# On the multi-dimensional chain-rule

Last version: `https://tales.mbivert.com/on-the-multi-dimensional-chain-rule.pdf`

M. Bivert

December 27, 2025

## Contents

**1 Introduction** **1**

**2 Preliminaries** **1**
2.1 Vector spaces, Linear maps . . . . . 1
2.2 Norms . . . . . 2
2.3 Continuity, limits . . . . . 3

**3 Multi-dimensional differentiation** **6**

**4 Chain-rule** **8**

# 1 Introduction

The goal here is to provide an in-depth version of a standard proof of the multi-dimensional differentiation chain-rule. A (very) generous amount of preliminaries is provided, in part to exercise, but also to better understand how various results and ideas articulate. You may want to skip to the last section for the chain-rule itself.

A good textbook reference is [**?**].

# 2 Preliminaries

## 2.1 Vector spaces, Linear maps

I'll assume familiarity with the definition of a *field*, of a *vector space*, of a *basis* and of the *dimension* of a vector space. But let's review *how linear maps* are defined, given their relevancy in the context of multi-dimensional differentiation: the derivative *is* the best linear map approximating a function in the vicinity of a point.

**Definition 1** (linear transformation)**.** A *linear transformation*, between two vector spaces over the same field is a vector space homomorphism: it's a transformation preserving the vector space structure.

More precisely, let $\phi : V \to W$ be a linear map between two vector spaces $(V, +_V, .V)$ and $(W, +_W, .W)$ over the same field $\mathbb{K}$. Then, $(\forall (u, v) \in V^2)$ and $(\forall \alpha \in \mathbb{K})$:

**Additivity** addition is preserved:

$$\boxed{\phi(u +_V v) = \phi(u) +_W \phi(v)}$$

**Homogeneity** Scalar multiplication is preserved:

$$\boxed{\phi(\alpha ._V u) = \alpha ._W \phi(u)}$$

**Remark 1.** To denote a linear function $\phi$ between two vector spaces $V$ and $W$ over the same field, we'll use the following notation:

$$\boxed{\phi : V \xrightarrow{\sim} W}$$

**Remark 2.** We also talk of *linear application*, *linear map* and *linear operator*. While in general those terms can have different meaning, I'll use them interchangeably here.

**Remark 3.** The set of linear maps between two vector spaces $E$ and $F$ is often denoted $\mathcal{L}(E, F)$, or $\mathrm{Hom}(E, F)$. The notation can be shortened to $\mathcal{L}(E)$ or $\mathrm{End}(E)$ when $E = F$.

We'll make frequent use of the following elementary result later on; I'll omit the proof to save space:

**Lemma 1.** Let $\phi : U \xrightarrow{\sim} V$ be a linear map between two vector spaces $(U, +_U, .U)$ and $(V, +_V, .V)$ over the same field $(\mathbb{K}, +, \times)$. Then, $\phi$ must preserves the additive neutral elements:

$$\boxed{\phi(0_U) = 0_V}$$

## 2.2 Norms

While the one-dimensional derivative can escape explicit mention of *norms*, they becomes mandatory in higher-dimensions. A norm is defined on a vector space, and there are subtleties tied to some qualities of the underlying field which can affect the definition of a norm.

For simplicity, we'll only take into consideration a non-problematic case that will be of interest for the multi-dimensional chain-rule, namely $\mathbb{R}$ or $\mathbb{C}$.

**Definition 2** (norm)**.** Let $V$ be a vector space over a field $\mathbb{K}$ being either $\mathbb{R}$ or $\mathbb{C}$. A *norm* on $V$ is a map $\|.\| : V \to \mathbb{R}$ satisfying:

**(i) The triangle inequality / sub-additivity** :

$$\boxed{(\forall (u, v) \in V^2),\ \|u + v\| \leq \|u\| + \|v\|}$$

**(ii) Absolute homogeneity**

$$\boxed{(\forall (u, k) \in V \times \mathbb{K}),\ \|ku\| = |k|\,\|u\|}$$

**(iii) Positive-definiteness**

$$\boxed{(\forall u \in V),\ \|u\| = 0 \Rightarrow u = 0_V}$$

**Definition 3** (Normed vector space)**.** A vector space $(V, +, .)$ equipped with a norm $\|.\|$ makes a *normed vector space*. It can be represented as a couple $(V, \|.\|)$.

**Remark 4.** It's customary, and I'll be often guilty of doing it, to use the same $\|.\|$ symbol for distinct norms. In theory, one can always infer the correct norm by looking at which set its argument is from.

For example, if we say that $(U, \|.\|)$ and $(V, \|.\|)$ are two normed vector spaces, what we usually mean is that we have $(U, \|.\|_U)$ and $(V, \|.\|_V)$ where $\|.\|_U$ and $\|.\|_V$ are generally distinct norms. We have the same problem with additions/scalar multiplication on different vector spaces, which are often denoted by the same symbols.

