Derivatives of Functions from $\mathbb{R}^n$ to $\mathbb{R}^m$

Marc H. Mehlman
Department of Mathematics
University of New Haven

20 October 2014

1 Matrices

Definition 1.1 A $m \times n$ matrix, $A$, is an array of numbers having $m$ rows and $n$ columns. The element of $A$ in the $i^{th}$ row and $j^{th}$ column is denoted by $A_{ij}$.

Scalar multiplication of a matrix is accomplished componentwise, i.e. $(kA)_{ij} = kA_{ij}$. Matrix addition (or subtraction) of two matrices of the same dimensions is also defined componentwise, i.e. $(A + B)_{ij} = A_{ij} + B_{ij}$. Matrix addition is not defined between matrices of different dimensions.

Matrix multiplication between an $m \times n$ matrix, $A$, and an $n \times p$ matrix, $B$, produces an $m \times p$ matrix, $AB$. The element of $AB$ in the $i^{th}$ row and $j^{th}$ column is obtained by taking the dot product of the $i^{th}$ row of $A$ and the $j^{th}$ column of $B$, i.e.,

$$[AB]_{ij} = \sum_{k=1}^{n} A_{ik}B_{kj} = (i^{th} \text{ row of } A) \cdot (j^{th} \text{ column of } B).$$

Example 1.2

$$\begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ -1 & 1 & 2 \\ 2 & 1 & 3 \end{bmatrix} = \begin{bmatrix} 5 & 7 & 14 \\ 3 & 0 & 5 \end{bmatrix}$$
While matrix multiplication need not be commutative, it is associative and it does obey the distributive laws.

**Convention 1.3** In the following discussions, vectors $\vec{x} \in \mathbb{R}^n$ will be considered as $n \times 1$ matrices, i.e., “column vectors”.

**Definition 1.4** A function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a linear transformation (sometimes referred to as just linear) iff for every $\vec{x}, \vec{y} \in \mathbb{R}^n$ and $c \in \mathbb{R}$,

1. $f(\vec{x} + \vec{y}) = f(\vec{x}) + f(\vec{y})$
2. $f(c\vec{x}) = cf(\vec{x})$

The following proposition is stated without proof, though the proof is not hard.

**Proposition 1.5** Let $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be a linear transformation. Then

1. $f(\cdot)$ is continuous and there exists $N > 0$ (depending on $f$) such that $\|f(\vec{x})\| \leq N\|\vec{x}\|$ for all $\vec{x} \in \mathbb{R}^n$.
2. Matrix multiplication is linear, i.e., if $A$ is an $m \times n$ matrix then $g(\vec{x}) \overset{\text{def}}{=} A\vec{x}$ is a linear transformation.
3. There exists an $m \times n$ matrix $A$ such that $f(\vec{x}) = A\vec{x}$. 

2 Derivatives

Assume \( m = n = 1 \), that is \( f : \mathbb{R} \to \mathbb{R} \) as in Calc I. Then

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

\[
\Leftrightarrow \lim_{x \to x_0} \left[ f'(x_0) - \frac{f(x) - f(x_0)}{x - x_0} \right] = 0
\]

\[
\Leftrightarrow \lim_{x \to x_0} \left[ \frac{f'(x_0)(x - x_0) - f(x) - f(x_0)}{x - x_0} \right] = 0
\]

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

This suggests:

**Definition 2.1** Let \( f : \mathbb{R}^n \to \mathbb{R}^m \). The Fréchet Derivative (Jacobian matrix or just Derivative) at \( \vec{x}_0 \) is the \( m \times n \) matrix, \( Df(\vec{x}_0) \) (if it exists), that satisfies,

\[
\lim_{\vec{x} \to \vec{x}_0} \frac{\|Df(\vec{x}_0)(\vec{x} - \vec{x}_0) - (f(\vec{x}) - f(\vec{x}_0))\|}{\|\vec{x} - \vec{x}_0\|} = 0. \tag{1}
\]

Actually the above definition is potentially flawed since in saying “the \( m \times n \) matrix, \( Df(\vec{x}_0) \) (if it exists)” it was implied that there can be at most one \( m \times n \) matrix that satisfies (1). Thus one needs the following theorem which states that a solution to (1) is unique.

