1. (**4 points**)

    (a) Show that the ℓ_1-, ℓ_2- and ℓ_∞-norms satisfy the three properties of a norm.

    (Non-negativity) For ℓ_1, ℓ_2, and ℓ_∞ norms, for any vector x, the norm $\|x\|_p \geq 0$ since it's a sum (or maximum) of non-negative values (absolute values of the components of x). Additionally, $\|x\|_p = 0$ if and only if $x = 0$, which is true for these norms because the sum (or maximum) of absolute values equals zero if and only if every component of x is zero.

    (Absolute Scalability) For ℓ_1, ℓ_2, and ℓ_∞ norms, absolute scalability holds. For any scalar α and vector x, $\|\alpha x\|_p = |\alpha|\|x\|_p$. This is true because scaling a vector by α scales each component by $|\alpha|$, and thus the sum or maximum of their absolute values is scaled by $|\alpha|$.

    (Triangle Inequality) For ℓ_1 norm: $\|x + y\|_1 = \sum |x_i + y_i| \leq \sum(|x_i| + |y_i|) = \|x\|_1 + \|y\|_1$. For ℓ_2 norm:, $\|x + y\|_2 \leq \|x\|_2 + \|y\|_2$. For ℓ_∞ norm: $\|x + y\|_\infty = \max_i |x_i + y_i| \leq \max_i(|x_i| + |y_i|) \leq \max_i |x_i| + \max_i |y_i| = \|x\|_\infty + \|y\|_\infty$.

    (b) The "zero-norm" is defined as $\|x\|_0 =$ number of non-zero elements in x. Show that this "norm" only satisfies some and not all of the properties of a norm.

    (Non-negativity) $\|x\|_0$ satisfies the property since $\|x\|_0 = 0$ iff x is the zero vector (i.e., all elements of x are zero). Therefore, $\|x\|_0 = 0$ implies that $x = 0$. Additionally, $\|x\|_0$, counts the number of non-zero elements in a vector x, the value is always non-negative. Thus, $\|x\|_0 \geq 0$ for any vector x.

    (Triangle Inequality) The zero-norm satisfies the triangle inequality. For vectors x and y, if an element in $x + y$ is non-zero, then the corresponding element in either x or y (or both) must be non-zero. Thus, the number of non-zero elements in $x + y$ cannot exceed the sum of the number of non-zero elements in x and y, leading to $\|x + y\|_0 \leq \|x\|_0 + \|y\|_0$.

    (Absolute Scalability) The zero-norm does not satisfy absolute scalability. For a scalar $\alpha \neq 0$ and a vector x, $\|\alpha x\|_0 = \|x\|_0$, as scaling x by a non-zero scalar does not change the number of non-zero elements. However, this is not equal to $|\alpha|\|x\|_0$ unless $|\alpha| = 1$. For example, with $\alpha = 2$ and $x = (1, 0)$, $\|\alpha x\|_0 = \|(2, 0)\|_0 = 1$, which is not equal to $|2|\|x\|_0 = 2$

    (c) Let $x = (-5, 3, 1, -1)$. Compute $\|x\|_1$, $\|x\|_2$, and $\|x\|_\infty$.

    $$\|x\|_1 = |-5| + |3| + |1| + |-1| = 5 + 3 + 1 + 1 = 10.$$

    $$\|x\|_2 = \sqrt{(-5)^2 + 3^2 + 1^2 + (-1)^2} = \sqrt{25 + 9 + 1 + 1} = \sqrt{36} = 6.$$

    $$\|x\|_\infty = \max\{|-5|, |3|, |1|, |-1|\} = \max\{5, 3, 1, 1\} = 5.$$

    (d) Suppose $x \in \mathbb{R}^2$. Show that $\|x\|_\infty \leq \|x\|_2 \leq \|x\|_1$.

    (For $\|x\|_\infty \leq \|x\|_2$) Let $x = (x_1, x_2)$, hence $\|x\|_\infty = \max\{|x_1|, |x_2|\}$. Let's assume $|x_1| \geq |x_2|$. Thus, $\|x\|_\infty = |x_1|$. Now, $\|x\|_2 = \sqrt{x_1^2 + x_2^2}$. Since $|x_1| \geq |x_2|$, we have $x_1^2 \geq x_2^2$, and thus

$$x_1^2 + x_2^2 \geq x_1^2. \text{ Therefore, } \|x\|_2 \geq \sqrt{x_1^2} = |x_1| = \|x\|_\infty.$$

(For $\|x\|_2 \leq \|x\|_1$) Let $x = (x_1, x_2)$, hence $\|x\|_1 = |x_1| + |x_2|$. Using the Cauchy-Schwarz inequality, we have $\|x\|_2^2 = x_1^2 + x_2^2 \leq (|x_1| + |x_2|)^2 = \|x\|_1^2$. Therefore, $\|x\|_2 \leq \|x\|_1$.

