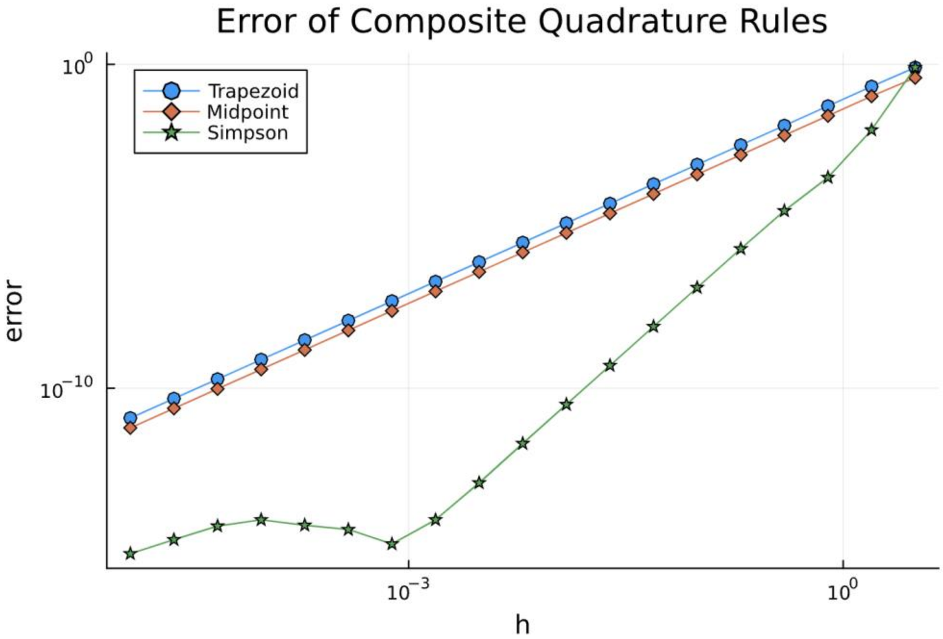
Cameron Beneteau (400274092)
beneteac@mcmaster.ca

Problem 1: Done in assignment3_handout.jl.

Problem 2:



a) Explain the slope of each line.

The above graph shows the error of composite quadrature rules on a log-log scale, which allows us to relate slopes to the error order of the method.

For the trapezoid and midpoint rules, these lines are parallel, which suggests they have the same order of error with respect to h. Furthermore, the slope of these lines can be interpreted as the order of the error rate. As seen in lecture, the error for the trapezoidal and midpoint rules is proportional to $O(h^2)$(second-order accuracy), which would imply a slope of 2 on the log-log plot. This is what we see on the graph above for the blue and red lines of these rules.

For Simpson's rule, this line has a steeper slope, indicating a higher order of error rate, which is $O(h^4)$ (fourth-order accuracy) as also seen in lecture. This corresponds to a slope of 4 on the log-log plot, seen above with much of the green line.

b) Explain the offset between the parallel lines.

The offset between the parallel lines for the trapezoidal and midpoint rules is due to the different coefficients in the leading term of their error expressions. While both rules have a quadratic error term of $O(h^2)$, the specific constant that multiplies h^2 in the error term depends on the quadrature rule. More specifically, this coefficient for the trapezoidal rule is $\frac{1}{12}$ while the coefficient for the midpoint rule is $\frac{1}{24}$. Note these values were derived in lecture. Analysing these numbers, the coefficient of the midpoint rule is half the magnitude of the trapezoidal rule, meaning it has a smaller error. This can also be seen graphically above, where red line of the midpoint rule lies below the blue line of the trapezoidal rule.

Cameron Beneteau (400274092)
beneteac@mcmaster.ca

c) Explain the roughly piecewise continuous behaviour of the third line.