**Theorem 2.2** A solution, \( Df(\vec{x}_0) \), to (1) is unique.

}\[
\]
Proof If $A$ is a $m \times n$ matrix that also satisfies (1) then,

$$0 \leq \lim_{\vec{x} \to \vec{x}_0} \frac{\|DF(\vec{x}_0)(\vec{x} - \vec{x}_0) - A(\vec{x} - \vec{x}_0)\|}{\|\vec{x} - \vec{x}_0\|}$$

$$= \lim_{\vec{x} \to \vec{x}_0} \frac{\|DF(\vec{x}_0)(\vec{x} - \vec{x}_0) - (f(\vec{x}) - f(\vec{x}_0)) - A(\vec{x} - \vec{x}_0) - (f(\vec{x}) - f(\vec{x}_0))\|}{\|\vec{x} - \vec{x}_0\|}$$

$$\leq \lim_{\vec{x} \to \vec{x}_0} \frac{\|DF(\vec{x}_0)(\vec{x} - \vec{x}_0) - (f(\vec{x}) - f(\vec{x}_0))\|}{\|\vec{x} - \vec{x}_0\|} + \lim_{\vec{x} \to \vec{x}_0} \frac{\|A(\vec{x} - \vec{x}_0) - (f(\vec{x}) - f(\vec{x}_0))\|}{\|\vec{x} - \vec{x}_0\|}.$$

$$= 0.$$

Thus for any $\vec{y} \in \mathbb{R}^n$ one observes that $(\vec{x}_0 + t\vec{y}) - \vec{x}_0 = t\vec{y}$ and $t \to 0 \iff \vec{x}_0 + t\vec{y} \to \vec{x}_0$. Thus

$$0 = \lim_{\vec{x} \to \vec{x}_0} \frac{\|DF(\vec{x}_0)(\vec{x} - \vec{x}_0) - A(\vec{x} - \vec{x}_0)\|}{\|\vec{x} - \vec{x}_0\|} = \lim_{t \to 0} \frac{\|DF(\vec{x}_0)(t\vec{y}) - A(t\vec{y})\|}{\|t\vec{y}\|}$$

$$= \lim_{t \to 0} \frac{\|DF(\vec{x}_0)\vec{y} - A\vec{y}\|}{\|\vec{y}\|}$$

$$= \frac{\|DF(\vec{x}_0)\vec{y} - A\vec{y}\|}{\|\vec{y}\|}.$$

This implies that $DF(\vec{x}_0)\vec{y} = A\vec{y}$ for all $\vec{y} \in \mathbb{R}^n$ which in turns implies that $DF(\vec{x}_0) = A$, which is what we wanted to show. 

Example 2.3 If $m = n = 1$, that is $f : \mathbb{R} \to \mathbb{R}$, then $DF(x_0)$ is just $f'(x_0)$ (the good old Calculus I derivative) since

$$\lim_{x \to x_0} \frac{|DF(x_0)(x - x_0) - (f(x) - f(x_0))|}{|x - x_0|} = \lim_{x \to x_0} \left| DF(x_0) - \frac{f(x) - f(x_0)}{x - x_0} \right| = 0.$$

which implies that $DF(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = f'(x_0).$ 

Lemma 2.4 Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable at $\vec{x}_0$. Then there exists a function $\lambda : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$f(\vec{x}) = f(\vec{x}_0) + DF(\vec{x}_0)(\vec{x} - \vec{x}_0) + \lambda(\vec{x})\|\vec{x} - \vec{x}_0\|$$

(2)

where $\lim_{\vec{x} \to \vec{x}_0} \lambda(\vec{x}) = \vec{0}$. 

Proof Define

$$\lambda(\vec{x}) \overset{\text{def}}{=} \frac{(f(\vec{x}) - f(\vec{x}_0)) - DF(\vec{x}_0)(\vec{x} - \vec{x}_0)}{\|\vec{x} - \vec{x}_0\|}.$$
Then $\lambda(\cdot)$ satisfies (2) and $\lim_{\vec{x} \to \vec{x}_0} \lambda(\vec{x}) = \vec{0}$ by (1).

Since the $\lambda(\vec{x})\|\vec{x} - \vec{x}_0\|$ is negligible near $\vec{x}_0$, local to $\vec{x}_0$ the function $f$ can be thought of as a linear transformation plus a transformation. If $\vec{x}_0 = \vec{0}$ and $f(\vec{0}) = \vec{0}$ there is no translation and for small $\vec{x}$, one sees that $f(\vec{x}) \approx Df(\vec{0})\vec{x}$, i.e., locally the function can be approximated with the linear transformation $\vec{x} \mapsto Df(\vec{0})\vec{x}$.

