Appendix A

General versions of the Chain Rule and the Leibniz Rule

At various points in the text, we will wish to have on hand explicit formulae for the Taylor series for compositions and products of smooth or holomorphic functions. These are messy inductive computations, and here we provide these for completeness. The formulae are stated in [Abraham, Marsden, and Ratiu 1988, Supplement 2.4A], and here we provide proofs. We shall work simultaneously with the real and complex cases. Thus $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. If $\mathbb{F} = \mathbb{R}$ then we wish to consider mappings that are of class C^∞ (with mappings of class C^ω being a subset of these) and if $\mathbb{F} = \mathbb{C}$ then we wish to consider mappings that are of class $\mathrm{C}^{\mathrm{hol}}$ (noting that these mappings are infinitely $\mathbb{F}$-differentiable by Theorem 1.1.23). Thus, throughout this section, we let $r = \infty$ if $\mathbb{F} = \mathbb{R}$ and $r = \mathrm{hol}$ if $\mathbb{F} = \mathbb{C}$.

A.1 The general Chain Rule

First we look at the Chain Rule, using the following notation. Let $r \in \mathbb{Z}_{>0}$ and let $r_1, \dots, r_k \in \mathbb{Z}_{\geq 0}$ satisfy $r_1 + \dots + r_k = r$. We denote by $\mathfrak{S}_{r_1,\dots,r_k}$ the subset of $\mathfrak{S}_r$ having the property that $\sigma \in \mathfrak{S}_{r_1,\dots,r_k}$ satisfies

$$\sigma(r_1 + \dots + r_j + 1) < \dots < \sigma(r_1 + \dots + r_j + r_{j+1}), \qquad j \in \{0, 1, \dots, k-1\},$$

with the understanding that $r_0 = 0$. Thus $\sigma \in \mathfrak{S}_{r_1,\dots,r_k}$ rearranges $\{1, \dots, r\}$ in such a way that order is preserved in the first r_1 entries, the next r_2 entries, and so on. Let us also denote by $\mathfrak{S}^<_{r_1,\dots,r_k}$ the subset of $\mathfrak{S}_{r_1,\dots,r_k}$ given by

$$\mathfrak{S}^<_{r_1,\dots,r_k} = \{\sigma \in \mathfrak{S}_{r_1,\dots,r_k} \mid \sigma(1) < \sigma(r_1 + 1) < \dots < \sigma(r_{k-1} + 1)\}.$$

In this case, the rearrangement by $\sigma \in \mathfrak{S}^<_{r_1,\dots,r_k}$ preserves order in the first place, the $(r_1 + 1)$st place, the $(r_2 + 1)$st place, and so on.

With this notation, we can state and prove the higher-order Chain Rule, a full proof being difficult to locate in the literature.

A.1.1 Lemma (Higher-order Chain Rule) *Let $\mathcal{U} \subseteq \mathbb{F}^n$ and $\mathcal{V} \subseteq \mathbb{F}^m$ be open, consider maps $\mathbf{g}\colon \mathcal{U} \to \mathcal{V}$ and $\mathbf{f}\colon \mathcal{V} \to \mathbb{F}^k$, and let $\mathbf{x} \in \mathcal{U}$. If $\mathbf{g}$ and $\mathbf{f}$ are of class C^r then $\mathbf{f} \circ \mathbf{g}$ is of class C^r and,* *moreover,*

$$\mathbf{D}^{\mathrm{r}}(\mathbf{f}\circ\mathbf{g})(\mathbf{x})\cdot(\mathbf{v}_1,\dots,\mathbf{v}_{\mathrm{r}}) = \sum_{\mathrm{j}=1}^{\mathrm{r}} \sum_{\substack{\mathrm{r}_1,\dots,\mathrm{r}_{\mathrm{j}}\in\mathbb{Z}_{>0}\\ \mathrm{r}_1+\dots+\mathrm{r}_{\mathrm{j}}=\mathrm{r}}} \sum_{\sigma\in\mathfrak{S}^{<}_{\mathrm{r}_1,\dots,\mathrm{r}_{\mathrm{j}}}} \mathbf{D}^{\mathrm{j}}\mathbf{f}(\mathbf{g}(\mathbf{x}))\cdot(\mathbf{D}^{\mathrm{r}_1}\mathbf{g}(\mathbf{x})\cdot(\mathbf{v}_{\sigma(1)},\dots,\mathbf{v}_{\sigma(\mathrm{r}_1)}),\dots, \mathbf{D}^{\mathrm{r}_{\mathrm{j}}}\mathbf{g}(\mathbf{x})\cdot(\mathbf{v}_{\sigma(\mathrm{r}_1+\dots+\mathrm{r}_{\mathrm{j}-1}+1)},\dots,\mathbf{v}_{\sigma(\mathrm{r})})) \quad (A.1)$$

for $\mathbf{v}_1,\dots,\mathbf{v}_{\mathrm{r}}\in\mathbb{F}^{\mathrm{n}}$.