2. (**5 points**)

   (a) Find a basis for $\mathbb{R}^3$.

   A basis for $\mathbb{R}^3$ can be the standard basis vectors: $\{(1,0,0),(0,1,0),(0,0,1)\}$. These vectors are linearly independent because no vector can be expressed as a linear combination of the others. Additionally, they span $\mathbb{R}^3$, as any vector in $\mathbb{R}^3$ can be expressed as a linear combination of these vectors.

   (b) Find a set of vectors that span $\mathbb{R}^3$ that is not a basis for $\mathbb{R}^3$.

   A set of vectors that spans $\mathbb{R}^3$ but is not a basis would be any set with more than three vectors, for example, $\{(1,0,0),(0,1,0),(0,0,1),(1,1,1)\}$. This set spans $\mathbb{R}^3$ but is not linearly independent, hence not a basis.

   (c) Find a set of linearly independent vectors in $\mathbb{R}^3$ that is not a basis for $\mathbb{R}^3$.

   A set of linearly independent vectors in $\mathbb{R}^3$ that is not a basis would be any set with fewer than three vectors, for example, $\{(1,0,0),(0,1,0)\}$. These vectors are linearly independent but do not span $\mathbb{R}^3$.

   (d) If $\{v_1, v_2, \cdots, v_n\}$ is a basis for $\mathcal{S}$, show that every vector in $\mathcal{S}$ is uniquely written as a linear combination of the v_i's.

   If $\{v_1, v_2, \ldots, v_n\}$ is a basis for S, every vector v in S can be expressed as a combination like $v = a_1 v_1 + a_2 v_2 + \ldots + a_n v_n$. To show this is unique, let's assume there is another way to write v. For example, $v = b_1 v_1 + b_2 v_2 + \ldots + b_n v_n$, subtracting these for v we get a sum of the basis vectors equal to zero. Since the basis vectors are linearly independent, all the differences in coefficients must be zero. Hence, every vector in S is uniquely written as a linear combination of the basis vectors.

   (e) If the set $\{v_1, v_2, \cdots, v_n\}$ is orthonormal, show that it is also linearly independent.

   To prove an orthonormal set $\{v_1, v_2, \ldots, v_n\}$ is linearly independent, assume a linear combination $a_1 v_1 + a_2 v_2 + \ldots + a_n v_n = 0$. Taking the dot product with v_1, we get a_1 (since $v_1 \cdot v_1 = 1$) and zeros for other terms (since v_1 is orthogonal to other vectors). This implies $a_1 = 0$. Repeating this for each vector shows all a_i are zero, proving linear independence.

3. (**3 points**) Let

$$f(x) = \begin{cases} x^2 \sin\left(\dfrac{1}{x}\right) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

   (a) Show that $f(x)$ is continuous at $x = 0$.

   To show continuity at $x = 0$, we need to prove that the limit of $f(x)$ as x approaches 0 equals $f(0)$. Since $f(0) = 0$, we examine the limit of $x^2 \sin\left(\frac{1}{x}\right)$ as x approaches 0. As $\sin\left(\frac{1}{x}\right)$ is bounded between -1 and 1, the product $x^2 \sin\left(\frac{1}{x}\right)$ approaches 0. Thus, $f(x)$ is continuous at $x = 0$.

(b) Show that $f(x)$ is differentiable at $x = 0$.

> To prove differentiability at $x = 0$, we need to find the derivative of $f(x)$ at 0. The derivative of $f(x)$ for $x \neq 0$ is $2x \sin\left(\frac{1}{x}\right) - \cos\left(\frac{1}{x}\right)$. As x approaches 0, the first term vanishes and the second term oscillates. We need the derivative from the definition of derivative, which is the limit of $\frac{f(x)-f(0)}{x-0}$. This is a known limit which is 0, hence $f(x)$ is differentiable at $x = 0$.

(c) Show that the derivative of $f(x)$ is not continuous at $x = 0$.

> The derivative of $f(x)$, is $2x \sin\left(\frac{1}{x}\right) - \cos\left(\frac{1}{x}\right)$ for $x \neq 0$, and 0 at $x = 0$. The term $-\cos\left(\frac{1}{x}\right)$ oscillates between -1 and 1 as x approaches 0, which means the derivative does not approach a single value as x approaches 0. Hence, the derivative of $f(x)$ is not continuous at $x = 0$.

