The Differential Lambda-Calculus

Thomas Ehrhard and Laurent Regnier
Institut de Mathématiques de Luminy,
C.N.R.S. U.P.R. 9016
`ehrhard@iml.univ-mrs.fr` and `regnier@iml.univ-mrs.fr`

May 29, 2001

Abstract

We present an extension of the lambda-calculus with differential constructions motivated by a model of linear logic discovered by the first author and presented in [Ehr01]. We state and prove some basic results (confluence, weak normalization in the typed case), and also a theorem relating the usual Taylor series of analysis to the linear head reduction of lambda-calculus.

Prerequisites. This paper assumes from the reader some basic knowledge in lambda-calculus and an elementary (but not technical) knowledge of differential calculus.

Notations. Following [Kri90], we denote by $(s)\,t$ the lambda-calculus application of s to t. If $a_1, \ldots, a_n$ are elements of some given set A, we denote by $[a_1, \ldots, a_n]$ the corresponding multi-set over A.

Introduction

A whole bunch of exciting perspectives. The differential lambda-calculus is a syntactical outcome of the first author's work on the Köthe space semantics of linear logic [Ehr01] (see appendix A for a short presentation) extending the usual (typed or untyped) lambda-calculus with differential operators.

From the viewpoint of analysis, the differential lambda-calculus is a formal definition of differentiation for higher order functionals[1] and the main achievement of the Köthe space semantics is to give reasonable topological grounds to this formalism. In particular it should be noticed that the differential lambda-calculus extends to functionals of all types the basic laws of differentiation establishing an unexpected bridge between two distinct areas of mathematics: analysis and lambda-calculus.

From the computer science point of view now, the differential operators have a very natural interpretation: $\mathrm{D}_1 t \cdot u$ should be read as the *linear* application of t to u and $\frac{\partial t}{\partial x} \cdot u$ should be understood as the *linear substitution* of u for x in t, that is the substitution of one *linear occurrence* of x in t by u. The word "linear" must be taken in the sense it has in lambda-calculus: the head occurrence of variable of a term is *linear* (and is the only occurrence of variable of a lambda-term that is linear independently on the reduction). The paper stresses that this notion of linearity coincides with the usual algebraic notion, considered as the central concept of differential calculus

[1] This is standard in mathematics: think of the formal derivatives of polynomes and formal series.

(the derivative is the best *linear* approximation of a function.). This interpretation of the differential operators yields a clean and general definition of linear substitution in lambda-calculus. It is worth noting that linear substitution is in general non-deterministic (as soon as the substituted variable has several occurrences) and this accounts for the $+$ construction on differential terms. The non-deterministic character of differential lambda-calculus might be an evidence of a link with process calculi; this idea is enforced by the existing relation with the resource calculi described below.

There is a natural notion of n-approximation of an application (of a term to another one) by an n-linear application constructed by means of differential operators. The Taylor formula states that any application is equal to the (infinite in general) sum of all its n-linear approximations. This interpretation of Taylor formula could shed some new lights on the theory of approximations. For example, if one fully develops each application in a term into its corresponding Taylor expansion, one expresses the term as an infinite sum of *purely differential terms* all of which are (multi)linear ...

Outline. The goal of the paper is to present the basics of differential lambda-calculus. Before going into details, let us mention that this work could as well have been carried out in the framework of linear logic where *differential proof-nets* can be naturally defined. The choice of lambda-calculus may however seem more natural as differentiation is traditionally understood as an operation on functions, and the lambda-calculus claims to be a general theory of functions.

The differential lambda-calculus is an extension of the usual lambda-calculus in two directions:

- Terms can be summed, and more generally, linearly combined (with coefficients taken in a semi-ring). This is necessary, because derivating with respect to a variable which is contracted leads to a sum. Keep in mind the usual equation of calculus $(uv)' = u'v + uv'$ where the sum is here because the derivative is taken with respect to a parameter on which both u and v can depend.

- A differential construction $\mathrm{D}_i t \cdot u$ is added which represents the derivative of a term t with respect to its i-th argument. This new term admits an additional parameter u, and is "linear" with respect to this parameter. Remember indeed that the derivative of a function at a given point is, in general, a linear function.

Associated to these new constructions, equations between terms must be added expressing the associativity and commutativity of addition, the neutrality of 0, the linearity of most syntactical constructions (for example the application of 0 to a term is equal to 0), and the commutation of differentiation with some other constructions, in the spirit of σ-equivalence [Reg94].

The most important addition is the reduction rule concerning differentiation:

$$\mathrm{D}_1(\lambda x\, t) \cdot u = \lambda x \left(\frac{\partial t}{\partial x} \cdot u \right)$$

This is a differential version of the β-rule (oriented from left to right). The term $\frac{\partial t}{\partial x} \cdot u$ is defined by induction on t, and can be seen as a kind of "substitution" of u for x in t, with the difference that it is *linear* in u (as a derivative must be[2]). The various cases in the inductive definition of $\frac{\partial t}{\partial x} \cdot u$ correspond to well-known elementary results of differential calculus (chain rule, derivative of a multi-linear function ...). This rewriting rule, together with the ordinary β-rule, will be the two reduction rules of our system.

[2]Remember that, in analysis, the derivative of a differentiable map from, e.g. a Banach space E to another, F, is a map from E to the space of linear continuous maps from E to F.

For these rules, we prove confluence (using an adaptation of the Tait-Martin-Löf technique) and normalization for the simply typed version (using an adaptation of the Tait reducibility method). Note that strong normalization seems also provable by the same technique but it involves a lot of technical work. These two results enforce the idea that the differential lambda-calculus can be considered as a reasonable logical system.

The Taylor Formula. The Köthe space semantics validates the usual Taylor formula:

$$(s)\,u = \sum_{n=0}^{\infty} \frac{1}{n!}\left(\mathrm{D}_1^n s \cdot u^n\right) 0$$

(where $\mathrm{D}_1^2 s \cdot u^2 = \mathrm{D}_1(\mathrm{D}_1 s \cdot u) \cdot u$ and so on) which may be syntactically understood as equating the application of s to u with the sum of all the n-linear applications of s to n occurrences of u. The question is : what does it mean for a term to contain n linear occurrences of a variable? We prove a simple theorem relating the Taylor formula to linear head reduction (see below) which answers this question in a (not so) particular case: if $s = \lambda x\, s_0$ and u are two *ordinary* lambda-terms such that $(s)\,u$ is β-equivalent to a given variable $\star$, then the number of occurrences of x in $(s)\,u$ is the number of times u is substituted in s along the head linear reduction of $(s)\,u$ to $\star$.

The sum. One of the most striking features of the differential lambda-calculus is the sum which is hardcoded in the syntax. It is worth noting that, due to the underlying Köthe space semantics, this sum has a strong algebraic content and is not a mere notation for non deterministic choice. This freedom of summing arbitrary (finite) families of terms is exactly equivalent to identifying the additive connectives & and $\oplus$ of linear logic (in a category which is enriched over commutative monoids, finite direct sums and products coincide). This will be made explicit in differential proof-nets.

Related work. Analysts have already extended smoothness and analyticity to “higher types”, defining various cartesian closed categories of smooth and analytic functions (see in particular [KM97] where the objects are particular locally convex topological vector spaces called “convenient”). The differential lambda-calculus is probably the internal language of such categories.

The idea that differentiation is a kind of linear substitution is a widespread idea. For example, it is central in the work of Conor McBride where linear substitution, or more precisely the notion of “one hole contexts”, is defined in terms of derivatives for a class of “regular” types which can be seen as generalized formal power series [McB00].

Various authors introduced notions of linear substitution in lambda-calculus. Let us quote two of them that carry intuitions similar to ours.

- The *head linear reduction* was introduced in [DHR96] as a constrained head beta-reduction where no other occurrence of variable than the head occurrence can be substituted. This reduction is implemented by Krivine’s abstract machine. As already noted, the head occurrence of variable of a term is the only one which is linear for sure.

- The *lambda-calculus with multiplicities* (or with *resources*) [Bou93, BCL99]. In this system, application is written $(s)\,T$ where T is not a term, but a *bag* of terms written $T = (t_1^{p_1} \mid \cdots \mid t_n^{p_n})$ with $p_j \in \{1, \infty\}$; the order on the elements of the bag is irrelevant. Such a bag specifies, for each of its elements (which are terms) whether it should be handled linearly during the reduction; the "∞" corresponds to the status of an argument in ordinary lambda-calculus: it satisfies $t^\infty = t \mid t^\infty$. The reduction of a redex $(\lambda x\, u)\, T$ consists in removing non-deterministically a term t from the bag T and substituting it "linearly" for some occurrence of x in u through an explicit substitution mecanism.

The differential lambda-calculus is very close to the lambda-calculus with resources: the term $(\mathrm{D}_1^n s \cdot (u_1, \ldots, u_n))\, t$ is similar to applying s to the bag $(u_1 \mid \cdots \mid u_n \mid t^\infty)$. However the lambda-calculus with resources equates the terms $(s)\,(t^\infty)$ and $(s)\,(t \mid t^\infty)$, whereas the corresponding terms $(s)\, t$ and $(\mathrm{D}_1 s \cdot t)\, t$ are distinct in differential lambda-calculus. A deeper difference between the two formalisms seems to lie in the central role played by the sum in differential lambda-calculus for which we see no equivalent in resource lambda-calculus.

1 Syntax

Let R be a commutative ring (a commutative semi-ring, like the set $\mathbb{N}$ of natural numbers, would do as well). Terms are built as follows.

Variable: if x is a variable, then x is a term;

Ordinary application: if s and t are terms, then $(s)\, t$ is a term;

Abstraction: if x is a variable and t is a term, then $\lambda x\, t$ is a term;

Differential application: if i is a *nonzero* integer and t and u are terms, then $\mathrm{D}_i t \cdot u$ is a term;

Zero: 0 is a term;

Linear combination: if s and t are terms and $a, b \in R$, then $as + bt$ is a term.

The intuition behind differential application is as follows. Consider that t is a function with n parameters, typically $t : A_1 \to \cdots \to A_n \to A$ if we were in a typed system, and that i is an index such that $1 \leq i \leq n$. Then $\mathrm{D}_i\, t$ is *the partial derivative of t with respect to its i-th argument*, that is a function with the same parameters as t, but with values in a space of *linear* mappings from A_i to B. In the term $\mathrm{D}_i t \cdot u$, u may be seen as an element of A_i and the term $\mathrm{D}_i t \cdot u$ as the linear application of this derivative to u. This will be made explicit in the typed version of this calculus, see section 3.

Structural congruence. Terms are considered up to a congruence relation, denoted by $\simeq$. It is the smallest congruence relation such that:

- if s and t are α-equivalent, then $s \simeq t$;
- for any three terms s, t and u, and any $i \in \mathbb{N}^+$, one has

  $$\mathrm{D}_i((s)\, t) \cdot u \simeq (\mathrm{D}_{i+1} s \cdot u)\, t\,. \tag{1}$$

  This expresses that, as differential application specifies the rank of the (linear) argument, it may commute with usual application modulo some easy bookkeeping of this rank;
- for any terms t and u, any variable x not free in u and any $i \in \mathbb{N}^+$, one has

  $$\mathrm{D}_{i+1}(\lambda x\, t) \cdot u \simeq \lambda x\, (\mathrm{D}_i t \cdot u)\,. \tag{2}$$

  This expresses that differential application may also commute with abstractions. However the differential argument u must not feed the λ, whence the $i + 1$;

- differential applications commute with each other, that is

$$\mathrm{D}_i(\mathrm{D}_j t \cdot v) \cdot u \simeq \mathrm{D}_j(\mathrm{D}_i t \cdot u) \cdot v\,. \tag{3}$$

  This corresponds to the well-known *Schwarz lemma*, a crucial property of the derivatives of sufficiently regular differentiable functions;

- ordinary application is linear *in the function* only, that is

$$\left(\sum_k a_k s_k\right) t \simeq \sum_k a_k\,(s_k)\,t\,; \tag{4}$$

- abstraction is linear, that is

$$\lambda x \sum_k a_k t_k \simeq \sum_k a_k \lambda x\, t_k\,; \tag{5}$$

- differential application is bilinear, that is

$$\mathrm{D}_i\left(\sum_k a_k t_k\right) \cdot \left(\sum_j b_j u_j\right) \simeq \sum_{k,j} a_k b_j \mathrm{D}_i t_k \cdot u_j\,; \tag{6}$$

- addition of terms is associative, commutative, admits 0 as neutral element (more precisely, the set of terms is an R-module).

