THE FAÀ DI BRUNO CONSTRUCTION

J.R.B. COCKETT AND R.A.G. SEELY

Abstract. In the context of Cartesian differential categories [BCS 09], the structure of the first-order chain rule gives rise to a fibration, the "bundle category". In the present paper we generalise this to the higher-order chain rule (originally developed in the traditional setting by Faà di Bruno in the nineteenth century); given any Cartesian differential category $\mathbb{X}$, there is a "higher-order chain rule fibration" $\mathsf{Faà}(\mathbb{X}) \to \mathbb{X}$ over it. In fact, $\mathsf{Faà}$ is a comonad (over the category of Cartesian left (semi-)additive categories). Our main theorem is that the coalgebras for this comonad are precisely the Cartesian differential categories. In a sense, this result affirms the "correctness" of the notion of Cartesian differential categories.

Introduction

Francesco Faà di Bruno (1825-1888) was an Italian of noble birth, a soldier, a mathematician, and a priest. In 1988 he was beatified by Pope John Paul II, apparently for his charitable work teaching young women mathematics. As a mathematician he studied with Cauchy in Paris. He was a tall man with a solitary disposition who spoke seldom and, when teaching class, not always successfully. Perhaps his most significant mathematical contribution concerned the combinatorics of the higher-order chain rules. These results were the cornerstone of "combinatorial analysis": a subject which never really took off. It is the combinatorics underlying the higher-order chain rule which is of interest to us here. [OR 97, L 09]

The paper is a sequel to our work on Cartesian differential categories [BCS 09]. In that paper we established that from any Cartesian left additive category one could construct a "bundle" fibration in which the fibres were additive, and where the composition had the same form as the chain rule; moreover the existence of a left additive section of this fibration provided an operator which already satisfies many of the axioms of a differential operator. We pointed out in that paper that all the axioms could be generated by a similar fibrational analysis involving the higher-order chain rule. Presenting the details of that claim is the main aim of the present paper.

After some combinatorial preliminaries, we give the definition of the Faà di Bruno category over a Cartesian left additive category $\mathbb{X}$. The most complicated aspect of the definition is the composition, which we shall see corresponds to the higher-order chain rule. Most of the paper is devoted to proving that the Faà di Bruno construction is in fact a comonad over the category of left additive categories. Finally, we show that the coalgebras of this comonad are exactly Cartesian differential categories.

1. Some combinatorics of symmetric trees

Counting trees of various shapes and sizes is a deeply combinatoric subject. It is also an issue which is of no small algebraic and analytic interest as these combinatoric numbers occur as coefficients in formulae such as the higher-order chain rule. Our current preoccupation is not so much with the bald numbers as with the structural relationships which cause these numbers to become important. We start this appreciation by introducing the combinatorial structures which underlie the chain rule.

1.1. Symmetric trees. A symmetric tree of height $n \geq 0$ and of width $m > 0$, in variables $V = \{x_1, \ldots, x_m\}$, is defined inductively by:

- The only symmetric tree of height 0 has width 1 and is a variable y;
- A symmetric tree of height $n \geq 1$, of width m, in the variables $\{x_1, \ldots, x_m\}$, is an expression $\bullet_r(t_1, \ldots, t_r)$ where each t_i is a symmetric tree of height $n-1$ in the variables V_i, where $\bigsqcup_{i=1}^{r} V_i = V$.

Here, the operation $\bullet$ is symmetric (so we are really considering equivalence classes), as indicated in the discussion below.

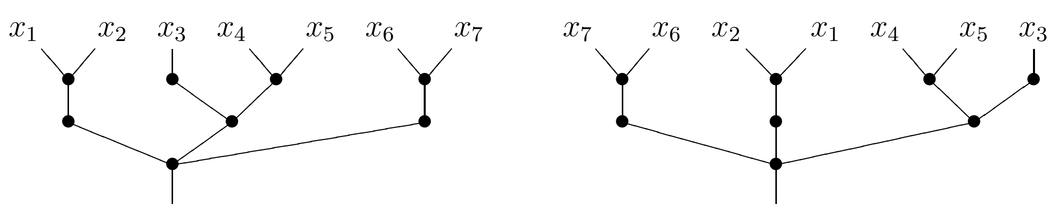
Note that the inductive step involves splitting the variables into r disjoint non-empty subsets. The combinatorics of this step are classical: Stirling numbers, of the second kind, are precisely the number of ways of partitioning a set with n elements into r non-empty partitions and are often written $S(n, r)$. Thus, we may already begin to see how combinatorics enters into our subject matter.

We may regard these trees in various ways. For what follows an important perspective will be to regard them as algebraic expressions: in this view the operations at the nodes are **symmetric**, or commutative, thus

$$\bullet_r(t_1, \ldots, t_r) = \bullet_r(t_{\sigma(1)}, \ldots, t_{\sigma(r)})$$

for any permutation $\sigma \in \mathcal{S}_r$. One can also regard these expressions as trees (in the graph theoretic sense) in which the leaves are uniquely labeled (as the root or by one of the variables) but in which no other node or edge is labeled. Two symmetric trees are the same if graph theoretically they are isomorphic in a way which respects the leaf labeling.

Here are two representations of the same symmetric tree:



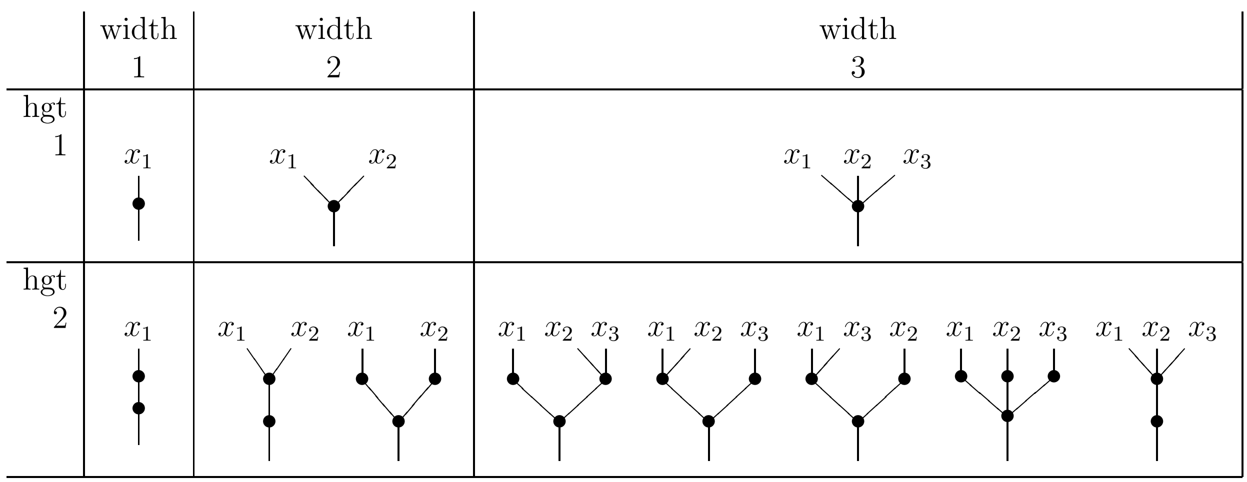
An important way to regard a symmetric tree of height n and width m is as a chain of surjective maps:

$$V = V_0 \longrightarrow V_1 \longrightarrow \ldots \longrightarrow V_{n-1} = 1$$

in which V is the set of variables, $|V| = m$ (the width of the tree), $|V_{n-1}| = 1$, and V_i represents the nodes at the i^{th}-level. This way of viewing a symmetric tree suggests a rather compact notation, representing the nodes as equivalence classes of the variables and then equivalence classes of these etc. Thus, a tree of height n can be represented as an element of $\mathcal{P}^n(V)$, where V is the (non-empty) set of variables and such that the iterated union is of these subsets is just the set of variables. This allows us to represent the height 3 tree above as:

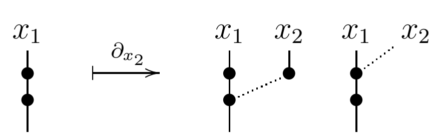
$$\{\{\{x_1, x_2\}\}, \{\{x_3\}, \{x_4, x_5\}\}, \{\{x_6, x_7\}\}\}$$

If one wishes to generate all the symmetric trees of a given height and width one is presented by a combinatoric problem as one must avoid generating trees which are already represented. In fact, as we shall see, there is a simple method for generating these trees. Meanwhile here is a classification of the first few symmetric trees separated by height and width:

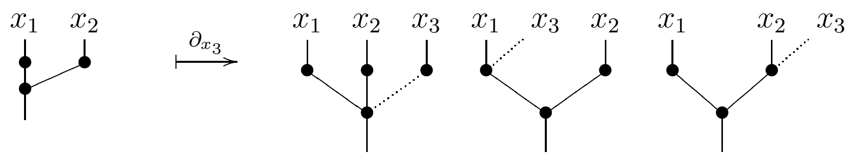


1.2. The differential of symmetric trees. The differential of a symmetric tree τ of height n and width r produces a bag of m trees of height n and width $r+1$, where m is the number of nodes of τ. The new trees of the differential are produced by picking a node and adding a "limb" to the new variable. The limb consists of a series of unary nodes applied to the new variable: these unary nodes are necessary in order to retain the uniform height of the tree.

For example the differential (introducing x_2) of the tree below is a pair of trees:



while the differential, introducing x_3, of the following tree (which is the first tree of the derivative above) is three trees:



1.2.1. Proposition. *Every symmetric tree of height h and width d can be obtained as a member of the d^{th}-derivative of the unary tree of the height h, denoted ι_h.*

Proof. To see this it is perhaps easier to think of the process in reverse: that is of *reducing* a tree by stripping out the limb of the tree whose leaf is the last introduced variable: that is the leaf and all the unary nodes below that leaf. Notice this process retains the height of the tree but at each step reduces the width by 1. Eventually any tree of height h can be reduced to the unique tree, $\iota_h^{x_1}$, of the height h and width 1 (on variable x_1). Reversing the process thus generates all the trees of the required width and height.

Furthermore, this process never generates the same tree twice. If at some step the same tree had been generated then stripping out the last limb shows that the previous step had also had a tree represented more than once. This means one can preserve the presence of repetition back to the starting point. As, in this case, we started with a single tree, $\iota_h^{x_1}$, we know there can be no repetitions. ■

Notation: Recall that $\iota_2^{x_1}$ is the (unique) height 2 width 1 tree with variable x_1; by

$$\mathcal{T}_2^V = \mathcal{T}_2^{x_1,\ldots,x_r} = \partial_{x_2,\ldots,x_r}(\iota_2^{x_1})$$

is meant the bag of trees obtained by formally deriving $\iota_2^{x_1}$ r-times with respect to the given variables $V = \{x_1, \ldots, x_r\}$. By the above discussion of the combinatorics of trees we know this is the set of *all* symmetric trees of height 2 and width r.

2. Faà di Bruno Bundle Categories

Given a Cartesian left additive category, $\mathbb{X}$, we shall construct two categories which we shall refer to as the Faà di Bruno (bundle) categories, $\mathsf{BFaà}(\mathbb{X})$ and a full subcategory $\mathsf{Faà}(\mathbb{X})$. The objects of the category $\mathsf{BFaà}(\mathbb{X})$ are pairs of objects of the original category (A, X). The objects of $\mathsf{Faà}(\mathbb{X})$ are those such pairs where $X = A$. We shall define the maps in a moment, but it might be worth pausing to remark on the choices of these objects. The more general pairs (A, X) will prove to be unsuitable when we come to our main theorem (characterizing Cartesian differential categories as coalgebras for the comonad $\mathsf{Faà}$), but otherwise the constructions to get us there work most transparently if we keep quite separate the two roles the object A plays (which is why we prefer to take the objects as pairs (A, A) rather than as simply objects A). The more general construction $\mathsf{BFaà}(\mathbb{X})$ seems actually to be more natural, and this is an indication that the construction has more general forms which we shall not explore here. In the meantime, at the very least it will help both writers and readers to keep track of variables in positions where they should be additive (where we use A) or not necessarily additive (where we shall use X). That will be apparent immediately, as we define the maps of the categories.

The maps or arrows $f\colon (A,X) \longrightarrow (B,Y)$ of the category consist of infinite sequences of maps

$$f = (f_*, f_1, f_2, \ldots)\colon (A,X) \longrightarrow (B,Y)$$

As the notation suggests the first map in this sequence is of a slightly different nature than the remainder. We require simply that $f_*\colon X \longrightarrow Y$ is a map in $\mathbb{X}$. For $r > 0$ we require

$$f_r\colon \underbrace{A \times \ldots \times A}_{r} \times X \longrightarrow B$$

is a symmetric form. This means that it is additive in each of the first r arguments and symmetric in these arguments. (The reader might think of this as $f_r\colon A^{\otimes^r}/r! \times X \longrightarrow B$, apart from the unfortunate fact that $\mathbb{X}$ need not have the tensor $\otimes$.) It will soon become apparent that the intended interpretation of f_n is they will be summands in the expression for higher order differential term. (See the remark following Proposition 2.2.3.)

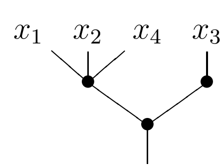
The difficulty is, of course, to define a composition for these arrows, which is where the Faà di Bruno convolution is used.

2.1. Faà di Bruno convolution. The description of the composition fg of two maps in $\mathsf{BFa\grave{a}}(\mathbb{X})$ is our next objective. The easy part is the composition in the first coordinate which is just, as expected, given by the composition in $\mathbb{X}$. The description of the composition on the remaining coordinates is more involved.