**Corollary 2.5** Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable at $\vec{x}_0$. Then there exists an $\delta > 0$ and an $M > 0$ so that

$$\|\vec{x} - \vec{x}_0\| < \delta \Rightarrow \frac{\|f(\vec{x}) - f(\vec{x}_0)\|}{\|\vec{x} - \vec{x}_0\|} < M.$$  

**Proof** Defining $\lambda(\cdot)$ as in (2), let $\delta > 0$ be such that $\|\vec{x} - \vec{x}_0\| < \delta_0 \Rightarrow \|\lambda(\vec{x})\| < 1$. Then for $\|\vec{x} - \vec{x}_0\| < \delta$

$$\frac{f(\vec{x}) - f(\vec{x}_0)}{\|\vec{x} - \vec{x}_0\|} = Df(\vec{x}) \left( \frac{\vec{x} - \vec{x}_0}{\|\vec{x} - \vec{x}_0\|} \right) + \lambda(\vec{x}).$$

Letting $N$ be as in Proposition 1.5,

$$\frac{\|f(\vec{x}) - f(\vec{x}_0)\|}{\|\vec{x} - \vec{x}_0\|} = \|Df(\vec{x}) \left( \frac{\vec{x} - \vec{x}_0}{\|\vec{x} - \vec{x}_0\|} \right) + \lambda(\vec{x})\| \leq \|Df(\vec{x}) \left( \frac{\vec{x} - \vec{x}_0}{\|\vec{x} - \vec{x}_0\|} \right)\| + \|\lambda(\vec{x})\| \leq N \left( \frac{\vec{x} - \vec{x}_0}{\|\vec{x} - \vec{x}_0\|} \right) + \delta = N + \delta.$$  

We now let $M = N + \delta$.

**Theorem 2.6 (Chain Rule)** If $g : \mathbb{R}^n \to \mathbb{R}^p$ is differentiable at $\vec{x}_0 \in \mathbb{R}^n$ and $f : \mathbb{R}^p \to \mathbb{R}^m$ is differentiable at $g(\vec{x}_0) \in \mathbb{R}^p$ then $(f \circ g)(\cdot)$ is differentiable at $\vec{x}_0$ and furthermore,

$$D(f \circ g)(\vec{x}_0) = Df(g(\vec{x}_0)) Dg(\vec{x}_0).$$  

□
Proof: Look at
\[
0 \leq \frac{\|Df(g(\bar{x}_0)) \ Dg(\bar{x}_0)(\bar{x} - \bar{x}_0) - (f(g(\bar{x})) - f(g(\bar{x}_0)))\|}{\|\bar{x} - \bar{x}_0\|}
\]
\[
= \frac{\|Df(g(\bar{x}_0)) [Dg(\bar{x}_0)(\bar{x} - \bar{x}_0) - (g(\bar{x}) - g(\bar{x}_0)) + (g(\bar{x}) - g(\bar{x}_0))] - (f(g(\bar{x})) - f(g(\bar{x}_0)))\|}{\|\bar{x} - \bar{x}_0\|}
\]
\[
\leq \frac{\|Df(g(\bar{x}_0)) [g(\bar{x}) - g(\bar{x}_0)] - (f(g(\bar{x})) - f(g(\bar{x}_0)))\|}{\|\bar{x} - \bar{x}_0\|}
\]
\[
+ \frac{\|Df(g(\bar{x}_0))(Dg(\bar{x}_0)(\bar{x} - \bar{x}_0)) - (g(\bar{x}) - g(\bar{x}_0))\|}{\|\bar{x} - \bar{x}_0\|}
\]

We now need to show that the two terms in the last line above go to zero as \( \bar{x} \to \bar{x}_0 \).

For the first term, by Corollary 2.5 one can find a \( \delta_0 > 0 \) and a \( M > 0 \) so that
\[
\|\bar{x} - \bar{x}_0\| < \delta_0 \Rightarrow \frac{\|g(\bar{x}) - g(\bar{x}_0)\|}{\|\bar{x} - \bar{x}_0\|} < M.
\]
Given an \( \epsilon > 0 \),
\[
\lim_{\tilde{y} \to g(\bar{x}_0)} \frac{\|Df(g(\bar{x}_0))(\tilde{y} - g(\bar{x}_0)) - (f(\tilde{y}) - f(g(\bar{x}_0)))\|}{\|\tilde{y} - f(g(\bar{x}_0))\|} = 0
\]
implies that there exists \( \delta_1 > 0 \) so that
\[
\|\tilde{y} - g(\bar{x}_0)\| < \delta_1 \Rightarrow \frac{\|Df(g(\bar{x}_0))(\tilde{y} - g(\bar{x}_0)) - (f(\tilde{y}) - f(g(\bar{x}_0)))\|}{\|\tilde{y} - f(g(\bar{x}_0))\|} < \frac{\epsilon}{M}
\]
\[
\Rightarrow \|Df(g(\bar{x}_0))(\tilde{y} - g(\bar{x}_0)) - (f(\tilde{y}) - f(g(\bar{x}_0)))\| < \frac{\epsilon}{M}\|\tilde{y} - f(g(\bar{x}_0))\|.
\]
Since \( g(\cdot) \) is differentiable, it is continuous so we can choose \( \delta_2 > 0 \) so that
\[
\|\bar{x} - \bar{x}_0\| < \delta_2 \Rightarrow \|g(\bar{x}) - g(\bar{x}_0)\| < \delta_1.
\]
Letting $\delta \overset{\text{def}}{=} \min\{\delta_0, \delta_2\}$ we can conclude that for $\|\vec{x} - \vec{x}_0\| < \delta$,

