The Total Derivative
Examples, Properties, The Mean Value Theorem

February 6, 2018
Contents

1 Review

2 Properties

3 The Mean Value Theorem

4 Sufficient Conditions for Differentiability
The Total Derivative

Definition

Let $S \subset \mathbb{R}^n$, $f : S \rightarrow \mathbb{R}^m$ and $a \in \text{int } S$. We say that $f$ is **differentiable** at the point $a$, if and only if there exists a **linear** transformation

$$T_a : \mathbb{R}^n \rightarrow \mathbb{R}^m$$

such that if we write

$$f(a + v) = f(a) + T_a v + \|v\|E(a, v),$$

for every $v$ such that $a + v \in S$, then

$$\lim_{v \rightarrow 0} E(a, v) = 0.$$
Remark

In this case the linear transformation $T_a$ is unique. It is usually denoted

$$Df(a),$$

and its value on any vector $v \in \mathbb{R}^n$ is given by the directional derivative:

$$Df(a)v = T_av = f'(a; v) \quad \text{for every } v \in \mathbb{R}^n.$$

It is called the **total derivative** of $f$ at $a$, or simply the **derivative** of $f$ at $a$. 
The Total Derivative

**Remark**

In this case the linear transformation $T_a$ is unique. It is usually denoted

$$Df(a),$$

and its value on any vector $v \in \mathbb{R}^n$ is given by the directional derivative:

$$Df(a)v = T_a v = f'(a; v) \quad \text{for every } v \in \mathbb{R}^n.$$

It is called the **total derivative** of $f$ at $a$, or simply the **derivative** of $f$ at $a$.

**Examples**
The Total Derivative

Remark

In this case the linear transformation $T_a$ is unique. It is usually denoted

$$Df(a),$$

and its value on any vector $v \in \mathbb{R}^n$ is given by the directional derivative:

$$Df(a)v = T_a v = f'(a; v) \quad \text{for every } v \in \mathbb{R}^n.$$

It is called the total derivative of $f$ at $a$, or simply the derivative of $f$ at $a$.

Examples

1. **Constant Function.** Any constant function $f : S \to \mathbb{R}^m$, $S \subset \mathbb{R}^n$ is differentiable (at every interior point $a$ in the domain) and $Df(a) = 0$. 
Examples

Indeed

\[ f(a + v) = f(a) = f(a) + 0 + \|v\| 0. \]

\[ T_a v \]

\[ E(a, v) \]
Examples

Indeed

\[ f(a + v) = f(a) = f(a) + 0 + \|v\|0 . \]

\[ T_a v \]

\[ E(a, v) \]
Examples

Indeed

\[ f(a + v) = f(a) = f(a) + \underbrace{0}_{T_a v} + \underbrace{0}_{\|v\| \cdot E(a,v)} \, . \]

2. **Linear Transformation.** Any linear transformation \( T : \mathbb{R}^n \to \mathbb{R}^m \) is differentiable, and \( DT(a) = T \), for every \( a \in \mathbb{R}^n \).
Examples

Indeed

\[ f(a + v) = f(a) = f(a) + \underbrace{0}_{T_a v} + \|v\| \underbrace{0}_{E(a, v)}. \]

2 Linear Transformation. Any linear transformation \( T : \mathbb{R}^n \to \mathbb{R}^m \) is differentiable, and \( DT(a) = T \), for every \( a \in \mathbb{R}^n \).

Indeed:

\[ T(a + v) = T(a) + T(v) = T(a) + T(v) + \underbrace{0}_{T_a v} + \|v\| \underbrace{0}_{E(a, v)}. \]
Examples

Scalar Product. For any \( n \in \mathbb{N} \) the scalar product \((,): \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}\) is differentiable at any point \( a = (a_1, a_2) \in \mathbb{R}^n \times \mathbb{R}^n \), and

\[
D(,)(a_1, a_2)(v_1, v_2) = (v_1, a_2) + (a_1, v_2),
\]

for every \( v = (v_1, v_2) \in \mathbb{R}^n \times \mathbb{R}^n \).
3 Scalar Product. For any $n \in \mathbb{N}$ the scalar product $(\ , \ ) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is differentiable at any point $a = (a_1, a_2) \in \mathbb{R}^n \times \mathbb{R}^n$, and

$$D(\ , \ )(a_1, a_2)(v_1, v_2) = (v_1, a_2) + (a_1, v_2),$$

for every $v = (v_1, v_2) \in \mathbb{R}^n \times \mathbb{R}^n$. Indeed:

$$(a_1 + v_1, a_2 + v_2) = (a_1, a_2) + (v_1, a_2) + (a_1, v_2) + \|v\| \left(\frac{v_1}{\|v\|}, \frac{v_2}{\|v\|}\right)_{E(a,v)},$$
Examples

and

\[ |E(a, v)| = \frac{|(v_1, v_2)|}{\|v\|} \leq C-S \frac{\|v_1\| \cdot \|v_2\|}{\|v\|} \leq \|v\| \to 0, \]

as \( v \to 0. \)
Examples

and

\[ |E(a, v)| = \frac{|(v_1, v_2)|}{\|v\|} \leq \frac{v_1 \cdot v_2}{\|v\|} \leq \|v\| \to 0, \]

as \( v \to 0 \).

