MAT257 Assignment 3

(Author's name here)
October 8, 2021

1. If $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at some $a \in \mathbb{R}^n$, we will prove that f is also continuous at a. First, by definition, we have $f(a+h) - f(a) - Df(a)(h) \in o(h)$, so:

$$\lim_{h\to 0} \frac{\left|f(a+h) - f(a) - Df(a)(h)\right|}{|h|} = 0. \qquad (*)$$

Now, to prove that f is continuous at a, suppose we are given any $\epsilon > 0$. Then, by $(*)$, there exists some $\delta' > 0$ such that all $h \in \mathbb{R}^n$ which satisfy $0 < |h - 0| < \delta'$ also satisfy:

$$\left| \frac{\left|f(a+h) - f(a) - Df(a)(h)\right|}{|h|} - 0 \right| < \epsilon,$$

which can be rewritten as:

$$\left|f(a+h) - f(a) - Df(a)(h)\right| < \epsilon|h|.$$

Then, by the triangle inequality, we obtain:

$$\epsilon|h| > \left|f(a+h) - f(a) - Df(a)(h)\right| \geq \left|f(a+h) - f(a)\right| - \left|Df(a)(h)\right|,$$

which can be rewritten as:

$$\left|f(a+h) - f(a)\right| < \epsilon|h| + \left|Df(a)(h)\right|.$$

Now, since $Df(a)$ is a linear map which acts on h, Assignment 1 Question 2 gives us some $M \in \mathbb{R}$ such that $\left|Df(a)(h)\right| \leq M|h|$ for all $h \in \mathbb{R}^n$. (We can assume without loss of generality that $M > 0$; otherwise, we could pick any $M' > 0$, and it would satisfy $\left|Df(a)(h)\right| \leq M|h| \leq M'|h|$ for all $h \in \mathbb{R}^n$.) Plugging $\left|Df(a)(h)\right| \leq M|h|$ into the above inequality, we obtain:

$$\left|f(a+h) - f(a)\right| < \epsilon|h| + M|h| = (M+\epsilon)|h|. \qquad (**)$$

Note that this result is true for all $h \in \mathbb{R}^n$ which satisfy $0 < |h - 0| < \delta'$.
Now, let us define $\delta := \min(\delta', \frac{\epsilon}{M+\epsilon}) > 0$. Then, for all $x \in \mathbb{R}^n$ which satisfy $0 < |x - a| < \delta$, we can define $h := x - a$, giving us $0 < |h| < \delta$. Since $0 < |h| < \delta \leq \delta'$, the inequality $(**)$ applies to h. We also have $|h| < \delta \leq \frac{\epsilon}{M+\epsilon}$. Plugging this into $(**)$, we obtain:

$$\left|f(x) - f(a)\right| = \left|f(a+h) - f(a)\right| < (M+\epsilon)|h| < (M+\epsilon)\frac{\epsilon}{M+\epsilon} = \epsilon.$$

Therefore, for all $\epsilon > 0$, we found $\delta > 0$ such that all $x \in \mathbb{R}^n$ which satisfy $|x - a| < \delta$ also satisfy $\left|f(x) - f(a)\right| < \epsilon$, so f is continuous at a, as required. $\square$

2. (Note: This question was not marked.)
We are given some continuous function $g : S^1 \to \mathbb{R}$ such that $g(0,1) = g(1,0) = 0$ and $g(-x) = -g(x)$. We also define $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(0) = 0$ and by $f(x) = |x|\, g(\frac{x}{|x|})$ for all nonzero x.
a) We will prove that given any $x_0 \in \mathbb{R}^2$, the function $h : \mathbb{R} \to \mathbb{R}$ defined by $h(t) := f(tx_0)$ is differentiable.
First, for the special case $x_0 = 0$, we have $h(t) = f(t \cdot 0) = 0$ for all $t \in \mathbb{R}$, so by Spivak's Theorem 2-3(1), h is differentiable with $Dh(a) = 0$ for all $a \in \mathbb{R}$.
Now, suppose that x_0 is nonzero. Then, we can scale x_0 to define the point $y := \frac{x_0}{|x_0|} \in S^1$. Next, we will prove for the following three cases that $h(t) = |x_0|\, g(y)t$:
Case 1: $t > 0$. Then, we have $|tx_0| = |t||x_0| = t|x_0|$, so:

$$h(t) = f(tx_0) = |tx_0|\, g(\frac{tx_0}{|tx_0|}) = t|x_0|\, g(\frac{tx_0}{t|x_0|}) = t|x_0|\, g(\frac{x_0}{|x_0|}) = |x_0|\, g(y)t.$$

