(1) Consider the function \( f(x, y, z) = xyz \), with \( y = x^2 \) and \( z = x^{1/3} \).

(a) Rewrite \( f \) as a function \( g : \mathbb{R} \to \mathbb{R} \) alone and compute \( g'(x) \). Using \( g' \), approximate the change in \( f \) when \( x \) increases by 0.1 units, starting from (8, 64, 2).

Ans:

\[
g(x) = f(x, x^2, x^{1/3}) = x \times x^2 \times x^{1/3} = x^{10/3} \quad \text{so that} \quad g'(x) = \frac{10}{3}x^{7/3}
\]

\[
g'(8) = \frac{10}{3} \times 8^{7/3} = \frac{10}{3} \times 2^7 = 426.6667
\]

\[
dg = df = 0.1 \times \frac{10}{3} \times 128 = 42.6.
\]

(b) Compute the total derivative of \( f \) with respect to \( x \). Using the total derivative, approximate the change in \( f \) when \( x \) increases by 0.1 units, starting from (8, 64, 2).

Ans:

\[
\frac{df}{dx} = f_x + f_y \frac{dy}{dx} + f_z \frac{dz}{dx} = yz + xz \times 2x + yx \times \frac{1}{3}x^{-2/3}
\]

\[
= x^{7/3} + x^{4/3} \times 2x + x^3 \times \frac{1}{3}x^{-2/3}
\]

\[
= x^{7/3}(1 + 2 + 1/3) = \frac{10}{3}x^{7/3}
\]

so that

\[
df = \frac{df(8)}{dx} = 0.1 \times \frac{10}{3} \times 8^{7/3} = 42.6
\]

(c) Write down the differential of \( f \) at (8, 64, 2). Using the differential, approximate the change in \( f \) when \( x \) increases by 0.1 units, starting from (8, 64, 2).

Ans:

\[
\nabla f(x, y, z) = \begin{bmatrix} yz & xz & yx \end{bmatrix} \quad \text{so that} \quad \nabla f(8, 64, 2) = \begin{bmatrix} 128 & 16 \end{bmatrix}
\]
\[
\frac{df}{dx} = \mathbf{\nabla} f(x, y, z) \left[ \begin{array}{c} dx \\ y'(x)dx \\ z'(x)dx \end{array} \right] = \left[ \begin{array}{rrr} 0.1 \\ 16 \times 0.1 \\ 0.0833 \times 0.1 \end{array} \right] = \left[ \begin{array}{rrr} 128 & 16 & 512 \end{array} \right] \left[ \begin{array}{c} dx \\ y'(x)dx \\ z'(x)dx \end{array} \right] = 12.8 + 16 \times 0.16 + 512 \times 0.025 = 42.7
\]

(d) Identify the direction \( h^* \) that \((x, y, z)\) moves in, starting from \((8, 64, 2)\), when \( x \) increases. Write down the directional derivative of \( f \) in the direction \( h^* \), i.e., \( f_{h^*}(\cdot, \cdot, \cdot) \), and evaluate this derivative at \((8, 64, 2)\). Using \( f_{h^*}(8, 64, 2) \), approximate the change in \( f \) when \( x \) increases by 0.1 units, starting from \((8, 64, 2)\).

**Ans:** When \( x \) increases by one, the vector \((x, y, z)\) increases in the direction \((dx, y'(x)dx, z'(x)dx)\). When \( x = 8 \) and \( dx = 1 \), therefore, \((x, y, z)\) increases in the direction \( h^* = (1, 16, 0.0833) \). The unit length vector pointing in this direction is \( h^*/||h^*|| = (0.0624, 0.9980, 0.0052) \). Therefore, using the differential to compute the directional derivative, \( f_{h^*}(8, 64, 2) = \left[ \begin{array}{rrr} 0.0624 \\ 0.9980 \\ 0.0052 \end{array} \right] = 26.613 \). Now a \( dx \) of 0.1 induces a shift in \( \mathbb{R}^3 \) of \( 0.1h^* \), which has length \( 0.1||h^*|| \). \( df = f_{h^*}(8, 64, 2) \times 0.1||h^*|| = 42.67 \)

(e) Check to see that all four of these distinct methods give you the same answer!