**Lemma 2.** Let $(V, \|.\|)$ be a normed vector space. Then

$$\boxed{(\forall u \in V),\ \|u\| \geq 0}$$

*Proof.* Indeed, let $u \in V$. Then:

$$\begin{aligned}
0 = &\ \|0_V\| && \text{(positive-definiteness)}\\
= &\ \|u + (-u)\| && \text{(additive inverses)}\\
\leq &\ \|u\| + \|-u\| && \text{(triangle inequality)}\\
\leq &\ \|u\| + |-1|\,\|u\| && \text{(absolute homogeneity)}\\
\leq &\ 2\,\|u\|
\end{aligned}$$

$$\Leftrightarrow \|u\| \geq 0$$

□

The following result will also be of use.

**Theorem 1** (Inverse triangular inequality)**.** . Let $(V, \|.\|)$ be a normed space. Then:

$$\boxed{(\forall (u,v) \in V^2),\ |\|u\| - \|v\|| \leq \|u - v\|}$$

*Proof.* Indeed, let $(u, v) \in V^2$. Then, using the *triangular inequality* and some clever "+0" trick:

$$\begin{aligned} \|u\| = \quad \|u - v + v\| \leq \quad \|u - v\| + \|v\| \quad &\Leftrightarrow \quad \|u\| - \|v\| \leq \quad \|u - v\| \\ \|v\| = \quad \|v - u + u\| \leq \quad \|v - u\| + \|u\| \quad &\Leftrightarrow \quad \|v\| - \|u\| \leq \quad \|v - u\| \end{aligned}$$

But, by *absolute homogeneity*:

$$\|v - u\| = \|(-1)(u - v)\| = |-1|\,\|u - v\| = \|u - v\|$$

And so we really have the following system of equations:

$$\begin{cases} \|u\| - \|v\| & \leq \|u - v\| \\ -(\|u\| - \|v\|) & \leq \|u - v\| \end{cases}$$

But this is exactly what we wanted to prove, considering the definition of the absolute value:

$$|\|u\| - \|v\|| = \begin{cases} \|u\| - \|v\| & \text{when } \|u\| - \|v\| \geq 0 \\ -(\|u\| - \|v\|) = \|v\| - \|u\| & \text{otherwise} \end{cases}$$

□

## 2.3 Continuity, limits

We now want to establish an important result, allowing to "swap" limits and function calls, as long as the function is continuous, and as long as the limit exist. As we'll see later, norms and linear functions in particular are continuous, and so we can swap limits with those.

For clarity, let's first recall how (multi-dimensional) limits are defined, in terms of $\epsilon - \delta$, and how multi-dimensional continuity is defined in terms of limit.

**Definition 4.** Let $f : U \to V$ where $U$ and $V$ are two normed vector spaces. We say that $f$ admits $L$ as a limit at $a \in X$ when:

$$(\forall \epsilon > 0),\ (\exists \delta > 0),\ (\forall u \in U), \quad \|u - a\| < \delta \quad \Rightarrow \quad \|f(u) - L\| < \epsilon$$

**Remark 5.** As we do in the single-dimensional case, we note:

$$\boxed{\lim_{u \to a} f(u) = L}$$

**Definition 5.** Let $f : U \to V$ be a function between two normed vector spaces $U$ and $V$.

1) We say that $f$ *is continuous at* $a \in X$ if:

$$\boxed{\lim_{x \to a} f(x) = f(a)}$$

i.e.:

$$(\forall \epsilon > 0),\ (\exists \delta > 0),\ (\forall u \in U), \quad \|u - a\| < \delta \quad \Rightarrow \quad \|f(u) - f(a)\| < \epsilon$$

2) If $f$ is continuous at every point of $X \subseteq U$, we say that $f$ *is continuous on* $X$.

**Theorem 2.** Let $f : V \to W$ and $g : U \to V$, where $U, V, W$ are three normed vector spaces. Furthermore, let $L \in V$ be the limit of $g$ as its input goes to $a$, and $f$ be continuous at least on $L$; then

$$\boxed{\lim_{u \to a} \Big( f(g(u)) \Big) = f\left( \lim_{u \to a} g(u) \right)}$$

*Proof.* We can rephrase the limit condition on $g$ as:

$$(\forall \epsilon' > 0),\ (\exists \delta_{\epsilon'} > 0),\ (\forall u \in U), \quad \|u - a\| < \delta_{\epsilon'} \quad \Rightarrow \quad \|g(u) - L\| < \epsilon'$$

And the continuity of $f$ at $L$ as:

$$(\forall \epsilon > 0),\ (\exists \delta' > 0),\ (\forall v \in V), \quad \|v - L\| < \delta' \quad \Rightarrow \quad \|f(v) - f(L)\| < \epsilon$$

