ELSEVIER

Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT®

Theoretical Computer Science 309 (2003) 1–41

Theoretical Computer Science

www.elsevier.com/locate/tcs

Fundamental Study

The differential lambda-calculus

Thomas Ehrhard*, Laurent Regnier

Institut de Mathématiques de Luminy, CNRS-UPR 9016, 163 Avenue de Luminy, F-13288 Marseille, France

Received 15 June 2001; received in revised form 19 March 2003; accepted 14 July 2003
Communicated by P.L. Curien

Abstract

We present an extension of the lambda-calculus with differential constructions. We state and prove some basic results (confluence, strong normalization in the typed case), and also a theorem relating the usual Taylor series of analysis to the linear head reduction of lambda-calculus.
© 2003 Elsevier B.V. All rights reserved.

Keywords: Lambda-calculus; Linear logic; Denotational semantics; Linear head reduction

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 [15], we denote by $(s)t$ the lambda-calculus application of s to t. The expression $(s)t_1 \dots t_n$ denotes the term $(\cdots(s)t_1 \cdots)t_n$ when $n \geqslant 1$, and s when $n = 0$. Accordingly, if $A_1, \dots, A_n$ and A are types, both expressions $A_1, \dots, A_n \to A$ and $A_1 \to \cdots \to A_n \to A$ denote the type $A_1 \to (\cdots(A_n \to A)\cdots)$. If $a_1, \dots, a_n$ are elements of some given set S, we denote by $[a_1, \dots, a_n]$ the corresponding multi-set over S. If x and y are variables, $\delta_{x,y}$ is equal to 1 if $x = y$ and to 0 otherwise. We denote by $\mathbb{N}^+$ the set of positive integers $\{1, 2, \dots\}$.

1. Introduction

1.1. Presentation

Denotational semantics vs. analysis. Denotational semantics usually interprets programs (lambda-terms) as partial functions. Partiality is necessary to account for the case of

* Corresponding author.
E-mail addresses: ehrhard@iml.univ-mrs.fr (T. Ehrhard), regnier@iml.univ-mrs.fr (L. Regnier).

doi:10.1016/S0304-3975(03)00392-X

divergent computations which arise as soon as recursion is available. This is achieved by considering the domains on which functions act as ordered sets with a least element that plays the role of the divergent program. The structure of domains further allows, as in Scott domains (we refer here to [2], as for all the standard lambda-calculus notions and results), to express some continuity properties of functions that account for the finite nature of computation. The main property is that a continuous function is always the limit (the least upper bound) of a sequence of *finite* functions (more precisely, of compact functions). In other words, the natural notion of approximation is by finite functions. This viewpoint on program approximation is reflected in the syntax by the theory of Böhm trees.

However this approach is orthogonal to one of the mainstream mathematical viewpoints where functions are total, defined on vector spaces rather than ordered sets and approximated by polynomials rather than by finite functions. In analysis, functions are approximated by (multi)linear maps through (iterated) differentiation.

Coherence semantics [12] suggests some way to conciliate these two viewpoints. The web of a coherence space is very similar to a basis of a vector space and the notion of stable linear map is similar to the usual notion of linear morphisms between vector spaces. This is reflected by the notation and terminology in use: tensor product, linear maps, direct product, etc. This correspondence has been further studied in [13], an attempt which however failed to fully account for the exponential connectives of linear logic.

As it turns out, it is possible to define models of the typed lambda-calculus (or equivalently of full linear logic) where types are interpreted as vector spaces and terms as functions defined by power series on these spaces (see [10]).[1] In these models all functions can be differentiated. However usual domain-theoretic intuitions are lost; in particular interpreting recursion becomes a delicate issue. A natural question to ask is then whether differentiation is a meaningful syntactic operation, and the present work provides a positive answer to this question by extending the lambda-calculus with differential operators and examining the obtained *differential lambda-calculus*. This gives also a syntactic solution to the recursion issue, since fix-point operators can be defined in the untyped differential lambda-calculus.

Linearity in the syntax. The syntactic concept of linearity has been introduced in lambda-calculus by the analysis of intuitionistic implication provided by linear logic [12]: in that sense the head variable occurrence in a term (i.e. the leftmost occurrence) is *linear* and is the only occurrence of variable in the term which is linear. This property of the head occurrence is in particular used in *linear head reduction*, a very natural notion in lambda-calculus that arises as soon as one wants to precisely evaluate reduction lengths. Linear head reduction has been considered by several authors (starting with De Bruijn [9] where it is called *minireduction*), and in particular, Danos and the second author, who related it to abstract machines (e.g. Krivine's machine [8]), game semantics and linear logic [7]. It is a kind of hyperlazy reduction strategy where at

[1] We give an example of such a model in the appendix of this paper. Interestingly enough, just as in Scott semantics, the basic idea of this *finiteness space* semantics is that the interaction between two agents should be finite. But the resulting models are completely different.

each step, only the head occurrence may be substituted. As a side effect, only subterms of the initial term are copied during a sequence of reductions.

It turns out that syntactic linearity can also be defined more abstractly through *non-deterministic choice*. Consider for instance three ordinary lambda-terms s, s' and t, and let $u = s+s'$ denote a non-deterministic choice between s and s'; we consider in the present discussion that this expression reduces to s or s', non-deterministically. Then it is reasonable to consider that $(u)t$ reduces to $(s)t + (s')t$ as the term u is used exactly once in the head reduction of $(u)t$. But we can certainly not say that $(t)u$ reduces to $(t)s + (t)s'$ as in the evaluation of $(t)u$, the argument u can be used several times leading to an arbitrary interleaving of uses of s and s', whereas the term $(t)s + (t)s'$ provides only the two trivial interleavings. We retrieve the fact, well known in linear logic, that application is linear in the function but not in the argument. So, as a first approximation at least, syntactic linearity can be defined as commutation to non-deterministic choice. This definition, quite satisfactorily, is completely similar to the usual algebraic definition of linearity.

Sums vs. non-determinism. Notice that in our axiomatization of sums in the differential lambda-calculus, the non-deterministic reduction rule $s+s' \to s$ will not be valid, but sums will nevertheless intuitively correspond to a version of non-deterministic choice where all the actual choice operations are postponed: the result of a term reduction will in general be a large formal sum of terms, and we can reduce the really "non-deterministic" part of the computation to a unique ultimate step consisting in choosing one term of that sum. This step is however a fiction and will not appear as a reduction rule of our calculus as otherwise essential properties such as confluence would obviously be lost. Another good reason for not considering this rule as an ordinary reduction rule is that, as suggested by the discussion above, it can be performed only when the "redex" stands in *linear* position.

Note that, from a logical viewpoint, introducing sums is not at all innocuous. In particular, as it is well known in category theory, it leads to the identification between finite categorical products and coproducts, the $\&$ and $\oplus$ connectives of linear logic.

Formal differentiation. Keeping in mind this syntactic viewpoint on linearity, we can give an account of differentiation within lambda-calculus. When $f(x)$ is a (sufficiently regular) function on a vector space E, its differential $f'(x)$ at each point $x \in E$ is a *linear map* on E; thus if $u \in E$ we can write $f'(x) \cdot u$ and read this as a *linear application* of $f'(x)$ to u. We therefore extend the lambda-calculus with a new syntactic construct: when t and u are terms, then $D_k t \cdot u$ is a term that may be read as the differential of t with respect to its kth argument, linearly applied to u. Similarly, the partial derivative $(\partial t/\partial x) \cdot u$, to be defined by induction on t, may be understood as a *linear substitution* of u for x in t, that is, a substitution of *exactly one linear occurrence* of x in t by u.

It is worth noting that linear substitution is a non-deterministic operation, as soon as the substituted variable has several occurrences: one has to choose a linear occurrence of the variable to be substituted and there are several possible such choices. This fundamental non-determinism of the 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.

Caution. Partial derivation is a linear substitution but does not behave exactly as one could expect from a substitution operation. Typically we have in general

$$\frac{\partial t}{\partial x} \cdot x \neq t.$$

If t is an ordinary lambda-term, the only case where the equation holds is when the variable x is linear in t, that is when its sole occurrence is in head position in t. In particular if x does not occur free in t then the left-hand side is 0.

1.2. 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 commutative semi-ring[2]) and a new term 0 is introduced. As already mentioned, this is necessary, because derivating with respect to a variable that occurs more than once 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 $D_i t \cdot u$ is added which represents the derivative of a term t with respect to its ith argument. This new term admits an additional parameter u, and is "linear" with respect to this parameter.

The most important addition is the reduction rule concerning differentiation:

$$D_1 \lambda x\, t \cdot u = \lambda x \left(\frac{\partial t}{\partial x} \cdot u \right).$$

This is a differential version of the β-rule (oriented from left to right). The term $(\partial t/\partial x) \cdot u$ is defined by induction on t and the various cases in this inductive definition correspond to well-known elementary results of differential calculus (chain rule, derivative of a multi-linear function...). This rewriting rule and the ordinary β-rule will be the two reduction rules of our system. As in ordinary lambda-calculus these rules are extended to all terms by context closure (a notion that needs some care to be well defined, as we shall see).

[2] A semi-ring is defined exactly like a ring, apart that one only requires addition to be a law of commutative monoid. A semi-ring will always be assumed to have a multiplicative unit. A typical example of semi-ring is $\mathbb{N}$ with $+$ and $\times$ as operations.

For these rules, we prove confluence (using an adaptation of the Tait–Martin–Löf technique) and strong normalization for the simply typed version[3] (using an adaptation of the Tait reducibility method). These two results enforce the idea that the differential lambda-calculus can be considered as a reasonable logical system.

For illustrating these new constructions, consider the term $t=(x)(x)y$ with two distinct free variables x and y. In this ordinary lambda-term, the only linear occurrence of a variable is the first occurrence of x; all the other occurrences are non-linear. The only occurrence of y in particular is non-linear as x could be instantiated by a non-linear function. Let u be another term, and let us examine what should be the value of the partial derivative $s=(\partial t/\partial x)\cdot u$. We have already isolated a linear occurrence of x, which is its first occurrence in t, so s will be of the shape $s=(u)(x)y+s'$ where s' will be obtained by substituting the other linear occurrences of x by u in t. But apparently, there are no other such occurrences. In the subterm $t_1=(x)y$, the occurrence of x is clearly linear, so u can be linearly substituted for x in t_1, leading to the term $s_1=(u)y$. The problem is that we cannot make the obvious choice for s', namely $s'=(x)(u)y$ as this term is not linear in u (if u is a variable, its occurrence in $(x)(u)y$ is not linear as it is not in head position). But the differential calculus gives us the solution: consider the first occurrence of x as a function of one argument (which is applied here to $(x)y$), and take the derivative $x'(z,h)=(D_1x\cdot h)z$ of this function with respect to its argument: this is a function of two parameters z and h, and it is linear with respect to its second parameter h. Now if in that term we replace h by $(u)y$ and z by $(x)y$, we obtain the term s' we were looking for. With our notations, $s'=(D_1x\cdot(u)y)(x)y$. To summarize,

$$\frac{\partial (x)(x)y}{\partial x}\cdot u=(u)(x)y+(D_1x\cdot(u)y)(x)y.$$

The term $D_1x\cdot h$ represents the function x where we have isolated one linear instance h of its parameter, that is, the function x to which we provide exactly one copy h of its parameter. This means that, if we replace x by an "actual function" $\lambda z\,t$ in that term, the resulting term should be equal to $\lambda z\,t'$, t' being t where z is replaced by h "exactly once", that is, linearly. In other terms, we should have $D_1\lambda z\,t\cdot h=\lambda z\,((\partial t/\partial z)\cdot h)$.

Of course, $D_1x\cdot h$ can be applied to another parameter which can be used an arbitrary number of times, in addition to the linearly provided parameter h which has to be used exactly once. This accounts for the presence of the "λz" which still appears in the expression $\lambda z\,((\partial t/\partial z)\cdot h)$.

Taylor formula. As expected, iterated differentiation yields a natural notion of multilinear approximation of the application of a term to another one. This notion relates to ordinary application through the Taylor formula:

$$(s)u=\sum_{n=0}^{\infty}\frac{1}{n!}\,(D_1^n s\cdot u^n)0,$$

[3] For strong normalization we require the coefficients to be natural numbers.

where $D_1^2 s \cdot u^2 = D_1(D_1 s \cdot u) \cdot u$ and so on. This equation is satisfied in the models we alluded to at the beginning of this introduction.

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 that 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 during the linear head reduction of $(s)u$ to $\star$.

More generally if one fully develops each application occurring in a term into its corresponding Taylor expansion, one expresses the term as an infinite sum of *purely differential terms* all of which contain only (multi)linear applications and applications to 0. Understanding the relation between the term and its full Taylor expansion might be the starting point of a renewing of the theory of approximations (usually based on Böhm trees).

1.3. Related work

Analysts have already extended smoothness and analyticity to "higher types", defining various cartesian closed categories of smooth and analytic functions (see e.g. [14] where 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 already occurred in different contexts. 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 [1,17].

Various authors introduced notions of linear substitution and reduction in the lambda-calculus. Let us quote one of them that carries intuitions similar to ours, the *lambda-calculus with multiplicities* (or with *resources*) [5,6]. In this system, application is written $(s)T$ where T is not a term, but a *bag* of terms written $T = (t_1^{p_1} \mid \cdots \mid t_n^{p_n})$ where the order on the elements of the bag is irrelevant. The p_i's are integers or ∞ and represent the number of time each t_i may be used during the reduction. The "∞" corresponds to the status of an argument in ordinary lambda-calculus and 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 (more precisely, decreasing its multiplicity, removing the term when its multiplicity reaches 0) and substituting it "linearly" for some occurrence of x in u through an explicit substitution mechanism.

Intuitions behind the differential lambda-calculus and the lambda-calculus with resources are very similar: the term $(D_1^n s \cdot (u_1, \ldots, u_n))t$ may be seen as the application of 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 $(D_1 s \cdot t)t$ are distinct in differential lambda-calculus. Also, the central role played by the sum in differential lambda-calculus seems to have no equivalent in resource lambda-calculus.

2. Syntax

Let R be a commutative semi-ring with unit; R can be for instance the set of natural numbers.[4] Given a set S, we denote by $R\langle S\rangle$ the free R-module generated by S, which can be described as the set of all R-valued functions defined on S which vanish for almost all values of their argument, with pointwise defined addition and scalar multiplication. As usual, an element t of $R\langle S\rangle$ will be denoted $\sum_{s\in S} a_s s$ where $s\mapsto a_s$ is the corresponding R-valued almost everywhere vanishing function (so that this sum is actually finite). We denote by $\mathrm{Supp}(t)$ the set $\{s\in S \mid a_s\neq 0\}$ (the *support* of t). Since R has a multiplicative unit 1, S can naturally be seen as a subset of $R\langle S\rangle$.