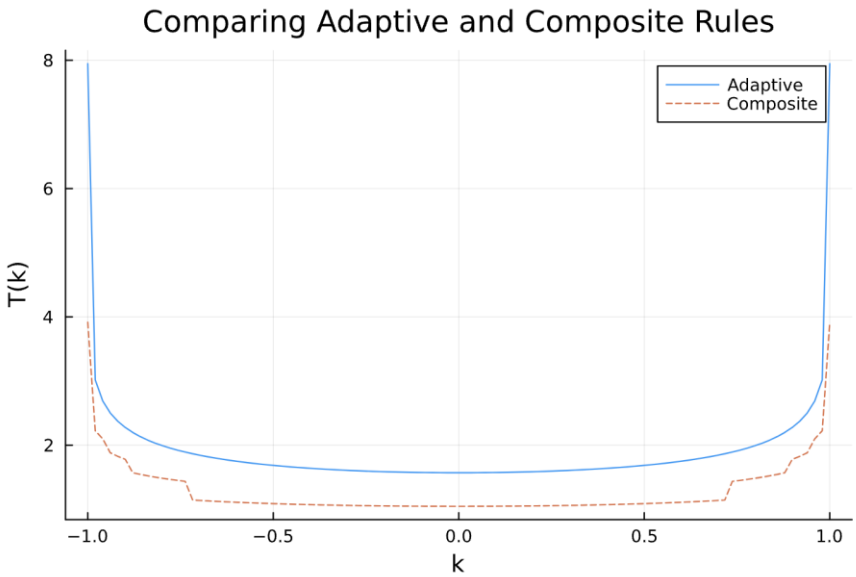
The third line representing the error of Simpson's rule on the plot shows piecewise linear behaviour that can be distinctly divided into two parts.

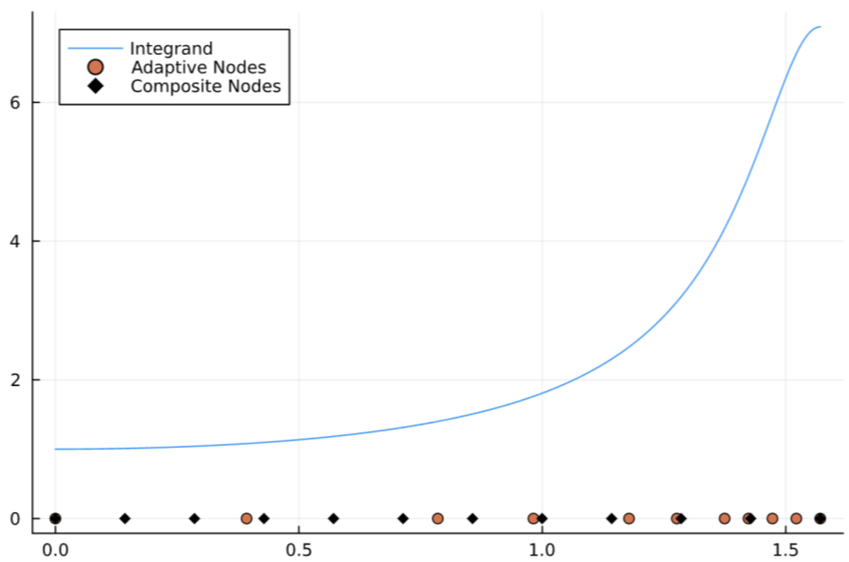
In the second part, the line looks to have a slope of 4, representative of what we anticipated based on the theoretical error term of $O(h^4)$ from earlier explanations. This is due to this error term being larger than other terms in the equation for Simpson's rule and as a result, dominating this result to give a slope of 4 for this part.

On the other hand, the first part does not follow this same trend with a slope of 4. The break in this trend signifies other factors may be overriding the theoretical $O(h^4)$ error term to dominate the error. Some examples of this could be numerical instabilities, such as round-off or truncation errors which may become larger as h becomes small. These errors can skew the true error of the integration, resulting in the inconsistent behaviour seen above.

Problem 3: Done in assignment3_handout.jl.

Problem 4:





a) Why do the integrals computed by the composite Simpson's rule differ from the adaptive solution? Which do you think is more accurate?

Composite Simpson's Rule: This method uses a fixed number of equally spaced nodes across the interval. It does not consider the function's behaviour and applies the same level of precision (number of nodes) uniformly across the interval.

Adaptive Simpson's Rule: This method dynamically adjusts the number of nodes based on the function's behaviour within subintervals. It uses a recursive strategy to place more nodes where the function has higher curvature (rapid changes), and fewer nodes where the function is flat or smoother.