So, if we're given an $\epsilon > 0$, the continuity condition of $f$ at $L$ yields:

$$(\exists \delta' > 0),\ (\forall v \in V),\ \|v - L\| < \delta' \quad \Rightarrow \quad \|f(v) - f(L)\| < \epsilon$$

As this is true $(\forall v \in V)$, so in particular, it's true for any $g(u)$, where $u \in U$:

$$(\exists \delta' > 0),\ (\forall u \in U),\ \|g(u) - L\| < \delta' \quad \Rightarrow \quad \|f(g(u)) - f(L)\| < \epsilon$$

Now observe that the limit condition on $g$ is true *for any* $\epsilon'$: in particular, it must be true for $\epsilon' = \delta'$:

$$(\exists \delta_{\delta'} > 0),\ (\forall u \in U),\ \|u - a\| < \delta_{\delta'} \quad \Rightarrow \quad \|g(u) - L\| < \delta'$$

What we have so far can be rewritten as:

$$(\forall \epsilon > 0),\ (\exists \delta' > 0),\ (\exists \delta_{\delta'} > 0),\ (\forall u \in U),\ \begin{cases} \|g(u) - L\| < \delta' \quad \Rightarrow \quad \|f(g(u)) - f(L)\| < \epsilon \\ \|u - a\| < \delta_{\delta'} \quad \Rightarrow \quad \|g(u) - L\| < \delta' \end{cases}$$

Or combining the two implications:

$$(\forall \epsilon > 0),\ (\exists \delta' > 0),\ (\exists \delta_{\delta'} > 0),\ (\forall u \in U),\ \|u - a\| < \delta_{\delta'} \Rightarrow \|g(u) - L\| < \delta' \Rightarrow \|f(g(u)) - f(L)\| < \epsilon$$

Which can be simplified to:

$$(\forall \epsilon > 0),\ (\exists \delta > 0),\ (\forall u \in U), \quad \|u - a\| < \delta \quad \Rightarrow \quad \|f(g(u)) - f(L)\| < \epsilon$$

Which is exactly the $\epsilon - \delta$ formulation of our result.

□

We now prove our two important special cases.

**Theorem 3.** Let $(V, \|.\|)$ be a normed vector space. Then $\|.\| : V \to \mathbb{R}$ is continuous, w.r.t. the topology induced by itself on $V$, and w.r.t. the standard topology on $\mathbb{R}$.

**Remark 6.** Observe that our present definition of continuity depends on two norms, one on the domain, one on the codomain of the functions being considered. A more general way to talk about continuity is to equip both sets with a different kind of mathematical structure: *topologies*. There are standard constructions allowing the construction of topologies from *metrics*, which themselves can easily be constructed from norms.

The theorem is simply being precise in this regard.

*Proof.* Let's rephrase what we want to prove in terms of $\epsilon - \delta$ (note that the norm on $\mathbb{R}$ is the absolute value)

$$(\forall a \in V),\ (\forall \epsilon > 0),\ (\exists \delta > 0),\ (\forall v \in V), \quad \|v - a\| < \delta \quad \Rightarrow \quad |\|v\| - \|a\|| < \epsilon$$

But recall the *inverse triangular inequality*:

$$(\forall (u, v) \in V^2),\ |\|u\| - \|v\|| \leq \|u - v\|$$

So in particular, if we pick an $a \in V$, then an arbitrary $\epsilon > 0$, then take $\delta = \epsilon$, we have:

$$(\forall v \in V),\ |\|v\| - \|a\|| \leq \|v - a\| < \delta = \epsilon$$

Which concludes the proof: as soon as we have $\|v - a\| < \epsilon$, $|\|v\| - \|a\|| < \epsilon$ naturally follows.

□

We're now ready to start tackling the case of the continuity of linear maps. Let's first recall the notion of the supremum of a subset of the real

**Definition 6** (supremum)**.** Let $E$ be a subset of the extended real line:

$$\overline{\mathbb{R}} := [-\infty, +\infty] = \mathbb{R} \cup \{-\infty, +\infty\}$$

Then the *supremum* of $E$ is the least upper bound of $E$ in $\overline{\mathbb{R}}$. It is denoted $\sup E$.

**Remark 7.** The notion of supremum can be defined in a more general context, in which case, it might not exists. But it's guaranteed to exists here.

This allows us to define something that is commonly referred to as the operator norm, which as the name indicates, *is* a norm, on the vector space made by equipping the set of all linear applications between two given vector spaces by the usual pointwise operations. However, we won't need this vector space structure, nor to establish that this is actually a norm, so I won't go into further details.

Saying it otherwise, for the purpose of those notes, we're just defining a symbol associated to a linear application.