First suppose τ is a symmetric tree of height 2 and width r on the variables $V = \{x_1, \ldots, x_r\}$. This means that $\tau \in \mathcal{T}_2^V$. Furthermore, suppose $f\colon (A,X) \longrightarrow (B,Y)$ and $g\colon (B,Y) \longrightarrow (C,Z)$ in $\mathsf{BFa\grave{a}}(\mathbb{X})$, then by the component of the composite at τ:

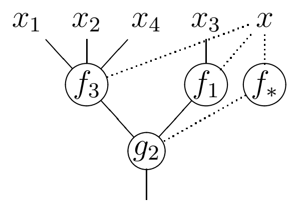
$$(f \star g)_\tau\colon \underbrace{A \times \ldots \times A}_{r} \times X \longrightarrow C$$

is meant the map obtained by substituting all the layer one nodes of arity i with f_i and the (one) node of τ at layer two with the function g_j, where j is the appropriate arity. Thus when τ is the following tree



then

$$(f \star g)_\tau = g_2(f_1(x_3, x), f_3(x_1, x_2, x_4, x), f_*(x))\colon \underbrace{A \times \ldots \times A}_{4} \times X \longrightarrow C.$$



We may replace the single tree τ by any bag T of trees in $\mathcal{T}_2^V$ in which case we simply sum the components

$$(f \star g)_T = \sum_{n\cdot\tau \in T} n \cdot (f \star g)_\tau$$

The Faà di Bruno convolution composition of f and g is then

$$(fg)_r = (f \star g)_{\mathcal{T}_2^{x_1,\ldots,x_r}}$$

which is well-defined as the result is clearly a symmetric form as permuting the variables of any $\tau \in \mathcal{T}_2^{x_1,\ldots,x_r}$ either leaves τ fixed or produces a new tree which still has the same height and therefore is in $\mathcal{T}_2^{x_1,\ldots,x_r}$. Thus we take this convolution to be the composition in the Faà di Bruno category. We have:

2.1.1. Proposition. *For any Cartesian left additive category* $\mathbb{X}$, $\mathsf{BFaà}(\mathbb{X})$ *(and so* $\mathsf{Faà}(\mathbb{X})$*) as defined above are Cartesian left additive categories.*

Proof. It is quite easily checked that this composition has as identity maps $1 = (1, \pi_0, 0, \ldots)\colon (A, X) \to (A, X)$. That the composition is associative follows immediately from the fact that the composition is given by summing over *all* the trees of height 2 (and appropriate width). A threefold composition is then just the sum over *all* trees of height 3 (and appropriate width) and so immediately associative.

The additive structure on the homsets is given pointwise:

$$f + g = (f_*, f_1, f_2, \ldots) + (g_*, g_1, g_2, \ldots) = (f_* + g_*, f_1 + g_1, f_2 + g_2, \ldots)$$

This composition is left additive as clearly, with respect to point-wise addition,

$$(f \star (g + g'))_\tau = (f \star g_1)_\tau + (f \star g_2)_\tau$$

as the underlying category is left additive and it immediately follows that this also holds for bags of such terms.

Furthermore the category has additive products: the product of (A, X) and (B, Y) is $(A \times B, X \times Y)$, given $f\colon (C, Z) \to (A, X)$ and $g\colon (C, Z) \to (B, Y)$ we define $\langle f, g\rangle$ as $(\langle f_i, g_i\rangle)_{i=0}^{\infty}$. The projections are $\pi_i = (\pi_i, \pi_0\pi_i, 0, 0, \ldots)$ for $i = 1, 2$ and the diagonal $\Delta = (\Delta, \pi_0\Delta, 0, 0, \ldots)$. These are additive as the first level of the Faà di Bruno category is the bundle fibration. ∎

Clearly we may truncate $\mathsf{BFa\grave{a}}(\mathbb{X})$ to any given number of levels to obtain a left additive category, $\mathsf{BFa\grave{a}}_n(\mathbb{X})$, as the composition at any given level always only involves maps from that and lower levels. As mentioned above $\mathsf{BFa\grave{a}}_1(\mathbb{X})$ is the more standard bundle fibration of $\mathbb{X}$ [BCS 09].

Our aim now is to show that the obvious functor

$$\varepsilon\colon \mathsf{BFa\grave{a}}(\mathbb{X}) \longrightarrow \mathbb{X} : \begin{array}{lcl} (A, X) & \mapsto & X \\ (f_*, f_1, f_2, \ldots) & \mapsto & f_* \end{array}$$

is also a fibration, (and similarly for $\mathsf{Fa\grave{a}}(\mathbb{X})$). Toward this end first note that ε is clearly a left additive functor which preserves the Cartesian structure on the nose. Now consider

$$\begin{tikzcd}
(C.Z) \arrow[dr, dotted, "{(f,f_1,f_2,\ldots)}"'] \arrow[drr, bend left, "{(fg,f_1,f_2,\ldots)}"] & & \\
& (A,X) \arrow[r, "{(g,\pi_0,0,0)}"'] & (B,Y) \\
& {} \arrow[d, mapsto, "\varepsilon"] & \\
& {} & \\
Z \arrow[r, "f"'] & X \arrow[r, "g"'] & Y
\end{tikzcd}$$

It is easily checked that the map sitting above f is uniquely determined. Thus we have:

2.1.2. Proposition. *For any Cartesian left additive category* $\mathbb{X}$

$$\varepsilon\colon \mathsf{BFa\grave{a}}(\mathbb{X}) \longrightarrow \mathbb{X}$$

is a strict Cartesian left additive functor which is a fibration. Similarly, $\varepsilon\colon \mathsf{Fa\grave{a}}(\mathbb{X}) \longrightarrow \mathbb{X}$ *is a strict Cartesian left additive functor and a fibration.*

The additive maps in a fiber are precisely those of the form $(1, f_1, 0, 0, \ldots)$ which above the second component vanish. This means that the fibers are far from being additive.

The maps in the Faà di Bruno category should be thought of as higher-order symmetric forms. When a differential is already present, as for ordinary differential equations, they will not, in general, be realizable as the differential forms of a function. However, what is true - and is the subject of the next section – is that a differentiable function gives rise such a form so that these forms should be viewed as abstract differentiable functions.

2.2. Functorial properties of the Faà di Bruno construction. It is worth noting that the Faà di Bruno construction is functorial on the category of Cartesian left additive categories, CLAdd, where the functors are taken to preserve addition and the product structure on the nose:

2.2.1. Lemma. *The Faà di Bruno constructions are functors.*

$$\mathsf{BFa\grave{a}}\colon \mathsf{CLAdd} \longrightarrow \mathsf{CLAdd} \qquad \mathsf{Fa\grave{a}}\colon \mathsf{CLAdd} \longrightarrow \mathsf{CLAdd}$$

Proof. That these are functors is clear provided we are assured that additivity in an argument is preserved by Cartesian left additive functors, as in that case forms are carried to forms. The fact that this is true relies on the fact that $f\colon A\times X \longrightarrow B$ is additive in its first argument in a left additive category if and only if $((\pi_0+\pi_1)\times 1)f = (\pi_0\times 1)f + (\pi_1\times 1)f$. This property is clearly preserved by the Cartesian left additive functors. We then define:

$$\mathsf{BFa\grave{a}}\colon \mathsf{CLAdd} \longrightarrow \mathsf{CLAdd}\colon \quad \begin{array}{c} \mathbb{X} \\ {\scriptstyle F}\big\downarrow \\ \mathbb{Y} \end{array} \quad \mapsto \quad \begin{array}{c} \mathsf{BFa\grave{a}}(\mathbb{X}) \\ \big\downarrow{\scriptstyle \mathsf{BFa\grave{a}}(F)} \\ \mathsf{BFa\grave{a}}(\mathbb{Y}) \end{array}$$

where $\mathsf{BFa\grave{a}}(F)(f_*, f_1, f_2, \ldots) = (F(f_*), F(f_1), F(f_2), \ldots)$. $\mathsf{Fa\grave{a}}$ is handled the same way (just restrict to the appropriate subcategories). ■

It is also clear that $\varepsilon = \varepsilon_{\mathbb{X}}\colon \mathsf{BFa\grave{a}}(\mathbb{X}) \longrightarrow \mathbb{X}$ is a natural transformation (natural in $\mathbb{X}$). However, more is true: there is a functor $\delta = \delta_{\mathbb{X}}\colon \mathsf{BFa\grave{a}}(\mathbb{X}) \longrightarrow \mathsf{BFa\grave{a}}(\mathsf{BFa\grave{a}}(\mathbb{X}))$, introduced below, natural in $\mathbb{X}$, such that:

2.2.2. Theorem. *$(\mathsf{BFa\grave{a}}, \varepsilon, \delta)$ is a comonad on CLAdd, as is the restriction to $\mathsf{Fa\grave{a}}$.*

The remainder of this section is given over to proving this theorem, which we shall do for $\mathsf{BFa\grave{a}}$, leaving the restriction to $\mathsf{Fa\grave{a}}$ to the reader. But first, in order to describe the functor $\delta := \delta_{\mathbb{X}}$ for some Cartesian left additive category $\mathbb{X}$, it will be useful to develop some conventions and fix notation. First for $f \in \mathsf{BFa\grave{a}}(\mathbb{X})$ we write $f = (f_*, f_1, f_2, \ldots)\colon (A, X) \longrightarrow (B, Y)$ where

$$\begin{aligned} f_*\colon X \longrightarrow Y \quad &; \quad x \mapsto f_*(x) \\ f_n\colon A^n \times X \longrightarrow B \quad &; \quad (a_{*1}, \ldots, a_{*n}, x) \mapsto f_n(x)\cdot a_{*1}\cdot \ldots \cdot a_{*n} \end{aligned}$$

Notice that we have "highlighted" the additive arguments by writing them as if they were simple multipliers. This should seem very familiar to the reader familiar with the term calculus for Cartesian differential categories from [BCS 09]; this is intended, of course.

We then define $\delta\colon \mathsf{BFa\grave{a}}(\mathbb{X}) \longrightarrow \mathsf{BFa\grave{a}}(\mathsf{BFa\grave{a}}(\mathbb{X}))$ as follows: on objects, δ takes (A, X) to $((A, A), (A, X))$. On arrows, $f \mapsto \delta(f) = (f, f^{[1]}, f^{[2]}, \ldots)$ by setting

$$\begin{aligned} f_*^{[n]}\colon A^n \times X \longrightarrow B \quad &: \quad (a_{*1}, \ldots, a_{*n}, x) \mapsto f_n(x)\cdot a_{*1}\cdot \ldots \cdot a_{*n} \\ f_r^{[n]}\colon (A^n \times A)^r \times (A^n \times X) \longrightarrow B \quad &: \end{aligned}$$

$$\left(\begin{array}{ccc|c} a_{11} & \ldots & a_{1n} & a_{1*} \\ \vdots & & & \vdots \\ a_{r1} & \ldots & a_{rn} & a_{r*} \\ \hline a_{*1} & \ldots & a_{*n} & x \end{array}\right) \mapsto \sum_{\substack{s \le n \,\&\, s \le r \\ \&\, \mathsf{ramp}^s_{r,n}(\alpha \mid \gamma)}} f_{r+n-s}(x)\cdot a_{\alpha_1 1}\cdot \ldots \cdot a_{\alpha_n n}\cdot a_{\gamma_1 *}\cdot \ldots \cdot a_{\gamma_{r-s}*}$$

where $\mathsf{ramp}^s_{r,n}(\alpha \mid \gamma)$ is a condition defined as follows.

α is an injection $\{1, 2, \ldots, n\} \longrightarrow \{*, 1, 2, \ldots, r\}$ (which we shall denote $n \longrightarrow r \cup *$ for brevity), with the property that the inverse image $\alpha^{-1}(\{*\})$ of $*$ has size $r - s$. γ is an injection $r - s \longrightarrow r$, whose image is disjoint from that of α (meaning no α_i can equal any γ_j). The effect of this is to assign s linear arguments of the form a_{ij}, $n - s$ arguments of the form a_{*j}, and $r - s$ arguments of the form a_{j*} in such a way that the selected linear arguments will include just one from each column and just one from each row (including the bottom row and the right column, but not $= x$, which is not linear).

Another way to express this is to say that there is an equality of bags:

$$\mathsf{ramp}^s_{r,n}(\alpha \mid \gamma) \text{ iff } \wr\, \alpha_1, \ldots, \alpha_n, \gamma_1, \ldots, \gamma_{r-s} \int = \wr (n - s) \cdot *, 1, \ldots, r \int$$

This is a little complicated, so we shall look at this several ways. First let us explicitly develop some of these terms. First consider $f^{[0]}$: this is just f as the ramp condition forces all the action into the indices for the a_{i*} and so there is no choice. We shall display this sequence in the term logic as follows:

$$\begin{array}{ccl}
\left(\begin{array}{|c}\hline x\end{array}\right) & \mapsto & f_*(x)\\
\left(\begin{array}{|c} a_{1*}\\ \hline x\end{array}\right) & \mapsto & f_1(x)\cdot a_{1*}\\
\left(\begin{array}{|c} a_{1*}\\ a_{2*}\\ \hline x\end{array}\right) & \mapsto & f_2(x)\cdot a_{1*}\cdot a_{2*}\\
& \cdots &
\end{array}$$

Now consider $f^{[1]}$:

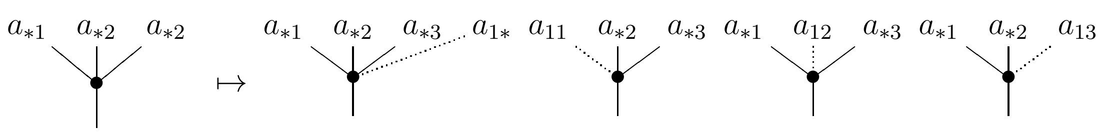
$$\begin{array}{ccl}
\left(\begin{array}{c|c}\hline a_{*1} & x\end{array}\right) & \mapsto & f_1(x)\cdot a_{*1}\\
\left(\begin{array}{c|c} a_{11} & a_{1*}\\ \hline a_{*1} & x\end{array}\right) & \mapsto & \begin{array}{l} f_2(x)\cdot a_{*1}\cdot a_{1*}\\ +f_1(x)\cdot a_{11}\end{array}\\
\left(\begin{array}{c|c} a_{11} & a_{1*}\\ a_{21} & a_{2*}\\ \hline a_{*1} & x\end{array}\right) & \mapsto & \begin{array}{l} f_3(x)\cdot a_{*1}\cdot a_{1*}\cdot a_{2*}\\ +f_2(x)\cdot a_{21}\cdot a_{1*}\\ +f_2(x)\cdot a_{11}\cdot a_{2*}\end{array}\\
& \cdots &
\end{array}$$

Notice how the domain is specified as a product with 2 components, then 4 components, then 6 components, ...