Proof The proof is by induction on r. For $r = 1$ the result is simply the usual Chain Rule [Abraham, Marsden, and Ratiu 1988, Theorem 2.4.3]. Assume the result is true for $r\in\{1,\dots,s\}$. We thus have

$$\boldsymbol{D}^{s}(\boldsymbol{f}\circ\boldsymbol{g})(\boldsymbol{x})\cdot(\boldsymbol{v}_2,\dots,\boldsymbol{v}_{s+1}) = \sum_{j=1}^{s} \sum_{\substack{s_1,\dots,s_j\in\mathbb{Z}_{>0}\\ s_1+\dots+s_j=s}} \sum_{\sigma\in\mathfrak{S}^{<}_{s_1,\dots,s_j}} \boldsymbol{D}^{j}\boldsymbol{f}(\boldsymbol{g}(\boldsymbol{x}))\cdot(\boldsymbol{D}^{s_1}\boldsymbol{g}(\boldsymbol{x})\cdot(\boldsymbol{v}_{\sigma(2)},\dots,\boldsymbol{v}_{\sigma(s_1+1)}),\dots, \boldsymbol{D}^{s_j}\boldsymbol{g}(\boldsymbol{x})\cdot(\boldsymbol{v}_{\sigma(s_1+\dots+s_{j-1}+2)},\dots,\boldsymbol{v}_{\sigma(s+1)}))$$

for every $\boldsymbol{v}_2,\dots,\boldsymbol{v}_{s+1}\in\mathbb{F}^n$, and where $\sigma\in\mathfrak{S}^{<}_{s_1,\dots,s_j}\subseteq\mathfrak{S}_s$ permutes the set $\{2,\dots,s+1\}$ in the obvious way.

Let us now make an observation about permutations. Let $j'\in\{1,\dots,s+1\}$, let $s'_1,\dots,s'_{j'}\in\mathbb{Z}_{>0}$ satisfy $s'_1+\dots+s'_{j'}=s+1$, and let $\sigma'\in\mathfrak{S}^{<}_{s'_1,\dots,s'_{j'}}$. For brevity denote $t'_l=s'_1+\dots+s'_l$ for $l\in\{1,\dots,j'\}$. We have two cases.

1. $s'_1=1$: In this case let $j=j'-1$, define $s_l=s'_{l+1}$ for $l\in\{1,\dots,j'-1\}$, and let $t_l=s_1+\dots+s_l$ for $l\in\{1,\dots,j\}$. We then have

$$((1),(\sigma'(t'_2-s'_2+1),\dots,\sigma'(t'_2)),\dots,(\sigma'(t'_{j'}-s'_{j'}+1),\dots,\sigma'(t'_{j'}))) = ((1),(\sigma(t_2-s_2+1),\dots,\sigma(t_2)),\dots,(\sigma(t_{j'}-s_{j'}+1),\dots,\sigma(t_{j'}))), \quad (A.2)$$

where $\sigma\in\mathfrak{S}^{<}_{s'_1,\dots,s'_j}\subseteq\mathfrak{S}_s$ permutes $\{2,\dots,s+1\}$ in the obvious way. Note that this uniquely specifies $s_1,\dots,s_j$ and σ.

2. $s'_1\neq 1$: Here we take $j=j'$, $s_1=s'_1-1$, $s_l=s'_l$ for $l\in\{2,\dots,j\}$. Let us denote $t_l=s_1+\dots+s_l$ for $l\in\{1,\dots,j\}$. Then there exist $l_0\in\{1,\dots,j\}$ giving the corresponding cycle $\tau\in\mathfrak{S}_j$ given by $\tau=(1\ \cdots\ l_0)$ and $\sigma\in\mathfrak{S}_{s_{\tau(1)},s_{\tau(2)},\dots,s_{\tau(j)}}$ such that

$$((\sigma'(t'_1-s'_1+1),\dots,\sigma'(t'_1)),\dots,(\sigma'(t'_{j'}-s'_{j'}+1),\dots,\sigma'(t'_{j'}))) = ((1,\sigma(t_{\tau(1)}-s_{\tau(1)}+1),\dots,\sigma(t_{\tau(1)})),\dots,(\sigma(t_{\tau(j)}-s_{\tau(j)}+1),\dots,\sigma(t_{\tau(j)}))), \quad (A.3)$$

where σ permutes $\{2,\dots,s+1\}$ in the obvious way. Note that this uniquely specifies $s_1,\dots,s_j,\tau$, and σ. Note that the cycle τ is necessary to ensure that $\sigma'(1)=1$, a necessary condition that $\sigma'\in\mathfrak{S}^{<}_{s'_1,\dots,s'_{j'}}$. The cycle serves to place the slot into which the "1" is inserted at the beginning of the slot list.

Conversely, let $j \in \{1, \dots, s\}$, let $s_1, \dots, s_j \in \mathbb{Z}_{>0}$ have the property that $s_1 + \dots + s_j = s$, and let $\sigma \in \mathfrak{S}^<_{s_1,\dots,s_k}$. Denote $t_l = s_1 + \dots + s_l$ for $l \in \{1, \dots, j\}$. Then we have two scenarios.

1. We take $j' = j + 1$, let $s'_1 = 1$ and $s'_l = s_{l-1}$ for $l \in \{2, \dots, s + 1\}$. Define $t_l = s_1 + \dots + s_l$. Then there exists $\sigma' \in \mathfrak{S}^<_{s'_1,\dots,s'_j}$ such that (A.2) holds. Moreover, this uniquely determines $s'_1, \dots, s'_{j'}$ and σ'.
2. We take $j = j'$ and let $l_0 \in \{1, \dots, j\}$. Then take $\tau \in \mathfrak{S}_j$ to be the cycle $(1 \ \cdots \ l_0)$. We then define $s'_1 = s_{\tau(1)} + 1$ and $s'_l = s_{\tau(l)}$ for $l \in \{2, \dots, j\}$. Then there exists $\sigma' \in \mathfrak{S}^<_{s'_1,\dots,s'_{j'}}$ such that (A.3) holds. Note that this uniquely specifies $s'_1, \dots, s'_{j'}$ and σ'.