The adaptive Simpson's rule is more accurate because of its efficiency in using more nodes in regions where the function has higher curvature and fewer nodes where the function is smoother. This specific allocation results in a better approximation with the same or even fewer total computations compared to the composite rule.

b) This script outputs a second plot which compares the "nodes" used by the adaptive and composite rules for k = 0.99. Describe the distribution of nodes for each method. How does the distribution of these nodes support your answer to part a)?

Adaptive Simpson's Rule Nodes: These nodes are denser in regions where the integrand changes rapidly or has high curvature. For the function T(k), the nodes are more concentrated around points of inflection or where the derivative of the function changes quickly.

Composite Simpson's Rule Nodes: These nodes are uniformly distributed across the interval, regardless of the function's behaviour.

These two observations from the above plots directly support the theory from part a). The adaptive rule's distribution of nodes being denser in regions of higher curvature shows its efficient use of nodes, which allows for higher accuracy compared to the composite rule. The composite rule's uniform distribution does not adapt to the function's behaviour, which leads to less accurate results, as shown in the example above. This is especially true for functions with multiple levels of curvature and smoothness across the interval of interest.

Cameron Beneteau (400274092)
beneteac@mcmaster.ca

Problem 5:

5) Finding the Jacobian of $F(x)$: $J(x) = D_x F$

$$F(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_n(x) \end{pmatrix} = \begin{pmatrix} \|x - p_1\| - d_1 \\ \|x - p_2\| - d_2 \\ \vdots \\ \|x - p_n\| - d_n \end{pmatrix} = 0 \qquad p(x) = \begin{pmatrix} | & | & & | \\ p_1 & p_2 & \cdots & p_n \\ | & | & & | \end{pmatrix}$$

Notation: $p_i(j)$ is i^{th} column, j^{th} row

$$f_i(x) = \|x - p_i\| - d_i = \left((x_1 - p_i(1))^2 + (x_2 - p_i(2))^2 + \cdots + (x_n - p_i(n))^2\right)^{1/2} - d_i$$

By definition:

$$J(x) = D_x F = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{pmatrix}$$

Finding $\frac{\partial f_i}{\partial x_j}$:

$$\frac{\partial f_i}{\partial x_j} = \frac{\partial}{\partial x_j}\left(\|x - p_i\| - d_i\right) = \frac{\partial}{\partial x_j}\left(\left((x_1 - p_i(1))^2 + (x_2 - p_i(2))^2 + \cdots + (x_n - p_i(n))^2\right)^{1/2} - d_i\right)$$

$$\frac{\partial f_i}{\partial x_j} = \frac{1}{2}\left((x_1 - p_i(1))^2 + (x_2 - p_i(2))^2 + \cdots + (x_n - p_i(n))^2\right)^{1/2} \cdot 2(x_j - p_i(j))$$

Note: Derivatives of all other $(x_k - p_i(k))^2$ term where $k \neq j$ are 0

Since $\|x - p_i\| = \left((x_1 - p_i(1))^2 + (x_2 - p_i(2))^2 + \cdots + (x_n - p_i(n))^2\right)^{1/2}$, simplify to get:

$$\frac{\partial f_i}{\partial x_j} = \frac{(x_j - p_i(j))}{\|x - p_i\|}$$

Filling the Jacobian matrix with this result:

$$\boxed{J(x) = D_x F = \begin{pmatrix} \frac{(x_1 - p_1(1))}{\|x - p_1\|} & \frac{(x_2 - p_1(2))}{\|x_2 - p_1\|} & \cdots & \frac{(x_n - p_1(n))}{\|x - p_1\|} \\ \frac{(x_1 - p_2(1))}{\|x - p_2\|} & \frac{(x_2 - p_2(2))}{\|x - p_2\|} & \cdots & \frac{(x_n - p_2(n))}{\|x - p_2\|} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{(x_1 - p_n(1))}{\|x - p_n\|} & \frac{(x_2 - p_n(2))}{\|x - p_n\|} & \cdots & \frac{(x_n - p_n(n))}{\|x - p_n\|} \end{pmatrix}}$$

Cameron Beneteau (400274092)
beneteac@mcmaster.ca

Problem 6: Done in assignment3_handout.jl.

