1. **Mean Value Theorems:**

(a) (10pts) Use the Mean Value Theorem from single variable calculus to prove the Mean Value Theorem below for scalar functions of several variables. A **convex set** \( A \subset \mathbb{R}^n \) is a set such that, for any two points in \( A \), the line segment between them is also in \( A \):

\[
x, y \in A \implies \{(1 - t)x + ty \mid 0 \leq t \leq 1\} \subset A
\]

**Theorem** (MVT). If \( f : A \subset \mathbb{R}^n \to \mathbb{R} \) is differentiable on a convex set \( A \), for any pair of points \( x, y \in A \), we have

\[
f(x) - f(y) = \nabla f(z) \cdot (x - y) = Df(z)(x - y)
\]

for some \( z \in \{(1 - t)x + ty \mid 0 \leq t \leq 1\} \subset A \).

**Proof.** Let \( c : \mathbb{R} \to \mathbb{R}^n \) be the curve \( t \mapsto (1 - t)x + ty = x + t(y - x) \), so the composition \( f \circ c : \mathbb{R} \to \mathbb{R} \) is a single variable function given by

\[
t \mapsto f((1 - t)x + ty).
\]

By the Mean Value Theorem in single variable calculus,

\[
(f \circ c)(0) - (f \circ c)(1) = \frac{d(f \circ c)}{dt}(t')(0 - 1), \quad \text{for some } t' \in [0, 1].
\]

However, evaluating this expression and using the chain rule, we have

\[
f(x) - f(y) = Df((c(t')))|_{t'(0-1)} = -Df(z)c'(t') = -Df(z)(y - x) = Df(z)(x - y)
\]

where \( z = (1 - t')x + t'y = c(t') \).

(Sorry about the minus signs. It would have been clearer to write the line between \( x \) and \( y \) as \( (tx + (1 - t)y) \mid 0 \leq t \leq 1 \)).

(b) (10pts) Why is the analogous statement (using the \( Df \) form of the right hand side) false in general for a vector valued function \( f : \mathbb{R}^n \to \mathbb{R}^m \) when \( m > 1 \)?

Since \( f(x) = (f_1(x), \ldots, f_m(x)) \), where the \( f_i(x) \) are scalar functions, we know from part (a) that there exist points \( z_1, \ldots, z_m \) on the line segment between \( x \) and \( y \) such that

\[
f_i(x) - f_i(y) = Df_i(z_i)(x - y),
\]

and certainly we have that

\[
f(x) - f(y) = \begin{bmatrix} f_1(x) - f_1(y) \\ \vdots \\ f_m(x) - f_m(y) \end{bmatrix} = \begin{bmatrix} Df_1(z_1) \\ \vdots \\ Df_m(z_m) \end{bmatrix} (x - y)
\]

where the \( Df_i(z_i) \) (which are \( 1 \times n \) matrices) provide the rows for a \( m \times n \) matrix. However, we cannot guarantee that \( z_1 = z_2 = \cdots = z_m \) in general, so this matrix is not generally equal to \( Df(z) \) for any \( z \).
2. (5pts) What is the derivative of a constant function \( f : \mathbb{R}^n \to \mathbb{R}^m \), (so \((x_1, \ldots , x_n) \mapsto (f_1, \ldots , f_m)\)), where each \( f_i \in \mathbb{R} \) is independent of \( x \)? Prove your answer using the definition of the derivative; i.e. that the derivative is the unique linear function \( T : \mathbb{R}^n \to \mathbb{R}^m \) such that

\[
\lim_{x \to x_0} \frac{\|f(x) - f(x_0) - T(x - x_0)\|}{\|x - x_0\|} = 0
\]

The derivative of such a function is identically 0 for all \( x \); that is,

\[ Df(x) = 0 \]

where \( 0 : \mathbb{R}^n \to \mathbb{R}^m \) is the linear function \( x \mapsto (0, \ldots , 0) \) for all \( x \), which is represented by the zero matrix

\[
0 = \begin{bmatrix}
0 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & 0
\end{bmatrix}
\]