Let be given a denumerable set of variables. We define by induction on k an increasing family of sets (Δ_k). We set $\Delta_0=\emptyset$ and Δ_{k+1} is defined as follows.
Monotonicity: if t belongs to Δ_k then t belongs to Δ_{k+1}.
Variable: if $n\in\mathbb{N}$, x is a variable, $i_1,\dots,i_n\in\mathbb{N}^+=\mathbb{N}\backslash\{0\}$ and $u_1,\dots,u_n\in\Delta_k$, then

$$D_{i_1,\dots,i_n}x\cdot(u_1,\dots,u_n)$$

belongs to Δ_{k+1}. This term is identified with all the terms of the shape $D_{i_{\sigma(1)},\dots,i_{\sigma(n)}}x\cdot(u_{\sigma(1)},\dots,u_{\sigma(n)})\in\Delta_{k+1}$ where σ is a permutation on $\{1,\dots,n\}$.
Abstraction: if $n\in\mathbb{N}$, x is a variable, $u_1,\dots,u_n\in\Delta_k$ and $t\in\Delta_k$, then

$$D_1^n\lambda x\,t\cdot(u_1,\dots,u_n)$$

belongs to Δ_{k+1}. This term is identified with all the terms of the shape $D_1^n\lambda x\,t\cdot(u_{\sigma(1)},\dots,u_{\sigma(n)})\in\Delta_{k+1}$ where σ is a permutation on $\{1,\dots,n\}$.
Application: if $s\in\Delta_k$ and $t\in R\langle\Delta_k\rangle$, then

$$(s)t$$

belongs to Δ_{k+1}.

Setting $n=0$ in the first two clauses, and restricting application by the constraint that $t\in\Delta_k\subseteq R\langle\Delta_k\rangle$, one retrieves the usual definition of lambda-terms which shows that differential terms are a superset of ordinary lambda-terms.

The permutative identification mentioned above will be called *equality up to differential permutation*. We also work up to α-conversion.

Terms and simple terms. We denote by Δ the union of all the sets Δ_k. We call *simple terms* the elements of Δ and *differential terms* or simply *terms* the elements of $R\langle\Delta\rangle$. Observe that $R\langle\Delta\rangle=\bigcup R\langle\Delta_k\rangle$. We write $\Delta(R)$ instead of Δ when we want to make explicit the underlying semi-ring.

Induction on terms. Proving a property by induction on terms means proving this property for each term t by induction on the least k such that $t\in R\langle\Delta_k\rangle$ (number which can be called the *height* of t).

[4] For expressing the Taylor formula one has to assume that R is a field. One could also take $R=\{0,1\}$ with $1+1=1$ in which case sum of terms is idempotent and the coefficients of the Taylor formula are all equal to 1.

Extending the syntactic constructs. Observe that if s is a simple term then $\lambda x\, s$ is a simple term, and if t is a term then $(s)t$ is a simple term. We extend these constructions by linearity to arbitrary terms s by setting $\lambda x \sum_{s\in\Delta} a_s s = \sum_{s\in\Delta} a_s \lambda x\, s$ and $(\sum_{s\in\Delta} a_s s)t = \sum_{s\in\Delta} a_s (s)t$. It is crucial that application is linear in the function but *not* linear in the argument.

Given simple terms s and u, we define by induction on s the simple term $D_i s\cdot u$ for each $i\in\mathbb{N}^+$.

$$D_i(D_{i_1,\dots,i_n} x\cdot(u_1,\dots,u_n))\cdot u = D_{i,i_1,\dots,i_n} x\cdot(u,u_1,\dots,u_n),$$

$$D_i(D_1^n \lambda x\, t\cdot(u_1,\dots,u_n))\cdot u = \begin{cases} D_1^{n+1}\lambda x\, t\cdot(u,u_1,\dots,u_n) & \text{if } i=1,\\ D_1^n\lambda x\,(D_{i-1}t\cdot u)\cdot(u_1,\dots,u_n) & \text{if } i>1,\end{cases}$$

$$D_i(t)v\cdot(u_1,\dots,u_n) = (D_{i+1}t\cdot(u_1,\dots,u_n))v.$$

Of course, in the second clause, we assume without loss of generality that x does not occur free in u.

We extend this construction to arbitrary terms s and u by bilinearity: $D_i(\sum_{s\in\Delta} a_s s)\cdot \sum_{u\in\Delta} b_u u = \sum_{s,u\in\Delta} a_s b_u D_i s\cdot u$.

This choice of notations introduces an ambiguity for expressions such as $D_{i_1,\dots,i_n} x\cdot (u_1,\dots,u_n)$ or $D_1^n\lambda x\, t\cdot(u_1,\dots,u_n)$ which now can be considered either as expression of the basic syntax of terms, or as expression in our extended setting. Obviously, no conflict can occur and this ambiguity is completely harmless.

An easy induction on s shows that

$$D_i(D_j s\cdot u)\cdot v = D_j(D_i s\cdot v)\cdot u.$$

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

$$D_{i_{\sigma(1)},\dots,i_{\sigma(k)}} t\cdot(u_{\sigma(1)},\dots,u_{\sigma(k)}) = D_{i_1,\dots,i_k} t\cdot(u_1,\dots,u_k).$$

When in particular the indices $i_1,\dots,i_k$ have a common value i, we write $D_i^k t\cdot(u_1,\dots,u_k)$ instead of $D_{i_1,\dots,i_k} t\cdot(u_1,\dots,u_k)$ If furthermore all the terms u_i are identical to a unique term u we simply write $D_i^k t\cdot u^k$.

Intuitions. We shall give a typing system for these terms which conveys basic intuitions on the differential lambda-calculus: the interested reader can immediately have a look at the first part of Section 3 where this system is presented and more intuitions are given.

2.1. Substitution operators

Ordinary substitution. Let x be a variable and let t be a term; we define by induction on the term s the substitution of x by t in s denoted $s\,[t/x]$.

$$(D_{i_1,\dots,i_n} y\cdot(u_1,\dots,u_n))\,[t/x] = \begin{cases} D_{i_1,\dots,i_n} t\cdot(u_1\,[t/x],\dots,u_n\,[t/x]) & \text{if } x=y,\\ D_{i_1,\dots,i_n} y\cdot(u_1\,[t/x],\dots,u_n\,[t/x]) & \text{otherwise,}\end{cases}$$

$$(D_1^n \lambda y\, v \cdot (u_1, \ldots, u_n))[t/x] = D_1^n \lambda y\, (v\,[t/x]) \cdot (u_1\,[t/x], \ldots, u_n\,[t/x]),$$

$$(v)w\,[t/x] = (v\,[t/x])w\,[t/x].$$

Given the definition at heights $\leqslant k$, these clauses define $s\,[t/x]$ for each simple term s of height $k+1$. We conclude the definition for arbitrary terms of height $k+1$ by setting

$$\left(\sum_{v \in \Delta_{k+1}} a_v v\right) [t/x] = \sum_{v \in \Delta_{k+1}} a_v v\,[t/x].$$

In the abstraction case one takes the usual precautions assuming without loss of generality thanks to α-conversion that $y \neq x$ and y does not occur free in t. Observe that substitution is linear in s but not in t.

Derivation. 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 of s with respect to x along u* denoted $(\partial s/\partial x) \cdot u$. As substitution, it is defined by induction on terms.

$$\frac{\partial D_{i_1,\ldots,i_n} y \cdot (u_1, \ldots, u_n)}{\partial x} \cdot u = \delta_{x,y} D_{i_1,\ldots,i_n} u \cdot (u_1, \ldots, u_n) + \sum_{i=1}^{n} D_{i_1,\ldots,i_n} y \cdot \left(u_1, \ldots, \frac{\partial u_i}{\partial x} \cdot u, \ldots, u_n\right)$$

$$\frac{\partial D_1^n \lambda y\, v \cdot (u_1, \ldots, u_n)}{\partial x} \cdot u = D_1^n \lambda y \left(\frac{\partial v}{\partial x} \cdot u\right) \cdot (u_1, \ldots, u_n) + \sum_{i=1}^{n} D_1^n\, \lambda y\, v \cdot \left(u_1, \ldots, \frac{\partial u_i}{\partial x} \cdot u, \ldots, u_n\right)$$

$$\frac{\partial (v)w}{\partial x} \cdot u = \left(\frac{\partial v}{\partial x} \cdot u\right) w + \left(D_1 v \cdot \left(\frac{\partial w}{\partial x} \cdot u\right)\right) w$$

$$\frac{\partial}{\partial x} \left(\sum_{v \in \Delta} a_v v\right) \cdot u = \sum_{v \in \Delta} a_v \frac{\partial v}{\partial x} \cdot u,$$

where, again, in the abstraction case, the usual precautions have to be respected: y should be different from x and should not occur free in the term u.

This definition says essentially that partial derivative distributes over syntactic constructs that are linear such as abstraction. The application case is the most involved one: partial derivative is safely applied to the function v because application is linear in the function, producing the term $((\partial v/\partial x) \cdot u)w$. But in order to apply it to the argument w which is not in linear position, we intuitively follow two steps: firstly we replace $(v)w$ by $(D_1 v \cdot w)w$ getting a *linear* copy of w; secondly we apply partial derivative to this copy. This can also be seen as a formal way of expressing the chain rule of differential calculus.

Note that, despite the intuition that $D_1 v \cdot w$ is a linear application, it is false that $(v)w = (D_1 v \cdot w)w$ even up to differential reduction (to be defined soon). Thus again partial derivative may be considered as a substitution operator only to a limited extent.

Lemma 1. *If* $i \in \mathbb{N}^+$ *and* $t, u, v \in R\langle \Delta \rangle$, *we have*

$$(D_i t \cdot u)[v/x] = D_i t[v/x] \cdot u\,[v/x]$$

and

$$\frac{\partial D_i t \cdot u}{\partial x} \cdot v = D_i \left(\frac{\partial t}{\partial x} \cdot v \right) \cdot u + D_i t \cdot \left(\frac{\partial u}{\partial x} \cdot v \right)$$

Both proofs are simple inductions on t.

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. This lemma allows us to write the parallel substitution of the terms u_i for the variables x_i in t, denoted $t\,[u_1, \ldots, u_n/x_1, \ldots, x_n]$, when none of the variables x_i is free in any of the terms u_i.

Lemma 3. *If* x *is not free in* t, *then* $(\partial t/\partial x) \cdot u = 0$. *For any 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. In particular the application case is true thanks to our linearization on the fly.

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

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

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

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

Proof. This is proven by an easy induction on t. We deal only with the cases "variable" and "application".

If $t = D_{i_1,\ldots,i_n} z \cdot (u_1, \ldots, u_n)$ we have

$$
\begin{aligned}
\frac{\partial}{\partial x}\left(\frac{\partial t}{\partial y} \cdot v\right) \cdot u = {} & \frac{\partial}{\partial x}\Bigg(\delta_{y,z} D_{i_1,\ldots,i_n} v \cdot (u_1, \ldots, u_n) \\
& + \sum_{i=1}^{n} D_{i_1,\ldots,i_n} z \cdot \left(u_1, \ldots, \frac{\partial u_i}{\partial y} \cdot v, \ldots, u_n\right)\Bigg) \cdot u \\
= {} & \delta_{y,z} D_{i_1,\ldots,i_n} \left(\frac{\partial v}{\partial x} \cdot u\right) \cdot (u_1, \ldots, u_n) && (1) \\
& + \delta_{y,z} \sum_{i=1}^{n} D_{i_1,\ldots,i_n} v \cdot \left(u_1, \ldots, \frac{\partial u_i}{\partial x} \cdot u, \ldots, u_n\right) && (2) \\
& + \delta_{x,z} \sum_{i=1}^{n} D_{i_1,\ldots,i_n} u \cdot \left(u_1, \ldots, \frac{\partial u_i}{\partial y} \cdot v, \ldots, u_n\right) && (3) \\
& + \sum_{i=1}^{n} \sum_{j=1}^{i-1} D_{i_1,\ldots,i_n} z \cdot \left(u_1, \ldots, \frac{\partial u_j}{\partial x} \cdot u, \ldots, \frac{\partial u_i}{\partial y} \cdot v, \ldots, u_n\right) && (4) \\
& + \sum_{i=1}^{n} \sum_{j=i+1}^{n} D_{i_1,\ldots,i_n} z \cdot \left(u_1, \ldots, \frac{\partial u_i}{\partial y} \cdot v, \ldots, \frac{\partial u_j}{\partial x} \cdot u, \ldots, u_n\right) && (5) \\
& + \sum_{i=1}^{n} D_{i_1,\ldots,i_n} z \cdot \left(u_1, \ldots, \frac{\partial}{\partial x}\left(\frac{\partial u_i}{\partial y} \cdot v\right) \cdot u, \ldots, u_n\right). && (6)
\end{aligned}
$$