Case 2: $t = 0$. Then, we have:

$$h(t) = f(0 \cdot x_0) = f(0) = 0 = |x_0|\, g(y) \cdot 0 = |x_0|\, g(y)t.$$

Case 3: $t < 0$. Then, we have $|tx_0| = |t||x_0| = -t|x_0|$, so:

$$h(t) = f(tx_0) = |tx_0|\, g(\frac{tx_0}{|tx_0|}) = t|x_0|\, (-g(\frac{tx_0}{-t|x_0|})) = t|x_0|\, (-g(-y)) = |x_0|\, g(y)t.$$

Therefore, in all three cases, we obtain $h(t) = |x_0|\, g(y)t$. Since $|x_0|\, g(y)$ is constant, it follows that h is a linear map with matrix $(|x_0|\, g(y))$. Then, Spivak's Theorem 2-3(2) states that h is differentiable, with $Dh(t) = h$ for all $t \in \mathbb{R}$. Overall, h is differentiable no matter whether x_0 is zero or nonzero, as required. $\square$
b) We will prove that f is differentiable at $(0,0)$ if and only if $g = 0$.
First, if $g = 0$, then we have $f(0) = 0$, as well as $f(x) = |x|\, g(\frac{x}{|x|}) = 0$ for all nonzero x. As a result, f is the constant function 0. Then, by Spivak's Theorem 2-3(1), f is differentiable at $(0,0)$ with $Df(0,0) = 0$, as desired.
Next, suppose that g is not zero everywhere, so there exists some $z \in S^1$ such that $g(z) \neq 0$. Then, assume for contradiction that f is differentiable at $(0,0)$. By definition, this gives us that the error function $f((0,0)+h) - f(0,0) - Df(0,0)(h)$ is in $o(h)$, so:

$$\lim_{h \to 0} \frac{\left|f((0,0)+h) - f(0,0) - Df(0,0)(h)\right|}{|h|} = 0.$$

Applying $f(0,0) = 0$, this limit can be rewritten as:

$$\lim_{h \to 0} \frac{\left|f(h) - Df(0,0)(h)\right|}{|h|} = 0.$$

In other words, for all $\epsilon > 0$, there exists $\delta > 0$ such that all $h \in \mathbb{R}^2$ which satisfy $0 < |h - 0| < \delta$ also satisfy:

$$\frac{\left|f(h) - Df(0,0)(h)\right|}{|h|} = \left|\frac{\left|f(h) - Df(0,0)(h)\right|}{|h|} - 0\right| < \epsilon. \qquad (*)$$

Now, consider $h_1 = (\frac{\delta}{2}, 0) \in \mathbb{R}^2$. First, we have:

$$0 < |h_1 - 0| = \sqrt{\left(\frac{\delta}{2}\right)^2 + 0^2} = \frac{\delta}{2} < \delta.$$

As a result, we can plug $h = h_1$ into $(*)$, so:

$$\begin{aligned}
\epsilon &> \frac{\left|f(\frac{\delta}{2}, 0) - Df(0,0)(\frac{\delta}{2}, 0)\right|}{\left|(\frac{\delta}{2}, 0)\right|} \\
&= \frac{\left|\left|(\frac{\delta}{2}, 0)\right| g(\frac{(\frac{\delta}{2},0)}{\left|(\frac{\delta}{2},0)\right|}) - Df(0,0)(\frac{\delta}{2}, 0)\right|}{\left|(\frac{\delta}{2}, 0)\right|} \\
&= \frac{\left|\frac{\delta}{2} g(1,0) - Df(0,0)(\frac{\delta}{2}, 0)\right|}{\frac{\delta}{2}} \\
&= \frac{\left|\frac{\delta}{2} g(1,0) - \frac{\delta}{2} Df(0,0)(1, 0)\right|}{\frac{\delta}{2}} && \text{(Since } Df(0,0) \text{ is a linear map)} \\
&= \left|g(1,0) - Df(0,0)(1,0)\right| \\
&= \left|0 - Df(0,0)(1,0)\right| \\
&= \left|Df(0,0)(1,0)\right|
\end{aligned}$$

Next, consider $h_2 = (0, \frac{\delta}{2}) \in \mathbb{R}^2$. First, we have:

$$0 < |h_2 - 0| = \sqrt{0^2 + \left(\frac{\delta}{2}\right)^2} = \frac{\delta}{2} < \delta.$$

As a result, we can plug $h = h_2$ into $(*)$, so:

$$\begin{aligned}
\epsilon &> \frac{\left|f(0, \frac{\delta}{2}) - Df(0,0)(0, \frac{\delta}{2})\right|}{\left|(0, \frac{\delta}{2})\right|} \\
&= \frac{\left|\left|(0, \frac{\delta}{2})\right| g(\frac{(0,\frac{\delta}{2})}{\left|(0,\frac{\delta}{2})\right|}) - Df(0,0)(0, \frac{\delta}{2})\right|}{\left|(0, \frac{\delta}{2})\right|} \\
&= \frac{\left|\frac{\delta}{2} g(0,1) - Df(0,0)(0, \frac{\delta}{2})\right|}{\frac{\delta}{2}} \\
&= \frac{\left|\frac{\delta}{2} g(0,1) - \frac{\delta}{2} Df(0,0)(0, 1)\right|}{\frac{\delta}{2}} && \text{(Since } Df(0,0) \text{ is a linear map)} \\
&= \left|g(0,1) - Df(0,0)(0,1)\right| \\
&= \left|0 - Df(0,0)(0,1)\right| \\
&= \left|Df(0,0)(0,1)\right|
\end{aligned}$$