We denote by $\mathrm{D}_{i_1,\dots,i_k} t \cdot (u_1, \dots, u_k)$ the expression $\mathrm{D}_{i_1}(\cdots(\mathrm{D}_{i_k} t \cdot u_k)\cdots) \cdot u_1$. By equation (3), for any permutation σ of $\{1, \dots, k\}$ one has

$$\mathrm{D}_{i_{\sigma(1)},\dots,i_{\sigma(k)}} t \cdot (u_{\sigma(1)}, \dots, u_{\sigma(k)}) \simeq \mathrm{D}_{i_1,\dots,i_k} t \cdot (u_1, \dots, u_k)\,.$$

When in particular the indices $i_1, \dots, i_k$ have a common value i, we write $\mathrm{D}_i^k t \cdot (u_1, \dots, u_k)$ instead of $\mathrm{D}_{i_1,\dots,i_k} t \cdot (u_1, \dots, u_k)$. If furthermore, all the u_i's are the same term u, we simply write $\mathrm{D}_i^k t \cdot u^k$.

Canonical and simple terms. Equations (1), (2), (4), (5) and (6) can be oriented from left to right, whereas equation (3) cannot be oriented. The other equations, dealing with α-conversion and algebraic properties of sums of terms and multiplication of terms by a scalar cannot be oriented either, but do not deserve particular mention. They will be considered as true equalities in the sequel, e.g. $u + v = v + u$.

Two terms which are equivalent with respect to the context-extension of rule (3) will be said to be equal up to *differential permutation*. We define by mutual induction the *canonical terms* and *simple terms*.

- If $t_1, \dots, t_n$ are simple terms which are pairwise distinct up to differential permutation and if $a_1, \dots, a_n \in R \setminus \{0\}$ then $\sum_{k=1}^n a_k t_k$ is a canonical term;
- if x is a variable, $n \geq 0$ and $i_1, \dots, i_n \geq 1$ and if $u_1, \dots, u_n$ are simple terms, then $\mathrm{D}_{i_1,\dots,i_n} x \cdot (u_1, \dots, u_n)$ is a simple term (in particular, when $n = 0$, this means that x is a simple term);
- if s is a simple term and t is a canonical term, then $(s)\,t$ is a simple term;
- if $n \geq 0$ and $s, u_1, \dots, u_n$ are simple terms, then $\mathrm{D}_1^n(\lambda x\, s) \cdot (u_1, \dots, u_n)$ is a simple term.

Up to differential permutation, canonical terms constitute the free R-module generated by simple terms. This means that, up to differential permutation, a canonical term may be written in a unique way as a linear combination of simple terms.

Canonical reduction. The *canonical reduction* is defined by the left-to-right orientation and context-extension of equations (1), (2), (4), (5) and (6).

One easily checks that a term t is canonical iff for any t' equal to t up to differential permutation, t' is normal for the canonical reduction. For instance, the term

$$t_0 = \mathrm{D}_1(\lambda x\, \mathrm{D}_1 x \cdot z_1) \cdot z_2$$

is in canonical form, whereas the term

$$t = \mathrm{D}_{2,1}\lambda x\, x \cdot (z_1, z_2)$$

is not because by differential permutation we have

$$t \simeq \mathrm{D}_{1,2}\lambda x\, x \cdot (z_2, z_1) = \mathrm{D}_1(\mathrm{D}_2 \lambda x\, x \cdot z_1) \cdot z_2$$

which reduces to t_0 by (2).

More generally, given two terms t and t_0, the term t_0 is a *canonical form* of t if t_0 is in canonical form and can be obtained from t by application of equation (3) and canonical reduction.

Lemma 1 *There exists a primitive recursive function which maps any term t to a canonical form $\langle t \rangle$ of t. If t_1 and t_2 are terms, and t'_1, t'_2 are canonical forms of t_1 and t_2 respectively then*

$$t_1 \simeq t_2 \quad \text{iff} \quad t'_1 =_3 t'_2$$

where $=_3$ denotes equality up to differential permutation.

Proof. The function $\langle t \rangle$ may be defined by mutual recursion with an auxiliary function $\Delta_{i_1,\dots,i_n} t \cdot (u_1, \dots, u_n)$ (defined on simple terms and yielding simple terms) as follows.

- $\langle x \rangle = x$;
- $\langle (s)\, t \rangle = \sum_k a_k \,(s_k)\, \langle t \rangle$ where the sum $\sum_k a_k s_k$ is the unique decomposition in simple terms of the canonical term $\langle s \rangle$, which we will simply write $\langle s \rangle = \sum_k a_k s_k$ in the sequel;
- $\langle \lambda x\, s \rangle = \sum_k a_k \lambda x\, s_k$ where $\langle s \rangle = \sum_k a_k s_k$;
- $\langle \mathrm{D}_i s \cdot u \rangle = \sum_{k,l} a_k b_l \Delta_i s_k \cdot u_l$ where $\langle s \rangle = \sum_k a_k s_k$ and $\langle u \rangle = \sum_l b_l u_l$;
- $\langle \sum_k a_k s_k \rangle = \sum_k a_k \langle s_k \rangle$;
- $\Delta_{i_1,\dots,i_n} x \cdot (u_1, \dots, u_n) = \mathrm{D}_{i_1,\dots,i_n} x \cdot (u_1, \dots, u_n)$, and this term is simple since the u_i's are assumed to be simple;
- $\Delta_{i_1,\dots,i_n} (s)\, t \cdot (u_1, \dots, u_n) = (\Delta_{i_1+1,\dots,i_n+1} s \cdot (u_1, \dots, u_n))\, t$;
- $\Delta_{i_1,\dots,i_n} (\lambda x\, s) \cdot (u_1, \dots, u_n) = \mathrm{D}_1^p(\lambda x\, \Delta_{i_{k_1}-1,\dots,i_{k_q}-1} s \cdot (u_{i_{k_1}}, \dots, u_{i_{k_q}})) \cdot (u_{i_{j_1}}, \dots, u_{i_{j_p}})$, where $1 \leq j_1 < \cdots < j_p \leq n$ are all the indexes $j \in \{1, \dots, n\}$ such that $i_j = 1$ and $1 \leq k_1 < \cdots < k_q \leq n$ are all the indexes $k \in \{1, \dots, n\}$ such that $i_k > 1$ (so that $p + q = n$);
- and last $\Delta_{i_1,\dots,i_n} (\mathrm{D}_i s \cdot u) \cdot (u_1, \dots, u_n) = \Delta_{i_1,\dots,i_n,i} s \cdot (u_1, \dots, u_n, u)$.

The rest of the lemma is immediate. □

The canonical convention. This shows that, up to differential permutation, the congruence class of any term has a unique canonical element. Thanks to this lemma, from now on we shall work up to structural congruence, using canonical terms as representatives of the congruence classes. In particular, all the constructions of the syntax are extended in the natural way to congruence classes, and are therefore considered as constructions yielding canonical terms when applied to canonical terms.

If Φ is any operation on simple terms yielding canonical terms, the *extension by linearity* of Φ is the operation defined on all canonical terms (i.e. congruence classes) by:

$$\Phi\left(\sum_j a_j s_j\right) = \sum_j a_j \Phi(s_j)\,.$$

Substitution. The basic operation of lambda-calculus is substitution of a term t for a variable x in another term s, the new term obtained is denoted by $s\,[t/x]$. This operation is defined here as in the usual lambda-calculus, by induction on canonical terms with the two new cases:

- $(\sum a_i s_i)\,[t/x] = \sum a_i s_i\,[t/x]$ where the s_i's are simple and pairwise distinct (up to differential permutation);
- $(\mathrm{D}_i s \cdot u)\,[t/x] = \mathrm{D}_i(s\,[t/x]) \cdot (u\,[t/x])$.

Derivative. We define now another operation which bears some similarities with substitution, but behaves in a *linear way* with respect to the "substituted" term: *partial derivative*. As substitution, it is defined by induction on canonical terms.

Linear combination.

$$\frac{\partial}{\partial x}\left(\sum_i a_i s_i\right) \cdot u = \sum_i a_i \frac{\partial s_i}{\partial x} \cdot u$$

Variable.

$$\frac{\partial y}{\partial x} \cdot u = \delta_{x,y} u \tag{7}$$

where $\delta_{x,y} = 1$ if $x = y$ and $\delta_{x,y} = 0$ if $x \neq y$.

Application.

$$\frac{\partial (s)\,t}{\partial x} \cdot u = \left(\frac{\partial s}{\partial x} \cdot u\right) t + \left(\mathrm{D}_1 s \cdot \left(\frac{\partial t}{\partial x} \cdot u\right)\right) t \tag{8}$$

Abstraction.

$$\frac{\partial \lambda y\, t}{\partial x} \cdot u = \lambda y \left(\frac{\partial t}{\partial x} \cdot u\right) \tag{9}$$

Differential.

$$\frac{\partial \mathrm{D}_i t \cdot v}{\partial x} \cdot u = \mathrm{D}_i\left(\frac{\partial t}{\partial x} \cdot u\right) \cdot v + \mathrm{D}_i t \cdot \left(\frac{\partial v}{\partial x} \cdot u\right) \tag{10}$$

In the application case, when x is not free in s, the first term vanishes (see lemma 3), and we can interpret this equation as the usual *chain rule* of differential calculus. When x occurs free in s and t, the sum appears because we are taking a derivative with respect to a contracted variable. This expression is coherent with the fact that application is linear in the function, but of course not in the argument.

In the abstraction case we suppose as usual that y is not free in u. This case is compatible with the linearity of abstraction. Similarly the differentiation case is compatible with the bilinearity in t and v of the expression $\mathrm{D}_i t \cdot v$.

These equations have as easy consequences the following ones

$$\begin{aligned}
\frac{\partial}{\partial x}\left(\mathrm{D}_{i_1,\dots,i_n} y \cdot (v_1,\dots,v_n)\right)\cdot u &= \delta_{x,y}\mathrm{D}_{i_1,\dots,i_n} u \cdot (v_1,\dots,v_n) \\
&\quad + \sum_{j=1}^{n} \mathrm{D}_{i_1,\dots,i_n} y \cdot \left(v_1,\dots,\frac{\partial v_j}{\partial x}\cdot u,\dots,v_n\right) \\
\frac{\partial}{\partial x}\left(\mathrm{D}_1^n \lambda y\, t \cdot (v_1,\dots,v_n)\right)\cdot u &= \mathrm{D}_1^n \lambda y \left(\frac{\partial t}{\partial x}\cdot u\right)\cdot (v_1,\dots,v_n) \\
&\quad + \sum_{j=1}^{n} \mathrm{D}_1^n \lambda y\, t \cdot \left(v_1,\dots,\frac{\partial v_j}{\partial x}\cdot u,\dots,v_n\right)
\end{aligned}$$

with for the second equation the proviso that y is not free in u and that $y \neq x$.

Differential reduction. We are now ready to give the reduction rules completing the definition of differential lambda-calculus. Informally there are two reduction rules:

Beta-reduction.

$$(\lambda x\, s)\, t \quad \text{reduces to} \quad s\,[t/x]\ ;$$

Differential reduction. When x is not free in u:

$$\mathrm{D}_1(\lambda x\, s)\cdot u \quad \text{reduces to} \quad \lambda x \left(\frac{\partial s}{\partial x}\cdot u\right)$$

This last rule is similar to β-reduction up to the fact that it only substitutes one linear occurrence of variable. For this reason the abstraction present in the differential redex stays. This rule is compatible with the intuitions behind differentiation of a function (D_1) and partial derivation with respect to a variable.

More formally, we define the relation β^1_{D} on canonical terms as the contextual closure of the reduction rules above. Precisely, β^1_{D} is given by the following inductive clauses.

- $\sum_i a_i t_i\ \beta^1_{\mathrm{D}}\ t'$ when $t' = \sum_i a_i t'_i$ with $t_j\ \beta^1_{\mathrm{D}}\ t'_j$ for exactly one j and $t'_i = t_i$ for $i \neq j$;
- $\mathrm{D}_{i_1,\dots,i_n} x \cdot (u_1,\dots,u_n)\ \beta^1_{\mathrm{D}}\ t'$ when $t' = \mathrm{D}_{i_1,\dots,i_n} x \cdot (u'_1,\dots,u'_n)$ with $u_j\ \beta^1_{\mathrm{D}}\ u'_j$ for exactly one $j \in \{1,\dots,n\}$ and $u_i = u'_i$ for $i \neq j$;
- $(s)\, u\ \beta^1_{\mathrm{D}}\ t'$ when $t' = (s')\, u$ with $s\ \beta^1_{\mathrm{D}}\ s'$, or $t' = (s)\, u'$ with $u\ \beta^1_{\mathrm{D}}\ u'$, or $s = \lambda x\, v$ and $t' = v\,[u/x]$;

- $\mathrm{D}_1^n(\lambda x\, s)\cdot(u_1,\ldots,u_n)\ \beta_{\mathrm{D}}^1\ t'$ when $t' = \mathrm{D}_1^n(\lambda x\, s')\cdot(u_1,\ldots,u_n)$ with $s\ \beta_{\mathrm{D}}^1\ s'$, or $t' = \mathrm{D}_1^n(\lambda x\, s)\cdot(u_1,\ldots,u_{j-1},u'_j,u_{j+1},\ldots,u_n)$ with $u_j\ \beta_{\mathrm{D}}^1\ u'_j$ (for some $j\in\{1,\ldots,n\}$), or

$$t' = \mathrm{D}_1^{n-1}\lambda x\left(\frac{\partial s}{\partial x}\cdot u_j\right)\cdot(u_1,\ldots,u_{j-1},u_{j+1},\ldots,u_n)$$

  for some $j\in\{1,\ldots,n\}$.

