573107479</td>
<td>0.0794291245948828</td>
</tr>
<tr>
<td>19</td>
<td>0.00437597572250069</td>
<td>0.0796920985415251</td>
</tr>
<tr>
<td>20</td>
<td>0.00415765262288223</td>
<td>0.0799288367892177</td>
</tr>
</tbody>
</table>
```
Solution 9 By Taylor’s Remainder Theorem, we have

$$|E_N| = \left| e^x - \sum_{n=0}^{N-1} \frac{x^n}{n!} \right| = \left| e^c x^N \frac{N!}{N!} \right|$$

for some $c$ between 0 and $x$.

Since $|x| \leq R$, we have

$$|E_N| \leq e^R \frac{R^N}{N!}.$$ 

By Stirling’s approximation, we know

$$N! \geq \sqrt{2\pi N} \left( \frac{N}{e} \right)^N.$$ 

Combining these and the fact that $\frac{e^R}{N} \leq \frac{1}{2}$ yields the absolute error bound

$$|E_N| \leq \frac{e^R}{\sqrt{2\pi N}} \left( \frac{eR}{N} \right)^N \leq e^R \left( \frac{1}{2N\sqrt{2\pi N}} \right).$$

Solution 10 Define

$$c_{jr} = \frac{(it_j)^{r-1}}{(r-1)!} \text{ and } d_{rk} = t_k^{r-1},$$

and form the $n \times 10$ matrix $C = (c_{jr})$ and the $10 \times n$ matrix $D = (d_{rk})$. Let $B = CD$ and thus

$$\text{rank}(B) \leq \min\{\text{rank}(C), \text{rank}(D)\} \leq 10.$$ 

For all $j$ and $k$, using the result from problem 9 (with $R = 1$),

$$|A_{jk} - B_{jk}| = \left| e^{it_j t_k} - \sum_{r=1}^{10} \frac{(it_j)^{r-1}}{(r-1)!} t_k^{r-1} \right|$$

$$= \left| e^{it_j t_k} - \sum_{r=0}^{9} \frac{(it_j t_k)^r}{r!} \right|$$

$$\leq \frac{e}{\sqrt{20\pi}} \left( \frac{e}{10} \right)^{10}$$

$$\approx 7.5535 \cdot 10^{-7}$$

$$< 10^{-6}.$$ 

Solution 11

**Summing from left to right** Define $a_k = \frac{1}{k^2}$ and $s_n = \sum_{k=1}^{n} a_k$, and let $s_n^*$ be the result for $s_n$ in floating point arithmetic when summing from left to right. Define $e_n$ by $s_n^* - s_n = e_n \varepsilon$, where $\varepsilon$ is machine precision. We note that $e_1 = 0$. 

5
Adding an additional term to the right gives

\[ s_{n+1}^* = \text{fl}(s_n^* + \text{fl}(a_{n+1})) \]
\[ = (s_n^* + a_{n+1}(1 + \varepsilon_1))(1 + \varepsilon_2), \]
\[ = s_n^* + a_{n+1} + s_n^* \varepsilon_2 + a_{n+1} \varepsilon_1 \varepsilon_2 \]
\[ = s_{n+1} + e_n \varepsilon + a_{n+1} \varepsilon_1 + s_n \varepsilon_2 + e_n \varepsilon_2 + a_{n+1} \varepsilon_2 + a_{n+1} \varepsilon_1 \varepsilon_2. \]

Thus

\[ s_{n+1}^* = s_{n+1} + e_n \varepsilon + a_{n+1} \varepsilon_1 + s_n \varepsilon_2 + a_{n+1} \varepsilon_2 + O(\varepsilon^2), \]

which indicates

\[ |s_{n+1}^* - s_{n+1}| \leq |e_n \varepsilon + a_{n+1} \varepsilon_1 + s_n \varepsilon_2 + a_{n+1} \varepsilon_2| \]
\[ \leq (|e_n| + a_{n+1} + s_{n+1}) \varepsilon, \]

that is,

\[ |e_{n+1}| \leq |e_n| + a_{n+1} + s_{n+1}. \]

Applying this inequality repeatedly and the estimate that

\[ s_n = \sum_{k=1}^{n} a_k \leq \sum_{k=1}^{\infty} a_k = \frac{\pi^2}{6} < 2 \]

to get

\[ |e_n| \leq |e_1| + \sum_{k=2}^{n} a_k + \sum_{k=2}^{n} s_k \]
\[ \leq s_n + \sum_{k=2}^{n} s_k \]
\[ \leq 2 + 2(n - 1) = 2n + 1. \]

Therefore the absolute error is bounded by \((2n + 1)\varepsilon\).