Let us denote by S_1, S_2, S_3, S_4, S_5 and S_6 the six summands of this expression. By inductive hypothesis we have $S_6 = S_7 + S_8$ where $S_7 = \sum_{i=1}^{n} D_{i_1,\ldots,i_n} z \cdot (u_1, \ldots, \partial/\partial y((\partial u_i/\partial x) \cdot u) \cdot v, \ldots, u_n)$ and $S_8 = \sum_{i=1}^{n} D_{i_1,\ldots,i_n} z \cdot (u_1, \ldots, (\partial u_i/\partial y) \cdot ((\partial v/\partial x) \cdot u), \ldots, u_n)$. But $S_1 + S_8 = (\partial t/\partial y) \cdot ((\partial v/\partial x) \cdot u)$ and, since $(\partial u/\partial y) \cdot v = 0$, we have by a similar computation

$$
\begin{aligned}
\frac{\partial}{\partial y}\left(\frac{\partial t}{\partial x} \cdot u\right) \cdot v = {} & \delta_{x,z} \sum_{i=1}^{n} D_{i_1,\ldots,i_n} u \cdot \left(u_1, \ldots, \frac{\partial u_i}{\partial y} \cdot v, \ldots, u_n\right) \\
& + \delta_{y,z} \sum_{i=1}^{n} D_{i_1,\ldots,i_n} v \cdot \left(u_1, \ldots, \frac{\partial u_i}{\partial x} \cdot u, \ldots, u_n\right) \\
& + \sum_{i=1}^{n} \sum_{j=1}^{i-1} D_{i_1,\ldots,i_n} z \cdot \left(u_1, \ldots, \frac{\partial u_j}{\partial y} \cdot v, \ldots, \frac{\partial u_i}{\partial x} \cdot u, \ldots, u_n\right) \\
& + \sum_{i=1}^{n} \sum_{j=i+1}^{n} D_{i_1,\ldots,i_n} z \cdot \left(u_1, \ldots, \frac{\partial u_i}{\partial x} \cdot u, \ldots, \frac{\partial u_j}{\partial y} \cdot v, \ldots, u_n\right) \\
& + \sum_{i=1}^{n} D_{i_1,\ldots,i_n} z \cdot \left(u_1, \ldots, \frac{\partial}{\partial y}\left(\frac{\partial u_i}{\partial x} \cdot u\right) \cdot v, \ldots, u_n\right) \\
= {} & S_3 + S_2 + S_5 + S_4 + S_7
\end{aligned}
$$

and we are done. □

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

$$
\begin{aligned}
&\frac{\partial}{\partial x}\left(\frac{\partial t}{\partial y}\cdot v\right)\cdot u\\
&\quad=\frac{\partial}{\partial x}\left(\left(\frac{\partial t_1}{\partial y}\cdot v\right)t_2+\left(D_1 t_1\cdot\left(\frac{\partial t_2}{\partial y}\cdot v\right)\right)t_2\right)\cdot u\\
&\quad=\left(\frac{\partial}{\partial x}\left(\frac{\partial t_1}{\partial y}\cdot v\right)\cdot u\right)t_2+\left(D_1\left(\frac{\partial t_1}{\partial y}\cdot v\right)\cdot\left(\frac{\partial t_2}{\partial x}\cdot u\right)\right)t_2\\
&\qquad+\left(D_1\left(\frac{\partial t_1}{\partial x}\cdot u\right)\cdot\left(\frac{\partial t_2}{\partial y}\cdot v\right)\right)t_2+\left(D_1 t_1\cdot\left(\frac{\partial}{\partial x}\left(\frac{\partial t_2}{\partial y}\cdot v\right)\cdot u\right)\right)t_2\\
&\qquad+\left(D_1^2 t_1\cdot\left(\frac{\partial t_2}{\partial y}\cdot v,\frac{\partial t_2}{\partial x}\cdot u\right)\right)t_2\\
&\quad=\left(D_1\left(\frac{\partial t_1}{\partial y}\cdot v\right)\cdot\left(\frac{\partial t_2}{\partial x}\cdot u\right)\right)t_2+\left(D_1\left(\frac{\partial t_1}{\partial x}\cdot u\right)\cdot\left(\frac{\partial t_2}{\partial y}\cdot v\right)\right)t_2\\
&\qquad+\left(D_1^2 t_1\cdot\left(\frac{\partial t_2}{\partial y}\cdot v,\frac{\partial t_2}{\partial x}\cdot u\right)\right)t_2\\
&\qquad+\left(\frac{\partial}{\partial y}\left(\frac{\partial t_1}{\partial x}\cdot u\right)\cdot v\right)t_2+\left(\frac{\partial t_1}{\partial y}\cdot\left(\frac{\partial v}{\partial x}\cdot u\right)\right)t_2\\
&\qquad+\left(D_1 t_1\cdot\left(\frac{\partial}{\partial y}\left(\frac{\partial t_2}{\partial x}\cdot u\right)\cdot v\right)\right)t_2+\left(D_1 t_1\cdot\left(\frac{\partial t_2}{\partial y}\cdot\left(\frac{\partial v}{\partial x}\cdot u\right)\right)\right)t_2
\end{aligned}
$$

by inductive hypothesis applied to t_1 and t_2. On the other hand, using the fact that $(\partial u/\partial y)\cdot v=0$, a similar computation leads to

$$
\begin{aligned}
&\frac{\partial}{\partial y}\left(\frac{\partial t}{\partial x}\cdot u\right)\cdot v\\
&\quad=\left(D_1\left(\frac{\partial t_1}{\partial y}\cdot v\right)\cdot\left(\frac{\partial t_2}{\partial x}\cdot u\right)\right)t_2+\left(D_1\left(\frac{\partial t_1}{\partial x}\cdot u\right)\cdot\left(\frac{\partial t_2}{\partial y}\cdot v\right)\right)t_2\\
&\qquad+\left(D_1^2 t_1\cdot\left(\frac{\partial t_2}{\partial x}\cdot u,\frac{\partial t_2}{\partial y}\cdot v\right)\right)t_2\\
&\qquad+\left(\frac{\partial}{\partial y}\left(\frac{\partial t_1}{\partial x}\cdot u\right)\cdot v\right)t_2+\left(D_1 t_1\cdot\left(\frac{\partial}{\partial y}\left(\frac{\partial t_2}{\partial x}\cdot u\right)\cdot v\right)\right)t_2
\end{aligned}
$$

and we conclude since $D_1^2 t_1\cdot((\partial t_2/\partial y)\cdot v,(\partial t_2/\partial x)\cdot u)=D_1^2 t_1\cdot((\partial t_2/\partial x)\cdot u,(\partial t_2/\partial y)\cdot v)$ and since

$$\left(\frac{\partial t_1}{\partial y}\cdot\left(\frac{\partial v}{\partial x}\cdot u\right)\right)t_2+\left(D_1 t_1\cdot\left(\frac{\partial t_2}{\partial y}\cdot\left(\frac{\partial v}{\partial x}\cdot u\right)\right)\right)t_2=\frac{\partial (t_1)t_2}{\partial y}\cdot\left(\frac{\partial v}{\partial x}\cdot u\right).$$

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 notation when the condition above on variables is not fulfilled).

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

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

$$\frac{\partial t[v/y]}{\partial x}\cdot u=\left(\frac{\partial t}{\partial x}\cdot u\right)[v/y]+\left(\frac{\partial t}{\partial y}\cdot\left(\frac{\partial v}{\partial x}\cdot u\right)\right)[v/y]. \tag{1}$$

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

$$\frac{\partial t\,[v/y]}{\partial x}\cdot u=\left(\frac{\partial t}{\partial x}\cdot u\right)[v/y]. \tag{2}$$

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:

$$\left(\frac{\partial t}{\partial x}\cdot u\right)[v/y]=\frac{\partial t}{\partial x}\cdot(u\,[v/y]). \tag{3}$$

This is a simple induction on t.

In the general case, let y' be a 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 $(\partial t/\partial x)\cdot u=((\partial t'/\partial x)\cdot u)[y/y']$. Then by Lemma 2 we have $((\partial t/\partial x)\cdot u)[v/y]=((\partial t'/\partial x)\cdot u)[v/y]\,[v/y']$ (because y' is not free in v). So by (3), since y does not occur in t', $((\partial t/\partial x)\cdot u)[v/y]=((\partial t'/\partial x)\cdot u\,[v/y])[v/y']$, and so $((\partial t/\partial x)\cdot u)[v/y]=(\partial t'\,[v/y']/\partial x)\cdot u\,[v/y]$ 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'.

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. *If the variables* $x_1,\ldots,x_k$ *do not occur free in the terms* $u_1,\ldots,u_k$, *the derivative* $(\partial^k D_i s\cdot t/\partial x_1\cdots\partial x_k)\cdot(u_1,\ldots,u_k)$ *is a finite sum of expressions*

$$D_i s'\cdot t',$$

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

$$s' = \frac{\partial^{r_0} s}{\partial y_1^{(0)}\cdots\partial y_{r_0}^{(0)}}\cdot(u_1^{(0)},\ldots,u_{r_0}^{(0)}),$$

$$t' = \frac{\partial^{r_1} t}{\partial y_1^{(1)}\cdots\partial y_{r_1}^{(1)}}\cdot(u_1^{(1)},\ldots,u_{r_1}^{(1)})$$

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

Lemma 8. *If the variables* $x_1,\ldots,x_k$ *do not occur free in the terms* $u_1,\ldots,u_k$, *the derivative* $(\partial^k(s)t/\partial x_1\cdots\partial x_k)\cdot(u_1,\ldots,u_k)$ *is a finite sum of expressions*

$$(D_1^q s'\cdot(t'_1,\ldots,t'_q))t$$

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

$$s' = \frac{\partial^{r_0} s}{\partial y_1^{(0)}\cdots\partial y_{r_0}^{(0)}}\cdot(u_1^{(0)},\ldots,u_{r_0}^{(0)})$$

$$t'_j = \frac{\partial^{r_j} t}{\partial y_1^{(j)}\cdots\partial y_{r_j}^{(j)}}\cdot(u_1^{(j)},\ldots,u_{r_j}^{(j)})$$

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

Both lemmas are proved by induction on k. □

3. Differential reduction

Let us begin by introducing some terminology and easy lemmas on relations between terms. We shall consider two kind of relations: relations from terms to terms which are subsets of $R\langle\Delta\rangle\times R\langle\Delta\rangle$ and relations from *simple* terms to terms which are subsets of $\Delta\times R\langle\Delta\rangle$.

Linear relations. A relation τ from terms to terms is *linear* if $0\,\tau\,0$ and $at+bu\,\tau\,at'+bu'$ as soon as $t\,\tau\,t'$ and $u\,\tau\,u'$. In the particular case where τ is a functional relation, this means that it is a linear map. As with linear maps the image of a linear subspace of $R\langle\Delta\rangle$ by τ is a linear subspace. However the image of the subspace $\{0\}$ is not in general reduced to $\{0\}$ (this is easily seen to be a necessary and sufficient condition for τ to be functional). Observe that as a subset of the linear space $R\langle\Delta\rangle\times R\langle\Delta\rangle = R\langle\Delta\rangle\oplus R\langle\Delta\rangle$, a relation is linear iff it is a linear subspace (just as for linear functions).

Contextual relations. A relation τ from terms to terms is *contextual* if it is reflexive, linear and satisfies the following conditions. Let x be a variable and s, t, s' and t' be terms such that $s\,\tau\,s'$ and $t\,\tau\,t'$; then

$$\lambda x\, s\,\tau\,\lambda x\, s',$$

$$(s)t\,\tau\,(s')t',$$

$$D_i s\cdot t\,\tau\,D_i s'\cdot t'.$$

Lemma 9. *Let τ be a contextual relation from terms to terms. If t, u and u' are terms such that $u\,\tau\,u'$, then $(\partial t/\partial x)\cdot u\,\tau\,(\partial t/\partial x)\cdot u'$ and $t\,[u/x]\,\tau\,t\,[u'/x]$*

This is proved by induction on t.

Extending a relation. Given a relation τ from simple terms to terms we shall define two relations $\bar\tau$ and $\tilde\tau$ from terms to terms by:

- $t\,\bar\tau\,t'$ if $t=\sum_{i=1}^n a_i s_i$ and $t'=\sum_{i=1}^n a_i s_i'$ where the terms s_i are simple, the terms s_i' are such that $s_i\,\tau\,s_i'$ for each i and the a_i are scalars;
- $t\,\tilde\tau\,t'$ if $t=as+u$ and $t'=as'+u$ where s is a simple term, s' is a term such that $s\,\tau\,s'$, u is a term and a is a nonzero scalar.

These two operations are monotone and ω-continuous in τ which means that given an increasing sequence of relations τ_n whose union is τ, the sequences $\bar\tau_n$ and $\tilde\tau_n$ are increasing and their respective unions are $\bar\tau$ and $\tilde\tau$.

Observe that $\tilde\tau$ is not linear. On the other hand $\bar\tau$ is the least linear relation from terms to terms which contains τ. In that sense it can be thought of as the extension by linearity of τ. Note that, given $t=\sum a_i s_i$ where the s_i are pairwise distinct simple terms and t' such that $t\,\bar\tau\,t'$, it is not true in general that t' may be written $t'=\sum a_i s_i'$ with $s_i\,\tau\,s_i'$. Typically if $t=0$ and s is any simple term such that $s\,\tau\,s'$ and $s\,\tau\,s''$ then we have $t\,\bar\tau\,s'-s''$ (assuming -1 belongs to R).

One-step reduction. We are now ready to give the reduction rules, completing the definition of the differential lambda-calculus. Informally the one step reduction β^1 is the smallest relation that is closed under syntactic constructions (e.g., if $s\,\beta^1\,s'$ then $(s)v\,\beta^1\,(s')v$) and that contains:

Beta-reduction.

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

Differential reduction. When x is not free in u:

$$D_1 \lambda x\, s \cdot u \quad \text{reduces to} \quad \lambda x \left(\frac{\partial s}{\partial x} \cdot u \right).$$

The last rule is similar to β-reduction up to the fact that it only substitutes one linear occurrence of variable. This is why the abstraction remains. This rule is compatible with the intuitions behind differentiation of a function (D_1) and partial derivation with respect to a variable.

Restricted to ordinary lambda-terms, this (to be precisely defined) reduction is just ordinary beta-reduction, which shows that the differential lambda-calculus is a conservative extension of the ordinary lambda-calculus.

The simplest definition of the one step reduction in ordinary lambda-calculus is by induction on the term to be reduced. By induction on (the size of) t, one defines the set of all terms t' such that t reduces to t':

- x never reduces to t';
- $\lambda x\, s$ reduces to t' if s reduces to s' and $t' = \lambda x\, s'$;
- $(s)u$ reduces to t' if s reduces to s' and $t' = (s')u$, or u reduces to u' and $t' = (s)u'$, or $s = \lambda x\, v$ and $t' = v\,[u/x]$.

This method is not available here because, when $-1 \in R$, we must accept that $(x)0$ reduces to $(x)(s' - s)$ as soon as s reduces to s', for an arbitrary term s. But specifying this in a definition of the reduction by induction on the size of the term to be reduced would require the size of s to be less than the size of $(x)0$, where s is an arbitrary term!