We denote by β_{D} the transitive closure of this relation on canonical terms.

1.1 Substitution and derivation lemmas

First, we observe that substitution is compatible with all syntactical constructions applied to canonical terms; for example we have

$$(\mathrm{D}_i t\cdot u)\,[v/x] = \mathrm{D}_i(t\,[v/x])\cdot u\,[v/x]\,.$$

whenever t, u and v are canonical. The definition of substitution states this property only when $\mathrm{D}_i t\cdot u$ is a simple term, as it is a definition by induction on the structure of canonical terms. Similar properties hold for partial derivatives, for example:

$$\frac{\partial \mathrm{D}_i t\cdot v}{\partial x}\cdot u = \mathrm{D}_i\left(\frac{\partial t}{\partial x}\cdot u\right)\cdot v + \mathrm{D}_i t\cdot\left(\frac{\partial v}{\partial x}\cdot u\right).$$

These properties will be used without mention in the sequel.

Lemma 2 *If x is not free in v, then we have*

$$t\,[u/x]\,[v/y] = t\,[v/y]\,[u\,[v/y]/x]\ .$$

The proof is a simple induction on t.

Lemma 3 *If x is not free in t, then $\frac{\partial t}{\partial x}\cdot u = 0$. For any canonical term t and variable x,*

$$\frac{\partial t}{\partial x}\cdot\left(\sum_j a_j u_j\right) = \sum_j a_j\frac{\partial t}{\partial x}\cdot u_j\ .$$

The proof is an easy induction on t.

Lemma 4 *If the variable y is not free in the term u, one has*

$$\frac{\partial}{\partial x}\left(\frac{\partial t}{\partial y}\cdot v\right)\cdot u = \frac{\partial}{\partial y}\left(\frac{\partial t}{\partial x}\cdot u\right)\cdot v + \frac{\partial t}{\partial y}\cdot\left(\frac{\partial v}{\partial x}\cdot u\right). \tag{11}$$

In particular, when moreover the variable x is not free in the term v, the following "syntactic Schwarz lemma" holds:

$$\frac{\partial}{\partial x}\left(\frac{\partial t}{\partial y}\cdot v\right)\cdot u = \frac{\partial}{\partial y}\left(\frac{\partial t}{\partial x}\cdot u\right)\cdot v\,. \tag{12}$$

Proof. This is the derivative analogue of equation (3). It is proven by an easy induction on t. We consider here only two cases.

Assume first that t is a variable z; there are two sub-cases.

- If $z = y$, then $\frac{\partial}{\partial x}\left(\frac{\partial t}{\partial y} \cdot v\right) \cdot u = \frac{\partial v}{\partial x} \cdot u = \frac{\partial t}{\partial y} \cdot \left(\frac{\partial v}{\partial x} \cdot u\right)$. We conclude because in that case $\frac{\partial}{\partial y}\left(\frac{\partial t}{\partial x} \cdot u\right) \cdot v = 0$ (if $y \neq x$, this holds because then $\frac{\partial t}{\partial x} \cdot u = 0$, and if $y = x$, this holds because $\frac{\partial t}{\partial x} \cdot u = u$ and y is not free in u).
- If $z \neq y$, then $\frac{\partial}{\partial x}\left(\frac{\partial t}{\partial y} \cdot v\right) \cdot u = 0$, and also $\frac{\partial t}{\partial y} \cdot \left(\frac{\partial v}{\partial x} \cdot u\right) = 0$. Moreover, the term $\frac{\partial}{\partial y}\left(\frac{\partial t}{\partial x} \cdot u\right) \cdot v$ vanishes also because either $z \neq x$ and then $\frac{\partial t}{\partial x} \cdot u = 0$ or $z = x$ and then we have $\frac{\partial u}{\partial y} \cdot v = 0$ since y is not free in u.

Assume now that t is an ordinary application, say $t = (t_1)\, t_2$.

$$
\begin{aligned}
\frac{\partial}{\partial x}\left(\frac{\partial t}{\partial y} \cdot v\right) \cdot u &= \frac{\partial}{\partial x}\left(\left(\frac{\partial t_1}{\partial y} \cdot v\right) t_2 + \left(\mathrm{D}_1 t_1 \cdot \left(\frac{\partial t_2}{\partial y} \cdot v\right)\right) t_2\right) \cdot u \\
&= \left(\frac{\partial}{\partial x}\left(\frac{\partial t_1}{\partial y} \cdot v\right) \cdot u\right) t_2 + \left(\mathrm{D}_1\left(\frac{\partial t_1}{\partial y} \cdot v\right) \cdot \left(\frac{\partial t_2}{\partial x} \cdot u\right)\right) t_2 \\
&\quad + \left(\mathrm{D}_1\left(\frac{\partial t_1}{\partial x} \cdot u\right) \cdot \left(\frac{\partial t_2}{\partial y} \cdot v\right)\right) t_2 + \left(\mathrm{D}_1 t_1 \cdot \left(\frac{\partial}{\partial x}\left(\frac{\partial t_2}{\partial y} \cdot v\right) \cdot u\right)\right) t_2 \\
&\quad + \left(\mathrm{D}_1^2 t_1 \cdot \left(\frac{\partial t_2}{\partial x} \cdot u, \frac{\partial t_2}{\partial y} \cdot v\right)\right) t_2 \\
&= \left(\mathrm{D}_1\left(\frac{\partial t_1}{\partial y} \cdot v\right) \cdot \left(\frac{\partial t_2}{\partial x} \cdot u\right)\right) t_2 + \left(\mathrm{D}_1\left(\frac{\partial t_1}{\partial x} \cdot u\right) \cdot \left(\frac{\partial t_2}{\partial y} \cdot v\right)\right) t_2 \\
&\quad + \left(\mathrm{D}_1^2 t_1 \cdot \left(\frac{\partial t_2}{\partial x} \cdot u, \frac{\partial t_2}{\partial y} \cdot v\right)\right) t_2 \\
&\quad + \left(\frac{\partial}{\partial y}\left(\frac{\partial t_1}{\partial x} \cdot u\right) \cdot v\right) t_2 + \left(\frac{\partial t_1}{\partial y} \cdot \left(\frac{\partial v}{\partial x} \cdot u\right)\right) t_2 \\
&\quad + \left(\mathrm{D}_1 t_1 \cdot \left(\frac{\partial}{\partial y}\left(\frac{\partial t_2}{\partial x} \cdot u\right) \cdot v\right)\right) t_2 + \left(\mathrm{D}_1 t_1 \cdot \left(\frac{\partial t_2}{\partial y} \cdot \left(\frac{\partial v}{\partial x} \cdot u\right)\right)\right) t_2
\end{aligned}
$$

by inductive hypothesis applied to t_1 and t_2 (and commutativity of addition). On the other hand,

$$
\begin{aligned}
\frac{\partial}{\partial y}\left(\frac{\partial t}{\partial x} \cdot u\right) \cdot v &= \left(\mathrm{D}_1\left(\frac{\partial t_1}{\partial y} \cdot v\right) \cdot \left(\frac{\partial t_2}{\partial x} \cdot u\right)\right) t_2 + \left(\mathrm{D}_1\left(\frac{\partial t_1}{\partial x} \cdot u\right) \cdot \left(\frac{\partial t_2}{\partial y} \cdot v\right)\right) t_2 \\
&\quad + \left(\mathrm{D}_1^2 t_1 \cdot \left(\frac{\partial t_2}{\partial x} \cdot u, \frac{\partial t_2}{\partial y} \cdot v\right)\right) t_2 \\
&\quad + \left(\frac{\partial}{\partial y}\left(\frac{\partial t_1}{\partial x} \cdot u\right) \cdot v\right) t_2 + \left(\mathrm{D}_1 t_1 \cdot \left(\frac{\partial}{\partial y}\left(\frac{\partial t_2}{\partial x} \cdot u\right) \cdot v\right)\right) t_2
\end{aligned}
$$

and we conclude since

$$
\left(\frac{\partial t_1}{\partial y} \cdot \left(\frac{\partial v}{\partial x} \cdot u\right)\right) t_2 + \left(\mathrm{D}_1 t_1 \cdot \left(\frac{\partial t_2}{\partial y} \cdot \left(\frac{\partial v}{\partial x} \cdot u\right)\right)\right) t_2 = \frac{\partial (t_1)\, t_2}{\partial y} \cdot \left(\frac{\partial v}{\partial x} \cdot u\right).
$$

□

If $x_1, \dots, x_k$ are variables not occurring free in the terms $u_1, \dots, u_k$ one has therefore, for any permutation σ of $\{1, \dots, k\}$,

$$\frac{\partial}{\partial x_1}\left(\cdots \frac{\partial t}{\partial x_k} \cdot u_k \cdots\right) \cdot u_1 = \frac{\partial}{\partial x_{\sigma(1)}}\left(\cdots \frac{\partial t}{\partial x_{\sigma(k)}} \cdot u_{\sigma(k)} \cdots\right) \cdot u_{\sigma(1)}$$

and we use the standard notation

$$\frac{\partial^k t}{\partial x_1 \cdots \partial x_k} \cdot (u_1, \dots, u_k)$$

for the common value of these expressions (we avoid this notations when the condition above on variables is not fulfilled).

Derivatives and substitutions. We shall now state a number of lemmas expressing the commutation between the derivative and the substitution operators.

Lemma 5 *If x and y are two distinct variables and y is not free in the terms u and v, one has*

$$\frac{\partial t\,[v/y]}{\partial x} \cdot u = \left(\frac{\partial t}{\partial x} \cdot u\right)[v/y] + \left(\frac{\partial t}{\partial y} \cdot \left(\frac{\partial v}{\partial x} \cdot u\right)\right)[v/y] \;. \tag{13}$$

In particular, if moreover x is not free in v, the following commutation holds:

$$\frac{\partial t\,[v/y]}{\partial x} \cdot u = \left(\frac{\partial t}{\partial x} \cdot u\right)[v/y] \;. \tag{14}$$

The proof is similar to the previous one. This lemma can also be seen as a version of the chain rule of differential calculus.

Lemma 6 *If the variable x is not free in the term v and if y is a variable distinct from x, we have*

$$\left(\frac{\partial t}{\partial x} \cdot u\right)[v/y] = \frac{\partial t\,[v/y]}{\partial x} \cdot (u\,[v/y]) \,.$$

Proof. We first prove the lemma when y is not free in t:

$$\left(\frac{\partial t}{\partial x} \cdot u\right)[v/y] = \frac{\partial t}{\partial x} \cdot (u\,[v/y]) \,. \tag{15}$$

This is a simple induction on t.

In the general case, let y' be a new fresh variable and let $t' = t\,[y'/y]$, so that $t = t'\,[y/y']$. Since y' does not occur in u, by lemma 5 we have $\frac{\partial t}{\partial x} \cdot u = \left(\frac{\partial t'}{\partial x} \cdot u\right)[y/y']$. Then by lemma 2 we have $\left(\frac{\partial t}{\partial x} \cdot u\right)[v/y] = \left(\frac{\partial t'}{\partial x} \cdot u\right)[v/y]\,[v/y']$ (because y' is not free in v). So by (15), since y does not occur in t', $\left(\frac{\partial t}{\partial x} \cdot u\right)[v/y] = \left(\frac{\partial t'}{\partial x} \cdot u\,[v/y]\right)[v/y']$, and so $\left(\frac{\partial t}{\partial x} \cdot u\right)[v/y] = \frac{\partial t'[v/y']}{\partial x} \cdot u\,[v/y]$ by lemma 5 because y' is not free in u and in v. We conclude by observing that $t'\,[v/y'] = t\,[v/y]$ by definition of t'. $\square$

Iterated derivatives. Iterating derivations leads to rather complicated expressions. However, one can easily prove the following lemmas which will be useful in the normalization proof.