Finally, consider $h_3 = \frac{\delta}{2}z \in \mathbb{R}^2$, where $z \in S^1$ was chosen above such that $g(z) \neq 0$. First, we have:

$$0 < |h_3 - 0| = \left|\frac{\delta}{2}z\right| = \left|\frac{\delta}{2}\right||z| = \frac{\delta}{2} \cdot 1 < \delta.$$

As a result, we can plug $h = h_3$ into $(*)$, so:

$$\begin{aligned}
\epsilon &> \frac{\left|f(\frac{\delta}{2}z) - Df(0,0)(\frac{\delta}{2}z)\right|}{\left|\frac{\delta}{2}z\right|} \\
&= \frac{\left|\left|\frac{\delta}{2}z\right| g(\frac{\frac{\delta}{2}z}{\left|\frac{\delta}{2}z\right|}) - Df(0,0)(\frac{\delta}{2}z)\right|}{\left|\frac{\delta}{2}z\right|} \\
&= \frac{\left|\frac{\delta}{2}g(\frac{\frac{\delta}{2}z}{\frac{\delta}{2}}) - Df(0,0)(\frac{\delta}{2}z)\right|}{\frac{\delta}{2}} \\
&= \frac{\left|\frac{\delta}{2}g(z) - \frac{\delta}{2}Df(0,0)(z)\right|}{\frac{\delta}{2}} && \text{(Since } Df(0,0) \text{ is a linear map)} \\
&= \left|g(z) - Df(0,0)(z)\right|
\end{aligned}$$

Overall, we proved that $\left|Df(0,0)(1,0)\right|$ and $\left|Df(0,0)(0,1)\right|$ are both less than ϵ for all $\epsilon > 0$. This implies that $\left|Df(0,0)(1,0)\right| = \left|Df(0,0)(0,1)\right| = 0$, so $Df(0,0)(1,0) = Df(0,0)(0,1) = 0$. Then, since $(1,0)$ and $(0,1)$ form a basis for $\mathbb{R}^2$, it follows that $Df(0,0)$ is the zero map. We also proved that $\left|g(z) - Df(0,0)(z)\right| < \epsilon$ for all $\epsilon > 0$, which implies $\left|g(z) - Df(0,0)(z)\right| = 0$, so $Df(0,0)(z) = g(z)$. Since $Df(0,0)$ is the zero map, this contradicts $g(z) \neq 0$. Thus, by contradiction, f is not differentiable at $(0,0)$ if g is not zero everywhere.

Overall, we proved that f is differentiable at $(0,0)$ if and only if g is zero everywhere, as required.

□

3. Given any function $f : \mathbb{R}^n \to \mathbb{R}$ such that $\left|f(x)\right| \leq |x|^2$, we will prove that f is differentiable at 0. In fact, we will prove that the differential of f at 0 is 0.
Let us define the linear transformation $L : \mathbb{R}^n \to \mathbb{R}$ by $L(h) = 0$ for all $h \in \mathbb{R}^n$. Then, to show that $Df(0) = L$, we need to show that the error function:

$$e(h) := f(0+h) - f(0) - L(h) = f(h) - f(0) - L(h)$$

is in $o(h)$. First, since $\left|f(0)\right| \leq |0|^2 = 0$, we obtain $f(0) = 0$ by the positive definiteness of the norm. We also know that $L(h) = 0$ for all h, so:

$$e(h) = f(h) - f(0) - L(h) = f(h) - 0 - 0 = f(h)$$

for all $h \in \mathbb{R}^n$. As a result, $e(0) = f(0) = 0$.
Now, we wish to show that $\lim_{h\to 0} \frac{\left|e(h)\right|}{|h|} = 0$. Let any $\epsilon > 0$ be given. Then, let us define $\delta := \epsilon > 0$. For all $h \in \mathbb{R}^n$ such that $0 < |h - 0| < \delta$, we obtain:

$$\begin{aligned} \left|\frac{\left|e(h)\right|}{|h|} - 0\right| &= \frac{\left|e(h)\right|}{|h|} \\ &\leq \frac{|h|^2}{|h|} \\ &= |h| \\ &< \delta \\ &= \epsilon. \end{aligned}$$