4 **Product.** The map \( f : \mathbb{R}^n \to \mathbb{R} \), defined by
\[ f(x) = x_1 \cdot x_2 \cdots x_n, \] for every \( x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \), is differentiable at every point \( a = (a_1, \ldots, a_n) \in \mathbb{R}^n \), and

\[ Df(a)v = v_1 a_2 \cdots a_n + a_1 v_2 a_3 \cdots a_n + \cdots + a_1 \cdots a_{n-1} v_n, \]

for every \( v = (v_1, \ldots, v_n) \in \mathbb{R}^n \).
Examples

Indeed:

\[(a_1 + v_1) \cdots (a_n + v_n) = a_1 \cdots a_n + \sum_i a_1 \cdots v_i \cdots a_n + \sum_{i<j} a_1 \cdots v_i \cdots v_j \cdots a_n + \cdots + \sum_{i_1<i_2<\cdots<i_k} a_1 \cdots v_{i_1} \cdots v_{i_2} \cdots v_{i_k} \cdots a_n + \cdots + v_1 \cdots v_n.\]
Examples

Indeed:

\[(a_1 + v_1) \cdots (a_n + v_n) = a_1 \cdots a_n + \sum_i a_1 \cdots v_i \cdots a_n + \]

\[\sum_{i<j} a_1 \cdots v_i \cdots v_j \cdots a_n + \cdots + \]

\[\sum_{i_1<i_2<\cdots<i_k} a_1 \cdots v_{i_1} \cdots v_{i_2} \cdots v_{i_k} \cdots a_n + \cdots + v_1 \cdots v_n.\]

So:
Examples

\[(a_1 + v_1) \cdots (a_n + v_n) = a_1 \cdots a_n + \sum_{i} a_1 \cdots v_i \cdots a_n + T_a v\]

\[+ \|v\| \left( \frac{1}{\|v\|} \sum_{k=2}^{n} \left( \sum_{i_1 < i_2 < \cdots < i_k} a_1 \cdots v_{i_1} \cdots v_{i_2} \cdots v_{i_k} \cdots a_n \right) \right),\]

\[E(a, v)\]

and, as before, \(E(a, v) \to 0\) as \(v \to 0\), since
Examples

\[(a_1 + v_1) \cdots (a_n + v_n) = a_1 \cdots a_n + \sum_i a_1 \cdots v_i \cdots a_n + \sum_{k=2}^{n} \left( \sum_{i_1 < i_2 < \cdots < i_k} a_1 \cdots v_{i_1} \cdots v_{i_2} \cdots v_{i_k} \cdots a_n \right) E(a,v) \]

and, as before, \( E(a,v) \to 0 \) as \( v \to 0 \), since

\[\sum_{i_1 < i_2 < \cdots < i_k} |a_1 \cdots v_{i_1} \cdots v_{i_2} \cdots v_{i_k} \cdots a_n| \leq \|v\|^k C_k(a),\]

so that
Examples

$$|E(a, v)| \leq C(a)\|v\|$$

for $$\|v\| \leq 1.$$
Examples

\[ |E(a, v)| \leq C(a)\|v\| \]

for \( \|v\| \leq 1. \)
Examples

|E(a, v)| \leq C(a)\|v\|

for \|v\| \leq 1.

5 Inverse. The map \( f : \mathbb{R} \setminus \{0\} \to \mathbb{R}, \) defined by \( f(x) = x^{-1}, \) is differentiable, and

\[ Df(a)v = -\frac{v}{a^2}, \]

for every \( a \in \mathbb{R} \) and every \( v \in \mathbb{R}. \)
Examples

\[ |E(a, v)| \leq C(a)\|v\| \]

for \( \|v\| \leq 1 \).

5 Inverse. The map \( f: \mathbb{R} \setminus \{0\} \to \mathbb{R} \), defined by \( f(x) = x^{-1} \), is differentiable, and

\[ Df(a)v = -\frac{v}{a^2}, \]

for every \( a \in \mathbb{R} \) and every \( v \in \mathbb{R} \).

- This is part of remark we made before, that:
The function $f : S \rightarrow \mathbb{R}^m$, $S \subset \mathbb{R}$ is differentiable at the point $a \in S$ if and only if

$$f'(a) = \lim_{v \rightarrow 0} \frac{f(a + v) - f(a)}{v} \in \mathbb{R}^m$$

exists, and

$$Df(a)v = f'(a) \cdot v$$

for every $v \in \mathbb{R}$.
Suppose $S_i \subset \mathbb{R}^{n_i}$, $f_i : S_i \rightarrow \mathbb{R}^{m_i}$ are functions and $a_i \in \text{int } S_i$, for every $1 \leq i \leq k$. Consider also the function

$$f \overset{\text{not.}}{=} f_1 \times \cdots \times f_k : S \overset{\text{not.}}{=} S_1 \times \cdots \times S_k \rightarrow \mathbb{R}^{m_1 + \cdots + m_k},$$

defined by

$$f(s_1, \ldots, s_k) = (f_1(s_1), \ldots, f(s_k)).$$

The function $f = f_1 \times \cdots \times f_k$ is differentiable at the point $a \overset{\text{not.}}{=} (a_1, \ldots, a_k) \in \text{int } S$, if and only if each function $f_i$ is differentiable at $a_i$.