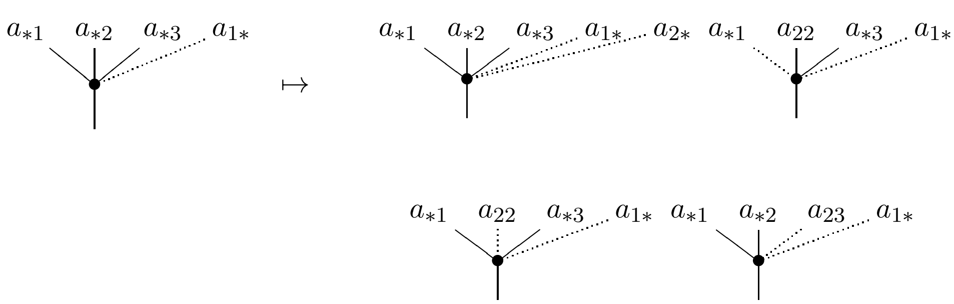
Here are the first few terms of $f^{[2]}$:

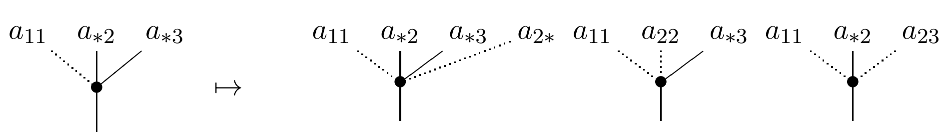
$$
\begin{aligned}
\left(\begin{array}{c|c} \overline{a_{*1}\, a_{*2}} & \overline{x} \end{array}\right) &\mapsto f_2(x)\cdot a_{*1}\cdot a_{*2} \\
\left(\begin{array}{c|c} a_{11}\, a_{12} & a_{1*} \\ \hline a_{*1} a_{*2} & x \end{array}\right) &\mapsto \begin{array}{l} f_3(x)\cdot a_{*1}\cdot a_{*2}\cdot a_{1*} \\ +f_2(x)\cdot a_{11}\cdot a_{*2} \\ +f_2(x)\cdot a_{*1}\cdot a_{12} \end{array} \\
\left(\begin{array}{c|c} a_{11}\, a_{12} & a_{1*} \\ a_{21}\, a_{22} & a_{2*} \\ \hline a_{*1}\, a_{*2} & x \end{array}\right) &\mapsto \begin{array}{l} f_4(x)\cdot a_{*1}\cdot a_{*2}\cdot a_{1*}\cdot a_{2*} \\ +f_3(x)\cdot a_{21}\cdot a_{*2}\cdot a_{1*} \\ +f_3(x)\cdot a_{*1}\cdot a_{22}\cdot a_{1*} \\ +f_3(x)\cdot a_{11}\cdot a_{*2}\cdot a_{2*} \\ +f_3(x)\cdot a_{*1}\cdot a_{12}\cdot a_{2*} \\ +f_2(x)\cdot a_{11}\cdot a_{22} \\ +f_2(x)\cdot a_{21}\cdot a_{12} \end{array} \\
&\cdots
\end{aligned}
$$

We may visualize this as a combinatoric process in which we start with a symmetric tree with one node with variables at the base level 0 and either adding a completely new dotted branch to an a at the current level or turn one of the existing solid branches into a dotted branch while attaching it to the new variable at the current level corresponding to its original variable:

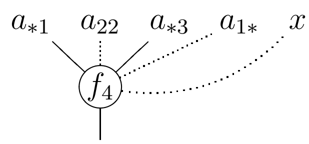


The next step then has two dotted branches entering each node. Here are the results of applying this process to two of the trees produced above:





As before we then clothe these trees with the particular forms:



to obtain the expressions for $f^{[n]}$ as described above.

The ramp condition can also be described by selecting partial isomorphisms from r to n, and thereby selecting certain entries from the following 'augmented' r by n variable base:

$$\left(\begin{array}{ccc|c} a_{11} & \dots & a_{1n} & a_{1*} \\ & \ddots & & \vdots \\ a_{r1} & \dots & a_{rn} & a_{r*} \\ \hline a_{*1} & \dots & a_{*n} & x \end{array}\right)$$

The selection process involves choosing a partial isomorphism σ between the top r rows and leftmost n columns, that is to say choosing places in the array subject to the following rules:

- At most one variable must be chosen from each of the top r rows (one cannot choose from the bottom row);
- At most one variable may be chosen from each of the leftmost n columns (one cannot choose from the rightmost column).

(We call this a "scatter set".) Here is an example when $n = 4$ and $r = 5$: σ is given by

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & a_{13} & a_{14} & a_{1*} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{2*} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{3*} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{4*} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{5*} \\ \hline a_{*1} & a_{*2} & a_{*3} & a_{*4} & x \end{array}\right) \overset{\sigma}{\rightsquigarrow} \left(\begin{array}{cccc|c} \boxed{a_{11}} & a_{12} & a_{13} & a_{14} & a_{1*} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{2*} \\ a_{31} & a_{32} & a_{33} & \boxed{a_{34}} & a_{3*} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{4*} \\ a_{51} & \boxed{a_{52}} & a_{53} & a_{54} & a_{5*} \\ \hline a_{*1} & a_{*2} & a_{*3} & a_{*4} & x \end{array}\right)$$

This selection can now can be "clothed" as above to represent the term f^σ:

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & a_{13} & a_{14} & a_{1*} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{2*} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{3*} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{4*} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{5*} \\ \hline a_{*1} & a_{*2} & a_{*3} & a_{*4} & x \end{array}\right) \mapsto \quad f^\sigma = f_6(x) \cdot a_{11} \cdot a_{52} \cdot a_{*3} \cdot a_{34} \cdot a_{2*} \cdot a_{4*}$$

The function f_6 is selected since if s coordinates are selected one needs $n + r - s$ linear arguments. In this case three coordinates were selected so $n + r - s = 4 + 5 - 3 = 6$ linear arguments are required. The linear arguments of f are determined by putting in the selected arguments and arguments from the bottom row and rightmost column corresponding to the rows and columns not containing a selected argument.

In general, given a partial isomorphism $\sigma\colon r \longrightarrow n$, let

$$\sigma_* = \sigma \cup \{(x,*) \mid x \in r \setminus \pi_1\sigma\} \cup \{(*,y) \mid y \in n \setminus \pi_2\sigma\}$$

Note that the set σ_* has $r+n-|\sigma|$ elements. Then we set $f^\sigma = f_m(x) \bullet_{(i,j)\,\in\,\sigma_*} a_{i,j}$, where $m = n+r-|\sigma|$.

Then it is clear that the term $f_r^{[n]}$ is then mapped under δ to the sum of the clothed scatter sets for the r by n augmented argument matrix with which we started:

2.2.3. Proposition.

$$f_r^{[n]} = \sum_{\sigma \in \mathrm{ParIso}(r,n)} f^\sigma$$

This perspective suggests a symmetry between $f_r^{[n]}$ and $f_n^{[r]}$ by transposition; of course this is only possible if the types A and X are the same.

Remark: It might help the reader if, at this stage, we recall that the intended interpretation of $f_r^{[n]}$ is the differential term

$$\frac{\mathsf{d}^r f(x) \cdot a_{1*} \cdot \cdots \cdot a_{n*}}{\mathsf{d}(x, a_{1*}, \ldots, a_{n*})}(a_{1*}, \ldots, a_{n*}, x) \cdot (a_{11}, \ldots, a_{1n}, a_{1*}) \cdot \cdots \cdot (a_{r1}, \ldots, a_{rn}, a_{r*})$$

We shall return to this point when we consider coalgebras later.

2.2.4. Lemma.

(i) $f_r^{[n]}$ *is an additive symmetric form in its first* r *arguments;*

(ii) $(f+g)_r^{[n]} = f_r^{[n]} + g_r^{[n]}$*;*

(iii) For any Cartesian left additive functor F, $\mathsf{BFaà}(F)(f^{[p]}) = (\mathsf{BFaà}(F)(f))^{[p]}$*;*

Proof.

(i) Recall that $f_r^{[n]}$ has the following form:

$f_r^{[n]}\colon (A^n \times A)^r \times (A^n \times X) \longrightarrow B$:

$$\left(\begin{array}{c|c} a_{11} \ldots a_{1n} & a_{1*} \\ \vdots & \vdots \\ a_{r1} \ldots a_{rn} & a_{r*} \\ \hline a_{*1} \ldots a_{*n} & x \end{array}\right) \mapsto \sum_{\substack{s \leq n \,\&\, s \leq r \\ \&\, \mathsf{ramp}^s_{r,n}(\alpha \mid \gamma)}} f_{r+n-s}(x) \cdot a_{\alpha_1 1} \cdot \ldots \cdot a_{\alpha_n n} \cdot a_{\gamma_{1*}} \cdot \ldots \cdot a_{\gamma_{r-s}*}$$

we must show that this term is additive in $(a_{i1} \ldots a_{in}\, a_{i*})$ $(i > 0)$. However, note that in each term there is only one occurrence of a variable from any i^{th}-level when $i > 0$. This makes the term additive in that variable and thus the whole term is additive in the whole level.

(ii) This is immediate as the "differential" in this second sense clearly preserves addition.

(iii) We have:

$$\begin{aligned} \mathsf{BFa\grave{a}}(F)(f^{[p]}) &= (F(f_*^{[p]}), F(f_1^{[p]}), F(f_2^{[p]}), \ldots) \\ &= (F(f)_*^{[p]}, F(f)_1^{[p]}, F(f)_2^{[p]}, \ldots) \\ &= \mathsf{BFa\grave{a}}(F)(f)^{[p]} \end{aligned}$$

where the penultimate step uses the fact that a left-additive functor preserves all the structure used to construct $f_r^{[n]}$.

■

Our first objective is to show why the following is true:

2.2.5. Proposition. $\delta\colon \mathsf{BFa\grave{a}}(\mathbb{X}) \longrightarrow \mathsf{BFa\grave{a}}(\mathsf{BFa\grave{a}}(\mathbb{X}))$ *is a functor which is (as a transformation* $\delta_{\mathbb{X}}$*) natural in* $\mathbb{X}$*.*

Of course, this is not so easy as δ itself involves some combinatorics and the composition in the Faà di Bruno categories is also combinatoric. In fact, just the composition in $\mathsf{BFa\grave{a}}(\mathsf{BFa\grave{a}}(\mathbb{X}))$ is a small challenge to understand. This is what we discuss next.

A map in $\mathsf{BFa\grave{a}}(\mathsf{BFa\grave{a}}(\mathbb{X}))$, $f\colon ((A,X),(A',X')) \longrightarrow ((B,Y),(B',Y'))$ is a doubly indexed array

$$\begin{array}{|llllll} \hline f_{*,*} & f_{1,*} & f_{2,*} & f_{3,*} & f_{4,*} & \cdots \\ f_{*,1} & f_{1,1} & f_{2,1} & f_{3,1} & f_{4,1} & \cdots \\ f_{*,2} & f_{1,2} & f_{2,2} & f_{3,2} & f_{4,2} & \cdots \\ f_{*,3} & f_{1,3} & f_{2,3} & f_{3,3} & f_{4,3} & \cdots \\ f_{*,4} & f_{1,4} & f_{2,4} & f_{3,4} & f_{4,4} & \cdots \\ \cdots & & & & & \end{array}$$

where the columns are maps in $\mathsf{BFa\grave{a}}(\mathbb{X})$:

$$\begin{aligned} f_* &= (f_{*,*}, f_{*,1}, \ldots)\colon (A',X') \longrightarrow (B',Y'); x \mapsto f_*(x) \\ f_n &= (f_{n,*}, f_{n,1}, \ldots)\colon (A,X)^n \times (A',X') = (A^n \times A', X^n \times X') \longrightarrow (B,Y); \\ & \quad\ (a_1, \ldots, a_n \mid x) \mapsto f_n(x) \cdot a_1 \cdot \ldots \cdot a_n \end{aligned}$$

so that the individual maps look like:

$$
\begin{array}{rcl}
f_{*,*} & : & X' \longrightarrow Y'; x \mapsto f_{*,*}(x) \\
f_{*,n} & : & A'^m \times X' \longrightarrow B'; (a_{*1}, \ldots, a_{*n} \mid x) \longrightarrow f_{*,n}(x) \cdot a_{*1} \cdot \ldots \cdot a_{*n} \\
f_{n,*} & : & X^n \times X' \longrightarrow Y; \left(\begin{array}{c} a_1 \\ \ldots \\ a_n \\ \hline x \end{array} \right) \mapsto f_{n,*} \left(\begin{array}{c} a_1 \\ \ldots \\ a_n \\ \hline x \end{array} \right) \\
f_{n,m} & : & (A^n \times A')^m \times (X^n \times X') \longrightarrow B; \\
 & & \left(\begin{array}{ccc|c} a_{11} & \ldots & a_{1m} & a_1 \\ & \ldots & & \\ a_{n1} & \ldots & a_{nm} & a_n \\ \hline a_{*1} & \ldots & a_{*m} & x \end{array} \right) \mapsto f_{n,m} \left(\begin{array}{c} a_1 \\ \vdots \\ a_n \\ \hline x \end{array} \right) \cdot \left(\begin{array}{c} a_{11} \\ \vdots \\ a_{n1} \\ \hline a_{*1} \end{array} \right) \cdot \ldots \cdot \left(\begin{array}{c} a_{1m} \\ \vdots \\ a_{nm} \\ \hline a_{*m} \end{array} \right)
\end{array}
$$

The first difficulty is to describe what a composition in $\mathsf{BFa\grave{a}}(\mathsf{BFa\grave{a}}(\mathbb{X}))$ looks like; to show that δ preserves composition we shall need a very concrete description of this composition. The composition in $\mathsf{BFa\grave{a}}(\mathsf{BFa\grave{a}}(\mathbb{X}))$ has two levels: the first step is a convolution over trees clothed from $\mathsf{BFa\grave{a}}(\mathbb{X})$: but the compositions in these trees must then, in turn, be translated out into convolutions. We shall now describe the effect of this in one step by describing $(fg)_{i,j}$ as

$$(fg)_{i,j} = \sum_{\tau \in \mathcal{T}_2^i} \sum_{\tau' \in \mathcal{T}_2^j} (f \star g)_{(\tau \times \tau')}$$

To explain this we must, in turn, unravel the notation $(f \star g)_{(\tau \times \tau')}$. To do this we first need to explain what the tree $\tau \times \tau'$ looks like. For this it is convenient to view a tree, τ, of height n, as being given by a chain of surjective maps:

$$X = X_0 \longrightarrow X_1 \longrightarrow \ldots \longrightarrow X_{n-2} \longrightarrow X_{n-1} = 1$$

where X_0 is the set of variables and X_i, for $0 < i < n$ are the symmetric operations. If τ' is another such tree of height n

$$Y = Y_0 \longrightarrow Y_1 \longrightarrow \ldots \longrightarrow Y_{n-2} \longrightarrow Y_{n-1} = 1$$

then their product, $\tau \times \tau'$ is the tree:

$$X \times Y = X_0 \times Y_0 \longrightarrow X_1 \times Y_1 \longrightarrow \ldots \longrightarrow X_{n-2} \times Y_{n-2} \longrightarrow X_{n-1} \times Y_{n-1} = 1$$

