

# The Partial Jacobian Matrix and the Chain Rule for Functions on Products of Real $n$ -Spaces

Amod Agashe

Department of Mathematics, Florida State University, Tallahassee, FL, USA

Email: amod.agashe@gmail.com

**How to cite this paper:** Agashe, A. (2025) The Partial Jacobian Matrix and the Chain Rule for Functions on Products of Real  $n$ -Spaces. *Applied Mathematics*, 16, 420-427. <https://doi.org/10.4236/am.2025.164022>

**Received:** December 24, 2024

**Accepted:** April 27, 2025

**Published:** April 30, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

The usual chain rule is for compositions of functions from  $\mathbf{R}^n$  to  $\mathbf{R}^m$ , for some  $m$  and  $n$ . We consider functions whose domains and codomains are *products* of various  $\mathbf{R}^n$ 's, and introduce the notion of a partial Jacobian matrix, using which we formulate a chain rule in this setting that should be useful for calculations as well as for some theoretical purposes.

## Keywords

Chain Rule, Partial Derivative, Jacobian Matrix, Partial Jacobian Matrix, Normed Vector Spaces, Manifolds

## 1. Introduction

The goal of this article is to fill some potential gaps in the literature and introduce some notation that should be useful in applications. To motivate the main result, which may look abstract, we consider a concrete example. The reader who just wants to see and use the main result of this paper may jump directly to Section 3.

Let  $f: \mathbf{R}^2 \rightarrow \mathbf{R}^2$  be given by  $x(u, t) = u + t$ ,  $y(u, t) = u - t$ , and  $g: \mathbf{R}^2 \rightarrow \mathbf{R}$  be given by  $z = x^2 + y$ . Thinking of  $z$  as a function of  $(u, t)$  via  $f$ , if one wants to find the rate of change of  $z$  with respect to  $u$  at the point  $(1, 1)$ , we can use the chain rule

$$\frac{\partial z}{\partial u} = \frac{\partial z}{\partial x} \frac{\partial x}{\partial u} + \frac{\partial z}{\partial y} \frac{\partial y}{\partial u}, \quad (1)$$

to get  $\frac{\partial z}{\partial u} = 2x \cdot 1 + 1 \cdot 1 = 2(u + t) + 1 = 2u + 2t + 1$ , which evaluated at  $(1, 1)$  gives the answer 5. Note that found the answer without computing the function  $g \circ f$ , which is  $(u + t)^2 + (u - t)$ , which we could have differentiated with respect to  $u$  to get the same answer. The chain rule is especially useful if one wants to find the

rate of change of  $z$  with respect to  $u$ , knowing the other rates of change in Equation (1); this happens often in real life situations, where one may not know the functions  $f$  and  $g$  explicitly, but knows the instantaneous rates of change on the right side of Equation (1). Such problems are called related rates problems.

Now suppose we are given functions  $f: \mathbf{R}^2 \times \mathbf{R}^3 \rightarrow \mathbf{R}^4 \times \mathbf{R}^5$  and  $g: \mathbf{R}^4 \times \mathbf{R}^5 \rightarrow \mathbf{R}$ . We can compose the two to get the function  $g \circ f: \mathbf{R}^2 \times \mathbf{R}^3 \rightarrow \mathbf{R}$ . Let  $z$  denote the coordinate of the codomain  $\mathbf{R}$ , and suppose we want to find the directional derivative of  $z$  relative to a change in the coordinates of the  $\mathbf{R}^2$  in the domain of  $g \circ f$  (keeping the coordinates of  $\mathbf{R}^3$  fixed) at some given point in the domain  $\mathbf{R}^2 \times \mathbf{R}^3$ . We could, of course, write down the function  $g \circ f$  in terms of the coordinates of  $\mathbf{R}^2 \times \mathbf{R}^3$  and then take partial derivatives, etc., but that is more work than needed. In this article, we derive a chain rule (formula (8) below) where the calculation can be done in a manner similar to the chain rule in the previous paragraph, thus making the calculation much easier (see Remark 3.3). Note that in the example above, we did not give formulas to define  $f$  and  $g$  since the number of variables is too much; the reader can easily make up an explicit example. Our chain rule is also useful for related rate problems in this context, either for computations or for theoretical use. In fact, the reason for the author to write up this article was that a colleague in the Economics department asked for a chain rule in a context similar to the one above, and neither of us could find it in the form that was needed in the literature. So it is hoped that this article fills a potential gap in the literature, and can be quoted for the sake of completeness in papers that need this version of the chain rule.