**Definition 7** (operator norm)**.** Let $(U, \|.\|_U)$ and $(V, \|.\|_V)$ be two normed vector spaces. Let $\phi : U \xrightarrow{\sim} V$ be a linear map. The following quantity is called the *operator norm* of $\phi$:

$$\boxed{\overline{\mathbb{R}} \ni \|\phi\| := \begin{cases} \sup\left\{ \frac{\|\phi(u)\|_V}{\|u\|_U} \;\middle|\; u \in U \backslash \{0_U\} \right\} & \text{if } U \neq \{0_U\}; \\ 0 & \text{otherwise.} \end{cases}}$$

From there, we can derive the notion of *bounded operator*:

**Definition 8** (bounded operator)**.** Let $(U, \|.\|_U)$ and $(V, \|.\|_V)$ be two normed vector spaces. Let $\phi : U \xrightarrow{\sim} V$ be a linear map (also called a linear operator). We say that:

$$\boxed{\phi \text{ is bounded} \quad :\Leftrightarrow \quad \|\phi\| < +\infty}$$

One last step before moving on to the continuity of linear maps:

**Theorem 4.** Let $(U, \|.\|_U)$ and $(V, \|.\|_V)$ be two normed vector spaces over a field $\mathbb{K}$, and where $U$ is finite-dimensional. Let $\phi : U \xrightarrow{\sim} V$ be a linear map. Then $\boxed{\phi \text{ is bounded}}$.

*Proof.* First, if $U = \{0_U\}$, then $\|\phi\| = 0 < +\infty$ and we're done.

So, consider the case where $U \neq \{0_U\}$. Let $n = \dim U$, take a basis $\{e_i\}_{i \in \{1,2,\ldots,n\}}$ for $U$. Let then $u \in U \backslash \{0_U\}$, and expand it in our basis:

$$u = \sum_{i=1}^{n} u^i e_i; \qquad \text{where } (\forall i \in \{1, 2, \ldots, n\}),\ u^i \in \mathbb{K}$$

Then:

$$\begin{aligned} \frac{\|\phi(u)\|}{\|u\|} &= \frac{1}{\|u\|} \left\| \phi(\sum_{i=1}^{n} u^i e_i) \right\| && \text{(expanding } u\text{)} \\ &= \frac{1}{\|u\|} \left\| \sum_{i=1}^{n} u^i \phi(e_i) \right\| && \text{(linearity of } \phi\text{)} \\ &\leq \underbrace{\frac{1}{\|u\|} \sum_{i=1}^{n} \left\| u^i \phi(e_i) \right\|}_{=:\alpha} && \text{(triangular inequality)} \end{aligned}$$

But $u \neq 0$ implies $\|u\| \neq 0$, and the numerator is a finite sum of real numbers, so $\alpha \in \mathbb{R}$. Furthermore, as $u$ was arbitrary, this implies by definition of the operator norm that $\|\phi\| \leq \alpha \in \mathbb{R}$, i.e. $\|\phi\| < +\infty$. □

Finally:

**Theorem 5.** Let $(U, \|.\|_U)$ and $(V, \|.\|_V)$ be two normed vector spaces. Then any bounded linear map $\phi : U \xrightarrow{\sim} V$ is continuous, w.r.t. the topologies induced by the norms on each space.

*Proof.* For clarity, let's rephrase what we want to prove in terms of $\epsilon - \delta$:

$$(\forall \epsilon > 0),\ (\exists \delta > 0),\ (\forall u \in U), \quad \|u - a\| < \delta \quad \Rightarrow \quad \|\phi(u) - \phi(a)\| < \epsilon$$

Let then $\epsilon > 0$.

First, if $U = \{0_U\}$, then pick say $\delta := \epsilon$.

Then we can only pick $u = a = 0_U$, as there are no other vectors in the space. By positive-definiteness of the norm on $U$, we always have $\|u - a\| = \|0_U\| = 0 < \delta$. Furthermore, $\phi(u) = \phi(a) = 0_V$, as $\phi$ is linear. And by positive-definiteness of the norm on $V$, $\|\phi(u) - \phi(a)\| = \|0_V\| = 0 < \epsilon$

Second, if $U \neq \{0_U\}$, i.e. $\|\phi\| \neq 0$. But $\phi$ is bounded, so we have $\|\phi\| < +\infty$, and thus we can define.

$$\delta := \frac{\epsilon}{\|\phi\|}$$

Then, if $u = a$, we're actually in the same case as before: $\|u - a\| = \|0_U\| = 0 < \delta$ always, by positive-definiteness of the norm on $U$. And $\|\phi(u) - \phi(a)\| = \|\phi(u - a)\| = \|\phi(0_U)\| = \|0_V\| = 0 < \epsilon$, by linearity of $\phi$, and because linear applications "send zero to zero", and by positive-definiteness of the norm on $V$.