Problem 7:

7) Finding the Gradient of $f(x)$: ∇f

$$f(x) = \sum_{i=1}^{m} |f_i(x)|^2 = |f_1(x)|^2 + |f_2(x)|^2 + \cdots + |f_m(x)|^2 = (f_1(x))^2 + (f_2(x))^2 + \cdots + (f_m(x))^2$$

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

Finding $\frac{\partial f}{\partial x_j}$:

Using $\frac{\partial f_i}{\partial x_j} = \frac{(x_j - p_i(j))}{\|x - p_i\|}$ from Q5

$$\frac{\partial f}{\partial x_j} = \frac{\partial}{\partial x_j}\left(|f_i(x)|^2\right) = 2\left(\|x - p_i\| - d_i\right) \cdot \frac{(x_j - p_i(j))}{\|x - p_i\|}$$

$$\frac{\partial f}{\partial x_j} = \frac{2(\|x-p_1\| - d_1)(x_1 - p_1(1))}{\|x - p_1\|} + \frac{2(\|x-p_2\| - d)(x_1 - p_2(1))}{\|x - p_2\|} + \cdots + \frac{2(\|x-p_m\| - d_m)(x_1 - p_m(1))}{\|x - p_m\|}$$

$$\frac{\partial f}{\partial x_j} = 2\sum_{i=1}^{m} \frac{(\|x-p_i\| - d_i)(x_j - p_i(j))}{\|x - p_i\|} = 2\sum_{i=1}^{m} \frac{f_i(x)(x_j - p_i(j))}{\|x - p_i\|}$$

Filling the Gradient matrix with this result:

$$\boxed{\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix} = \begin{pmatrix} 2\sum_{i=1}^{m} \frac{f_i(x)(x_1 - p_i(1))}{\|x - p_i\|} \\ 2\sum_{i=1}^{m} \frac{f_i(x)(x_2 - p_i(2))}{\|x - p_i\|} \\ \vdots \\ 2\sum_{i=1}^{m} \frac{f_i(x)(x_n - p_i(n))}{\|x - p_i\|} \end{pmatrix}}$$

Cameron Beneteau (400274092)
beneteac@mcmaster.ca

Finding the Hessian of $f(x)$: $\nabla^2 f$

$$\nabla^2 f = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \frac{\partial^2 f}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

Finding $\frac{\partial^2 f}{\partial x_j \partial x_K}$ (forget about $2\sum_{i=1}^{m}$ for now)

Define $\frac{f_i(x)(X_1 - P_i(1))}{\|X - P_i\|}$ as $\frac{f_i(x) \cdot g(x)}{h(x)}$

Using Quotient Rule and Chain Rule, we get

$$\frac{\partial}{\partial x}\left(\frac{f_i(x) \cdot g(x)}{h(x)}\right) = \frac{h(x)\left(f_i(x) g'(x) + g(x) f_i'(x)\right) - f_i(x)\, g(x)\, h'(x)}{h(x)^2}$$

$$\left(\nabla^2 f\right)_{j,K} = \frac{\partial^2 f}{\partial x_j \partial x_K} = \frac{\partial}{\partial x_j}\left(2\sum_{i=1}^{m} \frac{f_i(x)(X_K - P_i(K))}{\|X - P_i\|}\right)$$

$f_i(x) = \|X - P_i\| - d_i$, $\quad g(x) = X_K - P_i(K)$, $\quad h(x) = \|X - P_i\|$

$$\frac{\partial f_i}{\partial x_j} = \frac{(X_j - P_i(j))}{\|X - P_i\|} \qquad \frac{\partial g}{\partial x_j} = 0 \text{ when } j \neq K, \; = 1 \text{ when } j = K \qquad \frac{\partial h}{\partial x_j} = \frac{(X_j - P_i(j))}{\|X - P_i\|}$$

From Q5 $\qquad\qquad$ From Q5

when $j \neq K$ (non-diagonal element)

$$\left(\nabla^2 f\right)_{j,K} = 2 \sum_{i=1}^{m} \frac{\|x - p_i\| \left(f_i(x)(0) + (x_K - p_i(K)) \left(\frac{(x_j - p_i(j))}{\|x - p_i\|} \right) \right) - f_i(x)(x_K - p_i(K)) \frac{(x_j - p_i(j))}{\|x - p_i\|}}{\|x - p_i\|^2}$$

$$\left(\nabla^2 f\right)_{j,K} = 2 \sum_{i=1}^{m} \frac{(x_K - p_i(K))(x_j - p_i(j))}{\|x - p_i\|^2} - \frac{f_i(x)(x_K - p_i(K))(x_j - p_i(j))}{\|x - p_i\|^3}$$