$$\left\| Df(g(\vec{x}_0)) \left[ g(\vec{x}) - g(\vec{x}_0) \right] - (f(g(\vec{x})) - f(g(\vec{x}_0))) \right\| \leq \frac{\epsilon \left\| g(\vec{x}) - g(\vec{x}_0) \right\|}{\|\vec{x} - \vec{x}_0\|} \leq \frac{\epsilon M}{\|\vec{x} - \vec{x}_0\|} \leq \frac{\epsilon M}{M} = \epsilon. $$

Since $\epsilon$ was arbitrary, the first term goes to zero.

For the second term, let $N$ be as in Proposition \[1.5\] with $\vec{x}_0$ replaced with $g(\vec{x}_0)$. Then

$$\lim_{\vec{x} \to \vec{x}_0} \left\| \frac{Df(g(\vec{x}_0)) \left[ Dg(\vec{x}_0)(\vec{x} - \vec{x}_0) - (g(\vec{x}) - g(\vec{x}_0)) \right]}{\|\vec{x} - \vec{x}_0\|} \right\| \leq \lim_{\vec{x} \to \vec{x}_0} \frac{\|N [Dg(\vec{x}_0)(\vec{x} - \vec{x}_0) - (g(\vec{x}) - g(\vec{x}_0))]\|}{\|\vec{x} - \vec{x}_0\|} = N 0 = 0. \quad \blacksquare$$

**Example 2.7** Assume $m = n = p = 1$, that is $f, g : \mathbb{R} \to \mathbb{R}$. Then (4) and Example \[2.3\] gives us the ordinary chain rule, $(f \circ g)'(x_0) = f'(g(x_0))g'(x_0)$. \quad \square

**Definition 2.8** Any function $f : \mathbb{R}^n \to \mathbb{R}^m$ can be written as $f(\vec{x}) = (f_1(\vec{x}), \ldots, f_m(\vec{x}))$ where $f_j : \mathbb{R}^n \to \mathbb{R}$ for $1 \leq j \leq m$ are the *coordinate functions* of $f(\cdot)$. \quad \square

**Theorem 2.9** Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a function where $f(\vec{x}) = (f_1(\vec{x}), \ldots, f_m(\vec{x}))$ where $f_i : \mathbb{R}^n \to \mathbb{R}$ for $1 \leq i \leq m$. If all of the partial derivatives, $\frac{\partial f_i}{\partial x_j}(\vec{x})$ exist and are continuous at $\vec{x}_0$, then $Df(\vec{x}_0)$ exists and furthermore,

$$Df(\vec{x}_0) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\vec{x}_0) & \cdots & \frac{\partial f_1}{\partial x_n}(\vec{x}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\vec{x}_0) & \cdots & \frac{\partial f_m}{\partial x_n}(\vec{x}_0) \end{bmatrix}. \quad (5)$$

**Proof** First lets assume that $m = 1$, i.e., $f : \mathbb{R}^n \to \mathbb{R}$. Let $\vec{x}_0 = (x_0, \ldots, x_0n) \in \mathbb{R}^n$ and let $1 \leq j \leq n$. Define $g : \mathbb{R} \to \mathbb{R}^n$ by $g(x) = (x_0, \ldots, x_0(j-1), x, x_0(j+1), \ldots, x_0n)$. Then (4) and Example 2.3 gives us the ordinary chain rule, $(f \circ g)'(x_0) = f'(g(x_0))g'(x_0)$. \quad \square
where \( g(x) \) is \( \vec{x}_0 \) except for the \( j^{th} \) coordinate which is \( x \). Thus

\[
\frac{\partial f}{\partial x_j}(\vec{x}_0) = (f \circ g)'(x_{0j}) = Df(g(x_{0j})) Dg(x_{0j}) = Df(\vec{x}_0) \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = [Df(\vec{x}_0)]_{ij}
\]

and we have shown (5) for \( m = 1 \).