Otherwise, $(\forall u \in U \backslash \{a\})$

$$\begin{aligned}
\|u - a\| < \delta \Rightarrow \quad & \|u - a\| < \frac{\epsilon}{\|\phi\|} && \text{(definition of } \delta\text{)} \\
\Rightarrow \quad & \|\phi\| < \frac{\epsilon}{\|u - a\|} && (u \neq a \Rightarrow \|u - a\| \neq 0\text{, positive-definiteness)} \\
\Rightarrow \quad & \frac{\|\phi(u - a)\|}{\|u - a\|} \leq \|\phi\| < \frac{\epsilon}{\|u - a\|} && \text{(definition of } \|\phi\|\text{)} \\
\Rightarrow \quad & \|\phi(u - a)\| < \epsilon && (\|u - a\| \geq 0) \\
\Rightarrow \quad & \|\phi(u) - \phi(a)\| < \epsilon && (\phi \text{ is linear})
\end{aligned}$$

Which is what we wanted to prove.

□

**Remark 8.** We proved a slightly more general result that what is needed in the context of those notes. But it's not much more complicated, perhaps even simpler than having to rigorously deal with explicit matrices as is often the case, and will generalize better.

Hopefully, I didn't miss anything.

# 3 Multi-dimensional differentiation

Let's now recall how the derivative is defined in a multi-dimensional setting. This can be presented in different ways, for example via the Jacobian. But we can stick to the "best local linear approximation" type of definition, as it feels more explicit, avoids the need for partial and directional derivatives, and generalizes better[1].

I'll be considering the real case here, but this should work equally well on the complex field.

**Definition 9** (Total derivative, vector-valued real function)**.** Take a non-empty open $U \subseteq \mathbb{R}^m$, and let $f : U \to \mathbb{R}^n$ be a vector-valued function, where $\mathbb{R}^m$ and $\mathbb{R}^n$ are considered as normed vector spaces over $\mathbb{R}$.

1) If there exists a linear function $\lambda_{f,p} : \mathbb{R}^m \xrightarrow{\sim} \mathbb{R}^n$ such that, for a point $p \in \mathbb{R}^m$ we have:

$$\boxed{\lim_{\nu \to 0} \frac{\|f(p + \nu) - f(p) - \lambda_{f,p}(\nu)\|_{\mathbb{R}^n}}{\|\nu\|_{\mathbb{R}^m}} = 0}$$

[1] For completeness (pun intended), the usual generalization is the Fréchet derivative on normed spaces.

Then we say that $f$ is differentiable at $p$; $\lambda_{f,p}$ is the *total derivative around* $p$: it's the best linear-approximation of $f$ in a neighborhood of $p$.

2) If $f$ is differentiable at every point of $U$, we say that $f$ is differentiable in $U$; the following $df$ function is then called the *total derivative of* $f$ *in* $U$:

$$\boxed{df : \begin{pmatrix} \mathbb{R}^m & \to & (\mathbb{R}^m & \xrightarrow{\sim} & \mathbb{R}^n) \\ p & \mapsto & (q & \mapsto & \lambda_{f,p}(q)) \end{pmatrix}}$$

**Remark 9.** The *total derivative* is also called the *total differential*, or sometimes the *differential*, of $f$.

**Remark 10.** Let me emphasize that by contrast with the scalar-valued partial/directional derivatives, or what is commonly refered to as "derivative" in the one-dimensional case, the total derivative $df$ is *function*-valued: for a point in the domain, it returns a linear function, i.e. the best linear approximation of $f$ at that point:

$$\boxed{df : (U \subseteq \mathbb{R}^m) \to (\mathbb{R}^m \xrightarrow{\sim} \mathbb{R}^n)}$$

**Theorem 6** (Unicity of the derivative)**.** Let $f : (U \subseteq \mathbb{R}^m) \to \mathbb{R}^n$ be a function, differentiable at $p \in U$, for $U$ open. Then the derivate of $f$ at $p$, $\lambda_{f,p}$, is unique.

*Proof.* Indeed, suppose we have two such derivatives, say, $\lambda_{f,p}$ and $\sigma_{f,p}$, that we'll rename $\lambda$ and $\sigma$. If we want to prove that those two are equals, we need to prove that they are point-wise equal[2], i.e.:

$$\lambda = \sigma \quad :\Leftrightarrow \quad (\forall q \in \mathbb{R}^m),\ \lambda(q) = \sigma(q)$$

First, observe that because both are linear functions by definition, we must have:

$$\lambda(0) = \sigma(0) = 0$$

Then, pick $q \in \mathbb{R}^m \backslash \{0\}$:

$$\begin{aligned}
\|\lambda(q) - \sigma(q)\| = &\ \Big\| -\Big(f(p+q) - f(p) - \lambda(q)\Big) + \Big(f(p+q) - f(p) - \sigma(q)\Big) \Big\| && \text{(clever +0 trick)} \\
\leq &\ \Big\| -\Big(f(p+q) - f(p) - \lambda(q)\Big) \Big\| + \|f(p+q) - f(p) - \sigma(q)\| && \text{(triangular inequality)} \\
\leq &\ |-1| \, \|f(p+q) - f(p) - \lambda(q)\| + \|f(p+q) - f(p) - \sigma(q)\| && \text{(absolute homogeneity)} \\
\leq &\ \|f(p+q) - f(p) - \lambda(q)\| + \|f(p+q) - f(p) - \sigma(q)\|
\end{aligned}$$