Indeed, we have

\[
\lim_{x \to x_0} \frac{\|f(x) - f(x_0) - 0(x - x_0)\|}{\|x - x_0\|} = \lim_{x \to x_0} \frac{\|0 - 0(x - x_0)\|}{\|x - x_0\|} = \lim_{x \to x_0} \frac{\|0 - 0\|}{\|x - x_0\|} = 0.
\]

3. (5pts) What is the derivative of a linear function \( f : \mathbb{R}^n \to \mathbb{R}^m \), (so \( f(ax + by) = a f(x) + b f(y) \) for all \( x, y \in \mathbb{R}^n \), \( a, b \in \mathbb{R} \))? Prove your answer using the definition of the derivative.

The derivative of \( f \) at \( x \) is supposed to be the “best linear approximation to \( f \)” at \( x \). Since \( f \) is already linear, it is best approximated by \( f \) itself. Thus, \( Df(x) = f \) for all \( x \). Indeed,

\[
\lim_{x \to x_0} \frac{\|f(x) - f(x_0) - f(x - x_0)\|}{\|x - x_0\|} = \lim_{x \to x_0} \|f(x - x_0) - f(x - x_0)\| = 0
\]

using the property of linearity that \( f(x) - f(x_0) = f(x - x_0) \).

4. (10pts) A function \( f : A \subset \mathbb{R}^n \to B \subset \mathbb{R}^n \) is said to be **invertible** if there exists a function (called the inverse) \( f^{-1} : B \subset \mathbb{R}^n \to A \subset \mathbb{R}^n \) such that

\[
 f^{-1} \circ f = \text{Id} : A \to A \quad \text{and} \quad f \circ f^{-1} = \text{Id} : B \to B,
\]

where \( \text{Id} \) is the **identity function** which maps each point to itself,

\[
\text{Id} : \mathbb{R}^n \to \mathbb{R}^n, \quad x \mapsto x.
\]

(Note that the dimension of the domain and range must be the same, and note that \( f^{-1}(x) \) does not mean \( 1/f(x) \) unless \( n = 1 \), since division does not make sense for \( n > 1 \).)

Show that if \( f : A \subset \mathbb{R}^n \to B \subset \mathbb{R}^n \) is invertible, and if \( f \) is differentiable at \( x_0 \in A \), then \( f^{-1} \) is differentiable at \( y_0 = f(x_0) \) with derivative

\[
Df^{-1}(y_0) = (Df(x_0))^{-1}.
\]

(Hint: use the chain rule and your result from problem 3).

Since \((f^{-1} \circ f) = \text{Id} \) is differentiable at \( x_0 \), it follows that \( f^{-1} \) is differentiable at \( y_0 \). (It’s OK if you did not make this argument, but rather assumed that \( f^{-1} \) was differentiable at \( y_0 \)). What’s left is to show that \( Df^{-1}(y_0) \) is an inverse to \( Df(x_0) \), i.e. that

\[
ADf(x_0) = \text{Id}, \quad Df(x_0)A = \text{Id},
\]
where $A = \mathbf{D}(f^{-1})(y_0)$.

But using the fact that $\text{Id}$ is linear, so $\mathbf{D}(\text{Id})(x) = \text{Id}$, and the chain rule, we have

$$\text{Id} = \mathbf{D}(\text{Id})(x_0) = \mathbf{D}(f^{-1} \circ f)(x_0) = \mathbf{D}(f^{-1})(y_0) \mathbf{D} f(x_0)$$

and

$$\text{Id} = \mathbf{D}(\text{Id})(y_0) = \mathbf{D}(f \circ f^{-1})(y_0) = \mathbf{D} f(x_0) \mathbf{D}(f^{-1})(y_0),$$