Lemma 7 *The derivative* $\frac{\partial^k \mathrm{D}_i s \cdot t}{\partial x_1 \cdots \partial x_k} \cdot (u_1, \ldots, u_k)$ *is a finite sum of expressions*

$$\mathrm{D}_i s' \cdot t'$$

where s' *and* t' *have the shape:*

$$\begin{aligned} s' &= \frac{\partial^{r_0} s}{\partial y_1^0 \cdots \partial y_{r_0}^0} \cdot \left(u_1^0, \ldots, u_{r_0}^0\right) \\ t' &= \frac{\partial^{r_1} t}{\partial y_1^1 \cdots \partial y_{r_1}^1} \cdot \left(u_1^1, \ldots, u_{r_1}^1\right) \end{aligned}$$

with $r_0 + r_1 = k$, $[y_1^0, \ldots, y_{r_1}^1] = [x_1, \ldots, x_k]$ *and* $[u_1^0, \ldots, u_{r_1}^1] = [u_1, \ldots, u_k]$.

Lemma 8 *The derivative* $\frac{\partial^k (s) t}{\partial x_1 \cdots \partial x_k} \cdot (u_1, \ldots, u_k)$ *is a finite sum of expressions*

$$\left(\mathrm{D}_1^q s' \cdot (t'_1, \ldots, t'_q)\right) t$$

where s' *and* t'_j *have the shape:*

$$\begin{aligned} s' &= \frac{\partial^{r_0} s}{\partial y_1^0 \cdots \partial y_{r_0}^0} \cdot \left(u_1^0, \ldots, u_{r_0}^0\right) \\ t'_j &= \frac{\partial^{r_j} t}{\partial y_1^j \cdots \partial y_{r_j}^j} \cdot \left(u_1^j, \ldots, u_{r_j}^j\right) \end{aligned}$$

with $\sum r_j = k$, $[y_1^0, \ldots, y_{r_q}^q] = [x_1, \ldots, x_k]$ *and* $[u_1^0, \ldots, u_{r_q}^q] = [u_1, \ldots, u_k]$.

Both lemmas are proved by induction on k, using the following immediate property.

$$\begin{aligned} \frac{\partial}{\partial x} \left(\mathrm{D}_{i_1,\ldots,i_k} s \cdot (u_1, \ldots, u_k)\right) \cdot v = {} & \mathrm{D}_{i_1,\ldots,i_k} \left(\frac{\partial s}{\partial x} \cdot v\right) \cdot (u_1, \ldots, u_k) \\ & + \sum_{j=1}^{k} \mathrm{D}_{i_1,\ldots,i_k} s \cdot \left(u_1, \ldots, \frac{\partial u_j}{\partial x} \cdot v, \ldots, u_k\right). \end{aligned}$$

2 The Church-Rosser property

We prove the confluence theorem using the Tait-Martin-Löf technique, and following the presentation of [Kri90]. We first define, by induction on canonical terms, the *parallel reduction* relation.

1. $\sum_j a_j t_j \;\rho\; t'$ if $t' = \sum_j a_j t'_j$ where $t_j \;\rho\; t'_j$ for each j;
2. $\mathrm{D}_{i_1,\ldots,i_n} x \cdot (u_1, \ldots, u_n) \;\rho\; t'$ when $t' = \mathrm{D}_{i_1,\ldots,i_n} x \cdot (u'_1, \ldots, u'_n)$ with $u_j \;\rho\; u'_j$ for each j;
3. $(s)\, u \;\rho\; t'$ in one of the following situations:
    (a) $s \;\rho\; s'$, $u \;\rho\; u'$ and $t' = (s')\, u'$,

(b) $s = \mathrm{D}_1^n(\lambda x\, v) \cdot (w_1, \ldots, w_n)$, $v \;\rho\; v'$, $u \;\rho\; u'$, $w_j \;\rho\; w_j'$ for $j = 1, \ldots, n$ and

$$t' = \left(\frac{\partial^n v'}{\partial x^n} \cdot (w_1', \ldots, w_n') \right) [u'/x] \, ;$$

4. $\mathrm{D}_1^n(\lambda x\, s) \cdot (u_1, \ldots, u_n) \;\rho\; t'$ if $t' = \mathrm{D}_1^n(\lambda x\, s') \cdot (u_1', \ldots, u_n')$ with $s \;\rho\; s'$, $u_1 \;\rho\; u_1'$, $\ldots$,$u_n \;\rho\; u_n'$, or

$$t' = \mathrm{D}_1^{n-1} \lambda x \left(\frac{\partial s'}{\partial x} \cdot u_j' \right) \cdot (u_1', \ldots, u_{j-1}', u_{j+1}', \ldots, u_n')$$

for some $j \in \{1, \ldots, n\}$, with $s \;\rho\; s'$, $u_1 \;\rho\; u_1'$, $\ldots$,$u_n \;\rho\; u_n'$.

Notice that this relation is "contextual" in the sense that it commutes to all constructions of the syntax applied to canonical terms (cf. the canonical convention above). For instance if $t \;\rho\; t'$ and $u \;\rho\; u'$, then $\mathrm{D}_i t \cdot u \;\rho\; \mathrm{D}_i t' \cdot u'$.

Observe also that, for any canonical term t, one has $t \;\rho\; t$. It follows easily that $\beta_{\mathrm{D}}^1 \subseteq \rho$. A straightforward induction on terms shows that $\rho \subseteq \beta_{\mathrm{D}}$. Therefore the transitive closure of ρ is β_{D} so that, by standard reasoning, for proving confluence of β_{D} it suffices to prove confluence of ρ.

Lemma 9 *Let x be a variable and let t, u, t', and u' be canonical terms. If $t \;\rho\; t'$ and $u \;\rho\; u'$, then*

$$\frac{\partial t}{\partial x} \cdot u \;\; \rho \;\; \frac{\partial t'}{\partial x} \cdot u'$$

Proof. By induction on t. Assume that t is simple (the case where t is a linear combination of simple terms is trivial).

If $t = \mathrm{D}_{i_1,\ldots,i_n} y \cdot (u_1, \ldots, u_n)$, then

$$\begin{aligned} \frac{\partial t}{\partial x} \cdot u = {} & \delta_{x,y} \mathrm{D}_{i_1,\ldots,i_n} u \cdot (u_1, \ldots, u_n) \\ & + \sum_{j=1}^{n} \mathrm{D}_{i_1,\ldots,i_n} y \cdot \left(u_1, \ldots, \frac{\partial u_j}{\partial x} \cdot u, \ldots, u_n\right). \end{aligned}$$

Moreover, we know that $t' = \mathrm{D}_{i_1,\ldots,i_n} y \cdot (u_1', \ldots, u_n')$ with $u_j \;\rho\; u_j'$ for each j. For each j, we have, by inductive hypothesis $\frac{\partial u_j}{\partial x} \cdot u \;\rho\; \frac{\partial u_j'}{\partial x} \cdot u'$ and we conclude.

Assume now that $t = (s)\, w$. Then

$$\frac{\partial t}{\partial x} \cdot u = \left(\frac{\partial s}{\partial x} \cdot u \right) w + \left(\mathrm{D}_1 s \cdot \left(\frac{\partial w}{\partial x} \cdot u \right) \right) w \, .$$

If $t' = (s')\, w'$ with $s \;\rho\; s'$ and $w \;\rho\; w'$, then we conclude by inductive hypothesis. But it may also happen that $s = \mathrm{D}_1^n(\lambda y\, v) \cdot (u_1, \ldots, u_n)$, $v \;\rho\; v'$, $w \;\rho\; w'$, $u_j \;\rho\; u_j'$ for $j = 1, \ldots, n$, and

$$t' = \left(\frac{\partial^n v'}{\partial y^n} \cdot (u_1', \ldots, u_n') \right) [w'/y] \; .$$

In that situation, we have

$$\frac{\partial s}{\partial x} \cdot u = \mathrm{D}_1^n \left(\lambda y \, \frac{\partial v}{\partial x} \cdot u \right) \cdot (u_1, \ldots, u_n) + \sum_{j=1}^{n} \mathrm{D}_1^n(\lambda y\, v) \cdot \left(u_1, \ldots, \frac{\partial u_j}{\partial x} \cdot u, \ldots, u_n\right)$$

and by inductive hypothesis we have $\frac{\partial v}{\partial x}\cdot u\ \rho\ \frac{\partial v'}{\partial x}\cdot u'$ and $\frac{\partial u_j}{\partial x}\cdot u\ \rho\ \frac{\partial u'_j}{\partial x}\cdot u'$ for each j, and therefore, by the clause (3b) in the definition of ρ, we get

$$\left(\frac{\partial s}{\partial x}\cdot u\right) w\ \ \rho\ \ \left(\frac{\partial^n}{\partial y^n}\left(\frac{\partial v'}{\partial x}\cdot u'\right)\cdot(u'_1,\dots,u'_n)\right)[w'/y] + \sum_{j=1}^{n}\left(\frac{\partial^n v'}{\partial y^n}\cdot(u'_1,\dots,\frac{\partial u'_j}{\partial x}\cdot u',\dots,u'_n)\right)[w'/y]\ .$$

Similarly, we get

$$\begin{aligned}\left(\mathrm{D}_1 s\cdot\left(\frac{\partial w}{\partial x}\cdot u\right)\right) w &= \left(\mathrm{D}_1^{n+1}(\lambda y\, v)\cdot\left(\frac{\partial w}{\partial x}\cdot u, u_1,\dots,u_n\right)\right) w\\ &\rho\ \ \frac{\partial^{n+1}v'}{\partial y^{n+1}}\cdot\left(\frac{\partial w'}{\partial x}\cdot u', u'_1,\dots,u'_n\right)[w'/y]\ .\end{aligned}$$

Observe that we use here the possibility for n to take arbitrary values in the clause (3b) of the definition of ρ (and not only the value 0 as in the case of ordinary lambda-calculus). On the other hand, by iterating lemma 4 and by lemma 5, we get

$$\begin{aligned}\frac{\partial t'}{\partial x}\cdot u' &= \left(\frac{\partial}{\partial x}\left(\frac{\partial^n v'}{\partial y^n}\cdot(u'_1,\dots,u'_n)\right)\cdot u' + \frac{\partial}{\partial y}\left(\frac{\partial^n v'}{\partial y^n}\cdot(u'_1,\dots,u'_n)\right)\cdot\left(\frac{\partial w'}{\partial x}\cdot u'\right)\right)[w'/y]\\ &= \left(\frac{\partial^n}{\partial y^n}\left(\frac{\partial v'}{\partial x}\cdot u'\right)\cdot(u'_1,\dots,u'_n)\right)[w'/y]\\ &\quad+\sum_{j=1}^{n}\left(\frac{\partial^n v'}{\partial y^n}\cdot(u'_1,\dots,\frac{\partial u'_j}{\partial x}\cdot u',\dots,u'_n)\right)[w'/y]\\ &\quad+\frac{\partial^{n+1}v'}{\partial y^{n+1}}\cdot\left(\frac{\partial w'}{\partial x}\cdot u', u'_1,\dots,u'_n\right)[w'/y]\end{aligned}$$

and we are done, in this particular case.