Moreover

$$Df(a) = Df_1(a_1) \times \cdots \times Df_k(a_k).$$
Properties

Proof:

\[
f(a_1 + v_1, \ldots, a_k + v_k) = (f_1(a_1 + v_1), \ldots, f_k(a_k + v_k)) = \\
(f_1(a_1) + Df_1(a_1)v_1 + \|v_1\|E_1(a_1, v_1), \ldots \\
\ldots, f_k(a_k) + Df_k(a_k)v_k + \|v_k\|E_k(a_k, v_k)) = \\
(f_1(a_1), \ldots, f_k(a_k)) + (Df_1(a_1)v_1, \ldots, Df_k(a_k)v_k) + \\
\underbrace{T_{a,v}}_{E(a,v)} \\
\|v\|\left(\frac{\|v_1\|}{\|v\|}E_1(a_1, v_1), \ldots, \frac{\|v_k\|}{\|v\|}E_k(a_k, v_k)\right).
\]
Properties

Proof:

\[
\begin{aligned}
f(a_1 + v_1, \ldots, a_k + v_k) &= (f_1(a_1 + v_1), \ldots, f_k(a_k + v_k)) = \\
(f_1(a_1) + Df_1(a_1)v_1 + \|v_1\|E_1(a_1, v_1), \ldots \\
\ldots, f_k(a_k) + Df_k(a_k)v_k + \|v_k\|E_k(a_k, v_k)) = \\
(f_1(a_1), \ldots, f_k(a_k)) + (Df_1(a_1)v_1, \ldots, Df_k(a_k)v_k) + \\
\|v\|(\frac{\|v_1\|}{\|v\|}E_1(a_1, v_1)), \ldots, \frac{\|v_k\|}{\|v\|}E_k(a_k, v_k)).
\end{aligned}
\]

So, what we really have to prove is that \(E(a, v) \to 0\) as \(v \to 0\), if and only if

\[
E_i(a_i, v_i) \to 0, \text{ as } v_i \to 0,
\]

for every \(1 \leq i \leq k\).
Properties

⇒ If \( E(a, v) \to 0 \) as \( v \to 0 \), then in particular it does for vectors \( v \) of the form

\[
v = (0, \ldots, v_i, \ldots 0),
\]

but \( E(a, v) = E_i(a_i, v_i) \) in this case.
Properties

⇒ If $E(a, v) \to 0$ as $v \to 0$, then in particular it does for vectors $v$ of the form

$$v = (0, \ldots, v_i, \ldots 0),$$

but $E(a, v) = E_i(a_i, v_i)$ in this case.

⇐

$$\|E(a, v)\|^2 = \frac{\|v_1\|^2}{\|v\|^2} \|E_1(a_1, v_1)\|^2 + \cdots + \frac{\|v_k\|^2}{\|v\|^2} \|E_k(a_k, v_k)\|^2 \leq \|E_1(a_1, v_1)\|^2 + \cdots + \|E_k(a_k, v_k)\|^2,$$

since $\|v_i\| \leq \|v\|$, for every $1 \leq i \leq k$. 

QED

You see where we are going with this. Since we know that

1. Linear Transformations are differentiable,
2. The composition of differentiable maps is differentiable (The Chain Rule),

properties

⇒ If \( E(a, v) \to 0 \) as \( v \to 0 \), then in particular it does for vectors \( v \) of the form
\[
v = (0, \ldots, v_i, \ldots, 0),
\]
but \( E(a, v) = E_i(a_i, v_i) \) in this case.

⇐
\[
\| E(a, v) \|^2 = \frac{\| v_1 \|^2}{\| v \|^2} \| E_1(a_1, v_1) \|^2 + \cdots + \frac{\| v_k \|^2}{\| v \|^2} \| E_k(a_k, v_k) \|^2 \\
\leq \| E_1(a_1, v_1) \|^2 + \cdots + \| E_k(a_k, v_k) \|^2,
\]
since \( \| v_i \| \leq \| v \| \), for every \( 1 \leq i \leq k \).

QED
The Total Derivative

Review

Properties

The Mean Value Theorem

Sufficient Conditions for Differentiability

Properties

⇒ If \( E(a, v) \to 0 \) as \( v \to 0 \), then in particular it does for vectors \( v \) of the form
\[
v = (0, \ldots, v_i, \ldots, 0),
\]
but \( E(a, v) = E_i(a_i, v_i) \) in this case.

⇐

\[
\| E(a, v) \|^2 = \frac{\| v_1 \|^2}{\| v \|^2} \| E_1(a_1, v_1) \|^2 + \cdots + \frac{\| v_k \|^2}{\| v \|^2} \| E_k(a_k, v_k) \|^2 \\
\leq \| E_1(a_1, v_1) \|^2 + \cdots + \| E_k(a_k, v_k) \|^2,
\]
since \( \| v_i \| \leq \| v \| \), for every \( 1 \leq i \leq k \).

QED

You see where we are going with this. Since we know that
Properties

⇒ If $E(a, v) \to 0$ as $v \to 0$, then in particular it does for vectors $v$ of the form

$$v = (0, \ldots, v_i, \ldots 0),$$

but $E(a, v) = E_i(a_i, v_i)$ in this case.

⇐

$$\|E(a, v)\|^2 = \frac{\|v_1\|^2}{\|v\|^2} \|E_1(a_1, v_1)\|^2 + \cdots + \frac{\|v_k\|^2}{\|v\|^2} \|E_k(a_k, v_k)\|^2 \leq \|E_1(a_1, v_1)\|^2 + \cdots + \|E_k(a_k, v_k)\|^2,$$

since $\|v_i\| \leq \|v\|$, for every $1 \leq i \leq k$.