Now suppose that \( m \) is any positive integer and \( f : \mathbb{R}^n \to \mathbb{R}^m \). Using the case proved above, it suffices to show that

\[
\lim_{\vec{x} \to \vec{x}_0} \frac{\|\vec{x} - \vec{x}_0\|^2}{\|\vec{x} - \vec{x}_0\|} \left\| \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\vec{x}_0) & \ldots & \frac{\partial f_1}{\partial x_n}(\vec{x}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\vec{x}_0) & \ldots & \frac{\partial f_m}{\partial x_n}(\vec{x}_0) \end{bmatrix} (\vec{x} - \vec{x}_0) - (f(\vec{x}) - f(\vec{x}_0)) \right\|
\]

\[
= \lim_{\vec{x} \to \vec{x}_0} \frac{\|\vec{x} - \vec{x}_0\|^2}{\|\vec{x} - \vec{x}_0\|} \left\| \begin{bmatrix} Df_1(\vec{x}_0)(\vec{x} - \vec{x}_0) \\ \vdots \\ Df_m(\vec{x}_0)(\vec{x} - \vec{x}_0) \end{bmatrix} - \begin{bmatrix} f_1(\vec{x}) - f_1(\vec{x}_0) \\ \vdots \\ f_m(\vec{x}) - f_m(\vec{x}_0) \end{bmatrix} \right\|
\]

\[
= \lim_{\vec{x} \to \vec{x}_0} \frac{\|\vec{x} - \vec{x}_0\|^2}{\|\vec{x} - \vec{x}_0\|} \left\| \begin{bmatrix} Df_1(\vec{x}_0)(\vec{x} - \vec{x}_0) - (f_1(\vec{x}) - f_1(\vec{x}_0)) \\ \vdots \\ Df_m(\vec{x}_0)(\vec{x} - \vec{x}_0) - (f_m(\vec{x}) - f_m(\vec{x}_0)) \end{bmatrix} \right\|
\]

\[
\leq \lim_{\vec{x} \to \vec{x}_0} \sum_{j=1}^{m} \frac{\|Df_j(\vec{x}_0)(\vec{x} - \vec{x}_0) - (f_j(\vec{x}) - f_j(\vec{x}_0))\|}{\|\vec{x} - \vec{x}_0\|}
\]

\[
= \sum_{j=1}^{m} \lim_{\vec{x} \to \vec{x}_0} \frac{\|Df_j(\vec{x}_0)(\vec{x} - \vec{x}_0) - (f_j(\vec{x}) - f_j(\vec{x}_0))\|}{\|\vec{x} - \vec{x}_0\|}
\]

\[
= 0.
\]

**Example 2.10** Let \( f(x, y) = (x^2 + y, x \cos(y), yx^2) \) and \( g(x, y) = (xy, x - y^2) \). Then
since

\[
Df(x, y) = \begin{bmatrix}
2x & 1 \\
\cos(y) & -x \sin(y) \\
2xy & x^2
\end{bmatrix},
\]

one has

\[
D(f \circ g)(x, y) = Df(g(x, y)) Dg(x, y) = Df(xy, x - y^2) Dg(x, y)
\]

\[
= \begin{bmatrix}
2xy & 1 \\
\cos(x - y^2) & -xy \sin(x - y^2) \\
2xy(x - y^2) & x^2 y^2
\end{bmatrix}
\begin{bmatrix}
y & x \\
1 & -2y
\end{bmatrix}
\]

\[
= \begin{bmatrix}
2xy y^2 + 1 \\
y \cos(x - y^2) - xy \sin(x - y^2) & 2x^2 y - 2y \\
2xy^2(x - y^2) + x^2 y^2 & x \cos(x - y^2) - 2xy^2 \sin(x - y^2)
\end{bmatrix}
\]

\[
= \begin{bmatrix}
2x y & 1 \\
\cos(x - y^2) - xy \sin(x - y^2) \\
2xy(x - y^2) & x^2 y^2
\end{bmatrix}
\begin{bmatrix}
y & x \\
1 & -2y
\end{bmatrix}
\]

\[
= \begin{bmatrix}
2xy y^2 + 1 \\
y \cos(x - y^2) - xy \sin(x - y^2) & 2x^2 y - 2y \\
2xy^2(x - y^2) + x^2 y^2 & x \cos(x - y^2) - 2xy^2 \sin(x - y^2)
\end{bmatrix}
\]

3 Derivatives: Another Viewpoint

When considering the function \( f : \mathbb{R}^n \to \mathbb{R}^m \) consider a control panel with \( n \) levers in front of you. These levers control \( m \) cranes in the distance. Moving the \( j^{th} \) lever up can cause more than one crane to move. When the levers are in \( \vec{x}_0 \in \mathbb{R}^n \) position, \( \frac{\partial f_k}{\partial x_j}(\vec{x}_0) \) measures the movement in the \( k^{th} \) crane when pushing the \( j^{th} \) lever. When the levers are in \( \vec{x}_0 \) position, the dynamics of the entire apparatus is described by the set of partials

\[
\left\{ \frac{\partial f_k}{\partial x_j}(\vec{x}_0) : 1 \leq k \leq m \ & \& 1 \leq j \leq n \right\}.
\]