Using this observation, along with the usual Chain Rule and the symmetry of the derivatives of f of order up to s, we then compute

$$
\begin{aligned}
&\boldsymbol{D}^{s+1}(\boldsymbol{f} \circ \boldsymbol{g})(\boldsymbol{x}) \cdot (\boldsymbol{v}_1, \dots, \boldsymbol{v}_{s+1}) \\
&= \sum_{j=1}^{s} \sum_{\substack{s_1,\dots,s_j \in \mathbb{Z}_{>0} \\ s_1+\dots+s_j=s}} \sum_{\sigma \in \mathfrak{S}^<_{s_1,\dots,s_j}} \boldsymbol{D}^{j+1}\boldsymbol{f}(\boldsymbol{g}(\boldsymbol{x})) \cdot (\boldsymbol{D}\boldsymbol{g}(\boldsymbol{x}) \cdot \boldsymbol{v}_1, \\
&\quad \boldsymbol{D}^{s_1}\boldsymbol{g}(\boldsymbol{x}) \cdot (\boldsymbol{v}_{\sigma(2)}, \dots, \boldsymbol{v}_{\sigma(s_1+1)}), \dots, \\
&\quad \boldsymbol{D}^{s_j}\boldsymbol{g}(\boldsymbol{x}) \cdot (\boldsymbol{v}_{\sigma(s_1+\dots+s_{j-1}+2)}, \dots, \boldsymbol{v}_{\sigma(s+1)})) \\
&\quad + \boldsymbol{D}^{j}\boldsymbol{f}(\boldsymbol{g}(\boldsymbol{x})) \cdot (\boldsymbol{D}^{s_1+1}\boldsymbol{g}(\boldsymbol{x}) \cdot (\boldsymbol{v}_1, \boldsymbol{v}_{\sigma(2)}, \dots, \boldsymbol{v}_{\sigma(s_1+1)}), \dots, \\
&\quad \boldsymbol{D}^{s_j}\boldsymbol{g}(\boldsymbol{x}) \cdot (\boldsymbol{v}_{\sigma(s_1+\dots+s_{j-1}+2)}, \dots, \boldsymbol{v}_{\sigma(s+1)})) + \dots \\
&\quad + \boldsymbol{D}^{j}\boldsymbol{f}(\boldsymbol{g}(\boldsymbol{x})) \cdot (\boldsymbol{D}^{s_1}\boldsymbol{g}(\boldsymbol{x}) \cdot (\boldsymbol{v}_{\sigma(2)}, \dots, \boldsymbol{v}_{\sigma(s_1+1)}), \dots, \\
&\quad \boldsymbol{D}^{s_j}\boldsymbol{g}(\boldsymbol{x}) \cdot (\boldsymbol{v}_1, \boldsymbol{v}_{\sigma(s_1+\dots+s_{j-1}+2)}, \dots, \boldsymbol{v}_{\sigma(s+1)})) \\
&= \sum_{j'=1}^{s+1} \sum_{\substack{s'_1,\dots,s'_{j'} \in \mathbb{Z}_{>0} \\ s'_1+\dots+s'_{j'}=s+1}} \sum_{\sigma' \in \mathfrak{S}^<_{s'_1,\dots,s'_{j'}}} \boldsymbol{D}^{j'}\boldsymbol{f}(\boldsymbol{g}(\boldsymbol{x})) \cdot (\boldsymbol{D}^{s'_1}\boldsymbol{g}(\boldsymbol{x}) \cdot (\boldsymbol{v}_{\sigma'(1)}, \dots, \boldsymbol{v}_{\sigma'(s'_1)}), \\
&\quad \dots, \boldsymbol{D}^{s'_{j'}}\boldsymbol{g}(\boldsymbol{x}) \cdot (\boldsymbol{v}_{\sigma'(s'_1+\dots+s'_{j'-1}+1)}, \dots, \boldsymbol{v}_{\sigma'(s+1)})),
\end{aligned}
$$

as desired. ■

A.1.2 Remark (On the higher-order Chain Rule) In the single-variable case, the formula for the higher-order Chain Rule is due to Faà di Bruno [1855], an Italian mathematician and priest. The combinatorics of the formula arise in various places, including in the study of cumulants in probability theory. Generalisations of the Faà di Bruno formula to multiple-variables is enticing and have been studied by various people. Some useful formulae are given by Constantine and Savits [1996]. •

A.2 The general Leibniz Rule

Next we turn to the Leibniz Rule. To prove this, we first prove a result about derivatives of multilinear maps. Since derivatives are themselves multilinear maps, it will be useful to discriminate notationally between points in the domain of the map and points in the domain of the derivative of the map. Thus we shall write a point in $\mathbb{F}^{n_1} \times \cdots \times \mathbb{F}^{n_k}$ as $(x_1, \dots, x_k)$ when we mean it to be in the domain of the map L and we shall write a point in $\mathbb{F}^{n_1} \oplus \cdots \oplus \mathbb{F}^{n_k}$ as $(v_1, \dots, v_k)$ when we mean it to be an argument of the derivative. The argument of the rth derivative is an element of $(\mathbb{F}^{n_1} \oplus \cdots \oplus \mathbb{F}^{n_k})^r$ and will be written as

$$((v_{11}, \dots, v_{1k}), \dots, (v_{r1}, \dots, v_{rk})).$$

For $r \in \{1, \dots, k\}$ define

$$D_{r,k} = \{\{j_1, \dots, j_r\} \mid j_1, \dots, j_r \in \{1, \dots, k\} \text{ distinct}\}.$$