Therefore, for all $\epsilon > 0$, we found $\delta > 0$ such that all $h \in \mathbb{R}^n$ which satisfy $0 < |h - 0| < \delta$ also satisfy $\left|\frac{\left|e(h)\right|}{|h|} - 0\right| < \epsilon$, so $\lim_{h\to 0} \frac{\left|e(h)\right|}{|h|} = 0$. This, combined with $e(0) = 0$, proves that $e(h)$ is in $o(h)$. Therefore, it follows that f is differentiable at 0 with $Df(0) = L = 0$, as required. $\square$

4. We will evaluate f' for the following functions:
a) $f(x,y,z) := x^y$.
First, let $f_1 : \mathbb{R}^3 \to \mathbb{R}^2$, $f_2 : \mathbb{R}^2 \to \mathbb{R}$, and $f_3 : \mathbb{R} \to \mathbb{R}$ be defined by $f_1(x,y,z) := (\log(x), y)$, $f_2(x,y) := xy$, and $f_3(x) := e^x$. Then, we can write f as the following composition:

$$f(x,y,z) = x^y = (e^{\log(x)})^y = e^{\log(x)y} = f_3(f_2(f_1(x,y,z))) = (f_3 \circ f_2 \circ f_1)(x,y,z).$$

Now, from MAT157, we know that $f_3'(a) = (e^a)$ for all $a \in \mathbb{R}$. As a result:

$$f_3'(f_2(f_1(x,y,z))) = f_3'(f_2(\log(x), y)) = f_3'(\log(x)y) = (e^{\log(x)y}) = (x^y).$$

Next, from Spivak's Theorem 2-3(5), we know that $f_2'(a,b) = (b,a)$ for all $(a,b) \in \mathbb{R}^2$, so:

$$f_2'(f_1(x,y,z)) = f_2'(\log(x), y) = (y, \log(x)).$$

Finally, we wish to evaluate $f_1'(x,y,z)$. First, we can decompose f_1 further into $f_1 = (g_1, g_2)$, where the functions $g_1, g_2 : \mathbb{R}^3 \to \mathbb{R}$ are defined by $g_1(x,y,z) := \log(x)$ and $g_2(x,y,z) := y$. To evaluate $g_1'(x,y,z)$, we write g_1 as the composition $h_2 \circ h_1$, where $h_1 : \mathbb{R}^3 \to \mathbb{R}$ is defined by $h_1(x,y,z) := x$, and $h_2 : \mathbb{R} \to \mathbb{R}$ is defined by $h_2(x) := \log(x)$. Finally, we can find these derivatives. Since h_1 is the linear transformation with matrix $(1,0,0)$, Spivak's Theorem 2-3(2) tells us that $h_1'(x,y,z) = (1,0,0)$, and MAT157 also tells us that $h_2'(x) = \frac{1}{x}$. Thus, the Chain Rule gives us:

$$g_1'(x,y,z) = h_2'(h_1(x,y,z)) \cdot h_1'(x,y,z) = h_2'(x) \cdot (1,0,0) = (\frac{1}{x}) \cdot (1,0,0) = (\frac{1}{x}, 0, 0).$$

Meanwhile, since g_2 is a linear transformation with the matrix $(0,1,0)$, Spivak's Theorem 2-3(2) tells us that $g_2(x,y,z) = (0,1,0)$. Overall, applying Spivak's Theorem 2-3(3), we obtain:

$$f_1'(x,y,z) = \begin{pmatrix} g_1'(x,y,z) \\ g_2'(x,y,z) \end{pmatrix} = \begin{pmatrix} \frac{1}{x} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Finally, applying the Chain Rule twice on $f = f_3 \circ f_2 \circ f_1$, we obtain:

$$\begin{aligned} f'(x,y,z) &= (f_3 \circ f_2)'(f_1(x,y,z)) \cdot f_1'(x,y,z) \\ &= f_3'(f_2(f_1(x,y,z))) \cdot f_2'(f_1(x,y,z)) \cdot f_1'(x,y,z) \\ &= (x^y) \cdot (y, \log(x)) \cdot \begin{pmatrix} \frac{1}{x} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \\ &= (x^y) \cdot (\frac{y}{x}, \log(x), 0) \\ &= \boxed{(x^{y-1}y, x^y \log(x), 0)} \end{aligned}$$

□

b) $f(x,y,z) := (x^y, z)$.
First, we decompose f into $f = (f_1, f_2)$, where $f_1 : \mathbb{R}^3 \to \mathbb{R}$ is defined by $f_1(x,y,z) := x^y$ and $f_2 : \mathbb{R}^3 \to \mathbb{R}$ is defined by $f_2(x,y,z) := z$. Then, from part a), $f_1'(x,y,z) = (x^{y-1}y, x^y \log(x), 0)$. Moreover, f_2 is a linear transformation with the matrix $(0,0,1)$, so we know from Spivak's Theorem 2-3(2) that $f_2'(x,y,z) = (0,0,1)$. Therefore, applying Spivak's Theorem 2-3(3), we obtain:

$$f'(x,y,z) = \begin{pmatrix} f_1'(x,y,z) \\ f_2'(x,y,z) \end{pmatrix} = \boxed{\begin{pmatrix} x^{y-1}y & x^y \log(x) & 0 \\ 0 & 0 & 1 \end{pmatrix}}.$$