The matrix derivative,

\[
Df(\vec{x}_0) = \begin{bmatrix}
\frac{\partial f_1}{\partial x_1}(\vec{x}_0) & \cdots & \frac{\partial f_1}{\partial x_n}(\vec{x}_0) \\
\vdots & \ddots & \vdots \\
\frac{\partial f_m}{\partial x_1}(\vec{x}_0) & \cdots & \frac{\partial f_m}{\partial x_n}(\vec{x}_0)
\end{bmatrix}
\]

can be thought of a bookkeeper’s way of organizing the partials. The \( m \times n \) matrix derivative contains all the information given by (6). In addition, the bookkeepers way of organizing the partials is particularly fortunate in that \( D(f \circ g)(\vec{x}_0) \) corresponds
to $Df(g(\vec{x}_0)) Dg(\vec{x}_0)$, i.e., the one dimensional chain rule formula works here too, with matrix multiplication taking the place of ordinary multiplication.

This alternate way of view the Fréchet derivative is accurate as far as it goes, but it is not how most mathematicians picture the Fréchet derivative because this “bookkeeper’s viewpoint” does not encompasses all the important aspects of the Fréchet derivative. Lost here is the notion of the Fréchet derivative, $Df(\vec{x}_0)$, being the linear transformation that best approximates $f(\cdot)$ at $\vec{x}_0$. Indeed, one needs Definition 2.1 to obtain Lemma 2.4 used in the next section.

### 4 Differentials

Lemma 2.4 tells us that when $\vec{x} \in \mathbb{R}^n$ is close to $\vec{x}_0 \in \mathbb{R}^n$ then

$$Df(\vec{x}_0)(\vec{x} - \vec{x}_0) - (f(\vec{x}) - f(\vec{x}_0)) \approx 0$$

$$\Leftrightarrow$$

$$f(\vec{x}) \approx f(\vec{x}_0) + Df(\vec{x}_0)(\vec{x} - \vec{x}_0).$$ (7)

This “$\approx$” is quite good since the error, by Lemma 2.4 is just $\lambda(\vec{x}) \|\vec{x} - \vec{x}_0\|$, a quantity that goes to zero doubly fast as $\vec{x} \to \vec{x}_0$. If $f : \mathbb{R}^n \to \mathbb{R}$, one is thus led to:

**Definition 4.1** Given the points $\vec{x}_0 = \langle x_{01}, \ldots, x_{0n} \rangle \in \mathbb{R}^n$ and $\vec{x} = \langle x_1, \ldots, x_n \rangle \in \mathbb{R}^n$ and the differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, the quantity

$$Df(\vec{x}_0)(\vec{x} - \vec{x}_0) = f_{x_1}(\vec{x}_0)(x_1 - x_{01}) + \ldots + f_{x_n}(\vec{x}_0)(x_n - x_{0n})$$

is called the total differential of $f(\cdot)$ and is written as

$$df \overset{\text{def}}{=} f_{x_1}(\vec{x}_0)dx_1 + \ldots + f_{x_n}(\vec{x}_0)dx_n$$

where $dx_j \overset{\text{def}}{=} x_j - x_{0j}$.

If $f : \mathbb{R}^n \to \mathbb{R}$, (7) becomes

$$f(\vec{x}) \approx f(\vec{x}_0) + df.$$  

One can show that the above equation is just another way of saying that a function can locally be approximated with its tangent plane.
Equation (7) is very useful as seen in the following example.

**Example 4.2** Let \( f(x, y) = \sqrt{x^2 + y^2} \). Find an approximate answer for \( f(3.05, 3.96) \).

**Solution:** Let \((x_0, y_0) = (3, 4)\). Then

\[
\begin{align*}
 f_x(x, y) & = \frac{1}{2}(x^2 + y^2)^{-\frac{1}{2}}2x \\
 f_y(x, y) & = \frac{1}{2}(x^2 + y^2)^{-\frac{1}{2}}2y.
\end{align*}
\]

Thus using (7)

\[
 f(3.05, 3.96) \approx f(3, 4) + f_x(3, 4) \cdot x'(t) + f_y(3, 4) \cdot y'(t).
\]

Using a calculator one has \( f(3.05, 3.96) = 4.9984097 \ldots \), so our approximation is quite good!

## 5 Specific Instances of Chain Rule

The remaining theorems in this section are just, in truth, examples of the multivariate chain rule. They are mentioned here as theorems solely to be consistent with our book.