Accordingly we shall build β^1 by induction on the *depth* at which the redex is fired. We define an increasing sequence β^1_k of relations from simple terms to terms by: β^1_0 is the empty relation and

- $(s)u\, \beta^1_{k+1}\, t'$ if $t' = (s')u$ with $s\, \beta^1_k\, s'$, or $t' = (s)u'$ with $u\, \tilde{\beta}^1_k\, u'$, or $s = \lambda x\, v$ and $t' = [u/x]$;
- $D_{i_1,\dots,i_n} x \cdot (u_1, \dots, u_n)\, \beta^1_{k+1}\, t'$ if $t' = D_{i_1,\dots,i_n} x \cdot (u'_1, \dots, u'_n)$ with $u_j\, \beta^1_k\, u'_j$ for exactly one $j \in \{1, \dots, n\}$ and $u_i = u'_i$ for $i \neq j$;
- $D^n_1 \lambda x\, s \cdot (u_1, \dots, u_n)\, \beta^1_{k+1}\, t'$ if $t' = D^n_1 \lambda x\, s' \cdot (u_1, \dots, u_n)$ with $s\, \beta^1_k\, s'$, or $t' = D^n_1 \lambda x\, s \cdot (u'_1, \dots, u'_n)$ with $u_j\, \beta^1_k\, u'_j$ for exactly one $j \in \{1, \dots, n\}$ and $u_i = u'_i$ for $i \neq j$, or

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

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

We define $\beta^1 = \bigcup \beta^1_k$. Thanks to the ω-continuity of relation extensions, we have that $\tilde{\beta}^1 = \bigcup \tilde{\beta}^1_k$.

The one-step reduction is weakly contextual in the following sense:

Lemma 10. *We have the following*:
1. $(t)u\, \beta^1\, (t')u$ *if* t *is simple and* $t\, \beta^1\, t'$;
2. $(u)t\, \beta^1\, (u)t'$ *if* u *is simple and* $t\, \tilde{\beta}^1\, t'$;
3. $\lambda x\, t\, \beta^1\, \lambda x\, t'$ *if* t *is simple and* $t\, \beta^1\, t'$;

4. $D_i t \cdot u\, \beta^1\, D_i t' \cdot u$ *if* t, u *are simple and* $t\, \beta^1\, t'$;
5. $D_i u \cdot t\, \beta^1\, D_i u \cdot t'$ *if* t, u *are simple and* $t\, \beta^1\, t'$.

Proof. Each statement is separately proved. Statements 1 and 3 (application in function position and abstraction) are immediate by definition of β^1.

Statement 2 (application in argument position) results from the definition of $\tilde{\beta}^1$ and the continuity remark: from $t\, \tilde{\beta}^1\, t'$ we deduce that there is a k such that $t\, \tilde{\beta}^1_k\, t'$. Thus $(u)t\, \beta^1_{k+1}\, (u)t'$ and we are done.

Statement 5 is shown by induction on the simple term u.

Finally we prove statement 4 by induction on k showing that whenever $t\, \beta^1_k\, t'$ then $D_i t \cdot u\, \beta^1\, D_i t' \cdot u$. If $k = 0$ there is nothing to prove. In the case $k + 1$ we have the following possibilities:

- $t = (s)v$; then $D_i t \cdot u = (D_{i+1} s \cdot u)v$ is simple. If $t' = (s')v$ where $s\, \beta^1_k\, s'$ then by inductive hypothesis $D_{i+1} s \cdot u\, \beta^1\, D_{i+1} s' \cdot u$ and we conclude by applying statement 1.
  If $t' = (s)v'$ where $v\, \tilde{\beta}^1_k\, v'$ then we conclude by statement 2.
  If $s = \lambda x\, w$ and $t' = w\,[v/x]$ then $D_i t \cdot u = (\lambda x\, D_i w \cdot u)v$ (assuming without loss of generality that x is not free in u). Now w is simple because s is; so we have $(\lambda x\, D_i w \cdot u)v\, \beta^1_1 (D_i w \cdot u)\,[v/x] = D_i w\,[v/x] \cdot u = D_i t' \cdot u$ and we are done.
- $t = D_{i_1,\ldots,i_n} x \cdot (u_1, \ldots, u_n)$ and $t' = D_{i_1,\ldots,i_n} x \cdot (u_1', u_2, \ldots, u_n)$ with $u_1\, \beta^1_k\, u_1'$ (up to equivalence by permutations, we may suppose without loss of generality that the one step reduction takes place in the first differential argument). Then $D_i t \cdot u = D_{i,i_1,\ldots,i_n} x \cdot (u, u_1, \ldots, u_n)$ and $D_i t' \cdot u = D_{i,i_1,\ldots,i_n} x \cdot (u, u_1', u_2, \ldots, u_n)$; we conclude by definition of β^1.
- $t = D_1^n \lambda x\, s \cdot (u_1, \ldots, u_n)$ and $i = 1$. Then $D_i t \cdot u = D_1^{n+1} \lambda x\, s \cdot (u_1, \ldots, u_n, u)$. If the one-step reduction takes place in s or one of the terms u_i we reason as in the preceding case. Otherwise we have $t' = D_1^{n-1} \lambda x((\partial s/\partial x) \cdot u_1) \cdot (u_2, \ldots, u_n)$. Then $D_i t' \cdot u = D_1^n \lambda x ((\partial s/\partial x) \cdot u_1) \cdot (u_2, \ldots, u_n, u)$ and we are done.
- $t = D_1^n \lambda x\, s \cdot (u_1, \ldots, u_n)$ and $i > 1$. Then $D_i t \cdot u = D_1^n \lambda x\, (D_{i-1} s \cdot u) \cdot (u_1, \ldots, u_n)$. We have the following cases to consider, as to the reduction $t\, \beta^1_{k+1}\, t'$:
  - $s\, \beta^1_k\, s'$ and $t' = D_1^n \lambda x\, s' \cdot (u_1, \ldots, u_n)$; then $D_i t' \cdot u = D_1^n \lambda x\, (D_{i-1} s' \cdot u) \cdot (u_1, \ldots, u_n)$. By inductive hypothesis we have $D_{i-1} s \cdot u\, \beta^1\, D_{i-1} s' \cdot u$ and we are done.
  - $u_1\, \beta^1_k\, u_1'$ and $t' = D_1^n \lambda x\, s \cdot (u_1', u_2, \ldots, u_n)$; then $D_i t' \cdot u = D_1^n \lambda x\, (D_{i-1} s \cdot u) \cdot (u_1', u_2, \ldots, u_n)$. We conclude by definition of β^1 and the fact that as s is simple so is $D_{i-1} s \cdot u$.
  - $t' = D_1^{n-1} \lambda x((\partial s/\partial x) \cdot u_1) \cdot (u_2, \ldots, u_n)$. We thus have $D_i t' \cdot u = D_1^{n-1} \lambda x(D_{i-1}((\partial s/\partial x) \cdot u_1) \cdot u) \cdot (u_2, \ldots, un) = D_1^{n-1} \lambda x(\partial/\partial x(D_{i-1} s \cdot u) \cdot u_1) \cdot (u_2, \ldots, u_n)$ since x does not appear free in u.

Reduction. Let β be the reflexive and transitive closure of the relation $\tilde{\beta}^1$, and let us write $t\, \beta^+\, t'$ when t reduces to t' in at least one step (that is, $t\, \tilde{\beta}^1\, u\, \beta\, t'$ for some $u \in R\langle \Delta \rangle$).

Proposition 11. *The relation* β *is contextual.*

Proof. Linearity is clear and the other conditions result from linearity and from the preceding proposition. □

3.1. The Church–Rosser property

We prove confluence using the Tait–Martin–Löf technique, and following the presentation of [15]. We first define the *parallel reduction* relation ρ from simple terms to terms as the union of an increasing sequence (ρ_k) of relations: ρ_0 is the identity relation and ρ_{k+1} is given by

1. $(s)u\,\rho_{k+1}\,t'$ in one of the following situations:
   (a) $t' = (s')u'$ where $s\,\rho_k\,s'$ and $u\,\bar\rho_k\,u'$;
   (b) $t' = ((\partial^n v'/\partial x^n)\cdot(w'_1,\dots,w'_n))\,[u'/x]$ where $s = D_1^n\lambda x\,v\cdot(w_1,\dots,w_n)$, $v\,\rho_k\,v'$, $u\,\bar\rho_k\,u'$, and $w_j\,\rho_k\,w'_j$ for each j;
2. $D_{i_1,\dots,i_n}x\cdot(u_1,\dots,u_n)\,\rho_{k+1}\,t'$ if $t' = D_{i_1,\dots,i_n}x\cdot(u'_1,\dots,u'_n)$ where $u_j\,\rho_k\,u'_j$ for each j;
3. $D_1^n\lambda x\,s\cdot(u_1,\dots,u_n)\,\rho_{k+1}\,t'$ if $t' = D_1^{n-p}\lambda x((\partial^p s'/\partial x^p)\cdot u'_I)\cdot u'_J$ where I is a subset of $\{1,\dots,n\}$ of p elements, J is its complementary set, u'_K denotes the sequence $(u'_k)_{k\in K}$ for any $K\subseteq\{1,\dots,n\}$ (with the obvious order relation), $s\,\rho_k\,s'$ and $u_j\,\rho_k\,u'_j$ for each j.

Lemma 12. *The relation $\bar\rho$ is contextual.*

The proof is essentially the same as the proof of weak contextuality of β^1.

Lemma 13. $\beta^1\subseteq\rho\subseteq\beta$. *Thus the reflexive and transitive closure of $\bar\rho$ is β.*

Proof. For the first inclusion one proves by a straightforward induction on k that $\beta^1_k\subseteq\rho_k$ (using the obvious fact that ρ_k is reflexive).

The second is obtained again by induction using the fact that β is contextual.

As $\rho\subseteq\beta$ which is linear we also have $\bar\rho\subseteq\beta$ from which, together with the first inclusion, we derive the reflexive and transitive closure property. □

Lemma 14. *Let x be a variable and t, u, t', u' be terms. If $t\,\bar\rho\,t'$ and $u\,\bar\rho\,u'$, then*

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

Proof. We prove by induction on k that if $t\,\bar\rho_k\,t'$ and $u\,\bar\rho\,u'$, then $t\,[u/x]\,\bar\rho\,t'\,[u'/x]$. For $k=0$ we have $t=t'$ and we conclude by contextuality of $\bar\rho$ (applying Lemma 9).

Assume now that the property holds for k. By linearity of $\bar\rho$ and of substitution (the operation $(t,u)\mapsto t\,[u/x]$ is linear in t) we can reduce to the case where t is simple such that $t\,\rho_{k+1}\,t'$ and $u\,\bar\rho\,u'$.

Assume first that $t=(s)w$. Then $t\,[u/x] = (s\,[u/x])w\,[u/x]$. If $t'=(s')w'$ with $s\,\rho_k\,s'$ and $w\,\bar\rho_k\,w'$ we conclude directly by inductive hypothesis and contextuality of $\bar\rho$.

If $s = D_1^n\lambda y\,v\cdot(u_1,\dots,u_n)$, $v\,\rho_k\,v'$, $w\,\bar\rho_k\,w'$, $u_j\,\rho_k\,u'_j$ for each j, and

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

then we have $s\,[u/x] = D_1^n\lambda y\,(v\,[u/x])\cdot(u_1\,[u/x],\dots,u_n\,[u/x])$. Note that this term is not necessarily simple, because $v\,[u/x]$ is not simple in general. However since $v\,[u/x]\,\bar\rho$

$v'[u'/x]$ by inductive hypothesis, $v[u/x]$ is a linear combination of simple terms v_l and $v'[u'/x]$ is a linear combination with the same coefficients of terms v'_l such that $v_l\,\rho\,v'_l$ for each l. Thus by linearity of $\bar\rho$, of derivatives and of substitution, and by definition of ρ we have

$$\begin{aligned}
&t\,[u/x]\,\bar\rho\left(\frac{\partial^n v'[u'/x]}{\partial y^n}\cdot(u'_1\,[u'/x],\ldots,u'_n\,[u'/x])\right)[w'[u'/x]/y]\\
&=\left(\frac{\partial^n v'}{\partial y^n}\cdot(u'_1,\ldots,u'_n)\right)[u'/x]\,[w'\,[u'/x]/y]\quad\text{by Lemma 6}\\
&=\left(\frac{\partial^n v'}{\partial y^n}\cdot(u'_1,\ldots,u'_n)\right)[w'/y]\,[u'/x]\quad\text{by Lemma 2}\\
&=t'\,[u'/x]
\end{aligned}$$

since we may suppose that y is not free in u'.

The case $t=D_{i_1,\ldots,i_n}y\cdot(u_1,\ldots,u_n)$ is handled like in the proof of the next lemma.

Assume last that $t=D_1^n\lambda y\,s\cdot(u_1,\ldots,u_n)$ and that $t'=D_1^{n-p}\lambda y((\partial^p s'/\partial y^p)\cdot u'_I)\cdot u'_J$ where I is a subset of $\{1,\ldots,n\}$ of p elements, J is its complementary set, $s\,\rho_k\,s'$ and $u_j\,\rho_k\,u'_j$ for each j. Then $t\,[u/x]=D_1^n\lambda y\,(s\,[u/x])\cdot(u_1\,[u/x],\ldots,u_n\,[u/x])$. Therefore, by inductive hypothesis and by definition of ρ, we have $t\,[u/x]\,\bar\rho\,D_1^{n-p}\lambda y((\partial^p s''/\partial y^p)\cdot u''_I)\cdot u''_J$ where $s''=s'\,[u'/x]$ and $u''_l=u'_l\,[u'/x]$ for each l. But by Lemma 6 we have $(\partial^p s''/\partial y^p)\cdot u''_I=(\partial^p s'/\partial y^p\cdot u'_I)[u'/x]$ since we can assume that $y\neq x$ and that y does not occur free in u', and we are done.

Lemma 15. *Let x be a variable and let t, u, t', and u' be terms. If $t\,\bar\rho\,t'$ and $u\,\bar\rho\,u'$, then*

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

Proof. We prove by induction on k that if $t\,\bar\rho_k\,t'$ and if $u\,\bar\rho\,u'$, then $(\partial t/\partial x)\cdot u\,\bar\rho\,(\partial t'/\partial x)\cdot u'$. For $k=0$, since ρ_0 is the identity we have to show that $(\partial t/\partial x)\cdot u\,\bar\rho\,(\partial t/\partial x)\cdot u'$ which is consequence of Lemma 9 because $\bar\rho$ is contextual.

Assume now that the property holds for k. By linearity of $\bar\rho$ and of the partial derivative we can reduce to the case where t and u are simple such that $t\,\rho_{k+1}\,t'$ and $u\,\rho\,u'$.

Assume first that $t=(s)w$. Then

$$\frac{\partial t}{\partial x}\cdot u=\left(\frac{\partial s}{\partial x}\cdot u\right)w+\left(D_1 s\cdot\left(\frac{\partial w}{\partial x}\cdot u\right)\right)w.$$

If $t'=(s')w'$ with $s\,\rho_k\,s'$ and $w\,\bar\rho_k\,w'$, then by inductive hypothesis we have $(\partial s/\partial x)\cdot u\,\bar\rho\,(\partial s'/\partial x)\cdot u'$ and $(\partial w/\partial x)\cdot u\,\bar\rho\,(\partial w'/\partial x)\cdot u'$ and we conclude by contextuality of $\bar\rho$.