so by definition of invertibility, we have

$$\mathbf{D}(f^{-1})(y_0) = (\mathbf{D} f(x_0))^{-1}.$$  

5. (10pts) In class (and in Ch. 2.6 in the book), we discussed the product rule

$$\mathbf{D}(fg)(x_0) = g(x_0) \mathbf{D} f(x_0) + f(x_0) \mathbf{D} g(x_0)$$

for scalar functions $f, g : \mathbb{R}^n \to \mathbb{R}$, where the product $fg : \mathbb{R}^n \to \mathbb{R}$, $x \mapsto f(x)g(x)$ is ordinary multiplication in $\mathbb{R}$.

For vector valued functions $f, g : \mathbb{R}^n \to \mathbb{R}^m$, we can form the dot product function (note that it is scalar valued!)

$$f \cdot g : \mathbb{R}^n \to \mathbb{R}, \quad x \mapsto f(x) \cdot g(x) = \sum_{i=1}^m f_i(x)g_i(x)$$

where the product is the dot product in $\mathbb{R}^m$. Formulate and prove a product rule for the dot product. You may use the scalar version of the product rule in your proof, and any other properties of derivatives discussed in class. (Hint: in formulating your result in a concise manner, you may be interested in the transpose operation on matrices $A \mapsto A^T$. If $A$ is an $r \times s$ matrix, $A^T$ is a $s \times r$ matrix with $(A^T)_{ij} = A_{ji}$, that is, the rows and columns are swapped with one another.)

Since $f \cdot g$ is a function from $\mathbb{R}^n$ to $\mathbb{R}$, we expect $\mathbf{D}(f \cdot g)(x)$ to be a linear map from $\mathbb{R}^n$ to $\mathbb{R}$, that is, a $1 \times n$ matrix. Using the sum and product rules, we have

$$\mathbf{D}(f \cdot g)(x) = \mathbf{D} \left( \sum_{i=1}^m f_i(x)g_i(x) \right) = \sum_{i=1}^m (f_i(x) \mathbf{D} g_i(x) + g_i(x) \mathbf{D} f_i(x)).$$

Each term in the sum is a scalar, such as $f_i(x)$, times a $1 \times n$ matrix, such as $\mathbf{D} g_i(x)$, so the result is indeed a $1 \times n$ matrix. We can write this more economically by noting that

$$[\mathbf{D} g_i(x)]_{1j} = [\mathbf{D} g(x)]_{ij}$$

that is, the entry in the $j$th column of the $1 \times n$ matrix $\mathbf{D} g_i(x)$ (which is $\frac{\partial g_i(x)}{\partial x_j}$), is equal to the entry in the $i$th row, $j$th column entry of the $m \times n$ matrix $\mathbf{D} g(x)$.

Thus, using the formula $[AB]_{ik} = \sum_j A_{ij}B_{jk}$ for matrix multiplication, we have

$$[\mathbf{D}(f \cdot g)(x)]_{1j} = \sum_{i=1}^m \left( f_i(x) [\mathbf{D} g_i(x)]_{ij} + g_i(x) [\mathbf{D} f_i(x)]_{ij} \right) = [f(x)^T \mathbf{D} g(x) + g(x)^T \mathbf{D} f(x)]_{1j},$$

for the entry in each column, and so

$$\mathbf{D}(f \cdot g)(x) = f(x)^T \mathbf{D} g(x) + g(x)^T \mathbf{D} f(x)$$

Note that $f^T(x)$ is a $1 \times m$ matrix and $\mathbf{D} g(x)$ is an $m \times n$ matrix, (similarly for $g^T(x)$ and $\mathbf{D} f(x)$) so the result is a $1 \times n$ matrix, as expected.

This can equivalently be written as

$$\mathbf{D}(f \cdot g)(x) = (\mathbf{D} g(x))^T f(x) + (\mathbf{D} f(x))^T g(x).$$