For $\{j_1, \dots, j_r\} \in D_{r,k}$ let us denote by $\{j'_1, \dots, j'_{k-r}\}$ the complement of $\{j_1, \dots, j_r\}$ in $\{1, \dots, k\}$. Now, for $\{j_1, \dots, j_r\} \in D_{r,k}$ define

$$\boldsymbol{\lambda}_{j_1,\dots,j_r} \in \mathrm{L}((\mathbb{F}^{n_{j'_1}} \oplus \cdots \oplus \mathbb{F}^{n_{j'_{k-r}}}) \oplus (\mathbb{F}^{n_{j_1}} \oplus \cdots \oplus \mathbb{F}^{n_{j_r}}); \mathbb{F}^{n_1} \oplus \cdots \oplus \mathbb{F}^{n_k})$$

by asking that

$$\boldsymbol{\lambda}_{j_1,\dots,j_k}((x_1, \dots, x_{k-r}), (v_1, \dots, v_r))$$

be obtained by placing x_l in slot j'_l for $l \in \{1, \dots, k-r\}$ and by placing v_l in slot j_l for $l \in \{1, \dots, r\}$.

With this notation we have the following lemma.

A.2.1 Lemma (Derivatives of multilinear maps) *If* $\mathrm{L} \in \mathrm{L}(\mathbb{F}^{n_1} \oplus \cdots \oplus \mathbb{F}^{n_k}; \mathbb{F}^{m})$ *is a multilinear map then* L *is infinitely differentiable. Moreover, for* $\mathrm{r} \in \{1, \dots, \mathrm{k}\}$ *we have*

$$\begin{aligned}\mathbf{D}^{\mathrm{r}}\mathrm{L}(\mathbf{x}_1, \dots, \mathbf{x}_{\mathrm{k}}) \cdot ((\mathbf{v}_{11}, \dots, \mathbf{v}_{1\mathrm{k}}), \dots, (\mathbf{v}_{\mathrm{r}1}, \dots, \mathbf{v}_{\mathrm{rk}}))& \\ = \sum_{\sigma \in \mathfrak{S}_{\mathrm{r}}} \sum_{\{\mathrm{j}_1,\dots,\mathrm{j}_{\mathrm{r}}\} \in \mathrm{D}_{\mathrm{r,k}}} \mathrm{L} \circ \boldsymbol{\lambda}_{\mathrm{j}_1,\dots,\mathrm{j}_{\mathrm{r}}}((\mathbf{x}_{\mathrm{j}'_1}, \dots, \mathbf{x}_{\mathrm{j}'_{\mathrm{k-r}}}), (\mathbf{v}_{\sigma(1)\mathrm{j}_1}, \dots, \mathbf{v}_{\sigma(\mathrm{r})\mathrm{j}_{\mathrm{r}}}))&\end{aligned}$$

and for $\mathrm{r} > \mathrm{k}$ *we have* $\mathbf{D}^{\mathrm{r}}\mathrm{L}(\mathbf{x}_1, \dots, \mathbf{x}_{\mathrm{k}}) = \mathbf{0}$*.*

Proof We prove the result by induction on r. For $r = 1$ the lemma asserts that

$$\begin{aligned}\boldsymbol{D}L(x_{01}, \dots, x_{0k}) \cdot (v_1, \dots, v_k) = L(v_1, x_{02}, \dots, x_{0k})& \\ + L(x_{01}, v_2, \dots, x_{0k}) + \cdots + L(x_{01}, x_{02}, \dots, v_k).&\end{aligned}$$

To verify this we must show that

$$\begin{aligned}\lim_{\substack{(x_1,\dots,x_k) \\ \to (x_{01},\dots,x_{0k})}} \Big\| L(x_1, \dots, x_k) - L(x_{01}, \dots, x_{0k}) - L(x_1 - x_{01}, \dots, x_{0k})& \\ - L(x_{01}, \dots, x_k - x_{0k}) \Big\| / \|(x_1 - x_{01}, \dots, x_k - x_{0k})\| = 0.& \quad \text{(A.4)}\end{aligned}$$

We do this by induction on k. For $k = 1$ we have

$$L(x_1) - L(x_{01}) - L(x_1 - x_{01}) = \mathbf{0},$$

and so (A.4) holds trivially. Now suppose that (A.4) holds for $k = s \geq 2$ and let $L \in \mathrm{L}(\mathbb{F}^{n_1}, \dots, \mathbb{F}^{n_{s+1}}; \mathbb{F}^m)$. We first note that the numerator in the limit in (A.4) can be written as

$$L(x_1, \dots, x_s, x_{0(s+1)}) - L(x_{01}, \dots, x_{0s}, x_{0(s+1)}) + L(x_1, \dots, x_s, x_s - x_{0(s+1)}) \\ - L(x_1 - x_{01}, \dots, x_{0s}, x_{0(s+1)}) - \dots - L(x_{01}, \dots, x_s - x_{0s}, x_{0(s+1)}) \\ - L(x_{01}, \dots, x_{0s}, x_{s+1} - x_{0(s+1)}).$$