The organization of the rest of this article is as follows. In Section 2, we recall the classical setting alluded to two paragraphs above, and introduce a notation for the Jacobian matrix that may be new, and is very convenient. In Section 3, we consider a more general setting of the type discussed in the previous paragraph, and give the more general chain rule alluded to above, which involves a new object that we call a partial Jacobian matrix. Finally, in Section 4, we give the proof of our generalization of the chain rule.

## 2. The Classical Setting

Suppose  $f: \mathbf{R}^n \rightarrow \mathbf{R}^m$  is differentiable. Let  $\mathbf{x}$  denote the vector of coordinate functions on the domain (so  $x_1, \dots, x_n$ , the components of  $\mathbf{x}$ , are the coordinates of  $\mathbf{R}^n$ ) and let  $\mathbf{y}$  denote the vector of coordinate functions on the codomain.

A standard notation for the Jacobian matrix of  $f$  is  $\frac{\partial(f_1, \dots, f_m)}{\partial(x_1, \dots, x_n)}$  and  $\mathbf{J}_f$ ; let

us introduce another notation:

**Definition 2.1.** Let

$$\frac{d\mathbf{y}}{d\mathbf{x}}$$

denote the Jacobian matrix of  $f$ .

The notation above is clearly related to the earlier notation (our notation may

not be new, but we have not seen it before), and also makes notational sense: the  $(j, i)$ -th entry of  $\frac{dy}{dx}$  is  $\frac{\partial y_j}{\partial x_i}$ .

Let  $g: \mathbf{R}^m \rightarrow \mathbf{R}^k$  be differentiable, and let  $z$  denote the vector of coordinate functions on the codomain. Letting  $D$  denote the derivative as usual, the chain rule says that for  $\mathbf{a} \in \mathbf{R}^n$ , we have

$$D(g \circ f)(\mathbf{a}) = Dg(f(\mathbf{a})) \circ Df(\mathbf{a}). \quad (2)$$

This looks nice in theory and is very convenient for theoretical purposes, but has to be unraveled in applications and may be too abstract for people working in applied areas. A more concrete restatement is the matrix version:

$$\mathbf{J}_{g \circ f}(\mathbf{a}) = \mathbf{J}_g(f(\mathbf{a})) \mathbf{J}_f(\mathbf{a}), \quad (3)$$

where the product on the right is matrix multiplication. Equation (3) still has to be unraveled, as we shall do soon, but first note that in the notation of Definition 2.1, the equation becomes

$$\frac{dz}{dx} = \frac{dz}{dy} \frac{dy}{dx}, \quad (4)$$

where again the product on the right is the matrix product. The rule looks much nicer with the new notation, and moreover, if  $n = m = k = 1$ , then it becomes the usual chain rule in the Leibnitz notation, namely

$$\frac{dz}{dx} = \frac{dz}{dy} \frac{dy}{dx}.$$

As usual, for  $j = 1, \dots, m$ , let  $f_j$  denote the  $j$ -th component function of  $f$ ; thus  $f = (f_1, \dots, f_m)$ . Recall that for  $i = 1, \dots, n$ , and for  $\mathbf{a} \in \mathbf{R}^n$ , the symbol  $D_i f_j(\mathbf{a})$  denotes the partial derivative of  $f_j$  at  $\mathbf{a}$  with respect to the  $i$ -th coordinate; thus  $D_i f_j(\mathbf{a})$  is  $\frac{\partial y_j}{\partial x_i}$  evaluated at  $\mathbf{a}$ . The following version of the

chain rule is very explicit; it is well known, and the only reason that we are stating it as a result is that it does not seem to be easy to find the statement in the literature (for example, in ([1], Theorem~2.9) and ([2], pp. 61-62), it is only stated for  $k = 1$ , and in [3], it is only mentioned in the proof of Proposition 3.10).