Note that in this tree each operation $h = (h_1, h_2) \in X_i \times Y_i$ has a matrix of arguments whose two dimensions are the number of arguments of h_1 and the number of arguments of h_2. Consider the two height 2 trees:

$$\tau_1 = \bullet_2(\bullet_2(a_1, a_3), \bullet_2(a_2, a_4)) \quad \text{and} \quad \tau_2 = \bullet_2(\bullet_3(a_1, a_2, a_4), \bullet_1(a_3))$$

the product tree may be presented as

$$\tau_1 \times \tau_2 = \bullet_{2,2}\left(\begin{array}{cc} \bullet_{2,3}\begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,4}\\ a_{3,1} & a_{3,2} & a_{3,4}\end{pmatrix} & \bullet_{2,1}\begin{pmatrix} a_{1,3}\\ a_{3,3}\end{pmatrix} \\ \bullet_{2,3}\begin{pmatrix} a_{2,1} & a_{2,2} & a_{2,4}\\ a_{4,1} & a_{4,2} & a_{4,4}\end{pmatrix} & \bullet_{2,1}\begin{pmatrix} a_{2,3}\\ a_{4,3}\end{pmatrix}\end{array}\right)$$

As before we must clothe the tree to form $(f \star g)_{(\tau_1 \times \tau_2)}$ where this time $f, g \in \mathsf{BFaà}(\mathsf{BFaà}(\mathbb{X}))$. Recall that this means each of f and g are specified by infinite dimensional arrays and each $f_{i,j}$ has an $i+1$ by $j+1$ dimensional argument list (as opposed to i by j dimensional shown above). Thus we need to fill in the "fringes" which is accomplished as follows:

$$(f \star g)_{(\tau_1 \times \tau_2)} = g_{2,2}\left(\begin{array}{cc|c} f_{2,3}\left(\begin{array}{ccc|c} a_{1,1} & a_{1,2} & a_{1,4} & a_1\\ a_{3,1} & a_{3,2} & a_{3,4} & a_3\\ \hline a_{*,1} & a_{*,2} & a_{*,4} & x\end{array}\right) & f_{2,1}\left(\begin{array}{c|c} a_{1,3} & a_1\\ a_{3,3} & a_3\\ \hline a_{*,3} & x\end{array}\right) & f_{2,*}\left(\begin{array}{c} a_1\\ a_3\\ \hline x\end{array}\right) \\ f_{2,3}\left(\begin{array}{ccc|c} a_{2,1} & a_{2,2} & a_{2,4} & a_2\\ a_{4,1} & a_{4,2} & a_{4,4} & a_4\\ \hline a_{*,1} & a_{*,2} & a_{*,4} & x\end{array}\right) & f_{2,1}\left(\begin{array}{c|c} a_{2,3} & a_2\\ a_{4,3} & a_4\\ \hline a_{*,3} & x\end{array}\right) & f_{2,*}\left(\begin{array}{c} a_2\\ a_4\\ \hline x\end{array}\right) \\ \hline f_{*,3}\left(\begin{array}{ccc|c} a_{*,1} & a_{*,2} & a_{*,4} & x\end{array}\right) & f_{*,1}\left(\begin{array}{c|c} a_{*,3} & x\end{array}\right) & f_{*,*}\,(x)\end{array}\right)$$

This then explains what the composition is in $\mathsf{BFaà}(\mathsf{BFaà}(\mathbb{X}))$.

Proof (of 2.2.5). Our objective is to show that δ preserves composition, thus that $\delta(f)\delta(g) = \delta(fg)$. We show that each is the sum of the same set of composites; formally, the proof is given by the following equations.

$$
\begin{aligned}
\delta(f)\delta(g) &= \sum_{\tau_1,\tau_2} (\delta(f) \star \delta(g))_{\tau_1\times\tau_2} && (1)\\
&= \sum_{\tau_1,\tau_2} \left(\left(\sum_{\sigma: i \longrightarrow j} f^{\sigma} \right)_{ij} \star \left(\sum_{\sigma': k \longrightarrow l} g^{\sigma'} \right)_{kl} \right)_{\tau_1\times\tau_2} && (2)\\
&= \sum_{\tau_1,\tau_2} \left(\sum_{\sigma'} g^{\sigma'} \right) \left(\sum_{\sigma_{ij}: \alpha_i \longrightarrow \beta_j} f^{\sigma_{ij}} \right)_{ij} && (3)\\
&= \sum_{\tau_1,\tau_2} \sum_{\sigma'} g^{\sigma'} \left(\sum_{\sigma_{ij}} f^{\sigma_{ij}} \right)_{ij} && (4)\\
&= \sum_{\tau_1,\tau_2} \sum_{\sigma'} g^{\sigma'} \left(\sum_{\sigma_{ij}} f^{\sigma_{ij}} \right)_{ij\in\sigma'} && (5)\\
&= \sum_{\tau_1,\tau_2} \sum_{\sigma',\sigma_{ij}, ij\in\sigma'} g^{\sigma}(\ldots, f^{\sigma_{ij}}, \ldots) && (6)\\
&= \sum_{\sigma: n \longrightarrow m} \sum_{\tau\in\mathcal{T}_{\sigma_*}} (f\star g)^{\sigma}_{\tau} && (7)\\
&= \delta(fg) && (8)
\end{aligned}
$$

Equations (1 – 3), and (8) are definitional; (4 – 6) are consequences of additivity; the combinatorial heart of the proof rests with equation (7), which amounts to the following lemma.

2.2.6. Lemma. *There is an equivalence between the following two sets of data:*

- *Partitions* $\tau_1 = (\alpha_1, \ldots, \alpha_k), \tau_2 = (\beta_1, \ldots, \beta_l)$ *and partial isomorphisms* $\sigma' \colon k \longrightarrow l$ *and* $\sigma_{ij} \colon \alpha_i \longrightarrow \beta_j$ *for* $(i,j) \in \sigma'$
- *Partial isomorphism* $\sigma \colon n \longrightarrow m$ *and a partition* τ *of* σ_*

where n *is the set partitioned by* τ_1*,* m *the set partitioned by* τ_2*, and* σ *is the union of the* σ_{ij}*.*

Proof (of 2.2.6). In this proof, we shall represent a partial isomorphism as the set of pairs (i,j) where $i \mapsto j$. To start with, we suppose we are given partitions $\tau_1 = (\alpha_1, \ldots, \alpha_k), \tau_2 = (\beta_1, \ldots, \beta_l)$ and partial isomorphisms $\sigma' \colon k \longrightarrow l$ and $\sigma_{ij} \colon \alpha_i \longrightarrow \beta_j$ for $(i,j) \in \sigma'$ We must construct τ, a partition of σ_*.

Recall that for a partial isomorphism $\sigma \colon n \longrightarrow m$,

$$\sigma_* = \sigma \cup \{(x,*) \mid x \in n \setminus \pi_1\sigma\} \cup \{(*,y) \mid y \in m \setminus \pi_2\sigma\}$$

and that $|\sigma_*| = n + m - |\sigma|$. In addition, we shall write $\sigma_i = \bigcup_j \sigma_{ij}$ and $\sigma_j = \bigcup_i \sigma_{ij}$ (and similarly for σ_{i*}, σ_{j*}).

We define a partition τ on σ_* as

$$\tau = \{\sigma_{ij*}\}_{(i,j)\in\sigma'} \;\cup\; \{((\alpha_i \setminus \pi_1\sigma_i) \times \{*\}) \setminus \sigma_{i*}\}_{i\in k} \;\cup\; \{(\{*\} \times (\beta_j \setminus \pi_2\sigma_j)) \setminus \sigma_{j*}\}_{j\in l}$$

This means that pairs from the same σ_{ij*} end up in the same partition, and pairs with a $*$ end up in the same partition if the "other" elements come from the same α_i or β_j (and aren't already in some σ_{ij*}).

This completes one direction of the equivalence.

Example:

$$\begin{aligned}
&\tau_1 = ((1,3),(2,5),(4,6))\\
&\tau_2 = ((1,2,4),(3),(5)) \quad (\text{so } k = l = 3)\\
&\sigma' \colon 3 \longrightarrow 3 = \{(1,3),(3,1)\} \quad (\text{so } e.g.\ (2,2) \text{ is not in } \sigma)\\
&\sigma_{13} \colon \{1,3\} \longrightarrow \{5\} = \{(3,5)\}\\
&\sigma_{31} \colon \{4,6\} \longrightarrow \{1,2,4\} = \{(4,4),(6,1)\}
\end{aligned}$$

Then $\sigma = \bigcup_{ij} \sigma_{ij} \colon 6 \longrightarrow 5 = \{(3,5),(4,4),(6,1)\}$ and $n = 6, m = 5, |\sigma| = 3$

Furthermore we have

$$\begin{aligned}
&\sigma_* = \{(3,5),(4,4),(6,1),(1,*),(2,*),(5,*),(*,2),(*,3)\}\\
&\sigma_{13*} = \{(3,5),(1,*)\}\\
&\sigma_{31*} = \{(4,4),(6,1),(*,2)\}
\end{aligned}$$

and so

$$\tau = (((4,4),(6,1),(*,2)),((3,5),(1,*)),((2,*),(5,*)),((*,3)))$$

Before proceeding to the other direction, let's consider what's going on. The given partitions and partial isomorphisms amount to this selection from a variable base:

$$\left(\begin{array}{ccc}
\begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,4} \\ a_{3,1} & a_{3,2} & a_{3,4} \end{pmatrix} & \begin{pmatrix} a_{1,3} \\ a_{3,3} \end{pmatrix} & \boxed{\begin{pmatrix} a_{1,5} \\ \boxed{a_{3,5}} \end{pmatrix}} \\
\begin{pmatrix} a_{2,1} & a_{2,2} & a_{2,4} \\ a_{5,1} & a_{5,2} & a_{5,4} \end{pmatrix} & \begin{pmatrix} a_{2,3} \\ a_{5,3} \end{pmatrix} & \begin{pmatrix} a_{2,5} \\ a_{5,5} \end{pmatrix} \\
\boxed{\begin{pmatrix} a_{4,1} & a_{4,2} & \boxed{a_{4,4}} \\ \boxed{a_{6,1}} & a_{6,2} & a_{6,4} \end{pmatrix}} & \begin{pmatrix} a_{4,3} \\ a_{6,3} \end{pmatrix} & \begin{pmatrix} a_{4,5} \\ a_{6,5} \end{pmatrix}
\end{array}\right)$$

and it's clear that what both sets of data are defining is the following term from the sums that define $\delta(f)\delta(g)$ and $\delta(fg)$:

$$g_4(x) \cdot (f_3(x) \cdot a_{44} \cdot a_{61} \cdot a_{*2}) \cdot (f_2(x) \cdot a_{35} \cdot a_{1*}) \cdot (f_2(x) \cdot a_{2*} \cdot a_{5*}) \cdot (f_1(x) \cdot a_{*3})$$

Now we turn to consider the other direction of the equivalence:

Suppose we are given a partial isomorphism $\sigma\colon n \longrightarrow m$ and a partition τ of σ_*. We must construct partitions $\tau_1 = (\alpha_1, \ldots, \alpha_k), \tau_2 = (\beta_1, \ldots, \beta_l)$ and partial isomorphisms $\sigma'\colon k \longrightarrow l$ and $\sigma_{ij}\colon \alpha_i \longrightarrow \beta_j$ for $(i,j) \in \sigma'$, of appropriate sizes.

Since τ is a partition of a matrix, we easily obtain partitions τ_1, τ_2 of the rows and columns: define $\pi_i'\gamma = \pi_i\gamma \setminus \{*\}$, and let $\hat{\tau} = (\gamma_1, \ldots, \gamma_p)$; then define $\tau_1 = (\pi_1'\gamma_i)_i$ and $\tau_2 = (\pi_2'\gamma_i)_i$

We can also construct partial isomorphisms from τ, by ignoring the pairs with $*$s, and taking the remaining pairs from each partition: let $\tau_1 = (\alpha_1, \ldots, \alpha_k)$ and $\tau_2 = (\beta_1, \ldots, \beta_l)$ and then define $\sigma' = \{(i,j) \mid (\alpha_i \times \beta_j) \cap \sigma_* \neq \emptyset\}$ and, for $(i,j) \in \sigma'$, define $\sigma_{ij} = (\alpha_i \times \beta_j) \cap \sigma_*$. Note that by this construction, σ is the union of these partial isomorphisms, as required.

This completes the other direction of the equivalence; that these processes are inverse is clear from their construction.

Example: Let's take the σ of the previous example, with a new τ:

$$\begin{aligned}
&\sigma\colon 6 \longrightarrow 5 = \{(3,5),(4,4),(6,1)\} \\
&\text{so} \quad \sigma_* = \{(3,5),(4,4),(6,1),(1,*),(2,*),(5,*),(*,2),(*,3)\} \\
&\tau = (((3,5)),((4,4),(6,1)),((1,*),(2,*),(*,3)),((5,*),(*,2)))
\end{aligned}$$

Then we obtain

$$\tau_1 = ((3),(4,6),(1,2),(5)) \quad \text{and} \quad \tau_2 = ((5),(4,1),(3),(2))$$

(note $k = l = 4$, and $n = 6$, $m = 5$ as required).

Then $\sigma' = \{(1,1),(2,2)\}$ (since $\{(3,5)\}$ is a pair from σ_* coming from the first partition in τ_1 and the first partition in τ_2, and $\{(4,4),(6,1)\}$ are pairs in σ_* coming from the second partition in τ_1 and the second partition in τ_2). Also $\sigma_{11} = \{(6,1)\}$ and $\sigma_{22} = \{(4,4),(3,5)\}$, whose union is the $\sigma\colon 6 \longrightarrow 5 = \{(3,5),(4,4),(6,1)\}$ we started with.

Again, one might wonder what's going on.