$\square$

c) $f(x, y, z) := (x + y)^z$.
First, let us define the functions $f_1 : \mathbb{R}^3 \to \mathbb{R}^3$ and $f_2 : \mathbb{R}^3 \to \mathbb{R}$ by $f_1(x, y, z) := (x + y, z, 0)$ and $f_2(x, y, z) := x^y$. Then, we can write $f(x, y, z)$ as the following composition:

$$f(x, y, z) = (x + y)^z = f_2(x + y, z, 0) = (f_2 \circ f_1)(x, y, z).$$

Now, from part a), we know that $f_2'(x, y, z) = (x^{y-1}y, x^y \log(x), 0)$, so we obtain:

$$f_2'(f_1(x, y, z)) = f_2'(x + y, z, 0) = ((x + y)^{z-1}z, (x + y)^z \log(x + y), 0).$$

Additionally, since f_1 is a linear transformation with the matrix $\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$, Spivak's Theorem 2-3(2) tells us that $f_1'(x, y, z) = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$. Therefore, applying the Chain Rule on $f = f_2 \circ f_1$, we obtain:

$$\begin{aligned} f'(x, y, z) &= f_2'(f_1(x, y, z)) \cdot f_1'(x, y, z) \\ &= ((x + y)^{z-1}z, (x + y)^z \log(x + y), 0) \cdot \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \\ &= \boxed{((x + y)^{z-1}z, (x + y)^{z-1}z, (x + y)^z \log(x + y))}. \end{aligned}$$

$\square$

5. a) We will prove that if $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ is bilinear, then:

$$\lim_{(h,k)\to 0} \frac{\left|f(h,k)\right|}{\left|(h,k)\right|} = 0.$$

First, for all $1 \le i \le n$, let e_i be the point in $\mathbb{R}^n$ with its i^{th} coordinate equal to 1 and all other coordinates equal to 0. Then, let $g_i : \mathbb{R}^m \to \mathbb{R}^p$ be the linear map defined by $g_i(k) := f(e_i, k)$ for all $k \in \mathbb{R}^m$ – this map is linear because e_i is constant. Now, in Assignment 1 Question 2, we proved that there exists some $M_i \in \mathbb{R}$ such that $\left|g_i(k)\right| \le M_i|k|$ for all $k \in \mathbb{R}^m$. We can assume without loss of generality that $M_i > 0$; otherwise, we could pick any $M_i' > 0$, and it would satisfy $\left|g_i(k)\right| \le M_i|k| \le M_i'|k|$.
Next, for all $h = (h_1, \dots, h_n) \in \mathbb{R}^n$, we have $h = h_1e_1 + \cdots + h_ne_n$. Then, since f is bilinear, we obtain:

$$\begin{aligned}
\left|f(h,k)\right| &= \left|f(h_1e_1 + \cdots + h_ne_n, k)\right| \\
&= \left|h_1 f(e_1,k) + \cdots + h_n f(e_n,k)\right| \\
&\le \left|h_1 f(e_1,k)\right| + \cdots + \left|h_n f(e_n,k)\right| && \text{(Applying triangle inequality)} \\
&= |h_1|\left|f(e_1,k)\right| + \cdots + |h_n|\left|f(e_n,k)\right| \\
&= |h_1|\left|g_1(k)\right| + \cdots + |h_n|\left|g_n(k)\right| \\
&\le |h_1|\, M_1|k| + \cdots + |h_n|\, M_n|k| \\
&= (M_1|h_1| + \cdots + M_n|h_n|)|k|
\end{aligned}$$

Now, let us define $M := \max(M_1, \dots, M_n) > 0$. Then, we obtain:

$$\left|f(h,k)\right| \le (M_1|h_1| + \cdots + M_n|h_n|)|k| \le M(|h_1| + \cdots + |h_n|)|k|\,.$$

Next, let us define the point $z = (z_1, \dots, z_n) \in \mathbb{R}^n$ by $z_i = -1$ if h_i is negative, and $z_i = 1$ otherwise. In other words, $z_ih_i = |h_i|$. Then, by the Cauchy-Schwarz inequality, we obtain:

$$\begin{aligned}
\langle h, z\rangle &\le |h| \cdot |z| \\
h_1z_1 + \cdots + h_nz_n &\le |h| \cdot \sqrt{z_1^2 + \cdots + z_n^2} \\
|h_1| + \cdots + |h_n| &\le |h| \cdot \sqrt{(\pm 1)^2 + \cdots + (\pm 1)^2} \\
|h_1| + \cdots + |h_n| &\le \sqrt{n}|h|
\end{aligned}$$