By the induction hypothesis we have

$$\lim_{\substack{(x_1,\dots,x_s) \\ \to (x_{01},\dots,x_{0s})}} \Big\| L(x_1, \dots, x_s, x_{0(s+1)}) - L(x_{01}, \dots, x_{0s}, x_{0(s+1)}) \\ - L(x_1 - x_{01}, \dots, x_{0s}, x_{0(s+1)}) - L(x_{01}, \dots, x_s - x_{0s}, x_{0(s+1)}) \Big\| \\ / \|(x_1 - x_{01}, \dots, x_s - x_{0s})\| = 0.$$

Since

$$\|(x_1 - x_{01}, \dots, x_s - x_{0s})\| \leq \|(x_1 - x_{01}, \dots, x_s - x_{0s}, x_{s+1} - x_{0(s+1)})\|$$

this implies that

$$\lim_{\substack{(x_1,\dots,x_s,x_{s+1}) \\ \to (x_{01},\dots,x_{0s},x_{0(s+1)})}} \Big\| L(x_1, \dots, x_s, x_{0(s+1)}) - L(x_{01}, \dots, x_{0s}, x_{0(s+1)}) \\ - L(x_1 - x_{01}, \dots, x_{0s}, x_{0(s+1)}) - L(x_{01}, \dots, x_s - x_{0s}, x_{0(s+1)}) \Big\| \\ / \|(x_1 - x_{01}, \dots, x_s - x_{0s}, x_{s+1} - x_{0(s+1)})\| = 0. \quad \text{(A.5)}$$

We also have

$$\lim_{\substack{(x_1,\dots,x_s,x_{s+1}) \\ \to (x_{01},\dots,x_{0s},x_{0(s+1)})}} \Big\| L\Big(x_1, \dots, x_s, \tfrac{x_{s+1} - x_{0(s+1)}}{\|x_{s+1} - x_{0(s+1)}\|}\Big) - L\Big(x_{01}, \dots, x_{0s}, \tfrac{x_{s+1} - x_{0(s+1)}}{\|x_{s+1} - x_{0(s+1)}\|}\Big) \Big\| = 0$$

by continuity of L. Since

$$\|x_{s+1} - x_{0(s+1)}\| \leq \|(x_1 - x_{01}, \dots, x_s - x_{0s}, x_{s+1} - x_{0(s+1)})\|$$

this gives

$$\lim_{\substack{(x_1,\dots,x_s,x_{s+1}) \\ \to (x_{01},\dots,x_{0s},x_{0(s+1)})}} \Big\| L(x_1, \dots, x_s, x_{s+1} - x_{0(s+1)}) - L(x_{01}, \dots, x_{0s}, x_{s+1} - x_{0(s+1)}) \Big\| \\ / \|(x_1 - x_{01}, \dots, x_s - x_{0s}, x_{s+1} - x_{0(s+1)})\| = 0. \quad \text{(A.6)}$$

Combining (A.5) and (A.6) gives (A.4) for the case when $k = s + 1$ and so gives the conclusion of the lemma in the case when $r = 1$.

Now suppose that the lemma holds for $r \in \{1, \dots, s\}$ with $s < k$ and let $L \in \mathrm{L}(\mathbb{F}^{n_1}, \dots, \mathbb{F}^{n_k}; \mathbb{F}^m)$. Let us fix $\{j_1, \dots, j_s\} \in D_{s,k}$ and denote the complement of $\{j_1, \dots, j_s\}$ in $\{1, \dots, k\}$ by $\{j'_1, \dots, j'_{k-s}\}$, just as in our definitions before the statement of the lemma. Let us also fix $v_{j_l} \in \mathbb{F}^{n_{j_l}}$ for $l \in \{1, \dots, s\}$. Then define

$$P_{v_{j_1},\dots,v_{j_s}} \colon \mathbb{F}^{n_1} \times \dots \times \mathbb{F}^{n_k} \to (\mathbb{F}^{n_{j'_1}} \times \dots \times \mathbb{F}^{n'_{j_{k-s}}}) \times (\mathbb{F}^{n_{j_1}} \times \dots \times \mathbb{F}^{n_{j_s}}) \\ (x_1, \dots, x_k) \mapsto ((x_{j'_1}, \dots, x_{j'_{k-s}}), (v_{j_1}, \dots, v_{j_s})).$$

Now define $g_{v_{j_1},\dots,v_{j_s}}\colon \mathbb{F}^{n_1}\times\cdots\times\mathbb{F}^{n_k}\to\mathbb{F}^m$ by $g_{v_{j_1},\dots,v_{j_s}} = L\circ\lambda_{j_1,\dots,j_s}\circ P_{v_{j_1},\dots,v_{j_s}}$ and note that

$$g_{v_{j_1},\dots,v_{j_s}}(x_1,\dots,x_k) = L\circ\lambda_{j_1,\dots,j_s}((x_{j'_1},\dots,x_{j'_{k-s}}),(v_{j_1},\dots,v_{j_s})).$$