This time we have the following selection from the variable base:

$$\begin{pmatrix}
\boxed{\begin{pmatrix} \boxed{a_{3,5}} \end{pmatrix}} & \begin{pmatrix} a_{3,4} & a_{6,3} \end{pmatrix} & \begin{pmatrix} a_{3,3} \end{pmatrix} & \begin{pmatrix} a_{3,2} \end{pmatrix} \\
\begin{pmatrix} a_{4,5} \\ a_{6,5} \end{pmatrix} & \boxed{\begin{pmatrix} \boxed{a_{4,4}} & a_{4,1} \\ a_{6,4} & \boxed{a_{6,1}} \end{pmatrix}} & \begin{pmatrix} a_{4,3} \\ a_{6,3} \end{pmatrix} & \begin{pmatrix} a_{4,2} \\ a_{6,2} \end{pmatrix} \\
\begin{pmatrix} a_{1,5} \\ a_{2,5} \end{pmatrix} & \begin{pmatrix} a_{1,4} & a_{1,1} \\ a_{2,4} & a_{2,1} \end{pmatrix} & \begin{pmatrix} a_{1,3} \\ a_{2,3} \end{pmatrix} & \begin{pmatrix} a_{1,2} \\ a_{2,2} \end{pmatrix} \\
\begin{pmatrix} a_{5,5} \end{pmatrix} & \begin{pmatrix} a_{5,4} & a_{5,1} \end{pmatrix} & \begin{pmatrix} a_{5,3} \end{pmatrix} & \begin{pmatrix} a_{5,2} \end{pmatrix}
\end{pmatrix}$$

and the common function term corresponding to this is

$$g_6(x) \cdot (f_1(x) \cdot a_{35}) \cdot (f_2(x) \cdot a_{44} \cdot a_{61}) \cdot (f_2(x) \cdot a_{1*} \cdot a_{2*}) \cdot (f_1(x) \cdot a_{5*}) \cdot (f_1(x) \cdot a_{*3}) \cdot (f_1(x) \cdot a_{*2})$$

∎

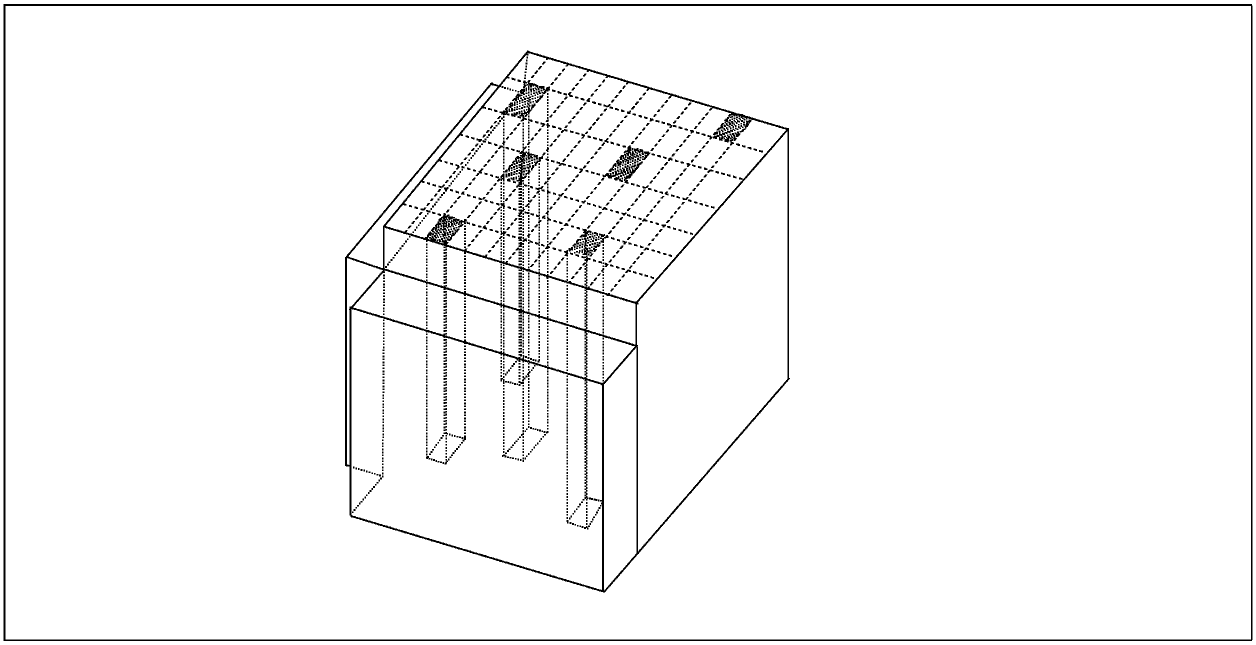


Figure 1: A 3-D scatter set

It is clear that δ preserves identities and that it is natural, as is evident from its combinatorial construction, completing our proof of 2.2.5. ∎

It is also easy to see that $\delta\varepsilon = F$ and $\delta \mathsf{BFa\grave{a}}(\varepsilon) = 1_{\mathsf{BFa\grave{a}}(\mathbb{X})}$. It remains to show that the comultiplication is associative:

Proof (of 2.2.2). To show it is a comonad we need to show that $\delta(\delta(f)) = \mathsf{BFa\grave{a}}(\delta)(\delta(f))$. We know that $\delta(\delta(f))$ is defined to be a "matrix" $(\delta(f)_r^{[n]})_{rn}$, where $\delta(f)_r^{[n]} = \sum_{\sigma : r \longrightarrow n} \delta(f)^\sigma$, in this case, the sum being a sum of "vectors"

$$\delta(f)^\sigma = f^{[|\sigma_*|]}(x) \bullet_{i \in \sigma} a_{i,\sigma(i)} \bullet_{i \notin \pi_1 \sigma} a_{i,*} \bullet_{j \notin \pi_2 \sigma} a_{*,j}$$

where the $a_{i,j}$ (*etc.*) are vectors of terms; notice that we have a 3-D variable base matrix here. Look at the m^{th} coordinate of $f^{[|\sigma_*|]}$: $f_m^{[|\sigma_*|]} = \sum_{\sigma' : (n+r-|\sigma|) \longrightarrow m} f^{\sigma'}$. Regard the $a_{i,j}, a_{i,*}, a_{*,j}$ as columns in a 3-D matrix. It is from these column-positions that σ picks out a scatter set, and σ' picks (from each column) a "height" (including "height 0" if it doesn't choose one). No two columns can have the same chosen "height" (other than 0). An illustration of such a situation may be seen in Figure 1, where the rows and columns with 0s in their entries are removed from the sides of the cube, and where the partial isomorphisms are indicated by shading the columns (σ) and indicating their heights (σ': note that two selected columns have 0 height, and so are not picked by σ').

With this intuition (in which we have replaced $*$ by 0), we can regard the situation as a three dimensional infinite array of maps $f_{r,n,m}$ where the domain of the map is given by a three dimensional matrix of variables:

$$(a_{i,j,k})_{\substack{i\,=\,0\,...\,p\\ j\,=\,0\,...\,q\\ k\,=\,0\,...\,r}}$$

The components of $f_{r,n,m}$ are given by choosing a "scattered" subset

$$\sigma \subset_{\mathrm{sc}} \{(i,j,k) \mid i = 0 \ldots r, j = 0 \ldots n, k = 0 \ldots m\}$$

where being scattered means that:

- For each $(i,j,k) \in \sigma$ at least two of i, j, and k are non-0 (*i.e.* not $*$);
- If $(i,j,k),(i',j',k') \in \sigma$ then either they are equal or $i \neq i'$, $j \neq j'$ and $k \neq k'$.

To each scattered subset we can associate a term f_σ where

$$f_\sigma = f_l(x) \bullet \left\{ a_{i,j,k} \;\middle|\; \begin{array}{l} (i,j,k) \in S \\ \vee(j = k = 0 \wedge (i,\_,\_) \notin S) \\ \vee(i = k = 0 \wedge (\_,j,\_) \notin S) \\ \vee(i = j = 0 \wedge (\_,\_,k) \notin S) \end{array} \right\}$$

(where $l = m + n + r - (\mid \sigma \mid + \mid \sigma' \mid))$, and then

$$f_{r,n,m} = (a_{i,j,k}) \mapsto \sum_\sigma f_\sigma$$

The point is that $\mathsf{BFaà}(\delta)(\delta(f))$ chooses exactly the same sort of 3-D scatter set, the term above being the natural 3-D generalization of σ_*, albeit possibly making the choice in a different order, and so produces the same sum, and so equals $\delta(\delta(f))$.

$\blacksquare$

We have arrived at the point now where we want to take the subfibration $\mathsf{Faà}$ as our main focus. We want to show that a coalgebra for this comonad, *i.e.* a Cartesian left additive category $\mathbb{X}$ with a functor $\mathcal{D}\colon \mathbb{X} \longrightarrow \mathsf{Faà}(\mathbb{X})$ such that $\mathcal{D}\varepsilon = 1$ and $\mathcal{D}\mathsf{Faà}(\mathcal{D}) = \mathcal{D}\delta$, in fact consists of a Cartesian left additive category $\mathbb{X}$ with a differential combinator D, making $\mathbb{X}$ a Cartesian differential category.

First let us observe what happens on objects. Setting $\mathcal{D}(X) = (\mathcal{D}_0(X), \mathcal{D}_1(X))$ and using the first coalgebra equation gives $X = \varepsilon(\mathcal{D}(X)) = \varepsilon(\mathcal{D}_0(X), \mathcal{D}_1(X)) = \mathcal{D}_1(X)$ so that we have $\mathcal{D}_1(X) = X$. Using the second coalgebra equation we have:

$$(\mathcal{D}\mathsf{Faà}(\mathcal{D}))(X) = \mathsf{Faà}(\mathcal{D})(\mathcal{D}(X)) = \mathsf{Faà}(\mathcal{D})(\mathcal{D}_0(X), X) = ((\mathcal{D}_0(\mathcal{D}_0(X)), \mathcal{D}_0(X))(\mathcal{D}_0(X), X))$$

$$(\mathcal{D}\delta)(X) = \delta(\mathcal{D}_0(X), X) = ((\mathcal{D}_0(X), \mathcal{D}_0(X)), (\mathcal{D}_0(X), X))$$

so that $\mathcal{D}_0(\mathcal{D}_0(X)) = \mathcal{D}_0(X)$, that is $\mathcal{D}_0$ is an idempotent on objects. But in $\mathsf{Faà}(\mathbb{X})$, the objects were pairs (A, A), rather than the more general (A, X). So in fact, since $\mathcal{D}_0(X) = \mathcal{D}_1(X) = X$, $\mathcal{D}_0$ is more than idempotent: it is the identity on objects.

The fact that we have a functor into the Faà di Bruno category means we have a functor into the ordinary bundle category and this immediately means that **[CD.1]**-**[CD.5]** [BCS 09] hold for the combinator

$$\frac{A \xrightarrow{f} B}{A \times A \xrightarrow[D[f]=f^{(1)}]{} B}$$

where we write $\mathcal{D}(f) = (f, f^{(1)}, f^{(2)}, \ldots)$, the first entry being forced to be the original function f by the unit equation for coalgebras. The second coalgebra equation, $\mathcal{D}\mathsf{Fa\grave{a}}(\mathcal{D}) = \mathcal{D}\delta$, then equates two different tables of functions:

$$\mathsf{Fa\grave{a}}(\mathcal{D})(\mathcal{D}(f)) = \begin{pmatrix} f & f^{(1)} & f^{(2)} & f^{(3)} & f^{(4)} & \ldots \\ f^{(1)} & (f^{(1)})^{(1)} & (f^{(2)})^{(1)} & (f^{(3)})^{(1)} & (f^{(4)})^{(1)} & \ldots \\ f^{(2)} & (f^{(1)})^{(2)} & (f^{(2)})^{(2)} & (f^{(3)})^{(2)} & (f^{(4)})^{(2)} & \ldots \\ f^{(3)} & (f^{(1)})^{(3)} & (f^{(2)})^{(3)} & (f^{(3)})^{(3)} & (f^{(4)})^{(3)} & \ldots \\ f^{(4)} & (f^{(1)})^{(4)} & (f^{(2)})^{(4)} & (f^{(3)})^{(4)} & (f^{(4)})^{(4)} & \ldots \\ \ldots \end{pmatrix}$$

$$\delta(\mathcal{D}(f)) = \begin{pmatrix} f & \mathcal{D}(f)_*^{[1]} & \mathcal{D}(f)_*^{[2]} & \mathcal{D}(f)_*^{[3]} & \mathcal{D}(f)_*^{[4]} & \ldots \\ f^{(1)} & \mathcal{D}(f)_1^{[1]} & \mathcal{D}(f)_1^{[2]} & \mathcal{D}(f)_1^{[3]} & \mathcal{D}(f)_1^{[4]} & \ldots \\ f^{(2)} & \mathcal{D}(f)_2^{[1]} & \mathcal{D}(f)_2^{[2]} & \mathcal{D}(f)_2^{[3]} & \mathcal{D}(f)_2^{[4]} & \ldots \\ f^{(3)} & \mathcal{D}(f)_3^{[1]} & \mathcal{D}(f)_3^{[2]} & \mathcal{D}(f)_3^{[3]} & \mathcal{D}(f)_3^{[4]} & \ldots \\ f^{(4)} & \mathcal{D}(f)_4^{[1]} & \mathcal{D}(f)_4^{[2]} & \mathcal{D}(f)_4^{[3]} & \mathcal{D}(f)_4^{[4]} & \ldots \\ \ldots \end{pmatrix}$$

If these are in fact equal, *i.e.* if $\mathbb{X}, \mathcal{D}$ form a coalgebra, then in particular $(f^{(1)})^{(1)} = \mathcal{D}(f)_1^{[1]}$, which when it is unravelled says:

$$\left(\begin{array}{c|c} a_{11} & a_1 \\ a_{*1} & x \end{array}\right) \mapsto (f^{(1)})^{(1)} \begin{pmatrix} a_{1*} \\ x \end{pmatrix} \cdot \begin{pmatrix} a_{11} \\ a_{*1} \end{pmatrix} = f^{(2)}(x) \cdot a_{*1} \cdot a_{1*} + f^{(1)}(x) \cdot a_{11}$$

which yields **[CD.6]** by setting $a_{*1} = 0$

$$(f^{(1)})^{(1)} \begin{pmatrix} a_{1*} \\ x \end{pmatrix} \cdot \begin{pmatrix} a_{11} \\ 0 \end{pmatrix} = f^{(1)}(x) \cdot a_{11}$$

and **[CD.7]** as

$$\begin{aligned} & (f^{(1)})^{(1)} \begin{pmatrix} a_{1*} \\ x \end{pmatrix} \cdot \begin{pmatrix} 0 \\ a_{*1} \end{pmatrix} \\ = \;& f^{(2)}(x) \cdot a_{*1} \cdot a_{1*} \\ = \;& f^{(2)}(x) \cdot a_{1*} \cdot a_{*1} \\ = \;& (f^{(1)})^{(1)} \begin{pmatrix} a_{*1} \\ x \end{pmatrix} \cdot \begin{pmatrix} 0 \\ a_{1*} \end{pmatrix} \end{aligned}$$

We have now proven:

2.2.7. Proposition. *Every coalgebra of the Faà di Bruno comonad* $\mathsf{Faà}$ *is a Cartesian differential category.*

It is worth noticing that we have in fact proved more: the coalgebra is in the following sense determined by the differential combinator of the Cartesian differential category: each $f^{(n)}$ is in fact determined by $D[f] = f^{(1)}$. For, as we have just seen,

$$\begin{array}{rcl} \mathcal{D}(f)_1^{[1]} \left(\begin{array}{c|c} 0 & a_1 \\ a_{*1} & x \end{array} \right) & = & (f^{(1)})^{(1)} \left(\begin{array}{c} a_{1*} \\ x \end{array} \right) \cdot \left(\begin{array}{c} 0 \\ a_{*1} \end{array} \right) \\ & = & f^{(2)}(x) \cdot a_{*1} \cdot a_{1*} + f^{(1)}(x) \cdot 0 \\ & = & f^{(2)}(x) \cdot a_{*1} \cdot a_{1*} + 0 \\ & = & f^{(2)}(x) \cdot a_{*1} \cdot a_{1*} \end{array}$$

In this manner, we can reconstruct $f^{(2)}$ from $(f^{(1)})^{(1)}$, and similarly (by induction) $f^{(n+1)}$ from $(f^{(n)})^{(1)}$.