If $s = D_1^n \lambda y\, v \cdot (u_1, \ldots, u_n)$, $v\, \rho_k\, v'$, $w\, \bar{\rho}_k\, w'$, $u_j\, \rho_k\, u_j'$ for each j, and

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

then we have

$$\frac{\partial s}{\partial x} \cdot u = D_1^n \lambda y \left(\frac{\partial v}{\partial x} \cdot u\right) \cdot (u_1, \ldots, u_n) + \sum_{j=1}^{n} D_1^n \lambda y\, v \cdot \left(u_1, \ldots, \frac{\partial u_j}{\partial x} \cdot u, \ldots, u_n\right)$$

and by inductive hypothesis we have $(\partial v/\partial x) \cdot u\, \bar{\rho}\, (\partial v'/\partial x) \cdot u'$ and $(\partial u_j/\partial x) \cdot u\, \bar{\rho}\, (\partial u_j'/\partial x) \cdot u'$ for each j. The property $(\partial v/\partial x) \cdot u\, \bar{\rho}\, (\partial v'/\partial x) \cdot u'$ means that we may write $(\partial v/\partial x) \cdot u$ and $(\partial v/\partial x) \cdot u$ as linear combinations $\sum a_l v_l$ and $\sum_l a_l v_l'$ where the terms v_l are simple, in such a way that $v_l\, \rho\, v_l'$ for each l. This together with the linearity of substitution operators and the definition of ρ entails that $(D_1^n \lambda y((\partial v/\partial x) \cdot u) \cdot (u_1, \ldots, u_n))w\, \bar{\rho}\, ((\partial^n/\partial y^n)((\partial v'/\partial x) \cdot u') \cdot (u_1', \ldots, u_n'))[w'/y]$. We proceed similarly for the other terms of the sum and apply the inductive hypothesis to get eventually:

$$\left(\frac{\partial s}{\partial x} \cdot u\right) w\, \bar{\rho} \left(\frac{\partial^n}{\partial y^n}\left(\frac{\partial v'}{\partial x} \cdot u'\right) \cdot (u_1', \ldots, u_n')\right)[w'/y]$$
$$+ \sum_{j=1}^{n} \left(\frac{\partial^n v'}{\partial y^n} \cdot \left(u_1', \ldots, \frac{\partial u_j'}{\partial x} \cdot u', \ldots, u_n'\right)\right)[w'/y].$$

Similarly, we get

$$\left(D_1 s \cdot \left(\frac{\partial w}{\partial x} \cdot u\right)\right) w = \left(D_1^{n+1} \lambda y\, v \cdot \left(\frac{\partial w}{\partial x} \cdot u, u_1, \ldots, u_n\right)\right) w$$
$$\rho\, \frac{\partial^{n+1} v'}{\partial y^{n+1}} \cdot \left(\frac{\partial w'}{\partial x} \cdot u', u_1', \ldots, u_n'\right)[w'/y].$$

On the other hand, by iterating Lemmas 4 and 5 (we can assume that y is not free in u'), we get

$$\begin{aligned}
\frac{\partial t'}{\partial x} \cdot u' &= \frac{\partial}{\partial x}\left(\left(\frac{\partial^n v'}{\partial y^n} \cdot (u_1', \ldots, u_n')\right)[w'/y]\right) \cdot u' \\
&= \left(\frac{\partial}{\partial x}\left(\frac{\partial^n v'}{\partial y^n} \cdot (u_1', \ldots, u_n')\right) \cdot u' \right. \\
&\quad \left. + \frac{\partial}{\partial y}\left(\frac{\partial^n v'}{\partial y^n} \cdot (u_1', \ldots, u_n')\right) \cdot \left(\frac{\partial w'}{\partial x} \cdot u'\right)\right)[w'/y] \\
&= \left(\frac{\partial^n}{\partial y^n}\left(\frac{\partial v'}{\partial x} \cdot u'\right) \cdot (u_1', \ldots, u_n')\right)[w'/y]
\end{aligned}$$

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

$$+\frac{\partial^{n+1}v'}{\partial y^{n+1}}\cdot\left(\frac{\partial w'}{\partial x}\cdot u',u'_1,\ldots,u'_n\right)[w'/y]$$

and we are done, in this particular case.

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

$$\frac{\partial t}{\partial x}\cdot u = \delta_{x,y} D_{i_1,\ldots,i_n} u\cdot(u_1,\ldots,u_n) + \sum_{j=1}^{n} D_{i_1,\ldots,i_n} y\cdot\left(u_1,\ldots,\frac{\partial u_j}{\partial x}\cdot u,\ldots,u_n\right).$$

Moreover, we know that $t' = D_{i_1,\ldots,i_n} y\cdot(u'_1,\ldots,u'_n)$ with $u_j\,\rho_k\,u'_j$ for each j. For each j, we have $(\partial u_j/\partial x)\cdot u\,\bar\rho\,(\partial u'_j/\partial x)\cdot u'$ by inductive hypothesis and we conclude by contextuality of $\bar\rho$.

Assume last that $t = D_1^n\lambda y\,s\cdot(u_1,\ldots,u_n)$ and that $t' = D_1^{n-p}\lambda y((\partial^p s'/\partial y^p)\cdot u'_I)\cdot u'_J$ where I is a subset of $\{1,\ldots,n\}$ of p elements, J is its complementary set, $s\,\rho_k\,s'$ and $u_j\,\rho_k\,u'_j$ for each j. Then, denoting by $[1,n]$ the set $\{1,\ldots,n\}$,

$$\frac{\partial t}{\partial x}\cdot u = D_1^n\lambda y\left(\frac{\partial s}{\partial x}\cdot u\right)\cdot(u_1,\ldots,u_n) + \sum_{l=1}^{n} D_1^n\lambda y\,s\cdot\left(\frac{\partial u_l}{\partial x}\cdot u, u_{[1,n]\setminus\{l\}}\right) \tag{4}$$

and

$$\frac{\partial t'}{\partial x}\cdot u' = D_1^{n-p}\lambda y\left(\frac{\partial}{\partial x}\left(\frac{\partial^p s'}{\partial y^p}\cdot u'_I\right)\cdot u'\right)\cdot u'_J$$

$$+\sum_{j\in J} D_1^{n-p}\lambda y\left(\frac{\partial^p s'}{\partial y^p}\cdot u'_I\right)\cdot\left(\frac{\partial u'_j}{\partial x}\cdot u', u'_{J\setminus\{j\}}\right)$$

$$= D_1^{n-p}\lambda y\left(\frac{\partial^p}{\partial y^p}\left(\frac{\partial s'}{\partial x}\cdot u'\right)\cdot u'_I\right)\cdot u'_J$$

$$+\sum_{i\in I} D_1^{n-p}\lambda y\left(\frac{\partial^p s'}{\partial y^p}\cdot\left(\frac{\partial u'_i}{\partial x}\cdot u', u'_{I\setminus\{i\}}\right)\right)\cdot u'_J$$

$$+\sum_{j\in J} D_1^{n-p}\lambda y\left(\frac{\partial^p s'}{\partial y^p}\cdot u'_I\right)\cdot\left(\frac{\partial u'_j}{\partial x}\cdot u', u'_{J\setminus\{j\}}\right),$$

using Lemma 4 which is possible since y can be assumed not to occur free in u'. By inductive hypothesis we have $(\partial s/\partial x)\cdot u\,\bar\rho\,(\partial s'/\partial x)\cdot u'$ and $(\partial u_l/\partial x)\cdot u\,\bar\rho\,(\partial u'_l/\partial x)\cdot u'$ for each l. Therefore (coming back to expression (4) of $(\partial t/\partial x)\cdot u$), we have $D_1^n\lambda y(\partial s/\partial x\cdot u)\cdot(u_1,\ldots,u_n)\,\bar\rho\,D_1^{n-p}\lambda y((\partial^p/\partial y^p)((\partial s'/\partial x)\cdot u')\cdot u'_I)\cdot u'_J$, and for each $l\in\{1,\ldots,n\}$:

- if $l\in I$, we have $D_1^n\lambda y\,s\cdot((\partial u_l/\partial x)\cdot u, u_{[1,n]\setminus\{l\}})\,\bar\rho\,D_1^{n-p}\lambda y((\partial^p s'/\partial y^p)\cdot((\partial u'_l/\partial x)\cdot u', u'_{I\setminus\{l\}}))\cdot u'_J$,

- and if $l \in J$, we have $D_1^n \lambda y\, s \cdot ((\partial u_l/\partial x) \cdot u, u_{[1,n]\setminus\{l\}})\, \bar{\rho}\, D_1^{n-p} \lambda y((\partial^p s'/\partial y^p) \cdot u'_I) \cdot ((\partial u'_l/\partial x) \cdot u', u'_{J\setminus\{l\}})$.

This concludes the proof. □

Multi-confluent pairs of relations. Let us say that a pair of binary relations (τ, φ) from terms to terms is *multi-confluent* if for any term t, any $m \in \mathbb{N}^+$ and any terms $t_1, \ldots, t_m$, if $t\, \tau\, t_i$ for each i, then there exists a term t' such that $t_i\, \varphi\, t'$ for each i.

Lemma 16. *Let τ be a relation from simple terms to terms and let φ be a linear relation from terms to terms. If the pair (τ, φ) is multi-confluent, then the pair $(\bar{\tau}, \varphi)$ is also multi-confluent.*

Proof. Let $t, t^1, \ldots, t^m$ be terms such that $t\, \bar{\tau}\, t^i$ for each i. Let us write as usual $t = \sum_{s\in\Delta} a_s s$. From $t\, \bar{\tau}\, t^i$ we deduce that for each simple term s and for each $i = 1, \ldots, m$, there is a finite set $I_i(s)$, some scalars $(a^i_{s,j})_{j\in I_i(s)}$ such that $a_s = \sum_{j\in I_i(s)} a^i_{s,j}$ and some terms $(U^i_j(s))_{j\in I_i(s)}$ such that $s\, \tau\, U^i_j(s)$ and $t^i = \sum_{s\in\Delta} \sum_{j\in I_i(s)} a^i_{s,j} U^i_j(s)$. We have $s\, \tau\, U^i_j(s)$ for each i and each $j \in I_i(s)$. But for each simple term s the set $\{U^i_j(s) \mid i = 1, \ldots, m$ and $j \in I_i(s)\}$ is finite.

If this set is empty, then each set $I_i(s)$ is empty and therefore $t = 0$ and $t_i = 0$ for each i. Therefore $t^i\, \varphi\, 0$ for each i since φ is linear.

If this set is nonempty, by multi-confluence of (τ, φ) there is a term $V(s)$, *depending only on* s, such that $U^i_j(s)\, \varphi\, V(s)$ for each i and each $j \in I_i(s)$. By linearity of φ we conclude that $t^i\, \varphi\, \sum_{s\in\Delta} \sum_{j\in I_i(s)} a^i_{s,j} V(s) = \sum_{s\in\Delta} a_s V(s)$.

Proposition 17. *The relation $\bar{\rho}$ is confluent.*

Proof. We prove by induction on k that the pair $(\bar{\rho}_k, \bar{\rho})$ is multi-confluent and this will clearly entail the confluence of $\bar{\rho}$. The base case $k = 0$ is trivial since $\bar{\rho}_0$ is just the identity relation. So let us assume that $(\bar{\rho}_k, \bar{\rho})$ is multi-confluent and let us prove that $(\bar{\rho}_{k+1}, \bar{\rho})$ is multi-confluent. For this purpose, by Lemma 16, it suffices to show that the pair $(\rho_{k+1}, \bar{\rho})$ is multi-confluent, what we do now. Let t be a simple term, and let $t^1, \ldots, t^m$ (with $m \geqslant 1$) be terms such that $t\, \rho_{k+1}\, t^i$ for each i.

Assume first that $t = (s)w$.

If for each i we have $t^i = (s^i)w^i$ with $s\, \rho_k\, s^i$ and $w\, \bar{\rho}_k\, w^i$, then the inductive hypothesis applies (thanks to Lemma 16 for the argument side of the application).

Otherwise, we have $s = D_1^n \lambda y\, v \cdot (u_1, \ldots, u_n)$, $v\, \rho_k\, v^i$, $w\, \bar{\rho}_k\, w^i$, $u_j\, \rho_k\, u^i_j$ for each i and each j, and for some $q \in \{1, \ldots, m+1\}$:

- if $1 \leqslant i < q$, $t^i = ((\partial^n v^i/\partial y^n) \cdot (u^i_1, \ldots, u^i_n))[w^i/y]$
- and if $q \leqslant i \leqslant m$, for some set $I_i \subseteq \{1, \ldots, n\}$ whose cardinality is p_i and whose complementary set is J_i, we have $t^i = (s^i)w^i$ where $s^i = D_1^{n-p_i} \lambda y((\partial^{p_i} v^i/\partial y^{p_i}) \cdot u^i_{I_i}) \cdot u^i_{J_i}$.

Observe that the first case must occur at least once, otherwise we are in the first situation for the application. By inductive hypothesis (invoking Lemma 16 for w), we can find terms v', w' and u'_j for each j such that for each i, $v^i\, \bar{\rho}\, v'$, $w^i\, \bar{\rho}\, w'$ and $u^i_j\, \bar{\rho}\, u'_j$ for $j = 1, \ldots, n$. For i such that $1 \leqslant i < q$, we apply Lemmas 14 and 15, and we get

$t^i\,\bar\rho\,((\partial^n v'/\partial y^n)\cdot(u'_1,\ldots,u'_n))[w'/y]$. For i such that $q\leqslant i\leqslant m$, we apply Lemma 15 as well as the definition of ρ and obtain $t^i\,\bar\rho\,((\partial^n v'/\partial y^n)\cdot(u'_1,\ldots,u'_n))[w'/y]$.

The case $t=D_{i_1,\ldots,i_n}x\cdot(u_1,\ldots,u_n)$ is straightforward.