QED

You see where we are going with this. Since we know that

1. Linear Transformations are differentiable,
Properties

⇒ If \( E(a, v) \to 0 \) as \( v \to 0 \), then in particular it does for vectors \( v \) of the form
\[
v = (0, \ldots, v_i, \ldots, 0),
\]
but \( E(a, v) = E_i(a_i, v_i) \) in this case.

\[\begin{align*}
\|E(a, v)\|^2 &= \frac{\|v_1\|^2}{\|v\|^2} \|E_1(a_1, v_1)\|^2 + \cdots + \frac{\|v_k\|^2}{\|v\|^2} \|E_k(a_k, v_k)\|^2 \\
&\leq \|E_1(a_1, v_1)\|^2 + \cdots + \|E_k(a_k, v_k)\|^2,
\end{align*}\]

since \( \|v_i\| \leq \|v\| \), for every \( 1 \leq i \leq k \).

QED

You see where we are going with this. Since we know that
\begin{enumerate}
\item Linear Transformations are differentiable,
\item The composition of differentiable maps is differentiable (The Chain Rule),
\end{enumerate}
Then this last theorem tells us that sums, products, scalar products, inverses of differentiable functions are differentiable.
Properties

3. Then this last theorem tells us that sums, products, scalar products, inverses of differentiable functions are differentiable.

4. In general any operation on functions that can be written as the composition

$$\Phi \circ f_1 \times \cdots \times f_k \circ \Delta,$$

with $\Phi$ differentiable will be differentiable at the points where all the $f_i'$s are differentiable.
Then this last theorem tells us that sums, products, scalar products, inverses of differentiable functions are differentiable.

In general any operation on functions that can be written as the composition

\[ \Phi \circ f_1 \times \cdots \times f_k \circ \Delta, \]

with \( \Phi \) differentiable will be differentiable at the points where all the \( f_i' \)'s are differentiable.

In particular we proved
Properties

3. Then this last theorem tells us that sums, products, scalar products, inverses of differentiable functions are differentiable.

4. In general any operation on functions that can be written as the composition

$$\Phi \circ f_1 \times \cdots \times f_k \circ \Delta,$$

with $\Phi$ differentiable will be differentiable at the points where all the $f_i'$s are differentiable.

In particular we proved

Proposition

Every rational function is differentiable, at every point in its domain of definition.
Then this last theorem tells us that sums, products, scalar products, inverses of differentiable functions are differentiable.

In general any operation on functions that can be written as the composition

$$\Phi \circ f_1 \times \cdots \times f_k \circ \Delta,$$

with $\Phi$ differentiable will be differentiable at the points where all the $f_i'$s are differentiable.

- In particular we proved

Proposition

Every rational function is differentiable, at every point in its domain of definition.

- Recall $f = \frac{P}{Q}$, with $P, Q$ polynomials in $n$ variables, defined on $\{x \mid Q(x) \neq 0\}$. 
Let me remind you that $f : S \to \mathbb{R}^m$ is the same as $m$ functions $f_i : S \to \mathbb{R}$, the components of the function $f$.

$$f(x) = (f_1(x), \ldots f_m(x)).$$
Properties

- Let me remind you that \( f : S \rightarrow \mathbb{R}^m \) is the same as \( m \) functions \( f_i : S \rightarrow \mathbb{R} \), the components of the function \( f \).

\[
f(x) = (f_1(x), \ldots, f_m(x)).
\]

Theorem

*The function \( f : S \rightarrow \mathbb{R}^m, (S \subset \mathbb{R}^n) \) is differentiable at the point \( a \in \text{int } S \), if and only if each component \( f_i : S \rightarrow \mathbb{R} \) is differentiable at \( a \).*

Moreover

\[
Df(a) = \begin{bmatrix}
Df_1(a) \\
\vdots \\
Df_i(a) \\
\vdots \\
Df_m(a)
\end{bmatrix}.
\]
Proof.
The proof is straightforward, but I am going to use the theorem we just proved.
Properties

Proof.

The proof is straightforward, but I am going to use the theorem we just proved.
Properties

Proof.
The proof is straightforward, but I am going to use the theorem we just proved.
⇒ If \( f \) is differentiable at \( a \in \text{int} \ S \), then \( f_i = p_i \circ f \), where \( p_i \) is the \( i \)'th projection map \( p_i(x) = x_i \). Since this is a linear map it is differentiable, so \( f_i \) is differentiable at \( a \).
Proof.
The proof is straightforward, but I am going to use the theorem we just proved.

⇒ If $f$ is differentiable at $a \in \text{int } S$, then $f_i = p_i \circ f$, where $p_i$ is the $i$'th projection map $p_i(x) = x_i$. Since this is a linear map it is differentiable, so $f_i$ is differentiable at $a$.

⇐ Note that

$$f = f_1 \times \cdots \times f_m \circ \Delta,$$

so it corresponds to $\Phi = Id$. 
The proof is straightforward, but I am going to use the theorem we just proved.

⇒ If \( f \) is differentiable at \( a \in \text{int } S \), then \( f_i = p_i \circ f \), where \( p_i \) is the i’th projection map \( p_i(x) = x_i \). Since this is a linear map it is differentiable, so \( f_i \) is differentiable at \( a \).

⇐ Note that

\[
f = f_1 \times \cdots \times f_m \circ \Delta,
\]

so it corresponds to \( \Phi = \text{id} \).
Gradient

**Definition**

\[ \nabla f(a)\]

and call it the gradient of \( f \) at \( a \).