3. Higher-Order Differentials and Chain Rules

Our objective is now to show that for any Cartesian differential category the differential supplies a coalgebra structure map into the Faà di Bruno category:

$$\mathcal{D}\colon \mathbb{X} \longrightarrow \mathsf{Faà}(\mathbb{X}); \begin{array}{lcl} A & \mapsto & (A, A) \\ f & \mapsto & (f, f^{(1)}, f^{(2)}, \ldots) \end{array}$$

which is a section of the fibration $\varepsilon\colon \mathsf{Faà}(\mathbb{X}) \longrightarrow \mathbb{X}$, where the $f^{(n)}$ are the higher-order derivatives defined below, based on the combinator D. The existence of this functor is the structural generalization of Faà di Bruno's original result on the higher differentials.

The technical details in this section require us to link the Faà di Bruno convolution composition to the higher-order chain rules. This involves some calculations for which we shall use the term logic [BCS 09].

In the notation of the previous section, our specific goal is to show, by induction on j, that $\mathcal{D}(f)_j^{[i]} = (f^{[i]})^{(j)}$, and hence that the two tables representing $\mathsf{Faà}(\mathcal{D})(\mathcal{D}(f))$ and $\delta(\mathcal{D}(f))$ are equal. (The other coalgebra equation is obvious from the definition of $\mathcal{D}(f)$.)

3.1. Higher-order derivatives. First we define the higher-order derivatives recursively by:

$$\begin{array}{rcl} \dfrac{\mathsf{d}^{(1)}t}{\mathsf{d}x}(p) \cdot u & = & \dfrac{\mathsf{d}t}{\mathsf{d}x}(p) \cdot u \\ \dfrac{\mathsf{d}^{(n)}t}{\mathsf{d}x}(p) \cdot u_1 \cdot \ldots \cdot u_n & = & \dfrac{\mathsf{d}\frac{\mathsf{d}^{(n-1)}t}{\mathsf{d}x}(x) \cdot u_1 \cdot \ldots \cdot u_{n-1}}{\mathsf{d}x}(p) \cdot u_n \end{array}$$

Note that it is immediate from this definition that the n^{th}-order derivative is additive in each $u_1, \ldots, u_n$ thus the higher differentials define symmetric forms.

In order understand higher-order derivatives it is useful to start with some basic observations:

3.1.1. Lemma.

(i) $\frac{\mathsf{d}t[x+s/y]}{\mathsf{d}x}(0)\cdot u = \frac{\mathsf{d}t}{\mathsf{d}y}(s)\cdot u$, *x not free in s*

(ii) $\frac{\mathsf{d}\frac{\mathsf{d}t}{\mathsf{d}x}(x)\cdot u_1}{\mathsf{d}x}(s)\cdot u_2 = \frac{\mathsf{d}\frac{\mathsf{d}t}{\mathsf{d}x}(x)\cdot u_2}{\mathsf{d}x}(s)\cdot u_1$ *that is* $\frac{\mathsf{d}^{(2)}t}{\mathsf{d}x}(s)\cdot u_1\cdot u_2 = \frac{\mathsf{d}^{(2)}t}{\mathsf{d}x}(s)\cdot u_2\cdot u_1$, *x not free in* u_1, u_2

(iii) $\frac{\mathsf{d}^{(n)}t}{\mathsf{d}x}(s)\cdot u_1\cdot\ldots\cdot u_n = \frac{\mathsf{d}^{(n)}t}{\mathsf{d}x}(s)\cdot u_{\sigma(1)}\cdot\ldots\cdot u_{\sigma(n)}$ *for any* $\sigma\in\mathcal{S}_n$.

Remark: Item (ii) is actually equivalent to **[Dt.7]** [BCS 09], and is a more direct translation of **[CD.7]**.

Proof.

(i)

$$\begin{aligned}
\frac{\mathsf{d}t[x+s/y]}{\mathsf{d}x}(0)\cdot u &= \frac{\mathsf{d}t}{\mathsf{d}y}(s)\cdot\left(\frac{\mathsf{d}(x+s)}{\mathsf{d}x}(0)\cdot u\right) \quad \text{because } x+s[0/x]=0+s=s\\
&= \frac{\mathsf{d}t}{\mathsf{d}y}(s)\cdot\left(\frac{\mathsf{d}x}{\mathsf{d}x}(0)\cdot u+\frac{\mathsf{d}s}{\mathsf{d}x}(0)\cdot u\right)\\
&= \frac{\mathsf{d}t}{\mathsf{d}y}(s)\cdot u
\end{aligned}$$

(ii) The problem in this calculation is that in order to swap u_1 and u_2 we need to differentiate with respect to variables not involved in the position of the differential. The first part of the lemma allows us to arrange this - but we have to use it twice:

$$\begin{aligned}
\frac{\mathsf{d}\frac{\mathsf{d}t}{\mathsf{d}x}(x)\cdot u_1}{\mathsf{d}x}(s)\cdot u_2 &= \frac{\mathsf{d}\frac{\mathsf{d}t}{\mathsf{d}x}(x')\cdot u_1}{\mathsf{d}x'}(s)\cdot u_2\\
&= \frac{\mathsf{d}\frac{\mathsf{d}t[x_1+x'/x]}{\mathsf{d}x_1}(0)\cdot u_1}{\mathsf{d}x'}(s)\cdot u_2\\
&= \frac{\mathsf{d}\frac{\mathsf{d}t[x_1+x'/x]}{\mathsf{d}x_1}(0)\cdot u_1[x_2+s/x']}{\mathsf{d}x_2}(0)\cdot u_2\\
&= \frac{\mathsf{d}\frac{\mathsf{d}t[x_1+x_2+s/x]}{\mathsf{d}x_1}(0)\cdot u_1}{\mathsf{d}x_2}(0)\cdot u_2\\
&= \frac{\mathsf{d}\frac{\mathsf{d}t[x_1+x_2+s/x]}{\mathsf{d}x_2}(0)\cdot u_2}{\mathsf{d}x_1}(0)\cdot u_1 \quad \text{by } \textbf{[Dt.7]}\\
&= \frac{\mathsf{d}\frac{\mathsf{d}t}{\mathsf{d}x}(x)\cdot u_2}{\mathsf{d}x}(s)\cdot u_1
\end{aligned}$$

(iii) The result is trivially true for $n=1$ and the previous result proves it for $n=2$. Assuming it true for $n+1$ we show it is true for $n+2$ by showing how we can exchange

u_{n+1} and u_{n+2}:

$$\begin{aligned}
\frac{\mathsf{d}^{(n+2)}t}{\mathsf{d}x}(s)\cdot u_1\cdot\ldots\cdot u_{n+1}\cdot u_{n+2} &= \frac{\mathsf{d}\frac{\mathsf{d}^{(n+1)}t}{\mathsf{d}x}(x)\cdot u_1\cdot\ldots\cdot u_{n+1}}{\mathsf{d}x}(p)\cdot u_{n+2}\\
&= \frac{\mathsf{d}\frac{\mathsf{d}\frac{\mathsf{d}^{(n)}t}{\mathsf{d}x}(x)\cdot u_1\cdot\ldots\cdot u_n}{\mathsf{d}x}(x)\cdot u_{n+1}}{\mathsf{d}x}(p)\cdot u_{n+2}\\
&= \frac{\mathsf{d}\frac{\mathsf{d}\frac{\mathsf{d}^{(n)}t}{\mathsf{d}x}(x)\cdot u_1\cdot\ldots\cdot u_n}{\mathsf{d}x}(x)\cdot u_{n+2}}{\mathsf{d}x}(p)\cdot u_{n+1}\\
&= \frac{\mathsf{d}^{(n+2)}t}{\mathsf{d}x}(s)\cdot u_1\cdot\ldots\cdot u_{n+2}\cdot u_{n+1}
\end{aligned}$$

∎

This demonstrates that the second-order differential gives rise to a symmetric form and whence that the n^{th}-order differential gives a symmetric form.

A useful fact to note is that differentiating a higher-differential with respect to one of its linear components satisfies the generalization of [**Dt.6**]:

3.1.2. COROLLARY. *In any Cartesian differential category*

$$\frac{\mathsf{d}\frac{\mathsf{d}^{(n)}t}{\mathsf{d}z}(s)\cdot u_1\cdot\ldots\cdot x\cdot\ldots\cdot u_n}{\mathsf{d}x}(s')\cdot u_r = \frac{\mathsf{d}^{(n)}t}{\mathsf{d}z}(s)\cdot u_1\cdot\ldots\cdot u_r\cdot\ldots\cdot u_n$$

PROOF. We use symmetry to move the variable into the last place at which point the linearity of the differential can be used:

$$\begin{aligned}
&\frac{\mathsf{d}\frac{\mathsf{d}^{(n)}t}{\mathsf{d}z}(s)\cdot u_1\cdot\ldots\cdot x\cdot\ldots\cdot u_n}{\mathsf{d}x}(s')\cdot u_r\\
&= \frac{\mathsf{d}\frac{\mathsf{d}^{(n)}t}{\mathsf{d}z}(s)\cdot u_1\cdot\ldots\cdot u_{r-1}\cdot u_{r+1}\ldots\cdot u_n\cdot x}{\mathsf{d}x}(s')\cdot u_r\\
&= \frac{\mathsf{d}\frac{\mathsf{d}\frac{\mathsf{d}^{(n-1)}s}{\mathsf{d}t}(u_1\cdot\ldots\cdot u_{r-1}\cdot u_{r+1}\ldots\cdot u_n)\cdot}{\mathsf{d}z}(z)\cdot x}{\mathsf{d}x}(s')\cdot u_r\\
&= \frac{\mathsf{d}\frac{\mathsf{d}^{(n-1)}s}{\mathsf{d}t}(u_1\cdot\ldots\cdot u_{r-1}\cdot u_{r+1}\ldots\cdot u_n)\cdot}{\mathsf{d}z}(z)\cdot u_r\\
&= \frac{\mathsf{d}^{(n)}t}{\mathsf{d}z}(s)\cdot u_1\cdot\ldots\cdot u_{r-1}\cdot u_{r+1}\cdot\ldots\cdot u_n\cdot u_r\\
&= \frac{\mathsf{d}^{(n)}t}{\mathsf{d}z}(s)\cdot u_1\cdot\ldots\cdot u_r\cdot\ldots\cdot u_n
\end{aligned}$$

∎

3.2. Chain rules. Let us start by explicitly calculating the second-order chain rule. We do this by first proving a lemma which tells us how to differentiate differentials with respect to variables not in the term. We shall then generalize this calculation to obtain the general form of the higher order chain rule in a Cartesian differential category.

3.2.1. Lemma. *When $y \notin t$ we have*

$$\frac{\mathsf{d}\frac{\mathsf{d}t}{\mathsf{d}x}(p)\cdot u}{\mathsf{d}y}(p')\cdot u' = \frac{\mathsf{d}^{(2)}t}{\mathsf{d}x}(p[p'/y])\cdot u[p'/y]\cdot\left(\frac{\mathsf{d}p}{\mathsf{d}y}(p')\cdot u'\right)+\frac{\mathsf{d}t}{\mathsf{d}x}(p[p'/y])\cdot\left(\frac{\mathsf{d}u}{\mathsf{d}y}(p')\cdot u'\right)$$

Proof.

$$\begin{aligned}
&\frac{\mathsf{d}\frac{\mathsf{d}t}{\mathsf{d}x}(p)\cdot u}{\mathsf{d}y}(p')\cdot u' \\
&= \frac{\mathsf{d}\frac{\mathsf{d}t}{\mathsf{d}x}(a)\cdot b}{\mathsf{d}(a,b)}(p[p'/y],u[p'/y])\cdot(\frac{\mathsf{d}p}{\mathsf{d}y}(p')\cdot u',\frac{\mathsf{d}u}{\mathsf{d}y}(p')\cdot u') \\
&= \frac{\mathsf{d}\frac{\mathsf{d}t}{\mathsf{d}x}(a)\cdot u[p'/y]}{\mathsf{d}a}(p[p'/y])\cdot\left(\frac{\mathsf{d}p}{\mathsf{d}y}(p')\cdot u'\right)+\frac{\mathsf{d}\frac{\mathsf{d}t}{\mathsf{d}x}(p[p'/y])\cdot b}{\mathsf{d}b}(u[p'/y])\cdot\left(\frac{\mathsf{d}u}{\mathsf{d}y}(p')\cdot u'\right) \\
&= \frac{\mathsf{d}^{(2)}t}{\mathsf{d}x}(p[p'/y])\cdot u[p'/y]\cdot\left(\frac{\mathsf{d}p}{\mathsf{d}y}(p')\cdot u'\right)+\frac{\mathsf{d}t}{\mathsf{d}x}(p[p'/y])\cdot\left(\frac{\mathsf{d}u}{\mathsf{d}y}(p')\cdot u'\right)
\end{aligned}$$