By Lemma A.1.1 we have

$$\begin{aligned}&\boldsymbol{D}g_{v_{j_1},\dots,v_{j_s}}(x_1,\dots,x_k)\cdot(u_1,\dots,u_k)\\ &\qquad= \boldsymbol{D}(L\circ\lambda_{j_1,\dots,j_r})(\boldsymbol{P}(x_1,\dots,x_k))\circ\boldsymbol{D}P_{v_{j_1},\dots,v_{j_s}}(x_1,\dots,x_k)\cdot(u_1,\dots,u_k).\end{aligned}$$

Note that since $P_{v_{j_1},\dots,v_{j_s}}$ is essentially a linear map (precisely, it is affine, meaning linear plus constant) we have

$$\boldsymbol{D}P_{v_{j_1},\dots,v_{j_s}}(x_1,\dots,x_k)\cdot(u_1,\dots,u_k) = ((u_{j'_1},\dots,u_{j'_{k-s}}),(\boldsymbol{0},\dots,\boldsymbol{0})).$$

Note that since $L\circ\lambda_{j_1,\dots,j_s}\in \mathrm{L}(\mathbb{F}^{n_{j'_1}},\dots,\mathbb{F}^{n_{j'_{k-s}}},\mathbb{F}^{n_{j_1}},\dots,\mathbb{F}^{n_{j_s}};\mathbb{F}^m)$ (as is readily verified), by the induction hypothesis,

$$\begin{aligned}&\boldsymbol{D}(L\circ\lambda_{j_1,\dots,j_s})(x_{j'_1},\dots,x_{j'_{k-s}},x_{j_1},\dots,x_{j_s})\cdot((u_{j'_1},\dots,u_{j'_{k-s}}),(u_{j_1},\dots,u_{j_s}))\\ &\qquad= L\circ\lambda_{j_1,\dots,j_s}((u_{j'_1},\dots,x_{j'_{k-s}}),(x_{j_1},\dots,x_{j_s}))+\dots\\ &\qquad\qquad+L\circ\lambda_{j_1,\dots,j_s}((x_{j'_1},\dots,x_{j'_{k-s}}),(x_{j_1},\dots,u_{j_s})).\end{aligned}$$

Therefore,

$$\begin{aligned}&\boldsymbol{D}g_{v_{j_1},\dots,v_{j_s}}(x_1,\dots,x_k)\cdot(u_1,\dots,u_k)\\ &\qquad= L\circ\lambda_{j_1,\dots,j_s}((u_{j'_1},\dots,x_{j'_{k-s}}),(v_{j_1},\dots,v_{j_s}))+\dots\\ &\qquad\qquad+L\circ\lambda_{j_1,\dots,j_s}((x_{j'_1},\dots,u_{j'_{k-s}}),(v_{j_1},\dots,v_{j_s})).\end{aligned}$$

Thus, for $v_j\in\mathbb{F}^{n_j}$, $j\in\{1,\dots,k\}$, we have

$$\begin{aligned}&\boldsymbol{D}g_{v_{j_1},\dots,v_{j_s}}(x_1,\dots,x_k)\cdot(v_1,\dots,v_k)\\ &\qquad= \sum_{j_{s+1}\notin\{j_1,\dots,j_s\}} L\circ\lambda_{j_1,\dots,j_s,j_{s+1}}((x_{j'_1},\dots,x_{j'_{k-(s+1)}}),(v_{j_1},\dots,v_{j_{s+1}})).\end{aligned}$$

Thus, using this relation along with linearity of the derivative, Lemma A.1.1, and the induction hypothesis, we compute

$$\begin{aligned}&\boldsymbol{D}^{s+1}f(x_1,\dots,x_k)\cdot((v_{11},\dots,v_{1k}),(v_{21},\dots,v_{2k}),\dots,\dots,(v_{(s+1)1},\dots,v_{(s+1)k}))\\ &\quad= \sum_{\sigma\in\mathfrak{S}_s}\sum_{\{j_2,\dots,j_{s+1}\}\in D_{s,k}}\boldsymbol{D}g_{v_{\sigma(2)j_2},\dots,v_{\sigma(s+1)j_{s+1}}}(x_1,\dots,x_k)\cdot(v_{11},\dots,v_{1k})\\ &\quad= \sum_{\sigma\in\mathfrak{S}_s}\sum_{\{j_2,\dots,j_{s+1}\}\in D_{s,k}}\sum_{j_1\notin\{j_2,\dots,j_{s+1}\}} L\circ\lambda_{j_1,\dots,j_s,j_{s+1}}((x_{j'_1},\dots,x_{j'_{k-(s+1)}}),\\ &\quad\quad(v_{j_1},v_{\sigma(2)j_2},\dots,v_{\sigma(s+1)j_{s+1}}))\\ &\quad= \sum_{\sigma\in\mathfrak{S}_{s+1}}\sum_{\{j_1,\dots,j_{s+1}\}\in D_{s+1,k}} L\circ\lambda_{\{j_1,\dots,j_{s+1}\}}((x_{j'_1},\dots,x_{j'_{k-(s+1)}}),\\ &\quad\quad(v_{\sigma(1)j_1},\dots,v_{\sigma(s+1)j_{s+1}})),\end{aligned}$$

where, in the second and third line, we define $\sigma \in \mathfrak{S}_s$ to be a bijection of $\{1, \dots, s+1\}$ by permutation of the last s elements.