And thus in particular, by definition of $\lambda$, $\sigma$ as differentials of $f$ at $p$, and because $q \neq 0 \Rightarrow \|q\| \neq 0$ (by positive-definiteness of the norm):

$$\begin{aligned}
\lim_{q \to 0} \left( \frac{\|\lambda(q) - \sigma(q)\|}{\|q\|} \right) \leq &\ \underbrace{\lim_{p \to 0} \left( \frac{\|f(p+q) - f(p) - \lambda(q)\|}{\|q\|} \right)}_{=0} + \underbrace{\lim_{q \to 0} \left( \frac{\|f(p+q) - f(p) - \sigma(q)\|}{\|q\|} \right)}_{=0} \\
\leq &\ 0
\end{aligned}$$

But then consider $q = kq'$ for $k \in \mathbb{R} \backslash \{0\}$ and $q' \in \mathbb{R}^m \backslash \{0\}$. Then:

$$\begin{aligned}
\lim_{q \to 0} \left( \frac{\|\lambda(q) - \sigma(q)\|}{\|q\|} \right) = &\ \lim_{k \to 0} \left( \frac{\|\lambda(kq') - \sigma(kq')\|}{\|kq'\|} \right) \\
= &\ \lim_{k \to 0} \left( \frac{\|k\lambda(q') - k\sigma(q')\|}{\|kq'\|} \right) && \text{(linearity)} \\
= &\ \lim_{k \to 0} \left( \frac{|k| \, \|\lambda(q') - \sigma(q')\|}{|k| \, \|q'\|} \right) && \text{(absolute homogeneity)} \\
= &\ \lim_{k \to 0} \left( \frac{\|\lambda(q') - \sigma(q')\|}{\|q'\|} \right) \\
= &\ \frac{\|\lambda(q') - \sigma(q')\|}{\|q'\|}
\end{aligned}$$

[2] This is what is usually meant when talking about function equality.

So by combining our two previous results, we've proved that for $q' \in \mathbb{R}^m \backslash \{0\}$,

$$\frac{\|\lambda(q') - \sigma(q')\|}{\|q'\|} \leq 0$$

But the norm is positive, so:

$$0 \leq \frac{\|\lambda(q') - \sigma(q')\|}{\|q'\|} \leq 0 \quad \Leftrightarrow \quad \frac{\|\lambda(q') - \sigma(q')\|}{\|q'\|} = 0$$

Now, $\|q'\|$ is finite, and thus we have:

$$\|\lambda(q') - \sigma(q')\| = 0$$

Finally, by positive definiteness, this is equivalent to saying:

$$\lambda(q') - \sigma(q') = 0 \quad \Leftrightarrow \quad \lambda(q') = \sigma(q')$$

And so the two differentials are indeed (pointwise) equal on all of $\mathbb{R}^m$.

□

The following little lemma offers a convenient alternative characterization of the total differential.

**Lemma 3** (Alternative characterization for the total derivative)**.** Take a non-empty open $U \subseteq \mathbb{R}^m$, and let $f : U \to \mathbb{R}^n$ be a vector-valued function, where $\mathbb{R}^m$ and $\mathbb{R}^n$ are considered as normed vector spaces over $\mathbb{R}$.

Then $f$ is *totally differentiable* function at $p \in \mathbb{R}^m$ iff:

$$\boxed{\left(\exists \mu \in \mathcal{C}^0(U \to \mathbb{R}^n), \lim_{\epsilon \to 0} \frac{\|\mu(\epsilon)\|}{\|\epsilon\|} = 0\right), \ f(p+\epsilon) - f(p) = \lambda_{f,p}(\epsilon) + \mu(\epsilon)}$$

*Proof.* We have:

$$\begin{aligned}
& \left(\exists \mu \in \mathcal{C}^0(U \to \mathbb{R}^n), \lim_{\epsilon \to 0} \frac{\|\mu(\epsilon)\|}{\|\epsilon\|} = 0\right), \ f(p+\epsilon) - f(p) = \lambda_{f,p}(\epsilon) + \mu(\epsilon) \\
\Leftrightarrow \quad & \left(\exists \mu \in \mathcal{C}^0(U \to \mathbb{R}^n), \lim_{\epsilon \to 0} \frac{\|\mu(\epsilon)\|}{\|\epsilon\|} = 0\right), \ \mu(\epsilon) = f(p+\epsilon) - f(p) - \lambda_{f,p}(\epsilon) \\
\Leftrightarrow \quad & \left(\exists \mu \in \mathcal{C}^0(U \to \mathbb{R}^n)\right), 0 = \lim_{\epsilon \to 0} \frac{\|\mu(\epsilon)\|}{\|\epsilon\|} = \lim_{\epsilon \to 0} \frac{\|f(p+\epsilon) - f(p) - \lambda_{f,p}(\epsilon)\|}{\|\epsilon\|} \\
\Leftrightarrow \quad & f \text{ differentiable at } p
\end{aligned}$$