**Summing from right to left**  Let

\[ b_k = \frac{1}{(n + 1 - k)^2} \]

for \(1 \leq k \leq n\). Define

\[ S_k = \sum_{j=1}^{k} b_j = \frac{1}{n^2} + \frac{1}{(n-1)^2} + \ldots + \frac{1}{(n-k+1)^2}, \]

let \( S_k^* \) be the result for \( S_k \) in floating point arithmetic summing from left to right in the above sum, and let \( E_k \) be defined by \( S_k^* - S_k = E_k \varepsilon \), where \( \varepsilon \) is machine precision.
Therefore \(|e_1| \leq b_1\), and
\[ S_{k+1}^* = \text{fl}(S_k^* + \text{fl}(b_{k+1})). \]

We use the bounds
\[ S_n \leq 2 \text{ and } S_k \leq (n - k + 1)b_k. \]

Working as in part a, we get
\[ |e_n| \leq |e_1| + \sum_{k=2}^{n} b_k + \sum_{k=2}^{n} S_k \]
\[ \leq S_n + \sum_{k=2}^{n} S_k \]
\[ \leq S_n + \sum_{k=2}^{n} \sum_{j=1}^{k} b_j. \]

We change the order of summation to get
\[ |e_n| \leq S_n + \sum_{k=2}^{n} b_1 + \sum_{j=2}^{n} \sum_{k=j}^{n} b_j \]
\[ = S_n + (n - 1)b_1 + \sum_{j=2}^{n} (n - j + 1)b_j \]
\[ \leq S_n + b_n + \sum_{j=2}^{n-1} (n - j + 1)b_j \]
\[ \leq 3 + \sum_{j=1}^{n-1} \frac{1}{n - j + 1} \]
\[ = 3 + \sum_{m=2}^{n} \frac{1}{m} \]
\[ \leq 3 + \sum_{m=2}^{n} \int_{m-1}^{m} \frac{1}{x} \, dx \]
\[ = 3 + \int_{1}^{n} \frac{1}{x} \, dx \]
\[ = 3 + \ln n. \]

Therefore the absolute error is bounded by \((3 + \ln n)e\).

**Solution 12** We have the following facts about IEEE floating point numbers:

**Fact 12.1.** If \(x\) is already a floating point number, then \(x = \text{fl}(x)\); otherwise, there exists a floating point number \(y\) and next floating point number \(z\) with \(y < x < z\) such that \(\text{fl}(x) = y\) whenever \(y < x < \frac{y + z}{2}\), and \(\text{fl}(x) = z\) whenever \(\frac{y + z}{2} \leq x < z\).

**Fact 12.2.** There is no floating point number between \(x\) and \(\text{fl}(x)\) for any real number \(x\).
We first prove the following lemma.

**Lemma 12.1.** If real numbers $a$ and $b$ satisfy $a < b$, then $\text{fl}(a) \leq \text{fl}(b)$.

**Proof.** Consider the following two cases.

Case 1: There is a floating point number $c$ such that $a \leq c \leq b$.

Since $a \leq c$, by Fact 12.1, we have $\text{fl}(a) \leq c$. Also, since $c \leq b$, by Fact 12.2, we have $c \leq \text{fl}(b)$. Combining these inequalities gives $\text{fl}(a) \leq \text{fl}(b)$.

Case 2: There is no floating point number in $[a, b]$; in particular, neither $a$ nor $b$ is a floating point number.

Let $d$ be the largest floating point number with $d < a$, and $e$ be the smallest floating point number with $b < e$. We have the following results according to Fact 12.1:

- If $a \geq \frac{d+e}{2}$, then $\text{fl}(a) = \text{fl}(b) = e$.
- If $b < \frac{d+e}{2}$, then $\text{fl}(a) = \text{fl}(b) = d$.
- If $a < \frac{d+e}{2} \leq b$, then $\text{fl}(a) = d < e = \text{fl}(b)$.

In both cases, we have $\text{fl}(a) \leq \text{fl}(b)$ whenever $a < b$. This proves the lemma. \qed

Since $a < b$, we have $a < \frac{a+b}{2}$; by Lemma 12.1, we get $\text{fl}(a) \leq \text{fl}(\frac{a+b}{2})$. Since $a$ is a floating point number, $a = \text{fl}(a)$.

Therefore $a \leq \text{fl}(\frac{a+b}{2})$. A similar argument gives $\text{fl}(\frac{a+b}{2}) \leq b$. Therefore

$$a \leq \text{fl}(\frac{a+b}{2}) \leq b.$$