As a result, we obtain:

$$\left|f(h,k)\right| \le M(|h_1| + \cdots + |h_n|)|k| \le M\sqrt{n}|h||k|\,.$$

Now, since $(|h| - |k|)^2 \ge 0$, we have $|h|^2 - 2|h||k| + |k|^2 \ge 0$, so:

$$|h||k| \le \frac{1}{2}(|h|^2 + |k|^2) = \frac{1}{2}\left|(h,k)\right|^2.$$

Finally, we obtain:

$$\frac{\left|f(h,k)\right|}{\left|(h,k)\right|} \le \frac{M\sqrt{n}|h||k|}{\left|(h,k)\right|} \le \frac{M\sqrt{n}}{2}\left|(h,k)\right|$$

for all nonzero (h,k).
Now, given any $\epsilon > 0$, let us define $\delta := \frac{2\epsilon}{M\sqrt{n}} > 0$. Then, for all $(h,k) \in \mathbb{R}^n \times \mathbb{R}^m$ satisfying $0 < \left|(h,k) - 0\right| < \delta$, we obtain:

$$\left|\frac{\left|f(h,k)\right|}{\left|(h,k)\right|} - 0\right| = \frac{\left|f(h,k)\right|}{\left|(h,k)\right|} \le \frac{M\sqrt{n}}{2}\left|(h,k)\right| < \frac{M\sqrt{n}}{2}\delta = \epsilon.$$

Therefore, for all $\epsilon > 0$, we found $\delta > 0$ such that all (h,k) which satisfy $0 < \left|(h,k) - 0\right| < \delta$ also satisfy $\left|\frac{\left|f(h,k)\right|}{\left|(h,k)\right|} - 0\right| < \epsilon$, so:

$$\lim_{(h,k)\to 0} \frac{\left|f(h,k)\right|}{\left|(h,k)\right|} = 0,$$

as required. □

b) We will prove that $Df(a,b)(x,y) = f(a,y) + f(x,b)$ whenever f is bilinear. In particular, for all points $(a,b) \in \mathbb{R}^n \times \mathbb{R}^m$, let us define the linear map $L(a,b) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ by $L(a,b)(x,y) := f(a,y) + f(x,b)$. (This map is linear if we treat a and b as constants because f is bilinear.) Then, by definition of differentiation, we need to show that:

$$e(h) := f((a,b) + h) - f(a,b) - L(a,b)(h) \in o(h).$$

First, if we decompose h as $h = (h_1, h_2)$, where $h_1 \in \mathbb{R}^n$ and $h_2 \in \mathbb{R}^m$, then we can use the fact that f is bilinear multiple times to obtain:

$$\begin{aligned}
e(h_1,h_2) &= f(a+h_1, b+h_2) - f(a,b) - L(a,b)(h_1,h_2) \\
&= f(a+h_1, b+h_2) - f(a,b) - f(a,h_2) - f(h_1,b) \qquad \text{(Applying definition of } L(a,b)\text{)} \\
&= (f(a,b+h_2) + f(h_1,b+h_2)) - f(a,b) - f(a,h_2) - f(h_1,b) \\
&= f(a,b+h_2) + f(h_1,b+h_2) - f(a,b+h_2) - f(h_1,b) \\
&= f(h_1,b+h_2) - f(h_1,b) \\
&= f(h_1,(b+h_2) - b) \\
&= f(h_1,h_2)
\end{aligned}$$

Now, if we plug in $h = (h_1,h_2) = 0$, since f is bilinear, we obtain:

$$e(0) = f(0,0) = f(0 \cdot 1, 0) = 0 \cdot f(1,0) = 0.$$

Moreover, since f is bilinear, part a) gives us that:

$$\lim_{(h_1,h_2)\to 0} \frac{\left|e(h_1,h_2)\right|}{\left|(h_1,h_2)\right|} = \lim_{(h_1,h_2)\to 0} \frac{\left|f(h_1,h_2)\right|}{\left|(h_1,h_2)\right|} = 0.$$

These two pieces of information prove that $e(h) \in o(h)$. Therefore, we obtain:

$$Df(a,b)(x,y) = L(a,b)(x,y) = f(a,y) + f(x,b),$$

as required. □

c) We will prove that the formula $Dp(a,b)(x,y) = bx + ay$, where $p : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is defined by $p(x,y) = xy$, is a special case of part b).
First, we will show that p is bilinear. If y is kept constant, then $p(x,y)$ is a linear transformation in

terms of x with matrix (y). Additionally, if x is kept constant, then $p(x, y)$ is a linear transformation in terms of y with matrix (x). Since p is linear in terms of each variable when the other variable is kept constant, p is bilinear, as desired.
Now, when we apply the formula in part b), we obtain:

$$Dp(a, b)(x, y) = p(a, y) + p(x, b) = ay + xb = bx + ay.$$

Therefore, the formula $Dp(a, b)(x, y) = bx + ay$ is a special case of part b), as required. $\square$

6. (Note: This question was not marked.)