$$\left(\nabla^2 f\right)_{j,K} = 2 \sum_{i=1}^{m} \frac{(x_K - p_i(K))(x_j - p_i(j))}{\|x - p_i\|^2} \left(1 - \frac{f_i(x)}{\|x - p_i\|} \right)$$

when $j = K$ (diagonal element)

$$\left(\nabla^2 f\right)_{j,j} = 2 \sum_{i=1}^{m} \frac{\|x - p_i\| \left(f_i(x)(1) + (x_j - p_i(j)) \left(\frac{(x_j - p_i(j))}{\|x - p_i\|} \right) \right) - f_i(x)(x_j - p_i(j)) \frac{(x_j - p_i(j))}{\|x - p_i\|}}{\|x - p_i\|^2}$$

$$\left(\nabla^2 f\right)_{j,K} = 2 \sum_{i=1}^{m} \frac{f_i(x)}{\|x - p_i\|^2} + \frac{(x_j - p_i(j))^2}{\|x - p_i\|^2} - \frac{f_i(x)(x_j - p_i(j))^2}{\|x - p_i\|^3}$$

Cameron Beneteau (400274092)
beneteac@mcmaster.ca

Filling the Hessian matrix with these results:

$$\boxed{\nabla^2 f = \begin{pmatrix} 2\sum_{i=1}^{m} \frac{f_i(x)}{\|x-p_i\|^2} + \frac{(x_1 - p_i(1))^2}{\|x-p_i\|^2} - \frac{f_i(x)(x_1 - p_i(1))^2}{\|x-p_i\|^3} & 2\sum_{i=1}^{m} \frac{(x_2 - p_i(2))(x_1 - p_i(1))}{\|x-p_i\|^2}\left(1 - \frac{f_i(x)}{\|x-p_i\|}\right) & \cdots & 2\sum_{i=1}^{m} \frac{(x_n - p_i(n))(x_1 - p_i(1))}{\|x-p_i\|^2}\left(1 - \frac{f_i(x)}{\|x-p_i\|}\right) \\ 2\sum_{i=1}^{m} \frac{(x_1 - p_i(1))(x_2 - p_i(2))}{\|x-p_i\|^2}\left(1 - \frac{f_i(x)}{\|x-p_i\|}\right) & 2\sum_{i=1}^{m} \frac{f_i(x)}{\|x-p_i\|^2} + \frac{(x_2 - p_i(2))^2}{\|x-p_i\|^2} - \frac{f_i(x)(x_2 - p_i(2))^2}{\|x-p_i\|^3} & \cdots & 2\sum_{i=1}^{m} \frac{(x_n - p_i(n))(x_2 - p_i(2))}{\|x-p_i\|^2}\left(1 - \frac{f_i(x)}{\|x-p_i\|}\right) \\ \vdots & \vdots & \ddots & \vdots \\ 2\sum_{i=1}^{m} \frac{(x_1 - p_i(1))(x_n - p_i(n))}{\|x-p_i\|^2}\left(1 - \frac{f_i(x)}{\|x-p_i\|}\right) & 2\sum_{i=1}^{m} \frac{(x_2 - p_i(2))(x_n - p_i(n))}{\|x-p_i\|^2}\left(1 - \frac{f_i(x)}{\|x-p_i\|}\right) & \cdots & 2\sum_{i=1}^{m} \frac{f_i(x)}{\|x-p_i\|^2} + \frac{(x_n - p_i(n))^2}{\|x-p_i\|^2} - \frac{f_i(x)(x_n - p_i(n))^2}{\|x-p_i\|^3} \end{pmatrix}}$$