So

\[ Df(a) = \nabla f(a) = \left( \frac{\partial f}{\partial x_1}(a), \ldots, \frac{\partial f}{\partial x_n}(a) \right), \]

and

\[ \nabla f(a)(x) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a)x_i. \]
The Mean Value Theorem

- Recall that for functions of one variable you know the
The Mean Value Theorem

- Recall that for functions of one variable you know the

Theorem: (The Mean Value Theorem)

Let $f : [a, b] \to \mathbb{R}$ be a function that is continuous on $[a, b]$ and differentiable on $(a, b)$.

Then there exists $c \in (a, b)$, such that

$$
\frac{f(b) - f(a)}{b - a} = f'(c).
$$
The Mean Value Theorem

Recall that for functions of one variable you know the

Theorem: (The Mean Value Theorem)

Let \( f : [a, b] \rightarrow \mathbb{R} \) be a function that is continuous on \([a, b]\) and differentiable on \((a, b)\).

Then there exists \( c \in (a, b) \), such that

\[
\frac{f(b) - f(a)}{b - a} = f'(c).
\]

You proved it as a consequence of Rolle’s Theorem, which is the particular case when

\( f(a) = f(b) \),

so that you get \( c \in (a, b) \) such that \( f'(c) = 0 \).
In this form the Theorem is obviously false for
\( f : [a, b] \rightarrow \mathbb{R}^m \), for \( m > 1 \).
In this form the Theorem is obviously false for $f : [a, b] \rightarrow \mathbb{R}^m$, for $m > 1$.

To see why, think moving particle, $f(t) =$ position at time $t$, $f'(t) =$ velocity at time $t \in (a, b)$. 
The Mean Value Theorem

- In this form the Theorem is obviously false for $f : [a, b] \rightarrow \mathbb{R}^m$, for $m > 1$.
- To see why, think moving particle, $f(t) =$ position at time $t$, $f'(t) =$ velocity at time $t \in (a, b)$.
- Then in dimension $m \geq 2$, you can end up in the same spot that you started from, without having to stop ($f'(c) = 0$) at any point.
The Mean Value Theorem

- In this form the Theorem is obviously false for $f : [a, b] \rightarrow \mathbb{R}^m$, for $m > 1$.
- To see why, think moving particle, $f(t) =$position at time $t$, $f'(t) =$velocity at time $t \in (a, b)$.
- Then in dimension $m \geq 2$, you can end up in the same spot that you started from, without having to stop ($f'(c) = 0$) at any point.
- What survives in higher dimensions is the inequality that says that if your speed is smaller than $C \geq 0$, then you cannot travel farther than $C(b - a)$. 
The Mean Value Theorem for $m \geq 2$.

Let $f[a, b] \to \mathbb{R}^m$ be a function that is continuous on $[a, b]$ and differentiable on $(a, b)$. Then

$$
\|f(b) - f(a)\| \leq \sup_{c \in (a,b)} \|Df(c)\|(b - a).
$$
The Mean Value Theorem for $m \geq 2$.

Let $f [a, b] \rightarrow \mathbb{R}^m$ be a function that is continuous on $[a, b]$ and differentiable on $(a, b)$. Then

$$
\|f(b) - f(a)\| \leq \sup_{c \in (a,b)} \|Df(c)\|(b - a).
$$

**Proof:** Let $v \in \mathbb{R}^m$ be a vector, and apply the usual ($m = 1$) Mean Value Theorem for the function

$$g_v(t) \equiv (f(t), v),$$

whose derivative at any point $c \in (a, b)$ is $g'_v(c) = (f'(c), v)$. It follows that
\[(f(b) - f(a), v) = (f'(c_v), v)(b - a)\]

for some \(c_v \in (a, b)\).
\[(f(b) - f(a), v) = (f'(c_v), v)(b - a)\]

for some \(c_v \in (a, b)\).

Now do this for the vector \(v = f(b) - f(a)\):

\[
\|f(b) - f(a)\|^2 = |(f'(c_v), f(b) - f(a))|(b - a) \leq_{C-S} \|f'(c_v)\| \|f(b) - f(a)\|(b - a)
\]
(f(b) − f(a), v) = (f′(c_v), v)(b − a)

for some c_v ∈ (a, b).

Now do this for the vector v = f(b) − f(a):

\[ \|f(b) − f(a)\|^2 = |(f′(c_v), f(b) − f(a))|(b − a) \leq C-S \]

\[ \|f′(c_v)\|\|f(b) − f(a)\|(b − a) \]

It follows that either \( \|f(b) − f(a)\| = 0 \) (in which case there is nothing to prove!), or

\[ \|f(b) − f(a)\| \leq \|f′(c_v)\|(b − a) \leq \sup_{c \in (a,b)}\|Df(c)\|(b − a). \]
\[(f(b) - f(a), v) = (f'(c_v), v)(b - a)\]

for some \(c_v \in (a, b)\).