**Proposition 2.2.** For  $i = 1, \dots, n$ , and  $\ell = 1, \dots, k$ , we have for  $\mathbf{a} \in \mathbf{R}^n$ ,

$$D_i(g \circ f)_\ell(\mathbf{a}) = \sum_{j=1}^m D_j g_\ell(f(\mathbf{a})) D_i f_j(\mathbf{a}), \text{ and} \quad (5)$$

$$\frac{\partial z_\ell}{\partial x_i} = \sum_{j=1}^m \frac{\partial z_\ell}{\partial y_j} \frac{\partial y_j}{\partial x_i}. \quad (6)$$

*Proof.* This follows from Equation (3), the definition of the Jacobian matrix, and the definition of matrix multiplication.  $\square$

Equation (6) above may be the most useful version of the chain rule in applications.

### 3. A More General Setting

For  $i = 1, \dots, n$  and  $j = 1, \dots, m$ , let  $X_i$  and  $Y_j$  each be  $\mathbf{R}^p$  for some  $p$  (so for example, we could have  $X_1 = \mathbf{R}^2$ ,  $Y_2 = \mathbf{R}^5$ , etc.), and let  $u_i$  and  $v_j$  denote the dimensions of  $X_i$  and  $Y_j$  respectively (so  $X_i = \mathbf{R}^{u_i}$  and  $Y_j = \mathbf{R}^{v_j}$ ). Let  $f : X_1 \times \dots \times X_n \rightarrow Y_1 \times \dots \times Y_m$  be differentiable (regarding the domain and the codomain as finite dimensional vector spaces in the natural way). Note that if each of  $X_i$  and  $Y_j$  are  $\mathbf{R}$ , then  $X_1 \times \dots \times X_n = \mathbf{R}^n$  and  $Y_1 \times \dots \times Y_m = \mathbf{R}^m$  as vector spaces, so we are in the classical setting of Section 2; thus our setup generalizes the classical setting. The function  $f$  above can be thought of as a function whose inputs and outputs, instead of being tuples of numbers, are tuples of vectors (not necessarily in the same vector space). Such functions arise often in applied areas such as Economics and Engineering.

For  $i = 1, \dots, n$  and  $j = 1, \dots, m$ , let  $\mathbf{x}_i$  denote the vector of coordinate functions on  $X_i$  (so  $x_{i,1}, \dots, x_{i,u_i}$ , the components of  $\mathbf{x}_i$ , are the coordinates of  $X_i$ ), and let  $\mathbf{y}_j$  denote the vector of coordinate functions on  $Y_j$ .

**Definition 3.1.** For each  $i, j$ , denote by

$$\frac{\partial \mathbf{y}_j}{\partial \mathbf{x}_i}$$

the matrix whose  $(r, s)$ -th entry for  $s = 1, \dots, u_i$  and  $r = 1, \dots, v_j$  is  $\frac{\partial y_r}{\partial x_s}$ ; we call it a partial Jacobian matrix.

The reason for this notation and terminology is that while this object looks like the Jacobian matrix in the classical setting of Section 2, which might suggest the notation

$\frac{d\mathbf{y}_j}{d\mathbf{x}_i}$  that was used for the Jacobian matrix therein, it plays the role of

a partial derivative in the chain rule analog of (6), as we shall soon see (Equation (8) below).