Assume last that $t=\mathrm{D}_1^n(\lambda y\, s)\cdot(u_1,\dots,u_n)$. Then

$$\begin{aligned}\frac{\partial t}{\partial x}\cdot u &= \mathrm{D}_1^n\lambda y\left(\frac{\partial s}{\partial x}\cdot u\right)\cdot(u_1,\dots,u_n)\\ &\quad+\sum_{j=1}^{n}\mathrm{D}_1^n\lambda y\, s\cdot(u_1,\dots,\frac{\partial u_j}{\partial x}\cdot u,\dots,u_n)\end{aligned}$$

We can have $t'=\mathrm{D}_1^n(\lambda y\, s')\cdot(u'_1,\dots,u'_n)$ with $s\ \rho\ s'$ and $u_j\ \rho\ u'_j$ for all j and in that case we conclude by direct application of the inductive hypothesis. But it may also happen that, for some given $k\in\{1,\dots,n\}$, one has (and then of course $n\geq 1$)

$$t'=\mathrm{D}_1^{n-1}\lambda y\left(\frac{\partial s'}{\partial y}\cdot u'_k\right)\cdot(u'_1,\dots,u'_{k-1},u'_{k+1},\dots,u'_n)$$

with $s\,\rho\,s'$ and $u_j\,\rho\,u'_j$ for all j. Then we have

$$\begin{aligned}\frac{\partial t'}{\partial x}\cdot u' = {}& \mathrm{D}_1^{n-1}\lambda y\left(\frac{\partial}{\partial y}\left(\frac{\partial s'}{\partial x}\cdot u'\right)\cdot u'_k\right)\cdot(u'_1,\ldots,u'_{k-1},u'_{k+1},\ldots,u'_n)\\
&+\mathrm{D}_1^{n-1}\lambda y\left(\frac{\partial s'}{\partial y}\cdot\left(\frac{\partial u'_k}{\partial x}\cdot u'\right)\right)\cdot(u'_1,\ldots,u'_{k-1},u'_{k+1},\ldots,u'_n)\\
&+\sum_{j=1}^{k-1}\mathrm{D}_1^{n-1}\lambda y\left(\frac{\partial s'}{\partial y}\cdot u'_k\right)\cdot(u'_1,\ldots,\frac{\partial u'_j}{\partial x}\cdot u',\ldots,u'_{k-1},u'_{k+1},\ldots,u'_n)\\
&+\sum_{j=k+1}^{n}\mathrm{D}_1^{n-1}\lambda y\left(\frac{\partial s'}{\partial y}\cdot u'_k\right)\cdot(u'_1,\ldots,u'_{k-1},u'_{k+1},\ldots,\frac{\partial u'_j}{\partial x}\cdot u',\ldots,u'_n)\end{aligned}$$

But by inductive hypothesis, we have

$$\begin{aligned}&\mathrm{D}_1^{n}\lambda y\left(\frac{\partial s}{\partial x}\cdot u\right)\cdot(u_1,\ldots,u_n)\\
&\qquad\rho\quad \mathrm{D}_1^{n-1}\lambda y\left(\frac{\partial}{\partial y}\left(\frac{\partial s'}{\partial x}\cdot u'\right)\cdot u'_k\right)\cdot(u'_1,\ldots,u'_{k-1},u'_{k+1},\ldots,u'_n)\end{aligned}$$

Furthermore, for $j=1,\ldots,k-1$, we have

$$\begin{aligned}&\mathrm{D}_1^{n}\lambda y\, s\cdot(u_1,\ldots,\frac{\partial u_j}{\partial x}\cdot u,\ldots,u_n)\\
&\qquad\rho\quad \mathrm{D}_1^{n-1}\lambda y\left(\frac{\partial s'}{\partial y}\cdot u'_k\right)\cdot(u'_1,\ldots,\frac{\partial u'_j}{\partial x}\cdot u',\ldots,u'_{k-1},u'_{k+1},\ldots,u'_n)\end{aligned}$$

and similarly for $j=k+1,\ldots,n$. Last,

$$\begin{aligned}&\mathrm{D}_1^{n}\lambda y\, s\cdot(u_1,\ldots,\frac{\partial u_k}{\partial x}\cdot u,\ldots,u_n)\\
&\qquad\rho\quad \mathrm{D}_1^{n-1}\lambda y\left(\frac{\partial s'}{\partial y}\cdot\left(\frac{\partial u'_k}{\partial x}\cdot u'\right)\right)\cdot(u'_1,\ldots,u'_{k-1},u'_{k+1},\ldots,u'_n)\end{aligned}$$

and hence $\frac{\partial t}{\partial x}\cdot u\,\rho\,\frac{\partial t'}{\partial x}\cdot u'$ as required, and the lemma is proved. □

Lemma 10 *Let x be a variable and let t, u, t', and u' be canonical terms. If $t\,\rho\,t'$ and $u\,\rho\,u'$, then*

$$t\,[u/x]\;\;\rho\;\; t'\,[u'/x]\ .$$

Proof. The case where t is a linear combination is immediate. We assume now that t is simple.

Assume first that $t=\mathrm{D}_{i_1,\ldots,i_n}y\cdot(u_1,\ldots,u_n)$ so that $t'=\mathrm{D}_{i_1,\ldots,i_n}y\cdot(u'_1,\ldots,u'_n)$; then $t\,[u/x]=\mathrm{D}_{i_1,\ldots,i_n}v\cdot(u_1\,[u/x]\,,\ldots,u_n\,[u/x])$ where $v=u$ if $y=x$ and $v=y$ otherwise. Thus $t'\,[u'/x]=\mathrm{D}_{i_1,\ldots,i_n}v'\cdot(u'_1\,[u'/x]\,,\ldots,u'_n\,[u'/x])$ with $v'=u'$ if $y=x$ and $v'=y$ otherwise, so that $v\,\rho\,v'$. By inductive hypothesis $u_j\,[u/x]\,\rho\,u'_j\,[u'/x]$, and we conclude.

Assume now that $t=(s)\,w$. Then $t\,[u/x]=(s\,[u/x])\,w\,[u/x]$. If $t'=(s')\,w'$ with $s\,\rho\,s'$ and $w\,\rho\,w'$, then we conclude by applying the inductive hypothesis. But it may also happen that $s=\mathrm{D}_1^n(\lambda y\,v)\cdot(u_1,\ldots,u_n)$, $v\,\rho\,v'$, $w\,\rho\,w'$, $u_j\,\rho\,u'_j$ for $j=1,\ldots,n$, and

$$t'=\left(\frac{\partial^n v'}{\partial y^n}\cdot(u'_1,\ldots,u'_n)\right)[w'/y]\ .$$

Then we have $s\,[u/x] = \mathrm{D}_1^n(\lambda y\, v\,[u/x]) \cdot (u_1\,[u/x]\,, \ldots, u_n\,[u/x])$ and hence, by inductive hypothesis,

$$\begin{aligned} t\,[u/x] \quad \rho \quad & \left(\frac{\partial^n v'\,[u'/x]}{\partial y^n} \cdot (u'_1\,[u'/x]\,, \ldots, u'_n\,[u'/x])\right)[w'\,[u'/x]\,/y] \\ = \quad & \left(\frac{\partial^n v'}{\partial y^n} \cdot (u'_1, \ldots, u'_n)\right)[u'/x]\,[w'\,[u'/x]\,/y] \end{aligned}$$

by lemma 6, since we can assume that y is not free in u'. Therefore, $t\,[u/x]\;\rho\;t'\,[u'/x]$ by lemma 2.

Assume last that $t = \mathrm{D}_1^n(\lambda y\, s)\cdot(u_1, \ldots, u_n)$. Then $t\,[u/x] = \mathrm{D}_1^n(\lambda y\, s\,[u/x])\cdot(u_1\,[u/x]\,, \ldots, u_n\,[u/x])$. The non-trivial case is when, for some given $k \in \{1, \ldots, n\}$, one has (and then of course $n \geq 1$)

$$t' = \mathrm{D}_1^{n-1}\lambda y \left(\frac{\partial s'}{\partial y} \cdot u'_k\right) \cdot (u'_1, \ldots, u'_{k-1}, u'_{k+1}, \ldots, u'_n)$$

with $s\;\rho\;s'$ and $u_j\;\rho\;u'_j$ for all j. Then

$$t\,[u/x]\;\rho\;\mathrm{D}_1^{n-1}\lambda y \left(\frac{\partial s''}{\partial y} \cdot u''_k\right) \cdot (u''_1, \ldots, u''_{k-1}, u''_{k+1}, \ldots, u''_n)$$

where $s'' = s'\,[u'/x]$ and $u''_j = u'_j\,[u'/x]$ for all j. We conclude as above, applying lemma 6, since y is not free in u'. □

Proposition 11 *The relation ρ is confluent.*

Proof. By induction on the canonical term t, we show that, for any two canonical terms t^1 and t^2, if $t\;\rho\;t^1$ and $t\;\rho\;t^2$, there exists a canonical term t' such that $t^1\;\rho\;t'$ and $t^2\;\rho\;t'$. The case where t is a linear combination is immediate.

If $t = \mathrm{D}_{i_1,\ldots,i_n} x \cdot (u_1, \ldots, u_n)$, then we must have $t^l = \mathrm{D}_{i_1,\ldots,i_n} x \cdot (u^l_1, \ldots, u^l_n)$ (for $l = 1, 2$) with $u_j\;\rho\;u^l_j$ for $l = 1, 2$ and $j = 1, \ldots, n$. By inductive hypothesis, there are canonical terms $u'_1, \ldots, u'_n$ such that $u^l_j\;\rho\;u'_j$ for $l = 1, 2$ and $j = 1, \ldots, n$, and we conclude, setting $t' = \mathrm{D}_{i_1,\ldots,i_n} x \cdot (u'_1, \ldots, u'_n)$.

Assume next that $t = (s)\,u$. The following cases are possible.

- $t^l = \left(s^l\right) u^l$ with $s\;\rho\;s^l$ and $u\;\rho\;u^l$ for $l = 1, 2$. One concludes immediately by inductive hypothesis.

- $s = \mathrm{D}_1^n(\lambda x\, v) \cdot (u_1, \ldots, u_n)$, $t^1 = \left(\frac{\partial^n v^1}{\partial x^n} \cdot \left(u^1_1, \ldots, u^1_n\right)\right)\left[u^1/x\right]$ and $t^2 = \left(s^2\right) u^2$, with $v\;\rho\;v^1$, $s\;\rho\;s^2$, $u\;\rho\;u^1$, $u\;\rho\;u^2$ and $u_j\;\rho\;u^1_j$ for $j = 1, \ldots, n$. Since $s\;\rho\;s^2$ and $s = \mathrm{D}_1^n(\lambda x\, v) \cdot (u^1_1, \ldots, u^1_n)$, we must consider two subcases.

  - First we may have

    $$s^2 = \mathrm{D}_1^n(\lambda x\, v^2) \cdot (u^2_1, \ldots, u^2_n)$$

    with $v\;\rho\;v^2$ and $u_j\;\rho\;u^2_j$ for $j = 1, \ldots, n$. By inductive hypothesis there are canonical terms v', u' and u'_j such that $u^l\;\rho\;u'$, $v^l\;\rho\;v'$ and $u^l_j\;\rho\;u'_j$ for $l = 1, 2$ and $j = 1, \ldots, n$. By definition of ρ, we have

    $$t^2 = \left(s^2\right) u^2\;\rho\;t' = \left(\frac{\partial^n v'}{\partial x^n} \cdot \left(u'_1, \ldots, u'_n\right)\right)\left[u'/x\right]\ .$$

    On the other hand, by lemma 9 and 10, we have $t^1\;\rho\;t'$.

– The other possibility is that, for some $k \in \{1, \ldots, n\}$,

$$s^2 = \mathrm{D}_1^{n-1} \lambda x \left(\frac{\partial v^2}{\partial x} \cdot u_k^2 \right) \cdot (u_1^2, \ldots, u_{k-1}^2, u_{k+1}^2, \ldots, u_n^2)$$

with $v \; \rho \; v^2$ and $u_j \; \rho \; u_j^2$ for $j = 1, \ldots, n$. Define v', u' and u_j' as above, using the inductive hypothesis. By definition of ρ and lemma 4, we have

$$\begin{array}{rcl} t^2 & \rho & \left(\dfrac{\partial^{n-1}}{\partial x^{n-1}} \left(\dfrac{\partial v'}{\partial x} \cdot u_k' \right) \cdot (u_1', \ldots, u_{k-1}', u_{k+1}', \ldots, u_n') \right) [u'/x] \\ & = & \left(\dfrac{\partial^n v'}{\partial x^n} \cdot (u_1', \ldots, u_n') \right) [u'/x] \end{array}$$

and we conclude as above.

- $s = \mathrm{D}_1^n(\lambda x \, v) \cdot (u_1, \ldots, u_n)$, $t^l = \left(\frac{\partial^n v^l}{\partial x^n} \cdot (u_1^l, \ldots, u_n^l) \right) [u^l/x]$ with $v \; \rho \; v^l$, $u \; \rho \; u^l$ and $u_j \; \rho \; u_j^l$ for $l = 1, 2$ and $j = 1, \ldots, n$. In that case, one concludes by inductive hypothesis, applying twice lemma 10 and 9.

Assume last that $t = \mathrm{D}_1^n(\lambda x \, s) \cdot (u_1, \ldots, u_n)$. The following situations may occur.

- For $l = 1, 2$, one has $t^l = \mathrm{D}_1^n(\lambda x \, s^l) \cdot (u_1^l, \ldots, u_n^l)$ with $s \; \rho \; s^l$, $u_1 \; \rho \; u_1^l$, … ,$u_n \; \rho \; u_n^l$. In that case, the inductive hypothesis applies straightforwardly.