**Ans:** Amazingly, they do!

(2) Recall that a function \( f : \mathbb{R}^n \to \mathbb{R}^m \) is nothing more than \( m \) functions, \( f^1, \ldots, f^m \), each mapping \( \mathbb{R}^n \to \mathbb{R} \), and stacked on top of each other.

(a) Using this fact, write down a formal definition of the directional derivative of \( f \) at \( x_0 \) in the direction \( h \in \mathbb{R}^n \), for a function \( f : \mathbb{R}^n \to \mathbb{R}^m \). Your definition should be of the form

\[
\text{blah, blah} = \lim_{k \to \infty} \frac{\text{blah}}{\text{blah, blah}}
\]
Ans: Definition: Given \( f : \mathbb{R}^n \to \mathbb{R}^m \) and \( h \in \mathbb{R}^n \), the directional derivative of \( f \) at \( x_0 \) in the direction \( h \) is given by, for \( i = 1, \ldots, m \),

\[
f^i_h = \lim_{|k| \to \infty} \frac{f^i(x_0 + h/k) - f^i(x_0)}{|h|/k}
\]

(b) Consider the function \( f : \mathbb{R}^2 \to \mathbb{R}^2 \) defined by, for \( i = 1, 2 \), \( f^i(x, y) = x^{i/3}y^{1-i/3} \). Using the formal definition in (a) above, compute \( f^i_{h^*}(27, 8) \), where \( h^* = (54, 16) \). (Hint: \((27, 8) + (54, 16)/k = (27, 8)(1 + 2/k)\)).

\[
\begin{align*}
f^1_{h^*}(27, 8) &= \lim_{|k| \to \infty} \frac{f^1((27, 8) + (54, 16)/k) - f^1(27, 8)}{|h^*|/k} \\
&= \lim_{|k| \to \infty} \frac{((3 \times 4)(1 + 2/k)) - (3 \times 4)}{|h^*|/k} = \frac{24}{||h^*||}. \\
f^2_{h^*}(27, 8) &= \lim_{|k| \to \infty} \frac{((9 \times 2)(1 + 2/k)) - (9 \times 2)}{|h^*|/k} = \frac{36}{||h^*||}.
\end{align*}
\]

Therefore

\[
f_{h^*}(27, 8) = \begin{bmatrix} 0.4261 \\ 0.6392 \end{bmatrix}
\]

(c) Now compute \( f^i_{h^*}(27, 8) \) using the differential of \( f \) at \((27, 8)\).

\[
\begin{align*}
\triangledown f^i(x, y) &= \begin{bmatrix} if^i(x, y) \quad (3-i) f^i(x, y) \end{bmatrix} \\
f^i_{h^*}(27, 8) &= \begin{bmatrix} if^i(x, y) \quad (3-i) f^i(x, y) \end{bmatrix} \begin{bmatrix} 54/||h^*|| \\ 16/||h^*|| \end{bmatrix} \\
&= \begin{bmatrix} \frac{12}{3 \times 27} \quad \frac{2 \times 12}{3 \times 8} \\ \frac{2 \times 18}{3 \times 27} \quad \frac{18}{3 \times 8} \end{bmatrix} \begin{bmatrix} 54/||h^*|| \\ 16/||h^*|| \end{bmatrix}
\end{align*}
\]

Therefore

\[
\begin{bmatrix} f^1_{h^*}(27, 8) \\ f^2_{h^*}(27, 8) \end{bmatrix} = \begin{bmatrix} 1 \quad 8 + 16 \\ \frac{1}{||h^*||} \quad 24 + 12 \end{bmatrix} = \begin{bmatrix} 0.4261 \\ 0.6392 \end{bmatrix}
\]
(d) Check to see that all of these three distinct methods give you the same answer!

**Ans:** Amazingly, they do!
(3) Consider the function \( f(x) = x_1^\rho + x_2^\rho \), where \( \rho \in (-\infty, 1] \). The whole point here is to use the differential of \( \nabla f \) to answer the following questions, i.e., to answer all parts of the question, approximate \( \nabla f(x + h) - \nabla f(x) \) using the differential of \( \nabla f \) at \( x \), evaluated at \( h \). There are lots of other ways to answer these questions, but the purpose of this question is to give you practice in using the differential of a vector-valued function.