For  $\ell = 1, \dots, k$ , let  $Z_\ell$  each be  $\mathbf{R}^p$  for some  $p$ , and let  $\mathbf{z}_\ell$  denote the vector of coordinate functions on  $Z_\ell$ . Let  $g : Y_1 \times \dots \times Y_m \rightarrow Z_1 \times \dots \times Z_k$  be differentiable. Thinking of each domain and codomain as real vector spaces in the natural way and letting  $D$  denote the derivative as usual, the chain rule again says that for  $\mathbf{a}$  in the domain of  $f$ ,

$$D(g \circ f)(\mathbf{a}) = Dg(f(\mathbf{a})) \circ Df(\mathbf{a}). \quad (7)$$

This equation is mostly of theoretical use, and is difficult to use in practice. Also, we are not distinguishing between the components of each domain (e.g., the components  $X_1, \dots, X_n$  of the domain of  $f$ ) in any way or treating an individual component special. A similar criticism holds for the analogs of equations (3) and (4), of which Equation (3) holds verbatim even in this situation.

We have the following version of the chain rule, which we prove in the next section:

**Theorem 3.2.** In the setup above, for  $\ell = 1, \dots, k$ , and  $i = 1, \dots, n$ , we have

$$\frac{\partial \mathbf{z}_\ell}{\partial \mathbf{x}_i} = \sum_{j=1}^m \frac{\partial \mathbf{z}_\ell}{\partial \mathbf{y}_j} \frac{\partial \mathbf{y}_j}{\partial \mathbf{x}_i}. \quad (8)$$

The equation above distinguishes between the components of each domain, unlike Equation (7) above, and is the analog of Equation (6) from the classical setting, which it agrees with when each of  $X_i$ ,  $Y_j$ , and  $Z_\ell$  is  $\mathbf{R}$ . Equation (8) may be the most concrete form of the chain rule in this general setting, and should be especially useful for calculations, but also for some theoretical purposes.

**Remark 3.3.** As an example of an application, if  $\mathbf{u} \in X_i$  is a unit vector, then the directional derivative<sup>1</sup> of the composite map

$X_i \rightarrow X_1 \times \cdots \times X_n \xrightarrow{g \circ f} Z_1 \times \cdots \times Z_k \rightarrow Z_\ell$  (where the first map is the natural inclusion, and the last map is the natural projection) at a point  $\mathbf{a} \in X_1 \times \cdots \times X_n$  in the direction of  $\mathbf{u}$  is

$$\frac{\partial \mathbf{z}_\ell}{\partial \mathbf{x}_i} \cdot \mathbf{u} = \left( \sum_{j=1}^m \frac{\partial \mathbf{z}_\ell}{\partial \mathbf{y}_j} \frac{\partial \mathbf{y}_j}{\partial \mathbf{x}_i} \right) \cdot \mathbf{u} = \sum_{j=1}^m \left( \frac{\partial \mathbf{z}_\ell}{\partial \mathbf{y}_j} \frac{\partial \mathbf{y}_j}{\partial \mathbf{x}_i} \cdot \mathbf{u} \right)$$

evaluated at  $\mathbf{a}$  (this follows from Theorem 3.2 and ([2], Theorem~5.1), for example).

Next we give a more theoretical version of Theorem 3.2, for which we need a generalization of the usual partial derivative.

**Definition 3.4.** For each  $j$ , let  $f_j$  denote the composite function

$$X_1 \times \cdots \times X_n \xrightarrow{f} Y_1 \times \cdots \times Y_m \rightarrow Y_j,$$

where the second map is the natural projection. Let  $\mathbf{a}$  be an element of  $X_1 \times \cdots \times X_n$ . For each  $i$ , let  $\iota_{\mathbf{a},i}$  denote the map  $X_i \rightarrow X_1 \times \cdots \times X_n$  that takes  $\mathbf{x} \in X_i$  to the element of  $X_1 \times \cdots \times X_n$  whose  $p$ -th component is the  $p$ -th component of  $\mathbf{a}$ , except for  $p = i$ , when it is  $\mathbf{x}$ . Denote by  $D_i f_j(\mathbf{a})$  the derivative at the  $i$ -th component of  $\mathbf{a}$  of the composite map

$$X_i \xrightarrow{\iota_{\mathbf{a},i}} X_1 \times \cdots \times X_n \xrightarrow{f} Y_1 \times \cdots \times Y_m \rightarrow Y_j,$$

where the last map is the natural projection; we call it the  $i$ -th partial derivative of  $f_j$  at  $\mathbf{a}$ .