□

# 4 Chain-rule

**Theorem 7** (Multi-dimensional chain rule)**.** Let $(U \subseteq \mathbb{R}^m) \xrightarrow{g} (V \subseteq \mathbb{R}^d) \xrightarrow{f} \mathbb{R}^n$.

1) Let $g$ be differentiable at $p \in U$, and $f$ be differentiable at $g(p) \in V$. Then $f \circ g$ is differentiable at $p$ and:

$$\boxed{\lambda_{f \circ g, p} = \lambda_{f, g(p)} \circ \lambda_{g,p}}$$

2) Of course, this naturally generalizes to open subsets: if $g$ is differentiable on an open $U \subseteq \mathbb{R}^m$ and $f$ is differentiable on an open $V \subseteq \mathbb{R}^d$, where $g(U) \subseteq V$, then:

$$\boxed{d(f \circ g)(p) = df(g(p)) \circ dg(p)}$$

*Proof.* The goal is to prove that $f \circ g$ is differentiable at $p$, and that the best linear approximation at $p$ is given by the composition of two linear functions: $df(g(p)) \circ dg(p)$.

By the definition of the differentiation at a point, this means we need to prove:

$$\lim_{\epsilon \to 0} \frac{\left\| (f \circ g)(p+\epsilon) - (f \circ g)(p) - \lambda_{(f \circ g),p}(\epsilon) \right\|}{\|\epsilon\|} = 0$$

Because of the norms positivity, this means we only need to prove that the limit is bounded above by zero, i.e.

$$\lim_{\epsilon \to 0} \frac{\left\| (f \circ g)(p+\epsilon) - (f \circ g)(p) - \lambda_{(f \circ g),p}(\epsilon) \right\|}{\|\epsilon\|} \leq 0$$

Or, by replacing $\lambda_{(f \circ g),p}$ with the expected result:

$$\lim_{\epsilon \to 0} \frac{\left\| (f \circ g)(p+\epsilon) - (f \circ g)(p) - ((\lambda_{f,g(p)} \circ \lambda_{g,p})(\epsilon) \right\|}{\|\epsilon\|} \leq 0$$

Now, by our previous characterization, the differentiability conditions on $f$ and $g$ at respectively $g(p)$ and $p$ can be expressed as:

$$\left( \exists \mu \in \mathcal{C}^0(U \to \mathbb{R}^d), \lim_{\epsilon' \to 0} \frac{\|\mu(\epsilon')\|}{\|\epsilon'\|} = 0 \right), \ f(g(p)+\epsilon') - f(g(p)) = \lambda_{f,g(p)}(\epsilon') + \mu(\epsilon')$$

$$\left( \exists \eta \in \mathcal{C}^0(V \to \mathbb{R}^m), \lim_{\epsilon \to 0} \frac{\|\eta(\epsilon)\|}{\|\epsilon\|} = 0 \right), \ g(p+\epsilon) - g(p) = \lambda_{g,p}(\epsilon) + \eta(\epsilon)$$

Pick $\epsilon \in U \subseteq \mathbb{R}^m \backslash \{0\}$, and let $\epsilon' := g(p+\epsilon) - g(p)$. Observe that then, as $\epsilon$ goes to zero, so is $\epsilon'$ We can now rewrite the numerator of limit involved the previously mentioned expected result as, modulo the norm:

$$\begin{aligned}
(f \circ g)(p+\epsilon) - (f \circ g)(p) - ((\lambda_{f,g(p)} \circ \lambda_{g,p})(\epsilon) &= f(g(p+\epsilon)) - f(g(p)) - \lambda_{f,g(p)}(\lambda_{g,p}(\epsilon)) \\
&= f(g(p)+\epsilon') - f(g(p)) - \lambda_{f,g(p)}(\lambda_{g,p}(\epsilon)) \\
&= \lambda_{f,g(p)}(\epsilon') + \mu(\epsilon') - \lambda_{f,g(p)}(\lambda_{g,p}(\epsilon)) \\
&= \lambda_{f,g(p)}\Big(\epsilon' - \lambda_{g,p}(\epsilon)\Big) + \mu(\epsilon') \text{ (linearity)} \\
&= \lambda_{f,g(p)}(\eta(\epsilon)) + \mu(\epsilon')
\end{aligned}$$