(a) Check that, up to a first order approximation, \(^1\) \( f \) is homothetic (cf the notes for lecture CALCULUS3\(^2\), specifically the second example in the subsection entitled Four Graphical Examples.\(^3\))

**Ans:** The following argument is not entirely rigorous, but can be made so. We need to show that if \( x \) and \( dx \) are colinear, (i.e., if \( dx = \lambda x \)), then \( \nabla f(x) \) and \( \nabla f(x + dx) \) are colinear, i.e., \( \nabla f(x + dx) = \delta \nabla f(x) \), for some \( \delta > 0 \).

Now \( \nabla f(x) = \begin{bmatrix} x_1^{\rho-1} & x_2^{\rho-1} \end{bmatrix} \) so that \( Hf(x) = \begin{bmatrix} \rho(x_1^{\rho-2}) & 0 \\ 0 & \rho(x_2^{\rho-2}) \end{bmatrix} \). Therefore, for \( dx = \lambda x \),

\[
\nabla f(x + dx) \approx \nabla f(x) + \begin{bmatrix} \rho(x_1^{\rho-2}) & 0 \\ 0 & \rho(x_2^{\rho-2}) \end{bmatrix} \begin{bmatrix} \lambda x_1 \\ \lambda x_2 \end{bmatrix} = \begin{bmatrix} \rho(x_1^{\rho-1}) + \lambda \rho(x_1^{\rho-2})x_2^{\rho-1} \\ \rho(x_2^{\rho-1}) + \lambda \rho(x_2^{\rho-2})x_1^{\rho-1} \end{bmatrix} = \rho(1 + \lambda(\rho - 1)) \begin{bmatrix} x_1^{\rho-1} \\ x_2^{\rho-1} \end{bmatrix}
\]

The argument is not entirely rigorous because of the approximation relationship above. To make it fully rigorous, we could use the 2-dimensional analog of the Taylor Lagrange theorem.

---

\(^1\) The qualifier “up to a first order approximation” means: you should pretend that the answer you get using the differential is exactly correct, even though in fact it is only approximately correct, and then only for small \( h \)’s, because there are non-zero higher order terms in the Taylor expansion of \( \nabla f \).

\(^2\) In the example in the notes, you don’t need the caveat about up to a first order approximation, because the higher order terms in the Taylor approximation are all zero. In this example they are not.

\(^3\) The lecture notes tend to change, and sometimes the problem sets don’t keep up. If this reference is no longer current, please notify Leo.
(b) When \( \rho > 0 \), does \( f \) exhibit increasing, constant or decreasing returns to scale? Is your answer true for all \( \rho \in (0, 1] \). (Again, your answer should be in terms of what happens to the gradient vector as you move out along a ray.)

\textbf{Ans:} We’ll show that \( f \) exhibits decreasing returns to scale whenever \( 0 < \rho < 1 \). When \( \rho = 0 \), the function is flat; when \( \rho < 0 \), the function decreases so that the idea of “returns to scale” doesn’t mean much. For \( \rho > 0 \), it is sufficient to show that as you increase scale, i.e., move from \( x \) to \( (1 + \lambda)x \), for any \( \lambda > 0 \), the length of the gradient vector shrinks. From the answer to the preceding part, note that for \( dx = \lambda x \), and \( \lambda > 0 \),

\[ \nabla f(x + dx) \approx (1 + \lambda(\rho - 1)) \nabla f(x). \]

For \( \rho < 1 \), \( (1 + \lambda(\rho - 1)) < 1 \), so the gradient vector indeed shrinks. When \( \rho = 1 \), \( (1 + \lambda(\rho - 1)) = 1 \), so the gradient vector remains constant, which is equivalent to constant returns to scale.

(c) Fix \( x = (\alpha, \alpha) \), and consider \( h = (-0.1, 0.1) \). Approximate \( \nabla f(x + h) \), for (i) \( \rho = 1/2 \); (ii) \( \rho = -1/2 \); (iii) \( \rho = -10 \).