The preceding argument gives the result when $r \in \{1, \dots, k\}$. For $r > k$ we argue as follows. We first note that

$$\boldsymbol{D}^k L((v_{11}, \dots, v_{1k}), \dots, (v_{k1}, \dots, v_{kk})) = \sum_{\sigma \in \mathfrak{S}_k} L(v_{\sigma(1)1}, \dots, v_{\sigma(k)k}). \tag{A.7}$$

It follows that $\boldsymbol{D}^r L = \boldsymbol{0}$ for $r > k$. ■

Now we can prove the desired Leibniz Rule.

A.2.2 Lemma (Higher-order Leibniz Rule) *Let $\mathcal{U} \subseteq \mathbb{F}^n$ be open, let $\mathbf{f}_j \colon \mathcal{U} \to \mathbb{F}^{n_j}$, $j \in \{1, \dots, k\}$, be r times differentiable at $\mathbf{x}_0 \in \mathcal{U}$, and let $L \in L(\mathbb{F}^{n_1}, \dots, \mathbb{F}^{n_k}; \mathbb{F}^m)$. If $\mathbf{f} \colon \mathcal{U} \to \mathbb{F}^m$ is defined by*

$$\mathbf{f}(\mathbf{x}) = L(\mathbf{f}_1(\mathbf{x}), \dots, \mathbf{f}_k(\mathbf{x}))$$

then $\mathbf{f}$ is r times differentiable at $\mathbf{x}_0$ and, moreover,

$$\mathbf{D}^r\mathbf{f}(\mathbf{x}_0) \cdot (\mathbf{v}_1, \dots, \mathbf{v}_r) = \sum_{\substack{r_1, \dots, r_k \in \mathbb{Z}_{\geq 0} \\ r_1 + \cdots + r_k = r}} \sum_{\sigma \in \mathfrak{S}_{r_1, \cdots, r_k}} L(\mathbf{D}^{r_1}\mathbf{f}_1(\mathbf{x}_0) \cdot (\mathbf{v}_{\sigma(1)}, \dots, \mathbf{v}_{\sigma(r_1)}), \dots, \mathbf{D}^{r_k}\mathbf{f}_k(\mathbf{x}_0) \cdot (\mathbf{v}_{\sigma(r_1 + \cdots + r_{k-1} + 1)}, \dots, \mathbf{v}_{\sigma(r)}))$$

for $\mathbf{v}_1, \dots, \mathbf{v}_r \in \mathbb{F}^n$.

Proof We prove the theorem by induction on r, noting that the case of $r = 1$ follows from Lemma A.1.1, and Lemma A.2.1, using the fact that $f = L \circ (f_1 \times \cdots \times f_k)$.

Assume the result is true for $r \in \{1, \dots, s\}$ and suppose that $f_1, \dots, f_k$ are of class C^{s+1}. Thus, for fixed $v_1, \dots, v_s \in \mathbb{F}^n$ the function

$$\begin{aligned} x \mapsto \boldsymbol{D}^s f(x) \cdot (v_2, \dots, v_{s+1}) \\ = \sum_{\substack{s_1, \dots, s_k \in \mathbb{Z}_{\geq 0} \\ s_1 + \cdots + s_k = s}} \sum_{\sigma \in \mathfrak{S}_{s_1, \cdots, s_k}} L(\boldsymbol{D}^{s_1} f_1(x) \cdot (v_{\sigma(2)}, \dots, v_{\sigma(s_1+1)}), \dots, \\ \boldsymbol{D}^{s_k} f_k(x) \cdot (v_{\sigma(s_1 + \cdots + s_{k-1} + 2)}, \dots, v_{\sigma(s+1)})), \end{aligned}$$

is differentiable at x_0, where we think of $\sigma \in \mathfrak{S}_s$ as a permutation of the set $\{2, \dots, s+1\}$ in the obvious way.

Let us now make an observation about permutations. Let $s'_1, \dots, s'_k \in \mathbb{Z}_{>0}$ have the property that $s'_1 + \cdots + s'_k = s + 1$ and let $\sigma' \in \mathfrak{S}_{s'_1, \dots, s'_k}$. For brevity denote $t'_j = s'_1 + \cdots + s'_j$ for $j \in \{1, \dots, k\}$. Then there exist unique $s_1, \dots, s_k \in \mathbb{Z}_{\geq 0}$ (denote $t_j = s_1 + \cdots + s_j$, $j \in \{1, \dots, k\}$), $\sigma \in \mathfrak{S}_{s_1, \dots, s_k}$, and $j_0 \in \{1, \dots, k\}$ such that

$$s_j = \begin{cases} s'_j, & j \neq j_0, \\ s'_j - 1, & j = j_0 \end{cases}$$

and

$$
\begin{aligned}
&((\sigma'(t'_1 - s'_1 + 1), \ldots, \sigma'(t'_1)), \ldots, (\sigma'(t'_{j_0} - s'_{j_0} + 1), \ldots, \sigma'(t'_{j_0})), \ldots, \\
&\qquad (\sigma'(t'_k - s'_k + 1) + \cdots + \sigma'(t'_k))) = ((\sigma(t_1 - s_1 + 1), \ldots, \sigma(t_1)), \ldots, \\
&\qquad\qquad (1, \sigma(t_{j_0} - s_{j_0}), \ldots, \sigma(t_{j_0} + 1)), \ldots, \\
&\qquad\qquad\qquad (\sigma(t_k - s_k), \ldots, \sigma(t_k + 1))), \quad \text{(A.8)}
\end{aligned}
$$