The last case is when $t=D_1^n\lambda x\,v\cdot(u_1,\ldots,u_n)$ and for some set $I_i\subseteq\{1,\ldots,n\}$ whose cardinality is p_i and whose complementary set is J_i, we have $t^i=D_1^{n-p_i}\lambda x\,((\partial^{p_i}v^i/\partial y^{p_i})\cdot u^i_{I_i})\cdot u^i_{J_i}$ (with $v\,\rho_k\,v^i$ and $u_j\,\rho_k\,u^i_j$), for each $i=1,\ldots,m$. Again, the inductive hypothesis provides us with terms v' and u'_j such that for each i, $v^i\,\bar\rho\,v'$ and $u^i_j\,\bar\rho\,u'_j$ for each j. Then applying for each i the definition of ρ (last case, with $I=J_i$ and $J=\emptyset$), we get $t^i\,\bar\rho\,\lambda x(\partial^n v'/\partial x^n)\cdot(u'_1,\ldots,u'_n)$ and we are done. □

Since the reflexive and transitive closure of $\bar\rho$ is β, we finally get the main result of this section.

Theorem 18. *The relation β over terms of the pure differential lambda-calculus enjoys the Church–Rosser property.*

Remember that any ordinary lambda-term is a differential lambda-term. The Church–Rosser result above enforces the observation that the differential lambda-calculus is a conservative extension of the ordinary lambda-calculus since it easily entails the following result.

Proposition 19. *If two ordinary lambda-terms are β-equivalent in the differential lambda-calculus, then they are β-equivalent in the ordinary lambda-calculus.*

Remark. We can easily derive from Lemmas 15 and 14 and from the inclusions $\beta^1\subseteq\rho\subseteq\beta$ (a direct proof would be possible as well) the two following lemmas, which will be useful in the sequel.

Lemma 20. *Let x be a variable and let t, u, t', and u' be terms. If $t\,\beta\,t'$ and $u\,\beta\,u'$, then*

$$\frac{\partial t}{\partial x}\cdot u\,\beta\,\frac{\partial t'}{\partial x}\cdot u'.$$

Lemma 21. *Let x be a variable and let t, u, t', and u' be terms. If $t\,\beta\,t'$ and $u\,\beta\,u'$, then*

$$t\,[u/x]\,\beta\,t'\,[u'/x].$$

4. Simply typed terms

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

$$\frac{}{\Gamma, x : A \vdash x : A} \quad (\mathit{Variable})$$

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

$$\frac{\Gamma, x : A \vdash s : B}{\Gamma \vdash \lambda x\, s : A \to B} \quad (\mathit{Abstraction})$$

$$\frac{\Gamma \vdash s : A_1, \dots, A_i \to B \qquad \Gamma \vdash u : A_i}{\Gamma \vdash D_i s \cdot u : A_1, \dots, A_i \to B} \quad (\mathit{Differential\ application})$$

$$\frac{}{\Gamma \vdash 0 : A} \quad (\mathit{Zero}) \qquad \frac{\Gamma \vdash s : A \qquad \Gamma \vdash t : A}{\Gamma \vdash as + bt : A} \quad (\mathit{Linear\ combination}),$$

where a and b are scalars.

The two last rules express 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 $(D_1 t \cdot u)s$ denotes. So $D_1 t \cdot u$ 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 ith parameter.

Let us say that a semi-ring is *positive* if $a + b = 0 \Rightarrow a = b = 0$ for all $a, b \in R$.

Lemma 22. *Under the assumption that R is positive, subject reduction holds, that is: if t and t' are canonical terms, if $\Gamma \vdash t : A$ and $t\, \beta\, t'$, 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 23. *If s and u are terms, if $\Gamma,\ x : A \vdash s : B$ and $\Gamma \vdash u : A$, then $\Gamma,\ x : A \vdash (\partial s / \partial x) \cdot u : B$.*

The proof is an easy induction on s.

The reason for the restriction on R is that we need the following property to hold: whenever we write a term t as a linear combination $\sum_{i=1}^{n} a_i s_i$ of simple terms with nonzero coefficients, each of the simple terms s_i belongs to $\mathrm{Supp}(t)$. Then we can show that if $\Gamma \vdash t : A$ is derivable, so are all the judgments $\Gamma \vdash s_i : A$.

For showing that the condition is necessary, assume that $a, b \in R \setminus \{0\}$ with $a + b = 0$. For any type A we have $\vdash 0 : A$ but $0 = a(I)t + b(I)t$ (where $I = \lambda x\, x$ and t is any nontypeable closed term). So $0\, \tilde{\beta}^1\, at + b(I)t$ and this latter term is not typeable since t is not typeable.

5. Strong normalization

We prove strong normalization for the simply typed differential lambda-calculus, using the Tait reducibility method, presented along the lines followed by Krivine in [15]. In this section, we assume that $R=\mathbb{N}$. The essential property of $\mathbb{N}$ that we shall use is that there are only finitely many ways of writing a natural number as the sum of two natural numbers.

Consider for instance the differential lambda-calculus with nonnegative rational coefficients ($R=\mathbb{Q}^+$), and let $I=\lambda x\, x$. Then we have the following infinite sequence of reductions, which shows that our restriction on coefficients is essential.

$$\begin{aligned}(I)I = \frac{1}{2}(I)I + \frac{1}{2}(I)I \,\tilde{\beta}^1\, \frac{1}{2}I + \frac{1}{2}(I)I &= \frac{1}{2}I + \frac{1}{4}(I)I + \frac{1}{4}(I)I\\ &\tilde{\beta}^1\, \frac{3}{4}I + \frac{1}{4}(I)I = \frac{3}{4}I + \frac{1}{8}(I)I + \frac{1}{8}(I)I\\ &\tilde{\beta}^1\, \frac{7}{8}I + \frac{1}{8}(I)I\\ &\vdots\end{aligned} \tag{5}$$

This problem has of course nothing to do with the differential part of the calculus and would already appear in ordinary lambda-calculus extended with linear combinations.

The module of strongly normalizing terms. Observe first that if $t\in\mathbb{N}\langle\Delta\rangle$ and if we write $t=as+u$ with s simple and $a\neq 0$, then necessarily $s\in\mathrm{Supp}(t)$. This is due to the fact that the scalars are natural numbers (more precisely: $\mathbb{N}$ is a positive semi-ring).

Lemma 24. *Let $t\in\mathbb{N}\langle\Delta\rangle$. There are only finitely many terms t' such that $t\,\tilde{\beta}^1\,t'$.*

Proof. By induction on the height of t (the least k such that $t\in\mathbb{N}\langle\Delta_k\rangle$). For $k=0$, $t=0$ and the property is trivial, so assume that it holds for heights $\leqslant k$. Using the inductive hypothesis, a simple inspection of the definition of β^1 shows that the property holds for $t\in\Delta_{k+1}$; here we use the fact that for t simple, $t\,\tilde{\beta}^1\,t'$ implies $t\,\beta^1\,t'$ thanks to our particular choice of scalars.[5] So let us prove the property for $t\in\mathbb{N}\langle\Delta_{k+1}\rangle$. Then $t=\sum_{s\in\Delta_{k+1}} a_s s$ and reducing t to t' means:

- choosing $s\in\Delta_{k+1}$ such that $a_s\neq 0$ (there are only finitely many such terms s);
- writing $a_s=b+c$ with $b\neq 0$ (there are only finitely many such decompositions of a_s since the scalars are natural numbers);
- choosing s' such that $s\,\beta^1\,s'$ (as we have seen, there are only finitely many such terms s')

and then $t'=bs'+cs+\sum_{u\in\Delta_{k+1}\setminus\{s\}} a_u u$. So there are only finitely many terms t' such that $t\,\tilde{\beta}^1\,t'$.

[5] This property would not hold if R were for instance the semi-ring of nonnegative rational numbers.

Therefore, by König's lemma, when t is a strongly normalizing term, there is a longest sequence of $\tilde{\beta}^1$-reductions of t to its normal form. We denote by $|t|$ the length of such a sequence. With general coefficients (for instance with $R=\mathbb{Z}$ or $R=\mathbb{Q}^+$) such a definition would not be possible.

We denote by $\mathcal{N}$ the set of all strongly normalizing simple terms. Given $t=\sum_{s\in\mathcal{N}} a_s s\in\mathbb{N}\langle\mathcal{N}\rangle$, we set $\|t\|=\sum_{s\in\mathcal{N}} a_s|s|$ and define in this way a linear operation from $\mathbb{N}\langle\mathcal{N}\rangle$ to $\mathbb{N}$.

Observe that if $u\,\tilde{\beta}^1\,u'$, then $u+t\,\tilde{\beta}^1\,u'+t$ for any term t. From this, it results easily that any strongly normalizing differential term belongs to $\mathbb{N}\langle\mathcal{N}\rangle$. We want now to prove the converse.

Lemma 25. *Let $s\in\mathcal{N}$ and let s' be such that $s\,\beta^1\,s'$. Then $\|s'\|<|s|$.*

Proof. Since the scalars are natural numbers, the term s' can be written as a sum of simple terms, $s'=\sum_{i=1}^m u_i$ and since s' is strongly normalizing, so are the terms u_i. We have by definition $\|s'\|=\sum_{i=1}^m |u_i|$. Now for each i we can find a reduction of length $|u_i|$ of u_i to its normal form, and concatenating these reductions, we get a reduction from s' of length $\|s'\|$, whence the result since s reduces to s' in one step.

Lemma 26. *The set of all strongly normalizing terms is $\mathbb{N}\langle\mathcal{N}\rangle$.*

Proof. We only have to prove that if $t=\sum_{s\in\mathcal{N}} a_s s\in\mathbb{N}\langle\mathcal{N}\rangle$, then t is strongly normalizing. We prove this by induction on $\|t\|$. Assume first that $\|t\|=0$. If $a_s\neq 0$, then $|s|=0$, that is s is normal. From this it results that t itself is normal. Indeed, if we can write $t=bs+u$ with s simple and $b\neq 0$, then necessarily $s\in\mathrm{Supp}(t)$ because our scalars are all positive numbers.

Inductive step: we have $\|t\|>0$ and we want to prove that t is strongly normalizing. So assume that $t\,\tilde{\beta}^1\,t'$. This means that we have found $s\in\mathrm{Supp}(t)$, $b\in\mathbb{N}$ such that $0<b\leqslant a_s$ and two terms s' and u such that $t=bs+u$, $t'=bs'+u$ and $s\,\beta^1\,s'$. We have $s\in\mathrm{Supp}(t)\subseteq\mathcal{N}$, hence s' is strongly normalizing, hence $s'\in\mathbb{N}\langle\mathcal{N}\rangle$ and therefore $t'\in\mathbb{N}\langle\mathcal{N}\rangle$ (because $u\in\mathbb{N}\langle\mathcal{N}\rangle$). We have $\|t'\|=b\|s'\|+\|u\|=b\|s'\|+\|t\|-b|s|=\|t\|-b(|s|-\|s'\|)$. Therefore we have $\|t'\|<\|t\|$ by Lemma 25. Now by inductive hypothesis t' is strongly normalizing and therefore t itself is strongly normalizing.

We conclude by observing the following easy fact which again results from our particular choice of scalars.

Lemma 27. *Let $s,t\in\mathbb{N}\langle\Delta\rangle$, $\mathcal{X}\subseteq\Delta$ and let $a\in\mathbb{N}\setminus\{0\}$. If $as+t\in\mathbb{N}\langle\mathcal{X}\rangle$, then $s\in\mathbb{N}\langle\mathcal{X}\rangle$.*

Proof. Since $a\neq 0$ and since all coefficients are positive we have $\mathrm{Supp}(s)\subseteq\mathrm{Supp}(as+t)\subseteq\mathcal{X}$. □

Redexes and contexts. An *$\mathcal{N}$-redex* is a simple term of the shape

$$t=(D_1^n\lambda x\,s\cdot(u_1,\ldots,u_n))v,$$

where $s, u_1, \ldots, u_n \in \mathcal{N}$ and $v \in \mathbb{N}\langle\mathcal{N}\rangle$. We denote by $\mathsf{Red}(t)$ the following set of terms, obtained by reducing this redex:

- if $n=0$, then $\mathsf{Red}(t) = \{s\,[v/x]\}$;
- otherwise, the elements of $\mathsf{Red}(t)$ are all the terms

$$\left(D_1^{n-1}\lambda x\left(\frac{\partial s}{\partial x}\cdot u_i\right)\cdot(u_1,\ldots,u_{i-1},u_{i+1},\ldots,u_n)\right)v$$

for $i=1,\ldots,n$.

An $\mathcal{N}$*-context* is a context C of the shape

$$C = (D_{j_1,\ldots,j_k}[\,]\cdot(w_1,\ldots,w_k))h_1\cdots h_p$$

where $w_1,\ldots,w_k \in \mathcal{N}$ and $h_1,\ldots,h_p \in \mathbb{N}\langle\mathcal{N}\rangle$. As usual, if t is a term, $C[t]$ is the term obtained by filling the hole $[\,]$ of C with t. Observe that $C[t]$ is simple if t is simple. Moreover $C[0]=0$ and $C[as+bt] = aC[s]+bC[t]$ (the hole is in linear position).

We shall say that a set $\mathcal{X}$ of simple terms is *closed under variable renamings* if $t\,[y/x] \in \mathcal{X}$ whenever $t \in \mathcal{X}$ and x, y are variables.

Saturated sets. A set $\mathcal{S}$ of simple terms is *saturated* if it satisfies the two following conditions.

1. For any $\mathcal{N}$-redex t and any $\mathcal{N}$-context C, if $C[t'] \in \mathbb{N}\langle\mathcal{S}\rangle$ for all $t' \in \mathsf{Red}(t)$, then $C[t] \in \mathcal{S}$.
2. $\mathcal{S}$ is closed under variable renamings.

These two saturation properties will be essential in the proof of the interpretation Lemma 34 (case of an abstraction), the most important being of course the first one and the second one being of a purely technical nature. In that proof we shall need a slightly different version of the first property, that we can prove by "iterating" the definition above. This is the object of the next lemma.

Lemma 28. *Let $\mathcal{S}$ be a saturated subset of $\mathcal{N}$ and let $n \in \mathbb{N}$. Let $s, u_1, \ldots, u_n \in \mathcal{N}$ and let $v \in \mathbb{N}\langle\mathcal{N}\rangle$. If*

1. *for all $I \subseteq \{1,\ldots,n\}$ one has (denoting by k the cardinality of I)*

$$\frac{\partial^k s}{\partial x^k}\cdot u_I \in \mathbb{N}\langle\mathcal{S}\rangle$$

2. *and if*

$$\frac{\partial^n s}{\partial x^n}\cdot(u_1,\ldots,u_n)[v/x] \in \mathbb{N}\langle\mathcal{S}\rangle,$$

then $(D_1^n\lambda x\,s\cdot(u_1,\ldots,u_n))v \in \mathcal{S}$.