■

We may apply this to get the second-order chain rule:

$$\begin{aligned}
\frac{\mathsf{d}^{(2)}f(g(x))}{\mathsf{d}x}(p)\cdot u_1\cdot u_2 &= \frac{\mathsf{d}\frac{\mathsf{d}f(g(y))}{\mathsf{d}y}(x)\cdot u_1}{\mathsf{d}x}(p)\cdot u_2 \\
&= \frac{\mathsf{d}\frac{\mathsf{d}f(z)}{\mathsf{d}z}(g(x))\cdot\left(\frac{\mathsf{d}g(y)}{\mathsf{d}y}(x)\cdot u_1\right)}{\mathsf{d}x}(p)\cdot u_2 \\
&= \frac{\mathsf{d}^{(2)}f(z)}{\mathsf{d}z}(g(p))\cdot\left(\frac{\mathsf{d}g(y)}{\mathsf{d}y}(p)\cdot u_1\right)\cdot\left(\frac{\mathsf{d}g(x)}{\mathsf{d}x}(p)\cdot u_2\right) \\
&\quad +\frac{\mathsf{d}f(x)}{\mathsf{d}x}(g(p))\cdot\left(\frac{\mathsf{d}^{(2)}g(x)}{\mathsf{d}x}(p)\cdot u_1\cdot u_2\right)
\end{aligned}$$

With each function $[x \mapsto f(x)]$ in a differential category we can associate a series of functions:

$$\begin{aligned}
f^{(1)}(z)\cdot u_1 &= \frac{\mathsf{d}f(x)}{\mathsf{d}x}(z)\cdot u_1 \\
f^{(2)}(z)\cdot u_1\cdot u_2 &= \frac{\mathsf{d}^{(2)}f(x)}{\mathsf{d}z}(u_1\cdot u_2)\cdot \\
&\cdots \\
f^{(n)}(z)\cdot u_1\cdot\ldots\cdot u_n &= \frac{\mathsf{d}^{(2)}f(x)}{\mathsf{d}z}(u_1\cdot\ldots\cdot u_n)\cdot \\
&\cdots
\end{aligned}$$

This allows us in a Cartesian differential category, given a sequence of functions

$$X_1 \xrightarrow{f_1} X_2 \xrightarrow{f_2} X_3 \to \ldots \xrightarrow{f_n} X_{n+1}$$

and a symmetric tree τ of height n and width r with variables $U = \{u_1, \ldots, u_r\}$ (which we should think of as the leaves of a tree) to define the meaning of $(f_1 \star \ldots \star f_n)_\tau(z)$ inductively by:

- Variables $(\_)_{\iota_0^{u_i}} = u_i$ become variables. (Recall $\iota_0^{u_i}$ is defined in Proposition 1.2.1 as the unique tree of height 0, width 1, and leaf or variable u_i.)

- Given a chain of one or more maps we define

$$\begin{aligned}&(f_1 \star \ldots \star f_{n+1})_{\bullet_m(\tau_1,\ldots,\tau_m)}(z)\\ &= \frac{\mathsf{d}^{(m)} f_{n+1}(x)}{\mathsf{d}x}(f_{n-1}(\ldots, f_1(z)\ldots)) \cdot (f_1 \star \ldots \star f_n)_{\tau_1}(z) \cdot \ldots \cdot (f_1 \star \ldots \star f_n)_{\tau_m}(z)\end{aligned}$$

For T a bag of trees we define:

$$(f_1 \star \ldots \star f_{n+1})_T(z) = \sum_{\tau \in T} (f_1 \star \ldots \star f_{n+1})_\tau(z)$$

Examining the second-order chain rule we see that it is:

$$f^{(2)}(z) \cdot (g^{(1)}(z) \cdot u_1) \cdot (g^{(1)}(z) \cdot u_2) + f^{(1)}(z) \cdot (g^{(2)}(z) \cdot u_1 \cdot u_2))$$

which is precisely $(f \star g)_{\mathcal{T}_2^2}(z)$ relating the higher-order chain rules to these symmetric trees.

We now have the following key result:

3.2.2. Proposition. *In any Cartesian differential category:*

(i) The first differential of a composite of functions is given by :

$$\frac{\mathsf{d} f_n(\ldots, f_1(x)\ldots)}{\mathsf{d}x}(z) \cdot u = (f_1 \star \ldots \star f_n)_{\iota_n^u}(z)$$

(ii) For any bag of trees, T, of height n and width r on variables $\{u_1, \ldots, u_r\}$

$$\frac{\mathsf{d}(f_1 \star \ldots \star f_n)_T(x)}{\mathsf{d}x}(z) \cdot u_{r+1} = (f_1 \star \ldots \star f_n)_{\partial_{u_{r+1}} T(x)}(z)$$

Proof.

(i) We may use the ordinary chain rule to prove the result. First note that for $n = 0$ the result holds as:

$$(\_)^{u}_{\iota_0}(z) = u = \frac{\mathsf{d}z}{\mathsf{d}z}(z) \cdot u$$

Suppose now the result holds for n then:

$$\begin{aligned}
\frac{\mathsf{d}f_{n+1}(f_n(\ldots\ f_1(x)\ldots)}{\mathsf{d}x}(z) \cdot u &= \frac{\mathsf{d}f_{n+1}(x)}{\mathsf{d}x}(f_n(\ldots\ f_1(z)\ldots)) \cdot \frac{\mathsf{d}f_n(\ldots\ f_1(z)\ldots)}{\mathsf{d}z}(z) \cdot u \\
&= \frac{\mathsf{d}f_{n+1}(x)}{\mathsf{d}x}(f_n(\ldots\ f_1(z)\ldots)) \cdot (f_1 \star \ldots \star f_n)_{\iota^u_n}(z) \\
&= (f_1 \star \ldots \star f_n \star f_{n+1})_{\iota^u_{n+1}}(z)
\end{aligned}$$

(ii) The differential of a variable $\frac{\mathsf{d}(\star)_{u_i}(x)}{\mathsf{d}z}(z) \cdot u_{r+1} = 0$.

The differential of $(f_1 \star \ldots \star f_{n+1})_{\bullet_r(\tau_1,\ldots,\tau_r)}(z)$ is

$$\begin{aligned}
&\frac{\mathsf{d}(f_1 \star \ldots \star f_{n+1})_{\bullet_r(\tau_1,\ldots,\tau_r)}(z)}{\mathsf{d}z}(z) \cdot u_{r+1} \\
=\ & \frac{\mathsf{d}\frac{\mathsf{d}^{(r)}f_{n+1}(x)}{\mathsf{d}x}(f_n(\ldots\ f_1(z)\ldots)) \cdot (f_1 \star \ldots \star f_n)_{\tau_1}(z) \cdot \ldots \cdot (f_1 \star \ldots \star f_n)_{\tau_m}(z)}{\mathsf{d}z}(z) \cdot u_{r+1} \\
=\ & \frac{\mathsf{d}\frac{\mathsf{d}^{(r)}f_{n+1}(x)}{\mathsf{d}x}(x) \cdot x_1 \cdot \ldots \cdot x_r}{\mathsf{d}(x, x_1, \ldots, x_r)} \begin{pmatrix} f_n(\ldots\ f_1(z)\ldots), \\ (f_1 \star \ldots \star f_n)_{\tau_1}(z), \\ \ldots, \\ (f_1 \star \ldots \star f_n)_{\tau_r}(z) \end{pmatrix} \cdot \begin{matrix} (\frac{\mathsf{d}f_n(\ldots\ f_1(z)\ldots)}{\mathsf{d}z}(z) \cdot u_{r+1}) \cdot \\ (\frac{\mathsf{d}(f_1 \star \ldots \star f_n)_{\tau_1}(z)}{\mathsf{d}z}(z) \cdot u_{r+1}) \cdot \\ \ldots \cdot \\ (\frac{\mathsf{d}(f_1 \star \ldots \star f_n)_{\tau_r}(z)}{\mathsf{d}z}(z) \cdot u_{r+1}) \end{matrix} \\
=\ & \frac{\mathsf{d}\frac{\mathsf{d}^{(r)}f_{n+1}(x)}{\mathsf{d}x}(x) \cdot x_1 \cdot \ldots \cdot x_r}{\mathsf{d}(x, x_1, \ldots, x_r)} \begin{pmatrix} f_n(\ldots\ f_1(z)\ldots), \\ (f_1 \star \ldots \star f_n)_{\tau_1}(z), \\ \ldots, \\ (f_1 \star \ldots \star f_n)_{\tau_r}(z) \end{pmatrix} \cdot \begin{matrix} ((f_1 \star \ldots \star f_n)_{\iota_n}(z) \cdot \\ ((f_1 \star \ldots \star f_n)_{\partial_{u_{r+1}}\tau_1}(z) \cdot \\ \ldots \cdot \\ ((f_1 \star \ldots \star f_n)_{\partial_{u_{r+1}}\tau_r}(z)) \end{matrix} \\
=\ & \frac{\mathsf{d}\frac{\mathsf{d}^{(r)}f_{n+1}(x)}{\mathsf{d}x}(x) \cdot (f_1 \star \ldots \star f_n)_{\tau_1}(z) \cdot \ldots \cdot (f_1 \star \ldots \star f_n)_{\tau_r}(z)}{\mathsf{d}x}(z) \cdot (f_1 \star \ldots \star f_n)_{\iota_n}(z) \\
&+ \frac{\mathsf{d}\frac{\mathsf{d}^{(r)}f_{n+1}(x)}{\mathsf{d}x}(f_n(\ldots f_1(z)\ldots)) \cdot x_1 \cdot (f_1 \star \ldots \star f_n)_{\tau_2}(z) \cdot \ldots \cdot (f_1 \star \ldots \star f_n)_{\tau_r}(z)}{\mathsf{d}x_1}(z) \cdot (f_1 \star \ldots \star f_n)_{\partial_{u_{r+1}}\tau_1}(z) \\
&+ \ldots \\
&+ \frac{\mathsf{d}\frac{\mathsf{d}^{(r)}f_{n+1}(x)}{\mathsf{d}x}(f_n(\ldots f_1(z)\ldots)) \cdot (f_1 \star \ldots \star f_n)_{\tau_1}(z) \cdot\ \ldots\ \cdot x_r}{\mathsf{d}x_r}(z) \cdot (f_1 \star \ldots \star f_n)_{\partial_{u_{r+1}}\tau_r}(z) \\
=\ & \frac{\mathsf{d}^{(r+1)}f_{n+1}(x)}{\mathsf{d}x}((f_1 \star \ldots \star f_n)_{\iota_n}(z)) \cdot (f_1 \star \ldots \star f_n)_{\tau_1}(z) \cdot \ldots \cdot (f_1 \star \ldots \star f_n)_{\tau_r}(z) \\
&+ \frac{\mathsf{d}^{(r)}f_{n+1}(x)}{\mathsf{d}x}(f_n(\ldots f_1(z)\ldots)) \cdot (f_1 \star \ldots \star f_n)_{\partial_{u_{r+1}}\tau_1}(z) \cdot (f_1 \star \ldots \star f_n)_{\tau_2}(z) \cdot \ldots \\
&\qquad\qquad \ldots \cdot (f_1 \star \ldots \star f_n)_{\tau_r}(z) \\
&+ \ldots \\
&+ \frac{\mathsf{d}^{(r)}f_{n+1}(x)}{\mathsf{d}x}(f_n(\ldots f_1(z)\ldots)) \cdot (f_1 \star \ldots \star f_n)_{\tau_1}(z) \cdot\ \ldots\ \cdot (f_1 \star \ldots \star f_n)_{\partial_{u_{r+1}}\tau_r}(z) \\
=\ & (f_1 \star \ldots \star f_n)_{\partial_{u_r}\tau}(z)
\end{aligned}$$

where we have used the first result and inductively the result on trees of lesser height.

■

We have (obviously) proved the generalization of Faà di Bruno's result:

3.2.3. Corollary. *In any Cartesian differential category:*

$$\frac{\mathsf{d}^{(n)} g(f(x))}{\mathsf{d}x}(z) \cdot u_1 \cdot \ldots \cdot u_n = (f \star g)_{\mathcal{T}_2^{\{u_1,\ldots,u_n\}}}(z)$$

Furthermore

$$\frac{\mathsf{d}^{(m)} f_n(f_{n-1}(\ldots(f(x))\cdots))}{\mathsf{d}x}(z) \cdot u_1 \cdots u_m = (f_1 \star f_2 \star \cdots \star f_n)_{\mathcal{T}_n^m}(z)$$

This means that for any Cartesian differential category $\mathbb{X}$ there is a functor

$$\mathcal{D}\colon \mathbb{X} \longrightarrow \mathsf{Faà}(\mathbb{X}); f \mapsto (f, f^{(1)}, f^{(2)}, \ldots)$$

We are now in a position to show this gives a coalgebra.

3.2.4. Theorem. *A Cartesian differential category is exactly a coalgebra for the Faà di Bruno comonad.*

Proof. We just need to show that these are equal:

$$\mathsf{Faà}(\mathcal{D})(\mathcal{D}(f)) = \begin{pmatrix} f & f^{(1)} & f^{(2)} & f^{(3)} & f^{(4)} & \ldots \\ f^{(1)} & (f^{(1)})^{(1)} & (f^{(2)})^{(1)} & (f^{(3)})^{(1)} & (f^{(4)})^{(1)} & \ldots \\ f^{(2)} & (f^{(1)})^{(2)} & (f^{(2)})^{(2)} & (f^{(3)})^{(2)} & (f^{(4)})^{(2)} & \ldots \\ f^{(3)} & (f^{(1)})^{(3)} & (f^{(2)})^{(3)} & (f^{(3)})^{(3)} & (f^{(4)})^{(3)} & \ldots \\ f^{(4)} & (f^{(1)})^{(4)} & (f^{(2)})^{(4)} & (f^{(3)})^{(4)} & (f^{(4)})^{(4)} & \ldots \\ \ldots & & & & & \end{pmatrix}$$

$$\delta(\mathcal{D}(f)) = \begin{pmatrix} f & \mathcal{D}(f)_*^{[1]} & \mathcal{D}(f)_*^{[2]} & \mathcal{D}(f)_*^{[3]} & \mathcal{D}(f)_*^{[4]} & \ldots \\ f^{(1)} & \mathcal{D}(f)_1^{[1]} & \mathcal{D}(f)_1^{[2]} & \mathcal{D}(f)_1^{[3]} & \mathcal{D}(f)_1^{[4]} & \ldots \\ f^{(2)} & \mathcal{D}(f)_2^{[1]} & \mathcal{D}(f)_2^{[2]} & \mathcal{D}(f)_2^{[3]} & \mathcal{D}(f)_2^{[4]} & \ldots \\ f^{(3)} & \mathcal{D}(f)_3^{[1]} & \mathcal{D}(f)_3^{[2]} & \mathcal{D}(f)_3^{[3]} & \mathcal{D}(f)_3^{[4]} & \ldots \\ f^{(4)} & \mathcal{D}(f)_4^{[1]} & \mathcal{D}(f)_4^{[2]} & \mathcal{D}(f)_4^{[3]} & \mathcal{D}(f)_4^{[4]} & \ldots \\ \ldots & & & & & \end{pmatrix}$$

which in turn essentially means we need to show that $\mathcal{D}(f)_j^{[i]} = (f^{(i)})^{(j)}$, which we shall do by induction on j. The case $j = 0$ is clear: $\mathcal{D}(f)_*^{[i]} = f^{(i)} = (f^{(i)})^{(0)}$. So we shall suppose that $\mathcal{D}(f)_{j-1}^{[i]} = (f^{(i)})^{(j-1)}$ is true, and prove the corresponding equation for j.