4. (**4 points**) Taylor expansion.

(a) Compute the Taylor expansion of $f(x) = \sin(x)$ around $x_0 = 0$.

> The Taylor expansion of $\sin(x)$ around $x_0 = 0$ can be obtained by calculating the derivatives of $\sin(x)$ at 0. The Taylor series is given by $f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \cdots$ The derivatives are $\sin(x)$, $\cos(x)$, $-\sin(x)$, $-\cos(x)$, and so on. Evaluating at 0, we get $\sin(0) = 0$, $\cos(0) = 1$, $-\sin(0) = 0$, etc. Therefore, the Taylor expansion of $\sin(x)$ around 0 is $x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots$

(b) Without knowing the exact value of $\sin(0.01)$, compute an estimate of $\sin(0.01)$ that is within 10^{-8} of the exact value.

> To estimate $\sin(0.01)$ within 10^{-8} of the exact value, we use the Taylor series expansion of $\sin(x)$ around 0. We truncate the series where the term's absolute value is less than 10^{-8}. For $x = 0.01$, the first few terms of the series are $0.01 - \frac{0.01^3}{3!} + \frac{0.01^5}{5!} - \frac{0.01^7}{7!} + \cdots$. Calculating these terms, we find that the terms beyond $\frac{0.01^5}{5!}$ are smaller than 10^{-8}. Therefore, an estimate of $\sin(0.01)$ is $0.01 - \frac{0.01^3}{3!} + \frac{0.01^5}{5!}$.

5. (**6 points**) Consider the functions

$$f(x) = x^2, \quad g(x) = x^3, \quad \text{and} \quad h(x) = x^4.$$

(a) What are the critical points for each function?

> For $f(x) = x^2$, the derivative is $f'(x) = 2x$. Setting $f'(x) = 0$ gives the critical point at $x = 0$.
> For $g(x) = x^3$, the derivative is $g'(x) = 3x^2$. Setting $g'(x) = 0$ gives the critical point at $x = 0$.
> For $h(x) = x^4$, the derivative is $h'(x) = 4x^3$. Setting $h'(x) = 0$ gives the critical point at $x = 0$.

(b) Which conditions of optimality do these critical points satisfy?

> For $f(x) = x^2$, the second derivative is $f''(x) = 2$, which is positive, indicating a minimum at $x = 0$.
> For $g(x) = x^3$, the second derivative is $g''(x) = 6x$, which is zero at $x = 0$. This does not provide conclusive information about optimality.
> For $h(x) = x^4$, the second derivative is $h''(x) = 12x^2$, which is nonnegative, indicating a minimum at $x = 0$.

(c) From this exercise, what can you conclude about the second-order a necessary condition for optimality?

This exercise shows that the 2nd order necessary condition for optimality helps identify potential minima. For $f(x)$ and $h(x)$, where the second derivatives at the critical points are nonnegative, we identify minima. However, for $g(x)$, where the second derivative at the critical point is zero, the condition is inconclusive. Thus, while the second-order necessary condition is useful, it may not always be sufficient to determine optimality.

6. (**4 points**) Convex sets and functions.

(a) Provide 3 examples of a convex set.

1. The set of all points on and inside a circle in the plane.
2. The set of all points in a closed interval on the real line
3. A solid cube in 3-dimensional space, including its interior and surface.

(b) Provide 2 examples of a convex function and 2 examples of functions that are not convex.

Examples of convex functions:
1. $f(x) = x^2$, which is a parabola opening upwards.
2. $g(x) = e^x$, the exponential function.
Examples of non-convex functions:
1. $h(x) = \sin(x)$, as it oscillates and does not satisfy the definition of convexity.
2. $j(x) = x^3$, which is not convex due to its inflection point at the origin.

7. (**6 points**) Compute the gradient and Hessian of the following functions:

(a) $f(x) = x_1 + x_2^2 + x_3^3$

The gradient of f is $\nabla f(\mathbf{x}) = \begin{pmatrix} 1 \\ 2x_2 \\ 3x_3^2 \end{pmatrix}$. The Hessian matrix of f is $\mathbf{H} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 6x_3 \end{pmatrix}$.

(b) $f(x) = x_1^2 + x_1 x_2 + x_2^2 + x_2 x_3 + x_3^2$

The gradient of f is $\nabla f(\mathbf{x}) = \begin{pmatrix} 2x_1 + x_2 \\ 2x_2 + x_1 + x_3 \\ 2x_3 + x_2 \end{pmatrix}$. The Hessian matrix of f is $\mathbf{H} = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}$.