\textbf{Ans:}

\[ \nabla f(x + h) \approx \rho \alpha^{\rho - 1} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} \rho(\rho - 1)\alpha^{\rho - 2} \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ \rho(\rho - 1)\alpha^{\rho - 2} \end{bmatrix} \begin{bmatrix} -0.1 \\ 0.1 \end{bmatrix} \]

\[ = \begin{cases} 2\alpha^{-1/2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 0.1 \begin{bmatrix} 2\alpha^{-1} \\ -2\alpha^{-1} \end{bmatrix} & \text{if } \rho = 1/2 \\ -2\alpha^{-3/2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 0.1 \begin{bmatrix} 2\alpha^{-1} \\ -2\alpha^{-1} \end{bmatrix} & \text{if } \rho = -1/2 \\ -10\alpha^{-11} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 0.1 \begin{bmatrix} 11\alpha^{-1} \\ -11\alpha^{-1} \end{bmatrix} & \text{if } \rho = -10 \end{cases} \]

(d) How does the curvature of the level sets of this function change as you move out along a ray through the origin. In particular, discuss the effect of the magnitude of \( \alpha \) on the rate of change in the direction of \( \nabla f \) as you add \( h = (-\beta, \beta) \) to \( x = (\alpha, \alpha) \).

\textbf{Ans:} Consider the answer to part (c): in each case, the magnitude of the rotation is determined by \( \alpha^{-1} \). When \( \alpha \) is very large, the rotation is miniscule, when \( \alpha \) is miniscule, the rotation is huge. Thus as you move out along a ray through the origin, the change in curvature generated by a given shift \( h = (-\beta, \beta) \) becomes less and less; that is, level sets become flatter and flatter as you move out along any ray.
(4) Suppose that $f : \mathbb{R} \rightarrow \mathbb{R}$ is $(n+1)$ times continuously differentiable.

(a) Show that a sufficient condition for $f$ to attain a strict (local) maximum at $x_0$ is that for some even number $n$, the derivatives $f^{(k)}(x_0)$ are zero for $k = 1...n - 1$, and $f^{(n)}(x_0)$ is negative

**Ans:**

(1) We are given that $n$ is even. Hence $(x - x_0)^n > 0$ for $x \neq x_0$

(2) Since $f^{(n)}(x_0) < 0$ and the first $n - 1$ terms in the Taylor expansion are all zero, we have that $T_n(f, x, dx) < 0$, for all $dx$.

(3) From the Taylor Young theorem, we know that if the $n'$th order Taylor expansion is nonzero for $dx$ sufficiently small, then for all $dx$ in some neighborhood $U$ of 0, the absolute value of the expansion dominates the absolute value of the remainder term.

(4) Conclude that for all $dx \in U$, $dx \neq 0$, $f(x + dx) < f(x)$.

(b) If $f^{(k)}(x_0)$ is zero for $k = 1...n - 1$ and $f^{(n)}(x_0)$ is non-zero, show that there exists an $\epsilon$-neighborhood around $x_0$ where the absolute value of the $n$th-order Taylor expansion is bigger than the absolute value of the remainder term $R_n(x)$.

**Ans:** The $n$th term in the expansion is $\frac{f^{(n)}(x_0)(x-x_0)^n}{n!}$. For some $\eta \in [x_0, x]$, the remainder term can be written as $\frac{f^{(n+1)}(\eta)(x-x_0)^{n+1}}{(n+1)!} = \frac{f^{(n+1)}(\eta)(x-x_0)^{n}(x-x_0)}{(n+1)n!}$. We need to show that for $x$ sufficiently close to $x_0$, $\left|\frac{f^{(n)}(x_0)(x-x_0)^n}{n!}\right| > \left|\frac{f^{(n+1)}(\eta)(x-x_0)^{n}(x-x_0)}{(n+1)n!}\right|$. If $f^{(n+1)}(\eta) = 0$, the inequality holds trivially. Assume therefore that $f^{(n+1)}(\eta) \neq 0$. Extracting the common term, the required inequality will hold if $\left|\frac{f^{(n)}(x_0)}{f^{(n+1)}(\eta)}\right| > \left|\frac{(x-x_0)}{(n+1)}\right|$. Since $f^{(n+1)}(\cdot)$ is continuous, and $[x, x_0]$ is a compact set, $\left|\frac{f^{(n)}(x_0)}{f^{(n+1)}(\cdot)}\right|$ attains a maximum on this set. (This is a consequence of Weierstrass theorem.) Let $\tilde{f}_{n+1}$ denote this maximum and pick $\epsilon < \frac{f^{(n)}(x_0)}{\tilde{f}_{n+1}}$. For $x$ such that $x - x_0 < \epsilon$,