We will prove that the function $f : \mathbb{R}^4 \to \mathbb{R}$ defined by $f(a,b,c,d) := \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$ is differentiable, and we will also compute its differential.

First, let us define the functions $f_1 : \mathbb{R}^4 \to \mathbb{R}^2$ and $f_2 : \mathbb{R}^2 \to \mathbb{R}$ by $f_1(a,b,c,d) := (ad, bc)$ and $f_2(a,b) := a-b$. Then, we can write f as the composition $f = f_2 \circ f_1$. Now, let us decompose f_1 further into $f_1 = (g_1, g_2)$, where $g_1 : \mathbb{R}^4 \to \mathbb{R}$ is defined by $g_1(a,b,c,d) := ad$ and $g_2 : \mathbb{R}^4 \to \mathbb{R}$ is defined by $g_2(a,b,c,d) := bc$. Then, g_1 can be written as a composition $g_1 = h_2 \circ h_1$, where $h_1 : \mathbb{R}^4 \to \mathbb{R}^2$ is defined by $h_1(a,b,c,d) = (a,d)$ and $h_2 : \mathbb{R}^2 \to \mathbb{R}$ is defined by $h_2(a,b) = ab$. Since h_1 is a linear transformation with matrix $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$, Spivak's Theorem 2-3(2) gives us that h_1 is differentiable and that $h_1'(a,b,c,d) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$. Additionally, Spivak's Theorem 2-3(5) states that h_2 is also differentiable and that $h_2'(a,b) = (b,a)$. Thus, applying the Chain Rule, $g_1 = h_2 \circ h_1$ is differentiable, and:

$$\begin{aligned} g_1'(a,b,c,d) &= h_2'(h_1(a,b,c,d)) \cdot h_1'(a,b,c,d) \\ &= h_2'(a,d) \cdot h_1'(a,b,c,d) \\ &= (d,a) \cdot \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\ &= (d,0,0,a). \end{aligned}$$

Similarly, g_2 can also be written as a composition $g_2 = h_2 \circ h_3$, where $h_3 : \mathbb{R}^4 \to \mathbb{R}^2$ is defined by $h_3(a,b,c,d) = (b,c)$. Since h_3 is a linear transformation with matrix $\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$, Spivak's Theorem 2-3(2) gives us that h_3 is differentiable and that $h_3'(a,b,c,d) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$. Thus, applying the Chain Rule, $g_2 = h_2 \circ h_3$ is differentiable, and:

$$\begin{aligned} g_2'(a,b,c,d) &= h_2'(h_3(a,b,c,d)) \cdot h_3'(a,b,c,d) \\ &= h_2'(b,c) \cdot h_3'(a,b,c,d) \\ &= (c,b) \cdot \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \\ &= (0,c,b,0). \end{aligned}$$

Now, by Spivak's Theorem 2-3(3), the function $f_1 = (g_1, g_2)$ is differentiable, and:

$$f_1'(a,b,c,d) = \begin{pmatrix} g_1'(a,b,c,d) \\ g_2'(a,b,c,d) \end{pmatrix} = \begin{pmatrix} d & 0 & 0 & a \\ 0 & c & b & 0 \end{pmatrix}.$$

Next, since the function $f_2(a,b) = a - b$ is a linear transformation with matrix $(1,-1)$, Spivak's Theorem 2-3(2) gives us that f_2 is differentiable and that $f_2'(a,b) = (1,-1)$. Finally, applying the Chain Rule, the function $f = f_2 \circ f_1$ is differentiable, and we obtain:

$$\begin{aligned} f'(a,b,c,d) &= f_2'(f_1(a,b,c,d)) \cdot f_1'(a,b,c,d) \\ &= (1,-1) \cdot \begin{pmatrix} d & 0 & 0 & a \\ 0 & c & b & 0 \end{pmatrix} \\ &= \boxed{(d,-c,-b,a)}. \end{aligned}$$

□

Notes on Intuition

Now, let's develop some intuition on how to approach these problems and find solutions for them. (Note: This section was not submitted on Crowdmark.)

1. We wish to prove that f is continuous at a. Since f is diffable at a, we know that $f(a+h)-f(a)$ equals a linear correction $L(h)$, plus a "tiny" error function $e(h)$. As a result, we can model our solution after A2 Q4 (where we proved that linear transformations are continuous). Due to the "tiny" error function, our solution also required some modifications where we place a linear bound (such as $\epsilon|h|$) on the "tiny" error function.