(c) $f(x) = x_1 \cos x_2 + x_1 x_2 x_3 + \sin x_3$

The gradient of f is $\nabla f(\mathbf{x}) = \begin{pmatrix} \cos x_2 + x_2 x_3 \\ -x_1 \sin x_2 + x_1 x_3 \\ x_1 x_2 + \cos x_3 \end{pmatrix}$. The Hessian matrix of f is complex and involves second derivatives of trigonometric functions.

(d) $f(x) = \frac{1}{2} x^T H x$, where H is an $n \times n$ matrix with constant entries and $x \in \mathbb{R}^n$

The gradient of f is $\nabla f(\mathbf{x}) = \mathbf{H}\mathbf{x}$. The Hessian matrix of f is $\mathbf{H}$ itself, as it is a constant matrix.

(e) $f(x) = b^T Ax - \frac{1}{2}x^T A^T Ax$, where A is an $m \times n$ matrix with constant entries, $b \in \mathbb{R}^m$ is a vector with constant entries, and $x \in \mathbb{R}^n$.

The gradient of f is $\nabla f(\mathbf{x}) = \mathbf{A}^T \mathbf{b} - \mathbf{A}^T \mathbf{A} \mathbf{x}$. The Hessian matrix of f is $-\mathbf{A}^T \mathbf{A}$, which is a constant matrix based on $\mathbf{A}$.

8. (**3 points**) Using the definition, determine whether each of the following matrices is positive definite, positive semidefinite, or neither.

$$(a) \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \qquad (b) \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix} \qquad (c) \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

For (a) To determine if the matrix is positive definite, positive semidefinite, or neither, we look at the principal minors. The first principal minor is 2, and the determinant of the matrix is $2 \times 2 - (-1) \times (-1) = 3$. Both are positive, indicating that the matrix is positive definite.

For (b) First, we check the determinant of the matrix: $2 \times 2 - (-1) \times 1 = 4 + 1 = 5$. However, the matrix is not symmetric, which is a requirement for being positive definite or positive semidefinite. Therefore, this matrix is neither positive definite nor positive semidefinite.

For (c) The first principal minor is 1, which is positive. The determinant of the matrix is $1 \times 1 - (-1) \times (-1) = 0$. Since the determinant is zero, the matrix is positive semidefinite, but not positive definite.

9. (**2 points**) Consider **Rosenbrock's function**, which is given by

$$f(x) = (1 - x_1)^2 + 100(x_2 - x_1^2)^2.$$

Check which conditions for optimality the point $x = (1, 1)$ satisfies.

To check for optimality, we compute the gradient and Hessian of $f(\mathbf{x})$ at $\mathbf{x} = (1, 1)$. The gradient is given by $\nabla f(\mathbf{x}) = \begin{pmatrix} -2(1 - x_1) - 400x_1(x_2 - x_1^2) \\ 200(x_2 - x_1^2) \end{pmatrix}$. Evaluating at $\mathbf{x} = (1, 1)$, we get $\nabla f(1, 1) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, showing that $\mathbf{x} = (1, 1)$ is a critical point.

The Hessian matrix of f is $\mathbf{H} = \begin{pmatrix} 2 - 400(x_2 - 3x_1^2) & -400x_1 \\ -400x_1 & 200 \end{pmatrix}$. Evaluating at $\mathbf{x} = (1, 1)$, we get $\mathbf{H}(1, 1) = \begin{pmatrix} 802 & -400 \\ -400 & 200 \end{pmatrix}$. The eigenvalues of this Hessian matrix are both positive, indicating that $\mathbf{x} = (1, 1)$ is a local minimum for the function.

10. (**3 points**) Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ be constant. Consider the function $F \colon \mathbb{R}^n \to \mathbb{R}^m$ given by

$$F(x) = Ax - b.$$

(a) What is $F'(x)$ equal to?

$F'(x)$ is the Jacobian matrix of F. Since $F(x) = Ax - b$ is a linear function, its derivative with respect to x is simply the matrix A. Therefore, $F'(x) = A$.

(b) When does a zero of F exist, i.e., there is some x such that $F(x) = 0$? If such a zero exists, when is it unique and when is it not unique?

If the equation $\mathrm{Ax} = \mathrm{b}$ can be solved, then there's a zero of F. This is true if b is part of A's column space. The solution is unique if all of A's columns are linearly independent from each other. But if A's columns are linearly dependent then there's more than one solution, actually, there are endless solutions.