An alternate definition, which is perhaps more conceptual, is given after Definition 4.1 in the next section. Note that the matrix of  $D_i f_j(\mathbf{a})$  in the standard

bases is  $\frac{\partial \mathbf{y}_j}{\partial \mathbf{x}_i}$  evaluated at  $\mathbf{a}$ . If each of  $X_i$  and  $Y_j$  are  $\mathbf{R}$ , then it is the usual

partial derivative  $D_i f_j(\mathbf{a})$ ; we remark that even in this situation, our conceptual definition above (or the equivalent one in the next section) in terms of maps seems to be missing in the literature (e.g., it is not there in [1] [2], and [3]), and could be useful for theoretical purposes (e.g., it is immediate that the partial derivative of a sum is the sum of the partial derivatives, since the partial derivative inherits that property of the derivative, from our definition). Note that in what follows, the definitions of  $D_j g_\ell$  and  $D_i (g \circ f)_\ell$  are analogous to that of  $D_i f_j$ .

<sup>1</sup>As defined on page 42 in ([2], Section 5), for example.

**Theorem 3.5.** In the setup above, for  $\ell = 1, \dots, k$ , and  $i = 1, \dots, n$ , we have for  $\mathbf{a}$  in  $X_1 \times \dots \times X_n$ ,

$$D_i(g \circ f)_\ell(\mathbf{a}) = \sum_{j=1}^m D_j g_\ell(f(\mathbf{a})) \circ D_i f_j(\mathbf{a}). \quad (9)$$

The equation above also distinguishes between the components of each domain, unlike Equation (7) above, and is the analog of Equation (5) from the classical setting, which it agrees with when each of  $X_i$ ,  $Y_j$ , and  $Z_\ell$  is  $\mathbf{R}$ .

**Remark 3.6.** As mentioned in the second proof of Theorem 3.5 in the next section, the proof does not use the fact that the  $X_i$ ,  $Y_j$ , and  $Z_k$  are finite dimensional vector spaces over the real numbers, just that they are normed vector spaces, so in particular Theorem 3.5 applies even if any of them are normed vector spaces of infinite dimension and/or over the complex numbers (over the complex numbers, the derivative is in the Fréchet sense, for example, as in ([4], Section VIII.1)).

**Remark 3.7.** Recall that given a smooth map  $f: X \rightarrow Y$  between two smooth manifolds, the derivative induces a map  $df: TX \rightarrow TY$  between the tangent bundles of the manifolds. The classical chain rule translates to a chain rule for such maps. Similarly our chain rule (9) applies to maps between products of manifolds, and gives a chain rule for products of manifolds, which we now state more precisely. For each  $i, j, \ell$  as above, let  $\mathcal{X}_i$ ,  $\mathcal{Y}_j$ , and  $\mathcal{Z}_\ell$  be manifolds<sup>2</sup> and let  $f: \mathcal{X}_1 \times \dots \times \mathcal{X}_n \rightarrow \mathcal{Y}_1 \times \dots \times \mathcal{Y}_m$  and  $g: \mathcal{Y}_1 \times \dots \times \mathcal{Y}_m \rightarrow \mathcal{Z}_1 \times \dots \times \mathcal{Z}_k$  be morphisms. Then the classical chain rule says that  $d_f(g \circ f) = dg \circ df$ . Now let  $d_i f_j$  denote the map induced by  $\mathcal{X}_i \rightarrow \mathcal{X}_1 \times \dots \times \mathcal{X}_n \rightarrow \mathcal{Y}_1 \times \dots \times \mathcal{Y}_m \rightarrow \mathcal{Y}_j$  (where the first map is the natural inclusion and the second map is the natural projection) on the corresponding tangent bundles; similar definitions apply to  $d_j g_\ell$  and  $d_i(g \circ f)_\ell$ . Then our discussion above gives the following application and analog of Theorem 3.5:  $d_i(g \circ f)_\ell = \sum_{j=1}^m d_j g_\ell \circ d_i f_j$ . In fact, by the previous remark, this will apply to certain infinite-dimensional manifolds, in particular to Banach manifolds.