with the convention that σ permutes the set $\{1, \ldots, t'_{j_0} - s'_{j_0}, t'_{j_0} - s'_{j_0} + 2, \ldots, s + 1\}$ in the obvious way. The point is that $\sigma'(t'_{j_0} - s'_{j_0} + 1) = 1$, and by definition of $\mathfrak{S}_{s'_1,\ldots,s'_k}$ this means that $\sigma'(t'_{j_0} - s'_{j_0} + 1)$ must appear at the beginning of one of the "slots" of length $s'_1, \ldots, s'_k$. Conversely, let $s_1, \ldots, s_k \in \mathbb{Z}_{\geq 0}$ be such that $s_1 + \cdots + s_k = s \geq 2$ and let $\sigma \in \mathfrak{S}_{s_1,\ldots,s_k}$. Denote $t_j = s_1 + \cdots + s_j$ for $j \in \{1, \ldots, k\}$. Then, for each $j_0 \in \{1, \ldots, k\}$ there exist unique $s'_1, \ldots, s'_k \in \mathbb{Z}_{\geq 0}$ (denote $t'_j = s'_1 + \cdots + s'_j$, $j \in \{1, \ldots, k\}$) such that

$$
s'_j = \begin{cases} s_j, & j \neq j_0, \\ s_j + 1, & j = j_0 \end{cases}
$$

and $\sigma' \in \mathfrak{S}_{s'_1,\ldots,s'_k}$ such that (A.8) holds.

Using this observation, and since the result holds for $r = 1$ and $r = s$, we get

$$
\begin{aligned}
\boldsymbol{D}^{s+1}\boldsymbol{f}(\boldsymbol{x}_0) \cdot (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{s+1}) &= (\boldsymbol{D}(\boldsymbol{D}^s\boldsymbol{f})(\boldsymbol{x}_0) \cdot (\boldsymbol{v}_2, \ldots, \boldsymbol{v}_{s+1})) \cdot \boldsymbol{v}_1 \\
&= \Big(\sum_{\substack{s_1,\ldots,s_k \in \mathbb{Z}_{\geq 0} \\ s_1+\cdots+s_k = s}} \sum_{\sigma \in \mathfrak{S}_{s_1,\cdots,s_k}} L(\boldsymbol{D}^{s_1+1}\boldsymbol{f}_1(\boldsymbol{x}_0) \cdot (\boldsymbol{v}_1, \boldsymbol{v}_{\sigma(2)}, \ldots, \boldsymbol{v}_{\sigma(s_1+1)}), \ldots, \\
&\qquad \boldsymbol{D}^{s_k}\boldsymbol{f}_k(\boldsymbol{x}_0) \cdot (\boldsymbol{v}_{\sigma(s_1+\cdots+s_{k-1}+2)}, \ldots, \boldsymbol{v}_{\sigma(s+1)}))\Big) + \ldots \\
&\qquad + \Big(\sum_{\substack{s_1,\ldots,s_k \in \mathbb{Z}_{\geq 0} \\ s_1+\cdots+s_k = s}} \sum_{\sigma \in \mathfrak{S}_{s_1,\cdots,s_k}} L(\boldsymbol{D}^{s_1}\boldsymbol{f}_1(\boldsymbol{x}_0) \cdot (\boldsymbol{v}_{\sigma(2)}, \ldots, \boldsymbol{v}_{\sigma(s_1+1)}), \ldots, \\
&\qquad \boldsymbol{D}^{s_k+1}\boldsymbol{f}_k(\boldsymbol{x}_0) \cdot (\boldsymbol{v}_1, \boldsymbol{v}_{\sigma(s_1+\cdots+s_{k-1}+2)}, \ldots, \boldsymbol{v}_{\sigma(s+1)}))\Big) \\
&= \sum_{\substack{s'_1,\ldots,s'_k \in \mathbb{Z}_{\geq 0} \\ s'_1+\cdots+s'_k = s+1}} \sum_{\sigma \in \mathfrak{S}_{s'_1,\cdots,s'_k}} L(\boldsymbol{D}^{s'_1}\boldsymbol{f}_1(\boldsymbol{x}_0) \cdot (\boldsymbol{v}_{\sigma(1)}, \ldots, \boldsymbol{v}_{\sigma(s'_1+1)}), \ldots, \\
&\qquad \boldsymbol{D}^{s'_k}\boldsymbol{f}_k(\boldsymbol{x}_0) \cdot (\boldsymbol{v}_{\sigma(s'_1+\cdots+s'_{k-1}+1)}, \ldots, \boldsymbol{v}_{\sigma(s+1)})),
\end{aligned}
$$

as desired. ■

Bibliography

Abraham, R., Marsden, J. E., and Ratiu, T. S. [1988] *Manifolds, Tensor Analysis, and Applications*, second edition, number 75 in Applied Mathematical Sciences, Springer-Verlag, ISBN 0-387-96790-7.

Constantine, G. M. and Savits, T. H. [1996] *A multivariate Faà di Bruno formula with applications*, Transactions of the American Mathematical Society, **348**(2), 503–520.

Faà di Bruno, C. F. [1855] *Note sur une nouvelle formule du calcul differentiel*, The Quarterly Journal of Mathematics. Oxford. Second Series, **1**, 359–360.