Proof. By induction on n. The term $t = (D_1^n\lambda x\,s\cdot(u_1,\ldots,u_n))v$ is an $\mathcal{N}$-redex and so, since $\mathcal{S}$ is saturated, it is sufficient to show that $t' \in \mathbb{N}\langle\mathcal{S}\rangle$, for all $t' \in \mathsf{Red}(t)$. If $n=0$, we have $t' = s\,[v/x]$ and our hypothesis (2) gives directly $t' \in \mathbb{N}\langle\mathcal{S}\rangle$. Assume now that $n>0$. Then

$$t' = \left(D_1^{n-1}\lambda x\left(\frac{\partial s}{\partial x}\cdot u_i\right)\cdot u_{[1,n]\setminus\{i\}}\right)v$$

for some $i \in \{1,\dots,n\}$. We can write $(\partial s/\partial x)\cdot u_i$ as a linear combination of simple terms with nonzero coefficients: $(\partial s/\partial x)\cdot u = \sum_{q=1}^{m} a_q s_q$. If we can show that, for each q, the data s_q, $u_{[1,n]\setminus\{i\}}$ and v satisfy conditions (1) and (2), then by inductive hypothesis, it will follow for each q that $(D_1^{n-1}\lambda x\, s_q \cdot u_{[1,n]\setminus\{i\}})v \in \mathscr{S}$, and hence that $t' \in \mathbb{N}\langle\mathscr{S}\rangle$.

So let $q \in \{1,\dots,m\}$ and let us check that s_q, $u_{[1,n]\setminus\{i\}}$ and v satisfy hypothesis (1), checking (2) being completely similar. Let $I \subseteq [1,n]\setminus\{i\}$ and let k be the cardinality of I. We have to show that $(\partial^k s_q/\partial x^k)\cdot u_I \in \mathbb{N}\langle\mathscr{S}\rangle$. But

$$\begin{aligned}\sum_{r=1}^{m} a_r \frac{\partial^k s_r}{\partial x^k} \cdot u_I &= \frac{\partial^k}{\partial x^k}\left(\frac{\partial s}{\partial x}\cdot u_i\right)\cdot u_I \\ &= \frac{\partial^{k+1} s}{\partial x^{k+1}} \cdot u_{I\cup\{i\}},\end{aligned}$$

where one should observe that $k+1$ is the cardinality of $I\cup\{i\}$ since $i \notin I$. By condition (1) satisfied by s, $(u_1,\dots,u_n)$ and v, we obtain $\sum_{r=1}^{m} a_r(\partial^k s_r/\partial x^k)\cdot u_I \in \mathbb{N}\langle\mathscr{S}\rangle$. Applying Lemma 27, we get finally $(\partial^k s_q/\partial x^k)\cdot u_I \in \mathbb{N}\langle\mathscr{S}\rangle$ since $a_q \neq 0$.

Lemma 29. *The set $\mathscr{N}$ is saturated.*

Proof. We prove property (1) of saturation. We use the notations above for an $\mathscr{N}$-redex t and an $\mathscr{N}$-context C. We set $|t|_0 = |s| + |v| + \sum_{i=1}^{n} |u_i|$ and $|C| = \sum_{j=1}^{k} |w_j| + \sum_{r=1}^{p} |h_r|$.

By induction on $|t|_0 + |C|$, we show that if

$$\forall t' \in \mathsf{Red}(t) \quad C[t'] \in \mathbb{N}\langle\mathscr{N}\rangle, \tag{6}$$

then

$$C[t] \in \mathscr{N}, \text{ or equivalently, if } C[t]\,\beta^1\, s', \text{ then } s' \in \mathbb{N}\langle\mathscr{N}\rangle.$$

There are several cases to consider as to the reduction $C[t]\,\beta^1\, s'$. The first case is when the redex fired in this reduction is t itself, and then $s' = C[t']$ where $t' \in \mathsf{Red}(t)$; we conclude applying directly our hypothesis (6).

In the other cases, the redex fired in the reduction $C[t]\,\beta^1\, s'$ is a subterm of C or of t. These cases can be subdivided in two categories:

- the cases where the reduction takes place in a subterm in linear position, that is in one of the terms s, u_i or w_i;
- and the cases where the reduction takes place in a subterm in non-linear position, that is in v or in one of the h_i.

We check only one case of each of these categories, the others being similar.

A non-linear case: Assume that $v\,\tilde{\beta}^1\, v'$ and $s' = C[t']$ with $t' = (D_1^n \lambda x\, s\cdot(u_1,\dots,u_n))v'$. Since t' is simple, it suffices to show that $C[t'] \in \mathscr{N}$. But v' is strongly normalizing since v is and hence t' is an $\mathscr{N}$-redex. Moreover $|t'|_0 < |t|_0$ and hence the inductive hypothesis applies to the pair (C, t'). Let $t'' \in \mathsf{Red}(t')$, it will be sufficient to show that $C[t''] \in \mathbb{N}\langle\mathscr{N}\rangle$.

If $n>1$, we have $t''=(D_1^{n-1}\lambda x((\partial s/\partial x)\cdot u_i)\cdot(u_1,\ldots,u_{i-1},u_{i+1},\ldots,u_n))v'$ for some i. The term $\hat{t}=(D_1^{n-1}\lambda x((\partial s/\partial x)\cdot u_i)\cdot(u_1,\ldots,u_{i-1},u_{i+1},\ldots,u_n))v$ belongs to $\mathsf{Red}(t)$ and hence $C[\hat{t}]\in\mathbb{N}\langle\mathcal{N}\rangle$ by our assumption (6) on (C,t). But $C[\hat{t}]\,\beta\,C[t'']$ by contextuality of β and hence $C[t'']$ is strongly normalizing, that is, by Lemma 26, $C[t'']\in\mathbb{N}\langle\mathcal{N}\rangle$ as required.

If now $n=0$, we have $t''=s\,[v'/x]$. We have $\hat{t}=s\,[v/x]\in\mathsf{Red}(t)$ and hence $C[\hat{t}]\in\mathbb{N}\langle\mathcal{N}\rangle$ by our assumption (6) on (C,t). By Lemma 21, we get $\hat{t}\,\beta\,t''$ and hence $C[\hat{t}]\,\beta\,C[t'']$ by contextuality of β. We conclude as before that $C[t'']\in\mathbb{N}\langle\mathcal{N}\rangle$.

A linear case: Assume that $n\geqslant 1$ and that $u_1\,\beta^1\,u_1'$ and $s'=C[t']$ with $t'=(D_1^n\lambda x\,s\cdot(u_1',u_2,\ldots,u_n))v$. We have $u_1\in\mathcal{N}$ and hence u_1' is strongly normalizing and thus belongs to $\mathbb{N}\langle\mathcal{N}\rangle$, that is, u_1' is a linear combination of strongly normalizing simple terms

$$u_1'=\sum_{q=1}^{m}a_q u_{1,q}'$$

with $a_q\neq 0$ for all q. Then s' is the linear combination $s'=\sum_{q=1}^{m}a_q C[t_q']$ of simple terms where

$$t_q'=(D_1^n\lambda x\,s\,\cdot\,(u_{1,q}',u_2\ldots,u_n))v$$

for $q=1,\ldots,m$. For each q, we show that $C[t_q']\in\mathcal{N}$ and this will show that $s'\in\mathbb{N}\langle\mathcal{N}\rangle$, as required.

We have $|t_q'|_0<|t|_0$ and hence the inductive hypothesis applies to the pair (C,t_q') (observe indeed that t_q' is an $\mathcal{N}$-redex since $u_{1,q}'$ is strongly normalizing). Therefore, it will be sufficient to show that for any $t''\in\mathsf{Red}(t_q')$, one has $C[t'']\in\mathbb{N}\langle\mathcal{N}\rangle$. There are two cases to consider as to the reduction of t_q' to t''.

Assume first that $t''=(D_1^{n-1}\lambda x((\partial s/\partial x)\cdot u_{1,q}')\cdot(u_2,\ldots,u_n))v$. Let $\hat{t}=(D_1^{n-1}\lambda x((\partial s/\partial x)\cdot u_1)\cdot(u_2,\ldots,u_n))v$, we have $\hat{t}\in\mathsf{Red}(t)$ and therefore $C[\hat{t}]\in\mathbb{N}\langle\mathcal{N}\rangle$ by our assumption (6) on (C,t). But $u_1\,\beta\,u_1'=\sum_{r=1}^{m}a_r u_{1,r}'$ and hence, by Lemma 20 and contextuality of β, we get

$$C[\hat{t}]\beta\sum_{r=1}^{m}a_r C[t_r''],$$

where $t_r''=(D_1^n\lambda x((\partial s/\partial x)\cdot u_{1,r}')\cdot(u_2,\ldots,u_n))v$ for $r=1,\ldots,m$ (so that $t_q''=t''$). Since $C[\hat{t}]\in\mathbb{N}\langle\mathcal{N}\rangle$ and $a_q\neq 0$, we deduce that $C[t'']\in\mathbb{N}\langle\mathcal{N}\rangle$ by Lemma 27.

The other case is: $t''=(D_1^n\lambda x\,((\partial s/\partial x)\cdot u_i)\cdot(u_{1,q}',u_2,\ldots,u_{i-1},u_{i+1},\ldots,u_n))v$ for some $i\in\{2,\ldots,n\}$, if $n>1$. It is handled similarly.

Property (2) of saturation is easy: it suffices to show that $t\,[y/x]\,\beta_k^1\,t'\,[y/x]\Rightarrow t\,\beta_k^1\,t'$ and this is done by induction on k.

Reducibility. Remember that if $t,u_1,\ldots,u_n$ are simple terms, then the term $D_{i_1,\ldots,i_n}t\cdot(u_1,\ldots,u_n)$ is always simple.

If $\mathcal{X}$ and $\mathcal{Y}$ are sets of simple terms, one defines $\mathcal{X}\to\mathcal{Y}\subseteq\Delta$ as

$$\mathcal{X}\to\mathcal{Y}=\{t\in\Delta\mid\forall p\in\mathbb{N},\forall s\in\mathbb{N}\langle\mathcal{X}\rangle,\forall u_1,\ldots,u_p\in\mathcal{X},$$

$$(D_1^p t\,\cdot\,(u_1,\ldots,u_p))s\in\mathcal{Y}\}.$$

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 34.

Lemma 30. *Let $\mathscr{X}_1,\dots,\mathscr{X}_i$ and $\mathscr{Y}$ be sets of simple terms. If $t\in\mathscr{X}_i\to\cdots\to\mathscr{X}_1\to\mathscr{Y}$ and $u\in\mathscr{X}_1$, then $D_i t\cdot u\in\mathscr{X}_i\to\cdots\to\mathscr{X}_1\to\mathscr{Y}$.*

Proof. By induction on $i\geqslant 1$. For $i=1$, it is an obvious consequence of the definition of $\mathscr{X}_1\to\mathscr{Y}$. Assume that the property holds for i, and take $t\in\mathscr{X}_{i+1}\to\mathscr{X}_i\to\cdots\to\mathscr{X}_1\to\mathscr{Y}$ and $u\in\mathscr{X}_1$. We must show that $D_{i+1}t\cdot u\in\mathscr{X}_{i+1}\to\mathscr{X}_i\to\cdots\to\mathscr{X}_1\to\mathscr{Y}$. So let $v_1,\dots,v_p\in\mathscr{X}_{i+1}$ and $s\in\mathbb{N}\langle\mathscr{X}_{i+1}\rangle$, we have to show that $(D_1^p(D_{i+1}t\cdot u)\cdot(v_1,\dots,v_p))s$ belongs to $\mathscr{X}_i\to\cdots\to\mathscr{X}_1\to\mathscr{Y}$.

By definition $(D_1^p t\cdot(v_1,\dots,v_p))s$ belongs to $\mathscr{X}_i\to\cdots\to\mathscr{X}_1\to\mathscr{Y}$ and thus, by inductive hypothesis, so does $D_i(D_1^p t\cdot(v_1,\dots,v_p))s\cdot u$. We conclude because this latter term is equal to $(D_1^p(D_{i+1}t\cdot u)\cdot(v_1,\dots,v_p))s$. $\square$

Lemma 31. *If $\mathscr{X}\subseteq\mathscr{X}'\subseteq\Delta$ and $\mathscr{Y}'\subseteq\mathscr{Y}\subseteq\Delta$, then $\mathscr{X}'\to\mathscr{Y}'\subseteq\mathscr{X}\to\mathscr{Y}$.*

The proof is immediate.

Lemma 32. *Let $\mathscr{S}\subseteq\Delta$ be saturated and let $\mathscr{X}\subseteq\mathscr{N}$ be closed under variable renamings. Then $\mathscr{X}\to\mathscr{S}$ is saturated.*

Proof. We prove first property (1) for the saturation of $\mathscr{X}\to\mathscr{S}$. So, with the notations of the definition of an $\mathscr{N}$-redex t and of an $\mathscr{N}$-context C, assume that $C[t']\in\mathbb{N}\langle\mathscr{X}\to\mathscr{S}\rangle$ for all $t'\in\mathsf{Red}(t)$, we have to show that $C[t]\in\mathscr{X}\to\mathscr{S}$. Let $w_{k+1},\dots,w_{k+q}\in\mathscr{X}$ and let $h_{p+1}\in\mathbb{N}\langle\mathscr{X}\rangle$; we must show that $s=(D_1^q(C[t])\cdot(w_{k+1},\dots,w_{k+q}))h_{p+1}\in\mathscr{S}$. But $s=C'[t]$ where C' is the $\mathscr{N}$-context

$$C'=(D_{j_1,\dots,j_{k+q}}[\,]\cdot(w_1,\dots,w_{k+q}))h_1\dots h_{p+1},$$

where we have set $j_l=1+p$ for $l=k+1,\dots,k+q$. The fact that C' is an $\mathscr{N}$-context results from our assumption $\mathscr{X}\subseteq\mathscr{N}$. Since $\mathscr{S}$ is saturated, it suffices therefore to show that $C'[t']\in\mathbb{N}\langle\mathscr{S}\rangle$ for all $t'\in\mathsf{Red}(t)$. But this results from our hypothesis that $C[t']\in\mathbb{N}\langle\mathscr{X}\to\mathscr{S}\rangle$ for all such t' and from the fact that $C'[t']=(D_1^q(C[t'])\cdot(w_{k+1},\dots,w_{k+q}))h_{p+1}$.