First

$$
\begin{aligned}
(f^{(i)})^{(j)} &= \frac{\mathsf{d}^{(j)}\frac{\mathsf{d}^{(i)}f(z)}{\mathsf{d}z}(x)\cdot a_1\cdots a_i}{\mathsf{d}(a_1,\ldots,a_i,x)}(a_{*1},\ldots,a_{*i},x)\cdot(a_{11},\ldots,a_{1i},x_1)\cdots(a_{j1},\ldots,a_{ji},x_j)\\
&= \frac{\mathsf{d}\frac{\mathsf{d}^{(j-1)}\frac{\mathsf{d}^{(i)}f(z)}{\mathsf{d}z}(x)\cdot a_1\cdots a_i}{\mathsf{d}(a_1,\ldots,a_i,x)}(a_1,\ldots,a_i,x)\cdot(a_{11},\ldots,a_{1i},x_1)\cdots(a_{j-1\,1},\ldots,a_{j-1\,i},x_{j-1})}{\mathsf{d}(a_1,\ldots,a_i,x)}\\
&\qquad (a_{*1},\ldots,a_{*i},x)\cdot(a_{j1},\ldots,a_{ji},x_j)\\
&= \frac{\mathsf{d}\mathcal{D}(f)^{[i]}_{j-1}\left(\begin{array}{ccc|c} a_{11} & \ldots & a_{1i} & a_1\\ \vdots & \vdots & \vdots & \vdots\\ a_{j-1\,1} & \ldots & a_{j-1\,i} & a_{j-1}\\ \hline a_1 & \ldots & a_i & x\end{array}\right)}{\mathsf{d}(a_1,\ldots,a_i,x)}(a_{*1},\ldots,a_{*i},x)\cdot(a_{j1},\ldots,a_{ji},x_j)\\
&= \frac{\mathsf{d}\mathcal{D}(f)^{[i]}_{j-1}\left(\begin{array}{ccc|c} a_{11} & \ldots & a_{1i} & a_1\\ \vdots & \vdots & \vdots & \vdots\\ a_{j-1\,1} & \ldots & a_{j-1\,i} & a_{j-1}\\ \hline a_1 & \ldots & a_{*i} & x\end{array}\right)}{\mathsf{d}a_1}(a_{*1})\cdot a_{j1}\\
&\qquad + \ \ldots\\
&\qquad \ldots + \frac{\mathsf{d}\mathcal{D}(f)^{[i]}_{j-1}\left(\begin{array}{ccc|c} a_{11} & \ldots & a_{1i} & a_1\\ \vdots & \vdots & \vdots & \vdots\\ a_{j-1\,1} & \ldots & a_{j-1\,i} & a_{j-1}\\ \hline a_{*1} & \ldots & a_i & x\end{array}\right)}{\mathsf{d}a_i}(a_{*i})\cdot a_{ji} \qquad\qquad (\star)\\
&\qquad + \frac{\mathsf{d}\mathcal{D}(f)^{[i]}_{j-1}\left(\begin{array}{ccc|c} a_{11} & \ldots & a_{1i} & a_1\\ \vdots & \vdots & \vdots & \vdots\\ a_{j-1\,1} & \ldots & a_{j-1\,i} & a_{j-1}\\ \hline a_{*1} & \ldots & a_{*i} & x\end{array}\right)}{\mathsf{d}x}(x)\cdot x_j
\end{aligned}
$$

where, in the sum $(\star)$, in each term the bottom row of the variable base matrix consists of "starred" variables a_{*k} except in one position where one has a_m, $m = 1,\ldots,j-1$, with the j^{th} term having x instead of x, (the sum being over all such terms).

To facilitate the calculation of $\mathcal{D}(f)^{[i]}_j$, we recall a bit of notation. In particular, Proposition 2.2.3 expressed δ in terms of partial isomorphisms (or "scatter sets"):

$$f^{[i]}_j = \sum_{\sigma\in\mathrm{ParIso}(j,i)} f^\sigma$$

We regard $\mathrm{ParIso}(j-1,i)\subseteq\mathrm{ParIso}(j,i)$, and for any $\sigma\in\mathrm{ParIso}(j,i)$, $\pi_1\sigma$ (respectively $\pi_2\sigma$) is the set of indices appearing in the domain (respectively codomain) of σ. Note that

$$\mathrm{ParIso}(j,i) = \mathrm{ParIso}(j-1,i)\cup\{\sigma'\cup(j,k)\ \mid\ \sigma'\in\mathrm{ParIso}(j-1,i), k\notin\pi_2\sigma'\}$$

and

$$\mathcal{D}f_j^{[i]} = \sum_{\sigma\in\mathrm{ParIso}(j,i)} Df^{\sigma} \quad \text{and} \quad \mathcal{D}f_{j-1}^{[i]} = \sum_{\sigma'\in\mathrm{ParIso}(j-1,i)} Df^{\sigma'}$$

so that

$$\begin{aligned}
\sum_{\sigma\in\mathrm{ParIso}(j,i)} Df^{\sigma} &= \sum_{\sigma'\in\mathrm{ParIso}(j-1,i)} Df^{\sigma'} + \sum_{\substack{\sigma' \in \mathrm{ParIso}(j-1,i) \\ 0 \le k \le i \\ k \notin \pi_2\sigma'}} \mathcal{D}f^{\sigma'\cup(j,k)} \\
&= \sum_{\sigma'} \frac{\mathsf{d}\mathcal{D}f^{\sigma'}[a_{*1}/a_1,\ldots,a_{*i}/a_i]}{\mathsf{d}x}(x)\cdot x_j \quad + \\
&\quad \sum_{\sigma',k} \frac{\mathsf{d}\mathcal{D}f^{\sigma'}[a_{*1}/a_1,\ldots,\widehat{a_{*i}/a_i},\ldots,x/x]}{\mathsf{d}a_k}(a_{*k})\cdot a_{jk}
\end{aligned}$$

(where as usual, the hat indicates a term suppressed).

3.2.5. Lemma. *For $\sigma' \in \mathrm{ParIso}(j-1,i)$,*

$$\frac{\mathsf{d}\mathcal{D}f^{\sigma'}[a_{*1}/a_1,\ldots,a_{*i}/a_i]}{\mathsf{d}x}(x)\cdot x_j = Df^{\sigma'}$$

$$\frac{\mathsf{d}\mathcal{D}f^{\sigma'}[a_{*1}/a_1,\ldots,\widehat{a_{*i}/a_i},\ldots,x/x]}{\mathsf{d}a_k}(a_{*k})\cdot a_{jk} = \begin{cases} 0 \ \textit{if } k \in \pi_2\sigma' \\ \mathcal{D}f^{\sigma'\cup(j,k)} \ \textit{otherwise} \end{cases}$$

Proof. (of the lemma) Some notation: for $\sigma' \in \mathrm{ParIso}(j-1,i)$, we shall write

$$\mathcal{D}f^{\sigma'} = \mathcal{D}f^{(i+s)}(x) \underset{r\,\notin\,\pi_1\sigma'}{\bullet} x_r \underset{(l,k)\,\in\,\sigma'}{\bullet} a_{lk} \underset{m\,\notin\,\pi_2\sigma'}{\bullet} a_m$$

Then note that:

$$\begin{aligned}
&\frac{\mathsf{d}\mathcal{D}f^{\sigma'}[a_{*1}/a_1,\ldots,a_{*i}/a_i]}{\mathsf{d}x}(x)\cdot x_j \\
&= \mathcal{D}f^{(i+s+1)}(x)\cdot x_j \cdot x_r \cdot a_{lk} \cdot a_{*m} \\
&= \mathcal{D}f^{\sigma'}
\end{aligned}$$

$$\begin{aligned}
&\frac{\mathsf{d}\mathcal{D}f^{\sigma'}[a_{*1}/a_1,\ldots,\widehat{a_{*i}/a_i},\ldots,x/x]}{\mathsf{d}a_k}(a_{*k})\cdot a_{jk} \\
&= \begin{cases} 0 \text{ if } k \in \pi_2\sigma' \quad \text{since } a_k \text{ does not occur} \\ \mathcal{D}f^{(i+s)}(x) \underset{r\,\notin\,\pi_1\sigma'}{\bullet} x_r \underset{(l,w)\,\in\,\sigma'}{\bullet} a_{lw} \underset{m\,\notin\,\pi_2\sigma',\, m\,\neq\, j}{\bullet} a_{*m}\cdot a_{jk} \\ \qquad = \mathcal{D}f^{\sigma'\cup(j,k)} \text{ otherwise} \end{cases}
\end{aligned}$$

which proves the lemma ∎

Finally, to complete the proof of Theorem 3.2.4, we just sum these terms to get the same sum ($\star$) for $\mathcal{D}(f)^{[i]}_j$ as we got for $(f^{(i)})^{(j)}$:

$$\begin{aligned}
\mathcal{D}(f)^{[i]}_j &= \sum_{\sigma \in \mathrm{ParIso}(j,i)} \mathcal{D}f^{\sigma} \\
&= \sum_{\sigma'} \frac{\mathsf{d}\mathcal{D}f^{\sigma'}[a_{*1}/a_1, \ldots, a_{*i}/a_i]}{\mathsf{d}x}(x) \cdot x_j \\
&\quad + \sum_{k=1}^{i} \sum_{\sigma'} \frac{\mathsf{d}\mathcal{D}f^{\sigma'}[a_{*1}/a_1, \ldots, \widehat{a_{*i}/a_i}, \ldots, x/x]}{\mathsf{d}a_k}(a_{*k}) \cdot a_{jk} \\
&= \frac{\mathsf{d}\mathcal{D}(f)^{[i]}_{j-1}\left(\begin{array}{ccc|c} a_{11} & \ldots & a_{1i} & a_1 \\ \vdots & \vdots & \vdots & \vdots \\ a_{j-1\,1} & \ldots & a_{j-1\,i} & a_{j-1} \\ \hline a_{*1} & \ldots & a_{*i} & x \end{array}\right)}{\mathsf{d}x}(x) \cdot x_j \\
&\quad + \sum_{k=1}^{i} \frac{\mathsf{d}\mathcal{D}(f)^{[i]}_{j-1}\left(\begin{array}{ccc|c} a_{11} & \ldots & a_{1i} & a_1 \\ \vdots & \vdots & \vdots & \vdots \\ a_{j-1\,1} & \ldots & a_{j-1\,i} & a_{j-1} \\ \hline a_{*1} & \ldots a_k \ldots & a_{*i} & x \end{array}\right)}{\mathsf{d}a_k}(a_{*k}) \cdot a_{jk}
\end{aligned}$$

■

Finally, these constructions establish the following equivalence.

3.2.6. THEOREM. *The category of coalgebras for the comonad* Faà *is equivalent to the category of Cartesian differential categories and Cartesian differential functors (Cartesian left additive functors which preserve the differential combinators).*

Conclusion

The Faà di Bruno construction was a complete surprise to us. It meant, in particular, that Cartesian differential categories were as common as left additive categories. For, given any left additive category $\mathbb{X}$, $\mathsf{Faà}(\mathbb{X})$ is always a Cartesian differential category and, furthermore, the cofree one on $\mathbb{X}$. Initially our motivation was simply to give ourselves assurance that the axiomatization of Cartesian differential categories was correct and we thought to add something about higher-order chain rules as an appendix to [BCS 09]. However, as the combinatorics of the higher-order chain rules began to unfold before us, it was clear that this was something of quite independent interest. Still we did not expect to be staring at a combinatorial comonad which governed differentiation.

Not only is this construction a rich potential source of models of differential algebras (such as differential combinatory algebras and λ-calculi) but also it is clear that this structure is an integral part of what it means to be differentiable. This poses questions not only for our original setting of differential categories [BCS 06], but also for the on-going development of notions of differentiability: for example the work (with Crutwell and Gallagher) on differential restriction categories (partial map categories) and on tangential structure for manifold categories. It also raises a broader question of whether there are other closely related comonads which determine interesting subvarieties of differential structure. For example, there is a subcomonad given by the sequences which eventually vanish: these abstractly capture "polynomial" differentiability. Is it possible to capture other notion of differentiability using the same coalgebraic techniques: *e.g.* can differentiability given by Taylor expansion be so described?

After distributing an earlier version of this paper, we learned of two earlier works treating the combinatorics of Faà di Bruno algebraically (as a Hopf algebra). Although they are not directly relevant to our approach, the reader might wish to consult them as well [EP 10, FG 05].

References

[BCS 06] R.F. Blute, J.R.B. Cockett, R.A.G. Seely. "Differential Categories". Mathematical Structures in Computer Science **16** (2006), 1049–1083.

[BCS 09] R.F. Blute, J.R.B. Cockett, R.A.G. Seely. "Cartesian differential categories". *Theory and Applications of Categories* **22** (2009), 622–672.

[CCG 10] G. Crutwell, J.R.B. Cockett, and J. Gallagher. "Differential restriction categories". (in preparation)

[EP 10] K. Ebrahimi-Fard and F. Patras. "Exponential Renormalization". Annales Henri Poincare **11** (2010) 943–971.

[FG 05] H. Figueroa, J.M. Gracia-Bondia. "Combinatorial Hopf algebras in quantum field theory I". Reviews in Mathematical Physics **17** (2005), 881–976.

[OR 97] J.J. O'Connor, E.F. Robinson. "Francesco Faà di Bruno". MacTutor History of Mathematics (1997), School of Mathematics and Statistics, University of St Andrews, Scotland. `http://www-history.mcs.st-andrews.ac.uk/Biographies/Faa_di_Bruno.html`

[L 09] Paul Linehan. "Francesco Faà di Bruno". The Catholic Encyclopedia **5**. New York: Robert Appleton Company, 1909. `http://www.newadvent.org/cathen/05740a.htm`