$$\left|\frac{f^{(n+1)}(\eta)(x-x_0)}{(n+1)}\right| < \frac{\epsilon f_{n+1}}{(n+1)} < \left|\frac{f^{(n)}(x_0)}{(n+1)}\right|$$

as required.
(c) Give a counter example to show that the result in part (a) would be false if the words “for some even n” were replaced with “for some n > 0”.

**Ans:** Take \( f(x) = -x^3 \) and note that \( f(0) = f(2) = 0 \) while \( f(3) = -6 \); This function exhibits the property that for so for the (odd) number \( n = 3 \), the derivatives \( f^{(k)}(x_0) \) are zero for \( k = 1 \ldots n - 1 \), and \( f^{(n)}(x_0) \) is negative. In this case, all of the conditions of the theorem are satisfied except that \( n \) is odd. And \( x^3 \) isn’t maximized at zero.

(d) Explain carefully, but in as few a words as possible, why the argument in (a) works for even \( n \) but not for odd \( n \).

**Ans:** If \( n \) is odd, then we cannot conclude, as we did in step (1) of our answer to (a) above, that \( (x - x_0)^{n} > 0 \) for \( x \neq x_0 \). Therefore we cannot conclude that the sign of the \( n' \)th term in the series is determined by the sign of \( f^{(n)} \).

(e) Show that the \( n^{th} \)-order Taylor expansion around any point \( x_0 \) of a polynomial of degree \( n \) (i.e. a function of the form \( f(x) = \sum_{k=0}^{n} a_k x^k \)) is perfectly accurate, regardless of the magnitude of \( dx \).

**Ans:** Recall that that for \( g(t) = \alpha t^k \)

\[
\begin{align*}
\text{If } n \leq k : & \quad g^{(n)}(t) = t \ast (t - 1) \ast \ldots \ast (t - n) \ast \alpha t^{k - n} \\
\text{If } n > k : & \quad g^{(n)}(t) = 0
\end{align*}
\]

Hence the \( (n + 1)^{th} \) derivative of a function \( f(x) = \sum_{k=0}^{n} a_k x^k \) is equal to zero and the remainder term \( R_n(x) = f^{(n+1)}(\eta) \frac{(x-x_0)^{n+1}}{(n+1)!} \) is zero. And \( R_n(x) = 0 \) implies that the expansion is perfectly accurate.

(f) Show that if \( f \) is an arbitrary polynomial of degree 2, i.e., \( f(x) = ax^2 + bx + c \), then for any point \( x_0 \), if you add to \( f(x_0) \) the \( 2^{nd} \)-order Taylor expansion around \( f \), the expression you get is precisely the original function \( f \). More precisely, show by writing out the Taylor expansion explicitly, that for arbitrary \( dx \),

\[
f(x_0 + dx) = f(x_0) + f'(x_0)dx + 0.5f''(x_0)dx^2.
\]
Ans: Let \( f(x) = ax^2 + bx + c \), so that the second order Taylor expansion of \( f \) around \( x_0 \) is

\[
\begin{align*}
    f'(x_0) \, dx + 0.5 f''(x_0) \, dx^2 \\
    &= (2ax_0 + b) \, dx + 0.5 \times 2a \, dx^2 \\
    &= (2ax_0 + b) \, dx + a \, dx^2
\end{align*}
\]

Therefore,

\[
\begin{align*}
    f(x_0 + dx) &= ax_0^2 + bx_0 + c + (2ax_0 + b) \, dx + a \, dx^2 \\
    &= a(x_0^2 + 2x_0 \, dx + dx^2) + b(x_0 + dx) + c \\
    &= a(x_0 + dx)^2 + b(x_0 + dx) + c
\end{align*}
\]

which is precisely the original function!