Now do this for the vector \(v = f(b) - f(a)\):

\[
\|f(b) - f(a)\|^2 = |(f'(c_v), f(b) - f(a))|(b - a) \leq C - S \\
\|f'(c_v)\|\|f(b) - f(a)\|(b - a)
\]

It follows that either \(\|f(b) - f(a)\| = 0\) (in which case there is nothing to prove!), or

\[
\|f(b) - f(a)\| \leq \|f'(c_v)\|(b - a) \leq \sup_{c \in (a, b)} \|Df(c)\|(b - a).
\]

QED
The Mean Value Theorem (both $n, m \geq 1$)

Suppose $S \subset \mathbb{R}^n$, $f : S \to \mathbb{R}^m$ and $a, b \in \text{int } S$, such that the whole line segment:

$$[a, b] \overset{\text{def}}{=} \{ x \mid x = (1 - t)a + tb, \ 0 \leq t \leq 1 \} \subset \text{int } S.$$  

If the function $f$ is differentiable on $[a, b]$, then

$$\|f(b) - f(a)\| \leq \sup_{c \in [a, b]} \|Df(c)(b - a)\|.$$
The Mean Value Theorem (both $n, m \geq 1$)

Suppose $S \subset \mathbb{R}^n$, $f : S \to \mathbb{R}^m$ and $a, b \in \text{int } S$, such that the whole line segment:

$$[a, b] \overset{\text{def}}{=} \{ x \mid x = (1 - t)a + tb, \ 0 \leq t \leq 1 \} \subset \text{int } S.$$

If the function $f$ is differentiable on $[a, b]$, then

$$\|f(b) - f(a)\| \leq \sup_{c \in [a,b]} \|Df(c)(b - a)\|.$$

**Proof:** This is simply the preceding theorem for the function

$$g(t) = f(a + t(b - a)).$$
The Mean Value Theorem (both $n, m \geq 1$)

Suppose $S \subset \mathbb{R}^n$, $f : S \rightarrow \mathbb{R}^m$ and $a, b \in \text{int } S$, such that the whole line segment:

$$[a, b] \overset{\text{def.}}{=} \{ x | x = (1 - t)a + tb, \ 0 \leq t \leq 1 \} \subset \text{int } S.$$

If the function $f$ is differentiable on $[a, b]$, then

$$\|f(b) - f(a)\| \leq \sup_{c \in [a, b]} \|Df(c)(b - a)\|.$$

**Proof:** This is simply the preceding theorem for the function

$$g(t) = f(a + t(b - a)).$$

QED
Continuous Partial Derivatives

- We want to prove the following result:
Continuous Partial Derivatives

We want to prove the following result:

**Proposition**

Suppose $S \subset \mathbb{R}^n$, $f : S \to \mathbb{R}^m$ and $a \in \text{int } S$. If there exists $r > 0$, such that
Continuous Partial Derivatives

- We want to prove the following result:

**Proposition**

Suppose $S \subset \mathbb{R}^n$, $f : S \to \mathbb{R}^m$ and $a \in \text{int } S$. If there exists $r > 0$, such that

1. $B(a, r) \subset S$, 

- All the partial derivatives $\frac{\partial f_j}{\partial x_i}$ exist on $B(a, r)$,
- All the partial derivatives $\frac{\partial f_j}{\partial x_i}$ are continuous at $a$.

Then $f$ is differentiable at $a$.
Continuous Partial Derivatives

- We want to prove the following result:

**Proposition**

Suppose $S \subset \mathbb{R}^n$, $f : S \to \mathbb{R}^m$ and $a \in \text{int } S$. If there exists $r > 0$, such that

1. $B(a, r) \subset S$,
2. All the partial derivatives $\frac{\partial f_j}{\partial x_i}$ exist on $B(a, r)$,
We want to prove the following result:

**Proposition**

Suppose $S \subseteq \mathbb{R}^n$, $f : S \rightarrow \mathbb{R}^m$ and $a \in \text{int } S$. If there exists $r > 0$, such that

1. $B(a, r) \subseteq S$,
2. All the partial derivatives $\frac{\partial f_j}{\partial x_i}$ exist on $B(a, r)$,
3. All the partial derivatives $\frac{\partial f_j}{\partial x_i}$ are continuous at $a$. 

Then $f$ is differentiable at $a$ and
Continuous Partial Derivatives

We want to prove the following result:

**Proposition**

Suppose $S \subset \mathbb{R}^n$, $f : S \to \mathbb{R}^m$ and $a \in \text{int } S$. If there exists $r > 0$, such that

1. $B(a, r) \subset S$,

2. All the partial derivatives $\frac{\partial f_j}{\partial x_i}$ exist on $B(a, r)$,

3. All the partial derivatives $\frac{\partial f_j}{\partial x_i}$ are continuous at $a$. 


Continuous Partial Derivatives

We want to prove the following result:

**Proposition**

Suppose $S \subset \mathbb{R}^n$, $f : S \rightarrow \mathbb{R}^m$ and $a \in \text{int } S$. If there exists $r > 0$, such that

1. $B(a, r) \subset S$,
2. All the partial derivatives $\frac{\partial f_j}{\partial x_i}$ exist on $B(a, r)$,
3. All the partial derivatives $\frac{\partial f_j}{\partial x_i}$ are continuous at $a$.