**Theorem 5.1** Let \( f(x, y) = z \) be a differentiable function. Let \( x : \mathbb{R} \rightarrow \mathbb{R} \) and \( y : \mathbb{R} \rightarrow \mathbb{R} \) also be differentiable functions. Then letting \( h(t) \overset{\text{def}}{=} f(x(t), y(t)) \) one has

\[
 h'(t) = f_x(x(t), y(t))x'(t) + f_y(x(t), y(t))y'(t).
\]

**Proof** Let \( g(t) \overset{\text{def}}{=} (x(t), y(t)) \). Then \( g : \mathbb{R} \rightarrow \mathbb{R}^2 \) and

\[
 h'(t) = (f \circ g)'(t) = D(f \circ g)(t) = Df(g(t)) \cdot Dg(t)
\]

\[
 = \begin{bmatrix} f_x(x(t), y(t)) & f_y(x(t), y(t)) \end{bmatrix} \begin{bmatrix} x'(t) \\ y'(t) \end{bmatrix}
\]

\[
 = f_x(x(t), y(t))x'(t) + f_y(x(t), y(t))y'(t).
\]
Theorem 5.2  Let \( f(x, y, z) = w \) be a differentiable function. Let \( x, y, z : \mathbb{R} \to \mathbb{R} \) be differentiable functions. Then letting \( h(t) \overset{\text{def}}{=} f(x(t), y(t), z(t)) \) one has
\[
h'(t) = f_x(x(t), y(t), z(t))x'(t) + f_y(x(t), y(t), z(t))y'(t) + f_z(x(t), y(t), z(t))z'(t).
\]

Proof  See the proof of Theorem 5.1.

Theorem 5.3  Let \( f : \mathbb{R}^2 \to \mathbb{R} \) and \( x, y : \mathbb{R}^2 \to \mathbb{R} \) be differentiable functions. Then letting \( h(s, t) \overset{\text{def}}{=} f(x(s, t), y(s, t)) \) one has
\[
h_s(s, t) = f_x(x(s, t), y(s, t))x_s(s, t) + f_y(x(s, t), y(s, t))y_s(s, t)
\]
\[
h_t(s, t) = f_x(x(s, t), y(s, t))x_t(s, t) + f_y(x(s, t), y(s, t))y_t(s, t).
\]

Proof  Let \( g(s, t) \overset{\text{def}}{=} (x(s, t), y(s, t)) \). Then \( g : \mathbb{R}^2 \to \mathbb{R}^2 \) and
\[
Dh(s, t) = D(f \circ g)(s, t)
\]
\[
= Df(g(s, t))Dg(s, t) = \begin{bmatrix} f_x(x(s, t), y(s, t)) & f_y(x(s, t), y(s, t)) \\ y_s(s, t) & y_t(s, t) \end{bmatrix} \begin{bmatrix} x_s(s, t) & x_t(s, t) \\ y_s(s, t) & y_t(s, t) \end{bmatrix}.
\]

After computing the above matrix multiplication, the theorem is proved by comparing the components of the first and last matrices above.

6  Summary

Definition 6.1  Given \( f : \mathbb{R}^n \to \mathbb{R}^m \), the \( m \times n \) matrix, \( Df(\vec{x}_0) \), that satisfies
\[
\lim_{\vec{x} \to \vec{x}_0} \frac{\|Df(\vec{x}_0)(\vec{x} - \vec{x}_0) - (f(\vec{x}) - f(\vec{x}_0))\|}{\|\vec{x} - \vec{x}_0\|} = 0,
\]
is called the Fréchet Derivative of \( f(\cdot) \) at \( \vec{x}_0 \).

Theorem 6.2 (Chain Rule)  If \( g : \mathbb{R}^n \to \mathbb{R}^p \) is differentiable at \( \vec{x}_0 \in \mathbb{R}^n \) and \( f : \mathbb{R}^p \to \mathbb{R}^m \) is differentiable at \( g(\vec{x}_0) \in \mathbb{R}^p \) then \( (f \circ g)(\cdot) \) is differentiable at \( \vec{x}_0 \) and furthermore,
\[
D(f \circ g)(\vec{x}_0) = Df(g(\vec{x}_0))Dg(\vec{x}_0).
\]
Theorem 6.3 Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a function where $f(\vec{x}) = (f_1(\vec{x}), \ldots, f_m(\vec{x}))$ where $f_i : \mathbb{R}^n \to \mathbb{R}$ for $1 \leq i \leq m$. If all of the partial derivatives, $\frac{\partial f_i}{\partial x_j}(\vec{x})$ exist and are continuous at $\vec{x}_0$, then $Df(\vec{x}_0)$ exists and furthermore,

$$Df(\vec{x}_0) = \begin{bmatrix}
\frac{\partial f_1}{\partial x_1}(\vec{x}_0) & \cdots & \frac{\partial f_1}{\partial x_n}(\vec{x}_0) \\
\vdots & \ddots & \vdots \\
\frac{\partial f_m}{\partial x_1}(\vec{x}_0) & \cdots & \frac{\partial f_m}{\partial x_n}(\vec{x}_0)
\end{bmatrix}.$$

From Lemma 2.4 we have the important fact that

$$f(\vec{x}) \approx f(\vec{x}_0) + Df(\vec{x}_0)(\vec{x} - \vec{x}_0).$$

This motivates us to present the following definition.