- There is $k \in \{1, \ldots, n\}$ such that

$$t^1 = \mathrm{D}_1^{n-1} \lambda x \left(\frac{\partial s^1}{\partial x} \cdot u_k^1 \right) \cdot (u_1^1, \ldots, u_{k-1}^1, u_{k+1}^1, \ldots, u_n^1)$$

and $t^2 = \mathrm{D}_1^n(\lambda x \, s^2) \cdot (u_1^2, \ldots, u_n^2)$ with $s \; \rho \; s^l$, $u_1 \; \rho \; u_1^l$, … ,$u_n \; \rho \; u_n^l$. By inductive hypothesis, we can find some canonical terms s', u_1', … ,u_n' such that $s^l \; \rho \; s'$, $u_1^l \; \rho \; u_1'$, … ,$u_n^l \; \rho \; u_n'$. By lemma 9, we have

$$t^1 \; \rho \; \mathrm{D}_1^{n-1} \lambda x \left(\frac{\partial s'}{\partial x} \cdot u_k' \right) \cdot (u_1', \ldots, u_{k-1}', u_{k+1}', \ldots, u_n') = t'$$

and by definition of ρ, we have $t^2 \; \rho \; t'$.

- In the last possible situation, for each $l = 1, 2$, there is exists $k^l \in \{1, \ldots, n\}$ such that

$$t^l = \mathrm{D}_1^{n-1} \lambda x \left(\frac{\partial s^l}{\partial x} \cdot u_{k^l}^l \right) \cdot (u_1^l, \ldots, u_{k^l-1}^l, u_{k^l+1}^l, \ldots, u_n^l)$$

with $s \; \rho \; s^l$, $u_1 \; \rho \; u_1^l$, … ,$u_n \; \rho \; u_n^l$. By inductive hypothesis, we can find some canonical terms s', u_1', … ,u_n' such that $s^l \; \rho \; s'$, $u_1^l \; \rho \; u_1'$, … ,$u_n^l \; \rho \; u_n'$. We distinguish two sub-cases. First, if $k^1 = k^2 = k$, applying lemma 9, we get that $t^1 \; \beta_{\mathrm{D}}^1 \; t'$ and $t^2 \; \beta_{\mathrm{D}}^1 \; t'$, where

$$t' = \mathrm{D}_1^{n-1} \lambda x \left(\frac{\partial s'}{\partial x} \cdot u_k' \right) \cdot (u_1', \ldots, u_{k-1}', u_{k+1}', \ldots, u_n') \, .$$

Next, assume without loss of generality that $k^1 < k^2$ (and so $n \geq 2$). Then by definition of ρ we have

$$t^1 \ \rho \ \mathrm{D}_1^{n-2} \lambda x \left(\frac{\partial^2 s'}{\partial x^2} \cdot (u'_{k^2}, u'_{k^1}) \right) \cdot (u'_1, \ldots, u'_{k^1-1}, u'_{k^1+1}, \ldots, u'_{k^2-1}, u'_{k^2+1}, \ldots, u'_n) = w^1$$

and

$$t^2 \ \rho \ \mathrm{D}_1^{n-2} \lambda x \left(\frac{\partial^2 s'}{\partial x^2} \cdot (u'_{k^1}, u'_{k^2}) \right) \cdot (u'_1, \ldots, u'_{k^1-1}, u'_{k^1+1}, \ldots, u'_{k^2-1}, u'_{k^2+1}, \ldots, u'_n) = w^2$$

and we conclude since $w^1 = w^2$ by lemma 4. □

As a corollary we get

Theorem 12 *The relation β_{D} over canonical terms of the pure differential lambda-calculus enjoys the Church-Rosser property.*

Observe that any ordinary lambda-term is a differential lambda-term. The Church-Rosser result above shows that the differential lambda-calculus is a conservative extension of the ordinary lambda-calculus.

Theorem 13 *If two ordinary lambda-terms are β_{D}-equivalent, then they are β-equivalent.*

Remark: We can easily derive from lemma 9 and 10 and from the inclusions $\beta_{\mathrm{D}}^1 \subseteq \rho \subseteq \beta_{\mathrm{D}}$ (a direct proof would be straightforward as well) the two following lemmas, which will be useful in the sequel.

Lemma 14 *Let x be a variable and let t, u, t', and u' be canonical terms. If $t \ \beta_{\mathrm{D}} \ t'$ and $u \ \beta_{\mathrm{D}} \ u'$, then*

$$\frac{\partial t}{\partial x} \cdot u \ \ \beta_{\mathrm{D}} \ \ \frac{\partial t'}{\partial x} \cdot u'$$

Lemma 15 *Let x be a variable and let t, u, t', and u' be canonical terms. If $t \ \beta_{\mathrm{D}} \ t'$ and $u \ \beta_{\mathrm{D}} \ u'$, then*

$$t\,[u/x] \ \ \beta_{\mathrm{D}} \ \ t'\,[u'/x]$$

3 Simply typed terms

We are given some atomic types $\alpha, \beta, \ldots$, and if A and B are types, then so is $A \to B$. The notion of typing context is the usual one, and the typing rules are as follows.

$$\frac{}{\Gamma, x : A \vdash x : A}$$

$$\frac{\Gamma \vdash s : A \to B \qquad \Gamma \vdash t : A}{\Gamma \vdash (s)\,t : B}$$

$$\frac{\Gamma, x : A \vdash s : B}{\Gamma \vdash \lambda x\, s : A \to B}$$

$$\frac{\Gamma \vdash s : A_1, \ldots, A_i \to B \qquad \Gamma \vdash u : A_i}{\Gamma \vdash \mathrm{D}_i s \cdot u : A_1, \ldots, A_i \to B}$$

$$\frac{}{\Gamma \vdash 0 : A} \qquad\qquad \frac{\Gamma \vdash s : A \qquad \Gamma \vdash t : A}{\Gamma \vdash as + bt : A}$$

where a and b are scalars.

The structural congruence is compatible with these typing rules in a restricted sense: if s and s' are structurally equivalent *canonical terms*, then $\Gamma \vdash s : A$ iff $\Gamma \vdash s' : A$. This is not quite true if s and s' are not canonical because of the congruence rules for 0; for example if $s = (0)\, t$ and $s' = 0$ then s' can be typed whereas s may not be typeable. Removing the equations for 0 leads to a structural congruence which is compatible with these typing rules. However, we shall not need this as our canonical convention settles this slight inadequacy.

The two last rules expresses that a type may be considered as an R-module.

Consider the differential application rule in the case $i = 1$: we are given a term t with $\Gamma \vdash t : A \to B$ that we may view as a function from A to B. The derivative of t should be a function t' from A to a space L of *linear* applications from A to B. So given $s : A$ and $u : A$, $t'(s)$ is a linear function from A to B that we may apply to u, getting a value $t'(s) \cdot u$ in B; this is precisely this value that the term $(\mathrm{D}_1 t \cdot u)\, s$ denotes. So $\mathrm{D}_1 t \cdot u$ denotes the function which maps $s : A$ to $t'(s) \cdot u : B$. When $i > 1$, the intuition is exactly the same, but in that case we do not derivate the function with respect to its first parameter, but with respect to its i-th parameter.

Lemma 16 *Subject reduction holds, that is: if t and t' are canonical terms, if $\Gamma \vdash t : A$ and $t\ \beta_{\mathrm{D}}\ t'$, then $\Gamma \vdash t' : A$.*

This is proven by a straightforward induction on the derivation of $\Gamma \vdash t : A$, with the help of the following “substitution” lemma (and of an ordinary substitution lemma that we do not state).

Lemma 17 *If s and u are canonical terms, if $\Gamma, x : A \vdash s : B$ and $\Gamma \vdash u : A$, then $\Gamma, x : A \vdash \frac{\partial s}{\partial x} \cdot u : B$.*

The proof is an easy induction on s.

4 Weak normalization

We prove weak normalization for the simply typed differential lambda-calculus, using the Tait reducibility method, presented along the lines followed by Krivine in [Kri90]. Proving strong normalization involves a more sophisticated notion of saturated sets leading to technically complex developments, that we postpone to future work. There is no doubt however that it holds.

We denote by Λ_{D} the R-module of all canonical terms (up to differential permutation). This is the free R-module generated by simple terms (or more precisely, by the classes of simple terms under differential permutation).

If $\mathcal{X}$ is a set of *simple* terms, we denote by $\overline{\mathcal{X}}$ its span, that is, the sub-module of Λ_{D} generated by $\mathcal{X}$. We denote by $\mathcal{N}$ the set of all weakly normalizable simple terms, so that $\overline{\mathcal{N}}$ is the set of all weakly normalizable canonical terms.

Saturated sets. A set $\mathcal{S}$ of simple terms is *saturated* if, for any simple terms t, $u_1, \dots, u_p$, $v_1, \dots, v_k$ and canonical terms s and $s_1, \dots, s_n$, whenever the term

$$w' = \left(\mathrm{D}_{i_1,\dots,i_k}\left(\frac{\partial^p t}{\partial x^p}\cdot(u_1,\dots,u_p)\right)[s/x]\cdot(v_1,\dots,v_k)\right) s_1\dots s_n$$

belongs to $\overline{\mathcal{S}}$, the simple term

$$w = (\mathrm{D}_{i_1,\dots,i_k}(\mathrm{D}_1^p(\lambda x\, t)\cdot(u_1,\dots,u_p))\, s\cdot(v_1,\dots,v_k))\, s_1\dots s_n$$

belongs to $\mathcal{S}$.

This definition of saturation does not preserve strong normalization: it might be that w' is strongly normalizable but not w. Take for instance $w' = 0$ and $w = \left(\mathrm{D}_1^2\lambda x\,(x)\,\delta\cdot(\delta,u)\right) v$ where δ is the usual $\lambda z\,(z)\,z$.

Observe that $\mathcal{N}$ is saturated since, with the notations of the definition above, one has $w\ \beta_{\mathrm{D}}\ w'$.

Reducibility. If $\mathcal{X}$ and $\mathcal{Y}$ are sets of simple terms, one defines $\mathcal{X}\to\mathcal{Y}\subseteq\Lambda_{\mathrm{D}}$ as

$$\mathcal{X}\to\mathcal{Y} = \{t \mid \forall p\in\mathbb{N},\ \forall s\in\overline{\mathcal{X}},\ \forall u_1,\dots,u_p\in\mathcal{X},\ (\mathrm{D}_1^p t\cdot(u_1,\dots,u_p))\, s\in\mathcal{Y}\}\,.$$

Observe that the term $(\mathrm{D}_1^p t\cdot(u_1,\dots,u_p))\, s$ is simple. This definition, which involves differential applications and not only ordinary applications, is motivated by the next lemma which will be essential in the proof of the interpretation lemma 21.

Lemma 18 *let $\mathcal{X}_1,\dots,\mathcal{X}_i$ and $\mathcal{Y}$ be sets of simple terms. If $t\in\mathcal{X}_i\to\cdots\to\mathcal{X}_1\to\mathcal{Y}$ and $u\in\mathcal{X}_1$, then $\mathrm{D}_i t\cdot u\in\mathcal{X}_i\to\cdots\to\mathcal{X}_1\to\mathcal{Y}$.*

Proof. By induction on $i\geq 1$. For $i=1$, it is an obvious consequence of the definition. Assume that the property holds for i, and take $t\in\mathcal{X}_{i+1}\to\mathcal{X}_i\to\cdots\to\mathcal{X}_1\to\mathcal{Y}$ and $u\in\mathcal{X}_1$. We must show that $\mathrm{D}_{i+1}t\cdot u\in\mathcal{X}_{i+1}\to\mathcal{X}_i\to\cdots\to\mathcal{X}_1\to\mathcal{Y}$. So let $v_1,\dots,v_p\in\mathcal{X}_{i+1}$ and $s\in\overline{\mathcal{X}_{i+1}}$, we have to show that $(\mathrm{D}_1^p(\mathrm{D}_{i+1}t\cdot u)\cdot(v_1,\dots,v_p))\, s$ belongs to $\mathcal{X}_i\to\cdots\to\mathcal{X}_1\to\mathcal{Y}$.