2. a) First, we will provide an intuitive explanation of the formula $f(x) := |x|\, g(\frac{x}{|x|})$. The expression $\frac{x}{|x|}$ rescales x to project it onto the unit circle S^1 so that we can input it into g. Then, we multiply the output by $|x|$ to scale it according to the size of $|x|$. Now, if we restrict f to any line through the origin, then the projection of this line onto S^1 consists of two diametrically opposite points, y and $-y$. This gives us lots of consistency as we input these projections into g. Then, $f(tx)$ is proportional to $\pm|tx|$, which is itself proportional to $\pm|t|$. Overall, these observations motivate us to show that h is linear, from which it follows that h is differentiable.
   b) When we consider all points h arbitrarily close to $(0,0)$, based on part a), we can learn about $f(h)$ by inputting the projection of h into g. These points h can get projected onto any point on S^1. Based on the textbook's hint, if we choose h which get projected horizontally or vertically onto S^1, then g will output 0, so f will also output 0. Due to linear algebra, this constrains $Df(0,0)$ to be zero. However, if we choose h which get projected onto a point $z \in S^1$ such that $g(z) \neq 0$, then f becomes a nonzero linear map when restricted along the direction of z. This contradicts $Df(0,0) = 0$, which allows us to complete our proof by contradiction.

3. The main idea for this problem is that if $\left|f(x)\right|$ is bounded by $|x|^2$, where $|x|^2$ becomes "tiny" as x approaches 0, then $f(x)$ itself should be tiny. Then, if we want to approximate $f(x)$ as a linear map where x is near 0, we should choose the zero map. Another way to see this is that $|x|^2$ has a parabolic shape near $x=0$, so $|x|^2$ is flat at $x=0$, and then $f(x)$ itself is flat at $x=0$ since it is bounded by $|x|^2$. This motivates us to choose the linear map L defined by $L(h)=0$. After considering the error function $e(h)$ and performing some computations to determine that $\frac{\left|e(h)\right|}{|h|} \leq |h|$, this motivates us to select $\delta = \epsilon$ in our delta-epsilon proof so that $|h| < \delta = \epsilon$.

4. a) Recall that, in lecture, we computed the differential of the function $g(x,y) = \frac{x}{y}$ by writing it in the exponential form $g(x,y) = e^{\log(x)-\log(y)}$. Our solution for $f(x,y,z) = x^y$ is heavily motivated by this proof technique. After writing f in the exponential form $f(x,y,z) = e^{\log(x)y}$, we finish the problem using the Chain Rule, very similarly to the computation of g' done in class.
   b) We immediately notice that the output of the function $f(x,y,z) = (x^y, z)$ contains x^y, whose differential we found in part a), and z, whose differential is relatively easy to find. This motivates us to decompose the output coordinate-wise, then combine the differentials of the coordinates x^y and z using Spivak's Theorem 2-3(3).
   c) We wish to use part a) to solve part c) directly, except the base and exponent of $(x+y)^z$ are slightly different. We account for this modification with a linear transformation of the coordinates. Then, we can finish this problem with the Chain Rule.

5. a) Since $f(h,k)$ is linear in terms of h, we expect $f(h,k)$ to be bounded proportionally to $|h|$ because of A1 Q2. By the same reasoning, $f(h,k)$ should also be bounded proportionally to k.

Overall, we wish to show that $f(h,k)$ is bounded proportionally to $|h||k|$. This product is of degree 2, so it becomes "tiny" as (h,k) approaches 0, which will help us show that f is "tiny".
Now, let us figure out how to prove that $f(h,k)$ is bounded proportionally to $|h||k|$. By defining the maps $g_i(k) := f(e_i, k)$, we use one of the components of f's bilinearity (i.e., the fact that f is linear in terms of k) to bound f proportionally to $|k|$. Next, we need to bound f proportionally to $|h|$ as well. To do this, we model our solution after A1 Q2 – note how the key Cauchy-Schwarz application reappears. As discussed above, once $f(h,k)$ is bounded proportionally to $|h||k|$, we are done.
b) The problem already gives us the candidate $Df(a,b)(x,y) = f(a,y) + f(x,b)$ for the differential of f, so we only need to verify that it satisfies the definition for differentials. When we work to verify this, the bilinearity of f allows us to make useful simplifications. Finally, when we compute that the error function is f itself, part a) (along with some technical details) confirms for us that f is "tiny", as desired.
c) This sub-problem was relatively straightforward with a couple of formal steps: Verify that $p(x,y) = xy$ is bilinear (so that we can apply part b)), then perform some computations after applying part b).

6. We observe that the determinant, $ad - bc$, is ultimately a difference of products. Then, this problem becomes relatively straightforward because we know how to differentiate products (e.g., by Spivak's Theorem 2-3(5)), and we also know how to differentiate differences.