And so:

$$\frac{\left\| (f \circ g)(p+\epsilon) - (f \circ g)(p) - ((\lambda_{f,g(p)} \circ \lambda_{g,p})(\epsilon) \right\|}{\|\epsilon\|} = \frac{\left\| \lambda_{f,g(p)}(\eta(\epsilon)) + \mu(\epsilon') \right\|}{\|\epsilon\|}$$

By the triangular inequality:

$$\begin{aligned}
\frac{\left\| (f \circ g)(p+\epsilon) - (f \circ g)(p) - ((\lambda_{f,g(p)} \circ \lambda_{g,p})(\epsilon) \right\|}{\|\epsilon\|} &\leq \frac{\left\| \lambda_{f,g(p)}(\eta(\epsilon)) \right\| + \|\mu(\epsilon')\|}{\|\epsilon\|} \\
&\leq \frac{\left\| \lambda_{f,g(p)}(\eta(\epsilon)) \right\|}{\|\epsilon\|} + \frac{\|\mu(\epsilon')\|}{\|\epsilon\|}
\end{aligned}$$

To establish our expected result, we need to study this as $\epsilon$ goes to 0: if we can prove that the limits of each term of the right hand side of the previous inequality goes to zero, then we can apply the sum rule for the limits, and conclude.

Consider first the second term of the right and side of the previous inequality: as we've already observed, by definition of $\epsilon'$ and by the differentiability of $f$ at $g(p)$:

$$\lim_{\epsilon \to 0} \frac{\|\mu(\epsilon')\|}{\|\epsilon\|} = \lim_{\epsilon' \to 0} \frac{\|\mu(\epsilon')\|}{\|\epsilon'\|} = 0$$

Consider now the first term. We're interested in studying:

$$\lim_{\epsilon \to 0} \frac{\left\| \lambda_{f,g(p)}(\eta(\epsilon)) \right\|}{\|\epsilon\|}$$

By absolute homogeneity of the norm, and the fact that $|\|\epsilon\|| = \|\epsilon\|$ as the norm is positive:

$$\frac{\left\|\lambda_{f,g(p)}(\eta(\epsilon))\right\|}{\|\epsilon\|} = \left|\frac{1}{\|\epsilon\|}\right| \left\|\lambda_{f,g(p)}(\eta(\epsilon))\right\| = \left\|\frac{1}{\|\epsilon\|}\lambda_{f,g(p)}(\eta(\epsilon))\right\| = \left\|\lambda_{f,g(p)}\left(\frac{1}{\|\epsilon\|}\eta(\epsilon)\right)\right\|$$

Where the last step relies on the linearity of the $\lambda$ functions. Now, we know that:

$$\lim_{\epsilon\to 0}\frac{\|\eta(\epsilon)\|}{\|\epsilon\|} = 0$$

Which means that $\|\eta(\epsilon)\|$ converges to zero as $\epsilon$ goes to zero. But because the norm is continuous, this implies that:

$$\left\|\lim_{\epsilon\to 0}\frac{\eta(\epsilon)}{\|\epsilon\|}\right\| = 0$$

Which implies, by positive-definiteness of the norm:

$$\lim_{\epsilon\to 0}\left(\frac{1}{\|\epsilon\|}\eta(\epsilon)\right) = 0$$

But then, because both norms and linear functions are continuous, we can "swap the limits":

$$\begin{aligned}
\lim_{\epsilon\to 0}\frac{\left\|\lambda_{f,g(p)}(\eta(\epsilon))\right\|}{\|\epsilon\|} &= \lim_{\epsilon\to 0}\left\|\lambda_{f,g(p)}\left(\frac{1}{\|\epsilon\|}\eta(\epsilon)\right)\right\| \\
&= \left\|\lim_{\epsilon\to 0}\left(\lambda_{f,g(p)}\left(\frac{1}{\|\epsilon\|}\eta(\epsilon)\right)\right)\right\| \\
&= \left\|\lambda_{f,g(p)}\left(\lim_{\epsilon\to 0}\left(\frac{1}{\|\epsilon\|}\eta(\epsilon)\right)\right)\right\| \\
&= \left\|\lambda_{f,g(p)}(0)\right\| \\
&= \|0\| \text{ (linear maps sends 0 to 0)} \\
&= 0 \text{ (norm's positiveness)}
\end{aligned}$$

And so indeed, our expected inequality holds:

$$\lim_{\epsilon\to 0}\frac{\left\|(f\circ g)(p+\epsilon) - (f\circ g)(p) - ((\lambda_{f,g(p)}\circ\lambda_{g,p})(\epsilon)\right\|}{\|\epsilon\|} \leq 0$$

By a previous result, we know that such local linear approximations are unique: we've found one, and it's thus *the* one. Which concludes the proof. □