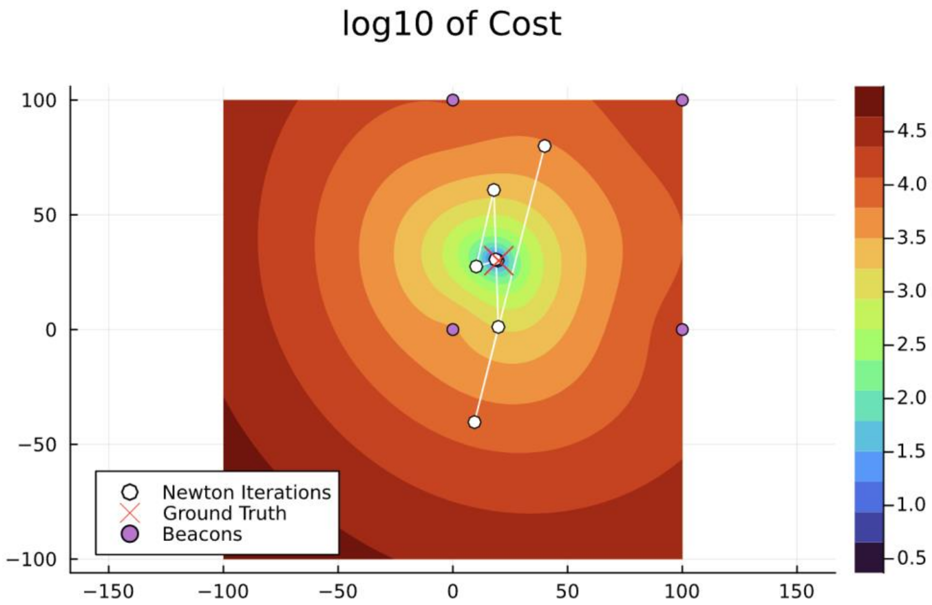
Note: Due to commutative multiplication, we can say:

$(x_k - p_i(k))(x_j - p_i(j)) = (x_j - p_i(j))(x_k - p_i(k))$ which also means $\nabla^2 f_{j,k} = \nabla^2 f_{k,j}$

By this property, we can say this Hessian matrix is symmetric.

Cameron Beneteau (400274092)
beneteac@mcmaster.ca

Problem 8:



a) Provide numerical evidence that your implementation of Newton's method has converged to a critical point.

The critical point at [20, 30] corresponds to a gradient of 0, indicating that the solution has reached this critical point. However, due to the computational limitations of this implementation of Newton's method, achieving a gradient norm of exactly 0 is not always possible. Instead, we can look for a gradient norm close to zero, less than our accepted tolerance, which indicates convergence toward the critical point over multiple iterations.

Iteration	Gradient	Norm of Gradient
1	[61.62263522567932, 172.72807967954768]	183.39121757036594
2	[-81.64075572677218, -282.51562746380785]	294.0753181702054
3	[-35.569361265574116, -90.41356973503092]	97.15859742231885
4	[6.578202636163592, 113.46285228172809]	113.65338357403918
5	[-19.09725180618199, -11.215246224213708]	22.146936005200835
6	[0.7852889432917264, -0.6020281639124377]	0.9895032261695852
7	[0.0019456888639375683, 0.0024505274508173705]	0.0031290238002386855
8	[-6.084824841906533e-9, -9.302745646727617e-9]	1.1116032112421791e-8

The above table shows the gradient and its norm for each iteration of Newton's method. Over the iterations, it can be seen the norm of the gradient is decreasing toward 0. By the 8th iteration, the norm of the gradient is close to zero, and less than our error tolerance, at 1.1116032112421791e-8. While not exactly 0, this value is extremely small and indicates that Newton's method implementation has converged to a critical point.

b) Provide numerical evidence that this critical point is a local minimum.

Critical points correspond to a gradient of 0, but a critical point is a local minimum only if its Hessian matrix is positive definite. For a matrix to be positive definite, the following two conditions must be met: it is symmetric, and all eigenvalues are positive. As outlined in the proof of Question 7 in this

Cameron Beneteau (400274092)
beneteac@mcmaster.ca

assignment, the Hessian matrix for this example is symmetric. The following table shows the properties of the Hessian matrix for the final iteration of Newton's method:

Property	**Value**
Hessian	[3.555001825882777 0.803876987166215; 0.803876987166215 4.336989501444659]
Hessian Eigenvalues	[3.0520748664573327, 4.8399164608701035]

As seen above, the Hessian matrix is symmetric since values at positions (1, 2) and (2, 1) are equal at 0.803876987166215 and both eigenvalues are positive at 3.0520748664573327 and 4.8399164608701035. These properties satisfy the conditions needed for a matrix to be positive definite. Since the Hessian matrix is positive definite, we can assume the critical point of [20, 30] is a local minimum.