Then $f$ is differentiable at $a$ and
The Total Derivative Review

Properties

The Mean Value Theorem

Sufficient Conditions for Differentiability

\[ Df(a) = \begin{bmatrix}
\frac{\partial f_1}{\partial x_1}(a) & \cdots & \frac{\partial f_1}{\partial x_i}(a) & \cdots & \frac{\partial f_1}{\partial x_n}(a) \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
\frac{\partial f_j}{\partial x_1}(a) & \cdots & \frac{\partial f_j}{\partial x_i}(a) & \cdots & \frac{\partial f_j}{\partial x_n}(a) \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
\frac{\partial f_m}{\partial x_1}(a) & \cdots & \frac{\partial f_m}{\partial x_i}(a) & \cdots & \frac{\partial f_m}{\partial x_n}(a)
\end{bmatrix}. \]

Proof: Obviously, we have to prove only differentiability! Then the (total) derivative has to be given by the above expression. Since it is enough to prove that the components are differentiable at \( a \in S \), it is enough to prove the case \( m = 1 \). (Scalar field.)
The Total Derivative

Properties

The Mean Value Theorem

Sufficient Conditions for Differentiability

\[
Df(a) = \begin{bmatrix}
\frac{\partial f_1}{\partial x_1}(a) & \cdots & \frac{\partial f_1}{\partial x_i}(a) & \cdots & \frac{\partial f_1}{\partial x_n}(a) \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
\frac{\partial f_j}{\partial x_1}(a) & \cdots & \frac{\partial f_j}{\partial x_i}(a) & \cdots & \frac{\partial f_j}{\partial x_n}(a) \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
\frac{\partial f_m}{\partial x_1}(a) & \cdots & \frac{\partial f_m}{\partial x_i}(a) & \cdots & \frac{\partial f_m}{\partial x_n}(a)
\end{bmatrix}.
\]

\textbf{Proof:} Obviously, we have to prove only differentiability!
\[ Df(a) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(a) & \ldots & \frac{\partial f_1}{\partial x_i}(a) & \ldots & \frac{\partial f_1}{\partial x_n}(a) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial f_j}{\partial x_1}(a) & \ldots & \frac{\partial f_j}{\partial x_i}(a) & \ldots & \frac{\partial f_j}{\partial x_n}(a) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(a) & \ldots & \frac{\partial f_m}{\partial x_i}(a) & \ldots & \frac{\partial f_m}{\partial x_n}(a) \end{bmatrix} \]

- **Proof**: Obviously, we have to prove only differentiability!
- Then the (total) derivative has to be given by the above expression.
The Total Derivative

Review
Properties
The Mean Value Theorem

Sufficient Conditions for Differentiability

\[ Df(a) = \begin{bmatrix}
\frac{\partial f_1}{\partial x_1}(a) & \cdots & \frac{\partial f_1}{\partial x_i}(a) & \cdots & \frac{\partial f_1}{\partial x_n}(a) \\
\vdots & & \vdots & & \vdots \\
\frac{\partial f_j}{\partial x_1}(a) & \cdots & \frac{\partial f_j}{\partial x_i}(a) & \cdots & \frac{\partial f_j}{\partial x_n}(a) \\
\vdots & & \vdots & & \vdots \\
\frac{\partial f_m}{\partial x_1}(a) & \cdots & \frac{\partial f_m}{\partial x_i}(a) & \cdots & \frac{\partial f_m}{\partial x_n}(a)
\end{bmatrix}. \]

- **Proof**: Obviously, we have to prove only differentiability!
- Then the (total) derivative has to be given by the above expression.
- Since it is enough to prove that the components are differentiable at \( a \in S \), it is enough to prove the case \( m = 1 \). (Scalar field.)
Let $v \in \mathbb{R}^n$, $\|v\| < \delta$. Then
Let $\mathbf{v} \in \mathbb{R}^n$, $\|\mathbf{v}\| < \delta$. Then

$$f(a_1 + v_1, a_2 + v_2, \ldots, a_n + v_n) = f(a_1, \ldots, a_n) +$$

$$f(a_1 + v_1, a_2 + v_2, \ldots, a_n + v_n) - f(a_1, a_2 + v_2, \ldots, a_n + v_n) +$$

$$f(a_1, a_2 + v_2, a_3 + v_3 \ldots, a_n + v_n) - f(a_1, a_2, a_3 + v_3 \ldots, a_n + v_n)$$

$$\vdots$$

$$f((a_1, a_2, \ldots, a_{n-1}, a_n + v_n) - f(a_1, a_2, \ldots, a_{n-1}, a_n)$$

Apply the Mean Value Theorem ($n = m = 1$) on each row to get $c_i \in [a_i, a_i + v_i]$ (line segment!) such that
The Total Derivative

Properties

The Mean Value Theorem

Sufficient Conditions for Differentiability

\[ f(a_1 + v_1, a_2 + v_2, \ldots, a_n + v_n) = f(a_1, \ldots, a_n) + \]
\[ \frac{\partial f}{\partial x_1}(c_1, a_2 + v_2, \ldots, a_n + v_n)v_1 + \]
\[ \frac{\partial f}{\partial x_2}(a_1, c_2, a_3 + v_3, \ldots, a_n + v_n)v_2 + \]
\[ \vdots \]
\[ \frac{\partial f}{\partial x_n}(a_1, a_2, \ldots a_{n-1}, c_n)v_n. \]