## 4. Proofs of Theorems 3.2 and 3.5

We continue using the notation of the previous section. First note that Theorem 3.2 and Theorem 3.5 follow from each other by considering the basis of a linear transformation in the standard basis. We now prove Theorem 3.2; later, we shall give an independent proof of Theorem 3.5.

Let  $p$  be an indexing variable for the entries of  $\mathbf{z}_\ell$  and  $q$  an indexing variable for the entries of  $\mathbf{x}_i$ . By definition, the  $(p, q)$ -th entry in the matrix on the left side of Equation (8) is  $\frac{\partial z_{\ell,p}}{\partial x_{i,q}}$ . Now  $z_{\ell,p}$  is a function of  $x_{i,q}$  via the intermediate variables  $y_{j,r}$  for  $j = 1, \dots, m$  and  $r = 1, \dots, v_j$  for each  $j$ . So by the classical version (6) of the chain rule, we see that the left side of Equation (8) is

<sup>2</sup>We are not using  $X_i$ , etc., to denote the manifolds, since the  $X_i$ 's correspond to local charts on the manifolds.

$$\sum_{j=1}^m \sum_{r=1}^{v_j} \frac{\partial z_{\ell,p}}{\partial y_{j,r}} \frac{\partial y_{j,r}}{\partial x_{i,q}}. \quad (10)$$

The  $(p, q)$ -th entry of the term  $\frac{\partial \mathbf{z}_{\ell}}{\partial \mathbf{y}_j} \frac{\partial \mathbf{y}_j}{\partial \mathbf{x}_i}$  on the right side of Equation (8) is the product of the  $p$ -th row of  $\frac{\partial \mathbf{z}_{\ell}}{\partial \mathbf{y}_j}$  and the  $q$ -th column of  $\frac{\partial \mathbf{y}_j}{\partial \mathbf{x}_i}$ . Now the  $p$ -th row of  $\frac{\partial \mathbf{z}_{\ell}}{\partial \mathbf{y}_j}$  is

$$\left[ \frac{\partial z_{\ell,p}}{\partial y_{j,1}} \dots \frac{\partial z_{\ell,p}}{\partial y_{j,v_j}} \right]$$

and the  $q$ -th column of  $\frac{\partial \mathbf{y}_j}{\partial \mathbf{x}_i}$  is

$$\begin{bmatrix} \frac{\partial y_{j,1}}{\partial x_{i,q}} \\ \vdots \\ \frac{\partial y_{j,v_j}}{\partial x_{i,q}} \end{bmatrix}.$$

So the product of the row and column becomes

$$\sum_{r=1}^{v_j} \frac{\partial z_{\ell,p}}{\partial y_{j,r}} \frac{\partial y_{j,r}}{\partial x_{i,q}}.$$

As we sum these terms over  $j$  on the right side of Equation (8), we get the expression (10), which recall is the left side of Equation (8). This proves that the two sides of Equation (8) are equal, which proves Theorem 3.2, and thus Theorem 3.5 as well.

We now give a second proof of Theorem 3.5, which gives another proof of Theorem 3.2 as well. In fact, the proof of Theorem 3.5 that we give below does *not* use the fact that  $X_i$ ,  $Y_j$ , and  $Z_{\ell}$  are finite dimensional, and so applies to normed vector spaces (even over the complex numbers) that need not be finite dimensional. Another reason for giving a second proof of Theorem 3.5 is to note how the second proof is conceptual and illustrates where the linearity of the derivative is used, while the first proof is computational and hides some of the ideas of the proof.

The following definitions will be convenient in the proof.