By definition $(\mathrm{D}_1^p t\cdot(v_1,\dots,v_p))\, s$ belongs to $\mathcal{X}_i\to\cdots\to\mathcal{X}_1\to\mathcal{Y}$ and thus, by inductive hypothesis, so does $\mathrm{D}_i(\mathrm{D}_1^p t\cdot(v_1,\dots,v_p))\, s\cdot u$. We conclude because this latter term is equal to $(\mathrm{D}_1^p(\mathrm{D}_{i+1}t\cdot u)\cdot(v_1,\dots,v_p))\, s$. □

Lemma 19 *If $\mathcal{X},\mathcal{S}\subseteq\Lambda_{\mathrm{D}}$ with $\mathcal{S}$ saturated, then $\mathcal{X}\to\mathcal{S}$ is also saturated. If $\mathcal{X}\subseteq\mathcal{X}'$ and $\mathcal{Y}'\subseteq\mathcal{Y}$, then $\mathcal{X}'\to\mathcal{Y}'\subseteq\mathcal{X}\to\mathcal{Y}$.*

Proof. We only prove the first part, the second statement being obvious. So, with the notations of the definition of saturated sets, assume that $w'\in\overline{\mathcal{X}\to\mathcal{S}}$, we have to show that $w\in\mathcal{X}\to\mathcal{S}$. Let $v_{k+1},\dots,v_{k+q}\in\mathcal{X}$ and let $s_{n+1}\in\overline{\mathcal{X}}$; we must show that $w_1 = (\mathrm{D}_1^q w\cdot(v_{k+1},\dots,v_{k+q}))\, s_{n+1}\in\mathcal{S}$. But

$$w_1 = \left(\mathrm{D}_{i_1,\dots,i_{k+q}}(\mathrm{D}_1^p(\lambda x\, t)\cdot(u_1,\dots,u_p))\, s\cdot(v_1,\dots,v_{k+q})\right) s_1\dots s_{n+1}$$

where we have set $i_j = 1+n$ for $j = k+1,\dots,k+q$. Since $\mathcal{S}$ is saturated, it suffices therefore to show that $w_1'\in\overline{\mathcal{S}}$, where

$$\begin{aligned} w_1' &= \left(\mathrm{D}_{i_1,\dots,i_{k+q}}\left(\frac{\partial^p t}{\partial x^p}\cdot(u_1,\dots,u_p)\right)[s/x]\cdot(v_1,\dots,v_{k+q})\right) s_1\dots s_{n+1} \\ &= \left(\mathrm{D}_1^q w'\cdot(v_{k+1},\dots,v_{k+q})\right) s_{n+1}\,. \end{aligned}$$

But we have assumed that $w' \in \overline{\mathcal{X} \to \mathcal{S}}$ and hence since $v_{k+1}, \dots, v_{k+q} \in \mathcal{X}$ and $s_{n+1} \in \overline{\mathcal{X}}$, we have $w'_1 \in \overline{\mathcal{S}}$ by definition of $\mathcal{X} \to \mathcal{S}$. □

Let $\mathcal{N}_0$ be the set of all simple terms of the shape

$$(\mathrm{D}_{i_1,\dots,i_k} x \cdot (u_1, \dots, u_k))\, s_1 \dots s_n$$

where $u_1, \dots, u_k \in \mathcal{N}$, $s_1, \dots, s_n \in \overline{\mathcal{N}}$ and x is a variable. It is clear that $\mathcal{N}_0 \subseteq \mathcal{N}$.

Lemma 20 *The following inclusions hold.*

$$\mathcal{N}_0 \subseteq \mathcal{N} \to \mathcal{N}_0 \subseteq \mathcal{N}_0 \to \mathcal{N} \subseteq \mathcal{N}\,.$$

Proof. The first inclusion immediately results from the definition of $\mathcal{N}_0$. The second inclusion results from lemma 19. For the last inclusion, take $t \in \mathcal{N}_0 \to \mathcal{N}$, and take a variable x which appears nowhere (free or bound) in t. Then $x \in \mathcal{N}_0$ and thus $(t)\, x \in \mathcal{N}$, that is, there is a finite sequence of canonical terms $s_0, \dots, s_n$ such that $s_0 = (t)\, x$, $s_i\ \beta^1_{\mathrm{D}}\ s_{i+1}$ for all $i < n$, and s_n is normal. We prove by induction on n that then t is weakly normalizing. The base case $n = 0$ is trivial since then s_0 itself is normal and hence t must be normal. Assume that the result holds for n and let us prove it for $n+1$. Consider the first reduction step $s_0\ \beta^1_{\mathrm{D}}\ s_1$. There are two cases.

- If $s_1 = (t_1)\, x$ with $t\ \beta_{\mathrm{D}}\ t_1$ and t_1 is a finite linear combination of simple terms, $t_1 = \sum_{k=1}^{m} a_k t_{1,k}$ with $a_k \neq 0$ for $k = 1, \dots, m$, then we know that, for each value of k, the term $(t_{1,k})\, x$ reduces to a normal form in at most n steps. Therefore, by inductive hypothesis, $t_{1,k}$ is weakly normalizing for each k, and hence t itself is weakly normalizing since $t\ \beta^1_{\mathrm{D}}\ t_1$.
- The other possibility is that $s_0 = (\lambda y\, u)\, x$ and $s_1 = u\,[x/y]$. Then, since s_1 is weakly normalizing, $\lambda x\, u\,[x/y]$ is weakly normalizing, but this latter term is α-equivalent to t by our assumption about x. □

To an arbitrary simple type A, we associate a saturated set A^* of simple terms by setting $\alpha^* = \mathcal{N}$ for all atomic types α, and $(A \to B)^* = A^* \to B^*$. Then combining lemmas 19 and 20, we get for all type A:

$$\mathcal{N}_0 \subseteq A^* \subseteq \mathcal{N}\,.$$

Lemma 21 (Interpretation) *Let t be a canonical term whose free variables are among the list (without repetitions) $x_1, \dots, x_n$, and assume that*

$$x_1 : A_1, \dots, x_n : A_n \vdash t : A$$

for some types $A_1, \dots, A_n, A$. Let $i_1, \dots, i_k \in \{1, \dots, n\}$ and let $u_1 \in A^*_{i_1}, \dots, u_k \in A^*_{i_k}$. Let also $s_1 \in \overline{A^*_1}, \dots, s_n \in \overline{A^*_n}$. Assume that the variables $x_1, \dots, x_n$ are not free in the terms $s_1, \dots, s_n$ and $u_1, \dots u_k$. Then

$$\left(\frac{\partial^k t}{\partial x_{i_1} \cdots \partial x_{i_k}} \cdot (u_1, \dots, u_k)\right) [s_1, \dots, s_n / x_1, \dots, x_n] \in \overline{A^*}\,.$$

Proof. By induction on the canonical term t. The case where t is a linear combination is immediate.

Variable. Assume first that t is a variable z.

If $k = 0$; if $z = x_i$ we conclude since we know that $s_i \in \overline{A_i^*}$, otherwise $z \notin \{x_1, \dots, x_n\}$ and we conclude since $z \in \mathcal{N}_0 \subseteq A^*$.

If $k = 1$ and $z = x_{i_1}$. Then $\left(\frac{\partial^k t}{\partial x_{i_1} \cdots \partial x_{i_k}} \cdot (u_1, \dots, u_k)\right) [s_1, \dots, s_n / x_1, \dots, x_n] = u_1$ since none of the variables x_i is free in u_1 and we are done, since we have assumed that $u_1 \in A_{i_1}^*$.

If $k \geq 2$ or $z \notin \{x_{i_1}, \dots, x_{i_k}\}$ then it is clear that $\frac{\partial^k t}{\partial x_{i_1} \cdots \partial x_{i_k}} \cdot (u_1, \dots, u_k) = 0$ and we are done since $0 \in \overline{A^*}$.

Application. Assume next that t is an ordinary application, $t = (s)\, w$ with $x_1 : A_1, \dots, x_n : A_n \vdash s : B \to A$ and $x_1 : A_1, \dots, x_n : A_n \vdash w : B$. By lemma 8, $\left(\frac{\partial^k t}{\partial x_{i_1} \cdots \partial x_{i_k}} \cdot (u_1, \dots, u_k)\right) [s_1, \dots, s_n / x_1, \dots, x_n]$ is a sum of terms of the shape $\left(\mathrm{D}_1^q s' \cdot (w'_1, \dots, w'_q)\right) (w\, [s_1, \dots, s_n / x_1, \dots, x_n])$ with

$$s' = \frac{\partial^p s}{\partial y_1 \cdots \partial y_p} \cdot (v_1, \dots, v_p)\, [s_1, \dots, s_n / x_1, \dots, x_n]$$

(with the variables y_j taken among $x_1, \dots, x_k$ and the terms v_j taken among $u_1, \dots, u_k$), and similarly

$$w'_r = \frac{\partial^{p_r} w}{\partial y_1^r \cdots \partial y_{p_r}^r} \cdot \left(v_1^r, \dots, v_{p_r}^r\right) [s_1, \dots, s_n / x_1, \dots, x_n]$$

(with the variables y_j^r taken among $x_1, \dots, x_k$ and the terms v_j^r taken among $u_1, \dots, u_k$). By inductive hypothesis, we know that $s' \in \overline{B^* \to A^*}$ and that $w'_1, \dots, w'_q \in \overline{B^*}$, and also that $w\, [s_1, \dots, s_n / x_1, \dots, x_n] \in \overline{B^*}$, and therefore

$$\left(\mathrm{D}_1^q s' \cdot (w'_1, \dots, w'_q)\right) (w\, [s_1, \dots, s_n / x_1, \dots, x_n]) \in \overline{A^*}$$

by definition of $B^* \to A^*$, and we conclude.

Differential application. Assume that t is a differential application, $t = \mathrm{D}_i s \cdot w$ with $x_1 : A_1, \dots, x_n : A_n \vdash s : B_1 \to \cdots \to B_i \to B = A$ and $x_1 : A_1, \dots, x_n : A_n \vdash w : B_i$. By lemma 7, the term $\left(\frac{\partial^k t}{\partial x_{i_1} \cdots \partial x_{i_k}} \cdot (u_1, \dots, u_k)\right) [s_1, \dots, s_n / x_1, \dots, x_n]$ is a sum of terms of the shape $\mathrm{D}_i s' \cdot w'$ with

$$s' = \frac{\partial^p s}{\partial y_1 \cdots \partial y_p} \cdot (v_1, \dots, v_p)\, [s_1, \dots, s_n / x_1, \dots, x_n]$$

(with the variables y_j taken among $x_1, \dots, x_k$ and the terms v_j taken among $u_1, \dots, u_k$), and similarly

$$w' = \frac{\partial^q w}{\partial z_1 \cdots \partial z_q} \cdot \left(v'_1, \dots, v'_q\right) [s_1, \dots, s_n / x_1, \dots, x_n]$$

(with the variables z_j taken among $x_1, \dots, x_k$ and the terms v'_j taken among $u_1, \dots, u_k$). By inductive hypothesis, we know that $s' \in \overline{A^*}$ and that $w' \in \overline{B_i^*}$. We conclude by lemma 18 that $\mathrm{D}_i s' \cdot w' \in \overline{A^*}$, as required.

Abstraction. Assume last that t is an abstraction, say $t = \lambda x\, s$. More precisely, we assume that the typing derivation of t ends with

$$\frac{x_1 : A_1, \ldots, x_n : A_n, x : B \vdash s : C}{x_1 : A_1, \ldots, x_n : A_n \vdash \lambda x\, s : B \to C}$$

and that $A = B \to C$. We must show that $\lambda x\, s' \in B^* \to C^*$, where

$$s' = \left(\frac{\partial^k s}{\partial x_{i_1} \cdots \partial x_{i_k}} \cdot (u_1, \ldots, u_k)\right) [s_1, \ldots, s_n / x_1, \ldots, x_n]$$

(we assume of course that x is different from all the variables x_i, and that x does not occur free in any of the terms u_j or s_i). So let $v_1, \ldots, v_p, w \in B^*$, we must show that

$$\left(\mathrm{D}_1^p(\lambda x\, s') \cdot (v_1, \ldots, v_p)\right) w \in B^*$$

and for this purpose, since B^* is saturated, it is sufficient to see that

$$\left(\frac{\partial^p s'}{\partial x^p} \cdot (v_1, \ldots, v_p)\right) [w/x] \in B^*$$

but this results from the inductive hypothesis on s and from lemma 5. □

Theorem 22 *The reduction relation β_{D} is weakly normalizing on typeable terms.*

Proof. Take first a closed term which is typeable of type A. Then by the interpretation lemma we have $t \in A^*$ but $A^* \subseteq \mathcal{N}$, so t is weakly normalizable. For a non closed term which is typeable in some typing context, any of its λ-closures is weakly normalizable, and so the term itself is weakly normalizable. □

Observe that saturated sets are closed under arbitrary intersections. Therefore, it is straightforward to adapt the proof above and show weak normalization of a second-oder version of the differential lambda-calculus.

A trivial size argument allows to prove strong normalization of the untyped *purely differential* version of the differential lambda-calculus, that is, the system where the only allowed ordinary applications are applications to 0 (observe that partial derivatives preserve this restriction on term construction).

5 Linear head reduction and the Taylor formula

The following version of Leibnitz formula will be useful in the sequel.