It follows that:
The Total Derivative Review Properties

The Mean Value Theorem Sufficient Conditions for Differentiability

\[ f(a_1 + v_1, \ldots, a_n + v_n) = f(a_1, \ldots, a_n) + \sum_{i=1}^{n} \left( \frac{\partial f}{\partial x_i}(a)v_i \right) \]

\[ + \|v\| \frac{1}{\|v\|} \sum_{i=1}^{n} \left( \frac{\partial f}{\partial x_i}(a_1, \ldots, c_i, \ldots, a_n + v_n) - \frac{\partial f}{\partial x_i}(a) \right)v_i. \]

\[ E(a,v) \]
The Total Derivative

Review
Properties
The Mean Value Theorem
Sufficient Conditions for Differentiability

\[ f(a_1 + v_1, \ldots, a_n + v_n) = f(a_1, \ldots, a_n) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a)v_i + \|v\| \frac{1}{\|v\|} \sum_{i=1}^{n} \left( \frac{\partial f}{\partial x_i}(a_1, \ldots, c_i, \ldots, a_n + v_n) - \frac{\partial f}{\partial x_i}(a) \right)v_i. \]

So we need to show that \( E(a, v) \to 0 \) as \( v \to 0 \).
The Total Derivative

Properties

The Mean Value Theorem

Sufficient Conditions for Differentiability

\[
f(a_1 + v_1, \ldots, a_n + v_n) = f(a_1, \ldots, a_n) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a)v_i + \left\| v \right\|_{\mathbb{R}^n} \sum_{i=1}^{n} \left( \frac{\partial f}{\partial x_i}(a_1, \ldots, c_i, \ldots, a_n + v_n) - \frac{\partial f}{\partial x_i}(a) \right)v_i.
\]

So we need to show that \( E(a, v) \to 0 \) as \( v \to 0 \).

Apply the Cauchy-Schwarz inequality to get that

\[
|E(a, v)| \leq \sqrt{\sum_{i=1}^{n} \left| \frac{\partial f}{\partial x_i}(a_1, \ldots, c_i, \ldots, a_n + v_n) - \frac{\partial f}{\partial x_i}(a) \right|^2},
\]

and note that by continuity of the partial derivatives
The Total Derivative

Review
Properties
The Mean Value Theorem
Sufficient Conditions for Differentiability

• at the point $a$: 

\[
\text{for every } \varepsilon > 0, \text{ there exists a } \delta > 0, \text{ such that } \\
x \in B(a, \delta) \implies |\frac{\partial f}{\partial x_i}(x) - \frac{\partial f}{\partial x_i}(a)| < \varepsilon \sqrt{n}, \text{ for every } 1 \leq i \leq n.
\]

But remark that if $\|v\| < \varepsilon$, then all the intermediate points $C_i = (a_1, \ldots, a_{i-1}, c_i, a_i + 1 + v_i + 1, \ldots, a_n + v_n) \in B(a, \delta)$, so that 

\[
|E(a, v)| \leq \sqrt{n} \sum_{i=1}^{n} |\frac{\partial f}{\partial x_i}(C_i) - \frac{\partial f}{\partial x_i}(a)|^2 < \varepsilon.
\]

QED
- at the point $a$:
- for every $\varepsilon > 0$, there exists a $\delta > 0$, $(\delta \leq r)$ such that $x \in B(a, \delta)$ implies that
  $$\left| \frac{\partial f}{\partial x_i}(x) - \frac{\partial f}{\partial x_i}(a) \right| < \frac{\varepsilon}{\sqrt{n}},$$
  for every $1 \leq i \leq n$, 

at the point $a$:

for every $\varepsilon > 0$, there exists a $\delta > 0$, $(\delta \leq r)$ such that $x \in B(a, \delta)$ implies that

$$|\frac{\partial f}{\partial x_i}(x) - \frac{\partial f}{\partial x_i}(a)| < \frac{\varepsilon}{\sqrt{n}},$$

for every $1 \leq i \leq n$,

But remark that if $\|v\| < \varepsilon$, then all the intermediate points

$$C_i = (a_1, \ldots, a_{i-1}, c_i, a_{i+1} + v_{i+1}, \ldots, a_n + v_n) \in B(a, \delta),$$
- at the point \( a \):
- for every \( \varepsilon > 0 \), there exists a \( \delta > 0 \), \((\delta \leq r)\) such that \( x \in B(a, \delta) \) implies that
\[
| \frac{\partial f}{\partial x_i}(x) - \frac{\partial f}{\partial x_i}(a) | < \frac{\varepsilon}{\sqrt{n}},
\]
for every \( 1 \leq i \leq n \),
- But remark that if \( \| v \| < \varepsilon \), then all the intermediate points
\[
C_i = (a_1, \ldots, a_{i-1}, c_i, a_{i+1} + v_{i+1}, \ldots, a_n + v_n) \in B(a, \delta),
\]
- so that
\[
|E(a, v)| \leq \sqrt{\sum_{i=1}^{n} | \frac{\partial f}{\partial x_i}(C_i) - \frac{\partial f}{\partial x_i}(a) |^2} < \varepsilon.
\]
at the point $a$:

for every $\varepsilon > 0$, there exists a $\delta > 0$, $(\delta \leq r)$ such that $x \in B(a, \delta)$ implies that

$$|\frac{\partial f}{\partial x_i}(x) - \frac{\partial f}{\partial x_i}(a)| < \frac{\varepsilon}{\sqrt{n}},$$

for every $1 \leq i \leq n$,

But remark that if $\|v\| < \varepsilon$, then all the intermediate points

$$C_i = (a_1, \ldots, a_{i-1}, c_i, a_{i+1} + v_{i+1}, \ldots, a_n + v_n) \in B(a, \delta),$$

so that

$$|E(a, v)| \leq \sqrt{\sum_{i=1}^{n} |\frac{\partial f}{\partial x_i}(C_i) - \frac{\partial f}{\partial x_i}(a)|^2} < \varepsilon.$$