**Definition 4.1.** If  $\mathbf{h} \in X_i$ , then let  $\mathbf{h}^i$  denote the vector in  $X_1 \times \cdots \times X_n$  whose components are all 0 except the  $i$ -component, which is  $\mathbf{h}$ ; so  $\mathbf{h}^i$  is the image of  $\mathbf{h}$  under the natural inclusion  $X_i \rightarrow X_1 \times \cdots \times X_n$ , and  $\mathbf{h} = \sum_i \mathbf{h}^i$ . Let  $\mathbf{a} \in X_1 \times \cdots \times X_n$ . If  $W$  is a vector space and  $\psi: X_1 \times \cdots \times X_n \rightarrow W$  is a linear map, then define  $D_i \psi(\mathbf{a}): X_i \rightarrow W$  by

$$D_i \psi(\mathbf{a})(\mathbf{h}) = D\psi(\mathbf{a})(\mathbf{h}^i); \quad (11)$$

it is a linear map.

Note that the definition of  $D_i\psi(\mathbf{a})$  above is consistent with Definition 3.4: when  $W = Y_j$  and  $\psi = f_j$ , we get  $D_i\psi(\mathbf{a}) = D_if_j(\mathbf{a})$ .

Let  $\mathbf{h} \in X_i$ . Then

$$\begin{aligned} D_i(g \circ f)_\ell(\mathbf{a})(\mathbf{h}) &= D(g \circ f)_\ell(\mathbf{a})(\mathbf{h}^i) \quad (\text{by equation (11)}) \\ &= (Dg_\ell(f(\mathbf{a})) \circ Df(\mathbf{a}))(\mathbf{h}^i) \quad (\text{by the usual chain rule (7)}) \\ &= Dg_\ell(f(\mathbf{a}))(Df(\mathbf{a})(\mathbf{h}^i)) \end{aligned} \quad (12)$$

Now the  $j$ -th component of  $Df(\mathbf{a})(\mathbf{h}^i)$  is  $Df_j(\mathbf{a})(\mathbf{h}^i)$ , so

$$Df(\mathbf{a})(\mathbf{h}^i) = \sum_j \left( Df_j(\mathbf{a})(\mathbf{h}^i) \right)^j \quad (\text{by the first sentence in Definition 4.1})$$

Putting this in Equation (12) above, we get

$$\begin{aligned} D_i(g \circ f)_\ell(\mathbf{a})(\mathbf{h}) &= Dg_\ell(f(\mathbf{a})) \left( \sum_j \left( Df_j(\mathbf{a})(\mathbf{h}^i) \right)^j \right) \\ &= \sum_j Dg_\ell(f(\mathbf{a})) \left( Df_j(\mathbf{a})(\mathbf{h}^i) \right)^j \quad (\text{since the derivative is linear}) \\ &= \sum_j D_j g_\ell(f(\mathbf{a})) \left( Df_j(\mathbf{a})(\mathbf{h}^i) \right) \quad (\text{by equation (11)}) \\ &= \sum_j D_j g_\ell(f(\mathbf{a})) \left( D_if_j(\mathbf{a})(\mathbf{h}) \right) \quad (\text{by equation (11) again}) \\ &= \sum_j \left( D_j g_\ell(f(\mathbf{a})) \circ D_if_j(\mathbf{a}) \right)(\mathbf{h}) \\ &= \left( \sum_j D_j g_\ell(f(\mathbf{a})) \circ D_if_j(\mathbf{a}) \right)(\mathbf{h}) \end{aligned}$$

Since  $\mathbf{h}$  was arbitrary, this shows that

$$D_i(g \circ f)_\ell(\mathbf{a}) = \sum_j D_j g_\ell(f(\mathbf{a})) \circ D_if_j(\mathbf{a}), \text{ as needed.}$$

## Acknowledgements

We are grateful to R. Vijay Krishna for several useful suggestions on this article.

## Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

## References

- [1] Spivak, M. (1965) *Calculus on Manifolds. A Modern Approach to Classical Theorems of Advanced Calculus*. W. A. Benjamin, Inc.
- [2] Munkres, J.R. (1991) *Analysis on Manifolds*. Addison-Wesley Publishing Company, Advanced Book Program.
- [3] Coleman, R. (2012) *Calculus on Normed Vector Spaces*. Universitext, Springer.
- [4] Dieudonné, J. (1969) *Foundations of Modern Analysis*, Vol. 10-I of Pure and Applied Mathematics. Academic Press.