Lemma 23 *Let t and u be canonical terms and let x, y and z be variables such that $x \neq y$, none of these variables occurring free in u. Assume also that z does not occur free in t. Then*

$$\frac{\partial^n t\,[z/x, z/y]}{\partial z^n} \cdot u^n = \sum_{p=0}^{n} \binom{n}{p} \left(\frac{\partial^n t}{\partial x^p \partial y^{n-p}} \cdot u^n\right) [z/x, z/y]$$

Proof. It suffices to consider the case $n = 1$. The general case is obtained using lemma 4 and reasoning like in the proof of the usual Leibnitz formula. So one has to show that

$$\frac{\partial t\,[z/x, z/y]}{\partial z} \cdot u = \left(\frac{\partial t}{\partial x} \cdot u\right)[z/x, z/y] + \left(\frac{\partial t}{\partial y} \cdot u\right)[z/x, z/y]$$

and this is done by a simple induction on t. □

Lemma 24 *Let x be a variable, let $\bar{t} = t_1 \dots t_k$ be a sequence of canonical terms and let u be a simple term. Let y be a a variable different from x and not occurring free in $\bar{t}$ and in u. Assume also that x does not occur free in u. Let $n \geq 1$. Then*

$$\frac{\partial^n (x)\,\bar{t}}{\partial x^n} \cdot u^n = n \frac{\partial^{n-1} (u)\,\bar{t}}{\partial x^{n-1}} \cdot u^{n-1} + \left(\frac{\partial^n (y)\,\bar{t}}{\partial x^n} \cdot u^n\right)[x/y]\ .$$

Consequently

$$\left(\frac{\partial^n (x)\,\bar{t}}{\partial x^n} \cdot u^n\right)[0/x] = n\left(\frac{\partial^{n-1} (u)\,\bar{t}}{\partial x^{n-1}} \cdot u^{n-1}\right)[0/x]\ .$$

Proof. By lemma 23, we have, since $(x)\,\bar{t} = ((y)\,\bar{t})\,[x/y]$,

$$\begin{aligned}\frac{\partial^n (x)\,\bar{t}}{\partial x^n} \cdot u^n &= \left(\frac{\partial^n (y)\,\bar{t}}{\partial x^n} \cdot u^n\right)[x/y] \\ &\quad + \sum_{p=1}^{n} \binom{n}{p} \left(\frac{\partial^{n-p}}{\partial x^{n-p}} \left(\frac{\partial^p (y)\,\bar{t}}{\partial y^p} \cdot u^p\right) \cdot u^{n-p}\right)[x/y]\ .\end{aligned}$$

Since y does not occur free in $\bar{t}$, one has, for $p \geq 1$,

$$\frac{\partial^p (y)\,\bar{t}}{\partial y^p} \cdot u^p = \left(\frac{\partial^p y}{\partial y^p} \cdot u^p\right)\bar{t} = \begin{cases} (u)\,\bar{t} & \text{if } p = 1 \\ 0 & \text{if } p > 1 \end{cases}$$

and this proves the first statement. The second statement is a clear consequence of the first one and of lemma 8. □

Let $\star$ be a distinguished variable.

Theorem 25 *Let s and u be terms of the ordinary lambda-calculus, and assume that $(s)\,u$ is β-equivalent to $\star$. Then there is exactly one integer n such that $(\mathrm{D}_1^n s \cdot u^n)\,0 \not\simeq_{\beta_{\mathrm{D}}} 0$, and for this value of n, one has*

$$\frac{1}{n!}\,(\mathrm{D}_1^n s \cdot u^n)\,0 \simeq_{\beta_{\mathrm{D}}} \star\,.$$

This means that the Taylor formula

$$(s)\,u = \sum_{n=0}^{\infty} \frac{1}{n!}\,(\mathrm{D}_1^n s \cdot u^n)\,0$$

holds in a rather trivial way in that particular case. This formula always holds, semantically, at least in the simply typed case (see [Ehr01]), but is not so easy to interpret in general.

Proof. If the term t is solvable (i.e. has a head normal form, i.e. has a finite head reduction), we call *the* head normal form of t the result of the head reduction of t. We recall the well-known lambda-calculus property that if t and v are any terms such that $(t)\,v$ (resp. $t\,[v/x]$) is solvable, then so is t. We denote by $t\ \tau^k\ t'$ the fact that t head reduces in k steps to t'. Another standard property of lambda-calculus that we shall also use without further mention is that if $t\ \tau^k\ t'$ then $t\,[v/x]\ \tau^k\ t'\,[v/x]$.

Assume s and u are as in the theorem. Thus s is solvable. For any term v we denote by v' the term $v\,[u/x]$. We define a number $\mathrm{L}(s,u)$ by induction on the length of the head reduction of $(s_0)\,u$ to $\star$ where s_0 is the head normal form of s. Without loss of generality we may assume that s is in head normal form. There are two cases:

- $s = \lambda x\,\star$ where $x \neq \star$;
- $s = \lambda x\,(x)\,\bar{t}$ for some sequence of terms $\bar{t}$ such that $(u)\,\bar{t}' \simeq_\beta \star$.

In the former case we set $\mathrm{L}(s,u) = 0$. In the latter case we define s^+ by:

$$s^+ = \lambda x\,(u)\,\bar{t}\,.$$

Note that s^+, u satisfy the theorem assumptions because $(s^+)\,u \simeq_\beta (u)\,\bar{t}' \simeq_\beta (s)\,u$. Let s_0^+ be the head normal form of s^+; then $s_0^+ = \lambda x\,v$ for some v. Let k be the length of the head reduction of s^+. With these notations we have $(u)\,\bar{t}\ \tau^k\ v$. Therefore $(s)\,u\ \tau^1\ (u)\,\bar{t}'\ \tau^k\ v'$. On the other hand $\left(s_0^+\right)u\ \tau^1\ v'$ so that the length of the head reduction to $\star$ of $(s)\,u$ is strictly greater than the length of the head reduction of $\left(s_0^+\right)u$ to $\star$ as soon as $k > 0$.

If $k > 0$ then by induction $\mathrm{L}(s^+,u)$ is defined and we set

$$\mathrm{L}(s,u) = \mathrm{L}(s^+,u) + 1\,.$$

If $k = 0$ then $(u)\,\bar{t}$ is a head normal form, thus u is a variable. Furthermore $(u)\,\bar{t}' \simeq_\beta \star$ entails that the sequence $\bar{t}$ is empty and that $u = \star$. From which we deduce that we must have $s = \lambda x\,x$ and $s^+ = \lambda x\,\star$. In this case we set $\mathrm{L}(s,u) = 1$. We note that, since from the first case of the induction $\mathrm{L}(s^+,u) = 0$, we still have $\mathrm{L}(s,u) = \mathrm{L}(s^+,u) + 1$.

We now prove the result by induction on $\mathrm{L}(s,u)$, which happens to be the announced value of n. If $\mathrm{L}(s,u) = 0$, this means that $s \simeq_\beta \lambda x\,\star$. Then

$$\left(\mathrm{D}_1^n s \cdot u^n\right) 0 \simeq_{\beta_\mathrm{D}} \left(\frac{\partial^n \star}{\partial x^n} \cdot u^n\right)[0/x] = \begin{cases} 0 & \text{if } n \neq 0 \\ \star & \text{if } n = 0 = \mathrm{L}(s,u) \end{cases}$$

since $x \neq \star$. Assume now that $\mathrm{L}(s,u) > 0$ so that $s \simeq_\beta \lambda x\,(x)\,\bar{t}$. If $n = 0$ then

$$\left(\mathrm{D}_1^n s \cdot u^n\right) 0 \simeq_{\beta_\mathrm{D}} (s)\,0 \simeq_\beta (0)\,\bar{t} = 0$$

Otherwise if $n \geq 1$ we have

$$\left(\mathrm{D}_1^n s \cdot u^n\right) 0 \simeq_{\beta_\mathrm{D}} \left(\frac{\partial^n (x)\,\bar{t}}{\partial x^n} \cdot u^n\right)[0/x] \simeq_{\beta_\mathrm{D}} n\left(\frac{\partial^{n-1} (u)\,\bar{t}}{\partial x^{n-1}} \cdot u^{n-1}\right)[0/x]$$

by lemma 24. So we have

$$\left(\mathrm{D}_1^n s \cdot u^n\right) 0 \simeq_{\beta_\mathrm{D}} n\left(\mathrm{D}_1^{n-1} s^+ \cdot u^{n-1}\right) 0\,.$$

But by inductive hypothesis

$$\left(\mathrm{D}_1^{n-1} s^+ \cdot u^{n-1}\right) 0 \simeq_{\beta_{\mathrm{D}}} \begin{cases} (n-1)!\star & \text{if } n-1 = \mathrm{L}(s^+, u) = \mathrm{L}(s, u) - 1 \\ 0 & \text{otherwise} \end{cases}$$

and the result is proved. □

The number $\mathrm{L}(s,u)$ counts the substitutions of the successive head variables of s in the *head linear reduction* of $(s)\,u$ ([DHR96]). The head variable of s is the only occurrence of variable in $(s)\,u$ which may be considered as linear. So $\mathrm{L}(s,u)$ may be viewed as counting the number of *linear substitutions* by u that are performed along the reduction. The theorem enforces the intuition that the derivation operator implements linear substitution in lambda-calculus.

References

[BCL99] Gérard Boudol, Pierre-Louis Curien, and Carolina Lavatelli. A semantics for lambda calculi with resource. *Mathematical Structures in Computer Science*, 9(4):437–482, 1999.

[Bou93] Gérard Boudol. The lambda calculus with multiplicities. Technical Report 2025, INRIA Sophia-Antipolis, 1993.

[DHR96] Vincent Danos, Hugo Herbelin, and Laurent Regnier. Games semantics and abstract machines. In *Proceedings of the eleventh Symposium on Logic in Computer Science*. IEEE Computer Society Press, 1996.

[Ehr01] Thomas Ehrhard. On Köthe sequence spaces and linear logic. *Mathematical Structures in Computer Science*, 2001. To appear.

[KM97] Andreas Kriegl and Peter W. Michor. *The convenient setting of global analysis*. Number 53 in Mathematical Surveys and Monographs. American Mathematical Society, 1997.

[Kri90] Jean-Louis Krivine. *Lambda-Calcul : Types et Modèles*. Études et Recherches en Informatique. Masson, 1990.

[McB00] Conor McBride. The Derivative of a Regular Type is its Type of One-Hole Contexts. Unpublished, electronically available, 2000.

[Reg94] Laurent Regnier. Une équivalence sur les lambda-termes. *Theoretical Computer Science*, 126:281–292, 1994.

A Short survey of the Köthe space semantics

In [Ehr01], the first author introduced a semantics of linear logic in a quite simple framework of locally convex topological vector spaces, in some sense similar to coherence spaces. Indeed, these objects can be seen as pairs $X = (|X|, E_X)$ where $|X|$ (the web of X) is a (at most countable) set and E_X is a subset of $\mathbf{R}^{|X|}$ ($\mathbf{R}$ is the field of real numbers, but complex numbers can be used as well), and elements of E_X can be seen as kind of real-valued "subsets of $|X|$". Given two such real-valued subsets x and x', replace the operation "cardinality of the intersection of x and x'" (which is the key operation for defining coherence spaces) by the operation $\sum_{a\in|X|} x_a x'_a$. The (necessarily absolute) convergence of this sum is a notion of "orthogonality" between elements of $\mathbf{R}^{|X|}$, and we only required that E_X be equal to its bi-orthogonal in that sense. Then E_X has automatically a structure of locally convex topological vector space. Taking linear continuous maps as morphisms between these spaces gives rise to a model of multiplicative-additive linear logic (a star-autonomous symmetric monoidal closed category with denumerable products). In this category, exponentials can also be interpreted: $!X$ is a space whose web is the set of all finite multi-sets of elements of $|X|$. The vector space E_X is defined as the orthogonal of the space of all families $(A_\mu)_{\mu\in|!X|}$ such that the series $\sum_{\mu\in|!X|} A_\mu x^\mu$ converges absolutely for all $x \in E_X$ (where $(x^\mu)_{\mu\in|!X|}$ is defined by $x^\mu = \prod_{a\in|X|} x_a^{\mu(a)}$), which itself can be identified with a space of real-valued analytic (actually entire) functions defined on E_X. These functions of course are extremely regular (though ... not continuous!), and can be differentiated at any point. This differentiation operation can be seen as a morphism $d : !X \otimes X \multimap !X$: given an entire map h from X to Y, which can be seen as a linear continuous map $f : !X \multimap Y$, the function $f' = f \circ d : !X \otimes X \multimap Y$ can be seen as an entire function from X to $X \multimap Y$, which maps $x \in E_X$ to the derivative of h at point x. This operation $d : !X \otimes X \multimap !X$ can be defined in terms of two operations: a map $X \multimap !X$ and a map $!X \otimes !X \multimap !X$, the first being a kind of "anti-dereliction" and the second a kind of "anti-contraction" (actually a convolution product). From this viewpoint, $!X$ can be seen as endowed with an algebraic structure (in addition to its usual co-algebraic structure), and actually as a commutative Hopf algebra.