Definition 6.4 Given the points $\vec{x}_0 = \langle x_{01}, \ldots, x_{0n} \rangle \in \mathbb{R}^n$ and $\vec{x} = \langle x_1, \ldots, x_n \rangle \in \mathbb{R}^n$ and the differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, the quantity

$$Df(\vec{x}_0)(\vec{x} - \vec{x}_0) = f_{x_1}(\vec{x}_0)(x_1 - x_{01}) + \ldots + f_{x_n}(\vec{x}_0)(x_n - x_{0n})$$

is called the total differential of $f(\cdot)$ and is written as

$$df \overset{\text{def}}{=} f_{x_1}(\vec{x}_0)dx_1 + \ldots + f_{x_n}(\vec{x}_0)dx_n$$

where $dx_j \overset{\text{def}}{=} x_j - x_{0j}$. \hfill \Box

And finally we have the three following special cases of the multivariate chain rule:

Theorem 6.5 Let $f(x, y) = z$ be a differentiable function. Let $x : \mathbb{R} \to \mathbb{R}$ and $y : \mathbb{R} \to \mathbb{R}$ also be differentiable functions. Then letting $h(t) \overset{\text{def}}{=} f(x(t), y(t))$ one has

$$h'(t) = f_x(x(t), y(t))x'(t) + f_y(x(t), y(t))y'(t).$$ \hfill \Box

Theorem 6.6 Let $f(x, y, z) = w$ be a differentiable function. Let $x, y, z : \mathbb{R} \to \mathbb{R}$ be differentiable functions. Then letting $h(t) \overset{\text{def}}{=} f(x(t), y(t), z(t))$ one has

$$h'(t) = f_x(x(t), y(t), z(t))x'(t) + f_y(x(t), y(t), z(t))y'(t) + f_z(x(t), y(t), z(t))z'(t).$$ \hfill \Box
Theorem 6.7 Let $f : \mathbb{R}^2 \to \mathbb{R}$ and $x, y : \mathbb{R}^2 \to \mathbb{R}$ be differentiable functions. Then letting $h(s, t) \overset{\text{def}}{=} f(x(s, t), y(s, t))$ one has
\[
\begin{align*}
    h_s(s, t) &= f_x(x(s, t), y(s, t))x_s(s, t) + f_y(x(s, t), y(s, t))y_s(s, t) \\
    h_t(s, t) &= f_x(x(s, t), y(s, t))x_t(s, t) + f_y(x(s, t), y(s, t))y_t(s, t). 
\end{align*}
\]

7 Exercises

Let $f : \mathbb{R}^2 \to \mathbb{R}^3$, let $g : \mathbb{R}^3 \to \mathbb{R}^2$, let $h : \mathbb{R}^2 \to \mathbb{R}$ and let $j : \mathbb{R} \to \mathbb{R}^2$ be such that
\[
\begin{align*}
    f(x, y) &\overset{\text{def}}{=} (x, xy, x \sin(y)) \\
    g(1, \pi/3, \sqrt{3}/2) &= (\pi/3, 3/4) \\
    Dg(x, y, z) &= \begin{bmatrix} 0 & 1 & 0 \\ z^2 & 0 & 2xz \end{bmatrix} \\
    h(x, y) &\overset{\text{def}}{=} x^2 + y + cx \\
    j(x) &\overset{\text{def}}{=} (2x - 1, 3 \cos(x)) \\
    \ell(x, y) &\overset{\text{def}}{=} (h \circ g \circ f)(x, y).
\end{align*}
\]

1. Letting $k(t) \overset{\text{def}}{=} h(2t - 1, 3 \cos(t))$, what is $k'(0)$?

2. What is $D(h \circ j)(0)$?

3. What is $Df(1, \pi/3)$?

4. What is $D(g \circ f)(1, \pi/3)$?

5. Given that $\ell_y(1, \pi/3) = 6$, what is $c$? Hint: look at $D\ell(1, \pi/3)$.

6. Find the approximate length of the largest stick that will fit in a box of length 2.01, width 1.97, and height 1.02 using differentials. Compare this to the actual length.
References