Now we prove that $\mathscr{X}\to\mathscr{S}$ is closed under variable renamings. Let $t\in\mathscr{X}\to\mathscr{S}$ and let x and y be variables. Let $w_1,\dots,w_n\in\mathscr{X}$ and $h\in\mathbb{N}\langle\mathscr{X}\rangle$. Let z be a variable which does not occur free in any of the terms $t,w_1,\dots,w_n$ and h. For $i=1,\dots,n$, let $w_i'=w_i\,[z/x]$ and let $h'=h\,[z/x]$. Since $\mathscr{X}$ is closed under variable renamings, we have $(D_1^n t\cdot(w_1',\dots,w_n'))h'\in\mathscr{S}$ and hence, since $\mathscr{S}$ is closed under variable renamings and x does not occur free in any of the terms w_i' and h', we get $(D_1^n t\,[y/x]\cdot(w_1',\dots,w_n'))h'\in\mathscr{S}$. Last using again the fact that $\mathscr{S}$ is closed under variable renamings we get (replacing z by x) $(D_1^n t\,[y/x]\cdot(w_1,\dots,w_n))h\in\mathscr{S}$ as required.

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

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

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

Lemma 33. *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 31. For the last inclusion, take $t\in\mathcal{N}_0\to\mathcal{N}$, and take a variable x. Then $x\in\mathcal{N}_0$ and thus $(t)x\in\mathcal{N}$. This clearly implies that $t\in\mathcal{N}$. □

With a 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 31 and 33, we get for all type A:

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

and A^* is saturated by Lemmas 29 and 32.

Lemma 34 (Interpretation). *Let t be a canonical term whose free variables belong to the list* (*without repetitions*) $x_1,\dots,x_n$, *and assume that the typing judgment*

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

is derivable, *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 $s_1\in\mathbb{N}\langle A^*_1\rangle,\dots,s_n\in\mathbb{N}\langle A^*_n\rangle$. Assume that the variables $x_1,\dots,x_n$ do not occur free in any of the terms $s_1,\dots,s_n$ and $u_1,\dots,u_k$. Then*

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

Proof. By induction on the typing derivation of the judgment $x_1:A_1,\dots,x_n:A_n\vdash t:A$. If the last rule is a zero rule or a linear combination rule, we conclude straightforwardly by linearity of the substitution and derivation operators. We consider the other possible last rules.

Variable. So $t=x_i$ for some i, and $A=A_i$.

If $k=0$ we conclude since we know that $s_i\in\mathbb{N}\langle A^*_i\rangle$.

If $k\geqslant 1$ and $i\notin\{i_1,\dots,i_k\}$ then $((\partial^k t/\partial x_{i_1}\cdots\partial x_{i_k})\cdot(u_1,\dots,u_k))[s_1,\dots,s_n/x_1,\dots,x_n]=0\in\mathbb{N}\langle A^*\rangle$.

If $k=1$ and $i=i_1$ then $((\partial^k t/\partial x_{i_1}\cdots\partial x_{i_k})\cdot(u_1,\dots,u_k))[s_1,\dots,s_n/x_1,\dots,x_n]=u_1$ since none of the variables x_j is free in u_1 and we are done, since we have assumed that $u_1\in A^*_{i_1}$.

If $k\geqslant 2$ and i is equal, say, to i_1 then we have $(\partial^k t/\partial x_{i_1}\cdots\partial x_{i_k})\cdot(u_1,\dots,u_k)=(\partial^{k-1}/\partial x_{i_2}\cdots\partial x_{i_k})(\partial t/\partial x_{i_1}\cdot u_1)\cdot(u_2,\dots,u_k)=(\partial^{k-1}u_1/\partial x_{i_2}\cdots\partial x_{i_k})\cdot(u_2,\dots,u_k)=0\in\mathbb{N}\langle A^*\rangle$ because the variables x_{i_j} do not occur free in u_1.

Application. So 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, the term $((\partial^k t/\partial x_{i_1}\cdots\partial x_{i_k})\cdot(u_1,\dots,u_k))[s_1,\dots,s_n/x_1,\dots,x_n]$ is a sum of terms of the shape $(D_1^q s'\cdot(w'_1,\dots,w'_q))(w[s_1,\dots,s_n/x_1,\dots,x_n])$ 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

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

(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 $s'\in\mathbb{N}\langle B^*\to A^*\rangle$ and that $w'_1,\dots,w'_q\in\mathbb{N}\langle B^*\rangle$, and also that $w[s_1,\dots,s_n/x_1,\dots,x_n]\in\mathbb{N}\langle B^*\rangle$, and therefore

$$(D_1^q s'\cdot(w'_1,\dots,w'_q))(w[s_1,\dots,s_n/x_1,\dots,x_n])\in\mathbb{N}\langle A^*\rangle$$

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

Differential application. So t can be written $t=D_i s\cdot w$ for some $i\geqslant 1$ 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, one can write the term $((\partial^k t/\partial x_{i_1}\cdots\partial x_{i_k})\cdot(u_1,\dots,u_k))[s_1,\dots,s_n/x_1,\dots,x_n]$ as a sum of terms of the shape $D_i s'\cdot w'$ 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

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

(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 $s'\in\mathbb{N}\langle A^*\rangle$ and that $w'\in\mathbb{N}\langle B_i^*\rangle$. We conclude by Lemma 30 that $D_i s'\cdot w'\in\mathbb{N}\langle A^*\rangle$, as required.

Abstraction. So $t=\lambda x\, s$, the typing derivation of t ends with

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

and we have $A=B\to C$. We must show that $\lambda x\, s'\in\mathbb{N}\langle B^*\to C^*\rangle$, where

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

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

$$(D_1^p \lambda x\, s' \cdot (v_1, \ldots, v_p))w \in \mathbb{N}\langle C^* \rangle$$

and for this purpose, since C^* is a saturated subset of $\mathcal{N}$, we are in position of applying Lemma 28 to s', $v_1, \ldots, v_p$ and w. Indeed, all these terms are strongly normalizing: let z be a variable different from $x, x_1, \ldots, x_n$ and not occurring in any of the terms u_j or s_i; by inductive hypothesis, since $z \in \mathcal{N}_0 \subseteq \mathbb{N}\langle B^* \rangle$, we have $s'[z/x] \in \mathbb{N}\langle C^* \rangle \subseteq \mathbb{N}\langle \mathcal{N} \rangle$ and hence $s' \in \mathbb{N}\langle \mathcal{N} \rangle$ since $\mathcal{N}$ is closed under variable renamings, and we also know that $v_1, \ldots, v_p \in \mathcal{N}$ and $w \in \mathbb{N}\langle \mathcal{N} \rangle$ since $B^* \subseteq \mathcal{N}$. So it suffices to show that

- for all $I \subseteq \{1, \ldots, p\}$ one has
$$\frac{\partial^m s'}{\partial x^m} \cdot v_I \in \mathbb{N}\langle C^* \rangle, \tag{7}$$
where m is the cardinality of I
- and
$$\frac{\partial^p s'}{\partial x^p} \cdot (v_1, \ldots, v_p)[w/x] \in \mathbb{N}\langle C^* \rangle. \tag{8}$$

Let us prove (8), the proof of (7) being completely similar. Let $z_1, \ldots, z_n$ be pairwise distinct variables, which are distinct from $x, x_1, \ldots, x_n$ and which do not occur free in any of the terms s, $u_1, \ldots, u_k$, $s_1, \ldots, s_n$, $v_1, \ldots, v_p$ and w. If r is a term, we denote by $\tilde{r}$ the term $r[z_1, \ldots, z_n/x_1, \ldots, x_n]$. Since B^* is closed under variable renamings, we have $\tilde{v}_1, \ldots, \tilde{v}_p \in B^*$ and $\tilde{w} \in \mathbb{N}\langle B^* \rangle$. Moreover, the variables $x, x_1, \ldots, x_n$ do not occur free in any of these terms. Therefore, by inductive hypothesis, we have

$$\left(\frac{\partial^{k+p} s}{\partial x_{i_1} \cdots \partial x_{i_k} \partial x^p} \cdot (u_1, \ldots, u_k, \tilde{v}_1, \ldots, \tilde{v}_p) \right) [s_1, \ldots, s_n, \tilde{w}/x_1, \ldots, x_n, x] \in \mathbb{N}\langle C^* \rangle,$$

but by our hypotheses on variables and by Lemma 5, this term is equal to $(\partial^p s'/\partial x^p) \cdot (\tilde{v}_1, \ldots, \tilde{v}_p)[\tilde{w}/x]$. Since $\mathbb{N}\langle C^* \rangle$ is closed under variable renamings, we get $((\partial^p s'/\partial x^p) \cdot (\tilde{v}_1, \ldots, \tilde{v}_p)[\tilde{w}/x])[x_1, \ldots, x_n/z_1, \ldots, z_n] \in \mathbb{N}\langle C^* \rangle$, and this latter term is equal to $(\partial^p s'/\partial x^p) \cdot (v_1, \ldots, v_p)[w/x]$ since the variables z_i are fresh.

Theorem 35. *The reduction relation β is strongly normalizing on typeable terms in $\Delta(\mathbb{N})$.*

Proof. Take first a closed term t which is typeable of type A. Then by the interpretation lemma we have $t \in \mathbb{N}\langle A^* \rangle$. But $A^* \subseteq \mathcal{N}$, so t is strongly normalizing. For a nonclosed term which is typeable in some typing context, any of its λ-closures is strongly normalizing, and so the term itself is strongly normalizing.

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

Weak normalization. The strong normalization theorem can be generalized to arbitrary commutative semi-rings R satisfying the following properties.

- If $ab=0$ then $a=0$ or $b=0$ (this integral domain property is used in the proof of Lemma 27).
- For all $a\in R$, there are only finitely many $b,c\in R$ such that $a=b+c$.
- If $a+b=0$ then $a=b=0$ (positivity).

In particular, given formal indeterminates $\xi_1,\ldots,\xi_n$, the semi-ring $\mathbb{N}[\xi_1,\ldots,\xi_n]$ of polynomials satisfies these properties, and strong normalization holds for terms whose coefficients belong to this semi-ring.

Now let R be an arbitrary commutative semi-ring and let $t\in\Delta(R)$. This term t contains a certain number of coefficients $a_1,\ldots,a_n\in R$ that we can replace by formal indeterminates $\xi_1,\ldots,\xi_n$, leading to a term $\hat{t}\in\Delta(\mathbb{N}[\xi_1,\ldots,\xi_n])$. This term strongly normalizes to a unique normal form $t_0\in\Delta(\mathbb{N}[\xi_1,\ldots,\xi_n])$. By replacing in t_0 each indeterminate ξ_i by its value $a_i\in R$ (and evaluating in R the coefficients of t_0 which are polynomials over the indeterminates ξ_i), we obtain a term $t_1\in\Delta(R)$. It is easy to check that t reduces to t_1. The term t_0, and a fortiori the term t_1, contains no redexes; indeed, when evaluating the polynomial coefficients in R, some subterms of t_0 can vanish, but certainly, no new non-normal term can appear.

Strictly speaking however, the term t_1 obtained in this way is not a normal form when R is not positive, for the reason we have already mentioned several times that $t_0=t_0+u-u$ for any non-normal term u, when $-1\in R$; and this kind of operation can be applied to any subterm of t_1 which is in argument position (of a standard application, that is, in non-linear position). We can nevertheless say that a restricted form of weak normalization holds, since any term obtained by reducing t_1 will reduce to t_1 (by the Church–Rosser property) and so t_1 can reasonably be considered as a kind of normal form.

When R is positive, this phenomenon cannot occur and t_1 is a "true" normal form. Therefore weak normalization holds in that case, but strong normalization does not: remember the example (5) with $R=\mathbb{Q}^+$.

6. Linear head reduction and the Taylor formula

We prove first a version of Leibniz formula which will be useful in the sequel.

Lemma 36. *Let t and u be terms and let x and y be distinct variables such that y not occur free in u. Then*

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

This an easy consequence of Lemma 5. The assumption $y\neq x$ is of course absolutely essential for the assumptions of Lemma 5 to be fulfilled (look at what happens when $t=x$ or $t=y$). This equation has a clear logical meaning in terms of cut elimination: it expresses how derivation behaves when interacting with a contraction. We generalize now this formula to iterated derivatives, and this leads to the announced Leibniz formula.

Lemma 37. *Let t and u be terms and let x and y be distinct variables such that y does not occur free in u. Then*

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

Proof. Induction on n, using the lemma above and the well known identity $\binom{n+1}{p+1} = \binom{n}{p+1} + \binom{n}{p}$. □

Lemma 38. *Let x be a variable, let $\bar{t} = t_1 \ldots t_k$ be a sequence of terms and let u be a simple term. Let y be 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 $n \geqslant 1$. 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 37, we have, since $(x)\bar{t} = ((y)\bar{t})\,[x/y]$,

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

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

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

and this proves the first statement. The second statement is a consequence of the first one and of an iterated use of Lemma 8 (if the length of the sequence of terms $\bar{t}$ is 1, then the lemma applies directly, otherwise some iteration on the length of this sequence is needed). A more interesting way of proving this statement is as follows.

Let us say that a term s is *linear* in a variable y if we have

$$\frac{\partial s}{\partial y} \cdot z = s\,[z/y]$$

where z is some (or equivalently, any) variable not occurring free in s. Applying Lemmas 4 and 5, one shows that if s is linear in y and y does not occur free in the term u, then $(\partial s/\partial x)\cdot u$ is linear in y, as soon as x is a variable different from y.

We conclude using the fact that if s is linear in y, then $s\,[0/y] = (\partial s/\partial y)\cdot 0 = 0$ and the fact that $(y)\bar{t}$ is linear in y (indeed, y is linear in y and if s is linear in y and y does not occur free in t, then $(s)t$ is linear in y). $\square$

Let $\star$ be a distinguished variable.

Theorem 39. *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 $(D_1^n s \cdot u^n)0 \not\simeq_\beta 0$, and for this value of n, one has*

$$(D_1^n s \cdot u^n)0 \simeq_\beta n!\star.$$

This means that the Taylor formula

$$s(u) = \sum_{n=0}^{\infty} \frac{1}{n!}(D_1^n s \cdot u^n)0$$

holds in a rather trivial way in that particular case. This formula always holds, semantically, at least in the simply typed case (see [10]), 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 write $t\,\tau^k\,t'$ when t head reduces in k steps to t'. Another standard lambda-calculus property 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 $L(s,u)$ 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 $L(s,u) = 0$. In the latter case we define s^+ by

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

Note that s^+, u satisfy the assumptions of the theorem 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 $(s_0^+)u\,\tau^1\,v'$ so that the length of the head reduction to $\star$ of $(s)u$ is strictly greater than the length of the head reduction of $(s_0^+)u$ to $\star$ as soon as $k > 0$.

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

$$L(s,u) = L(s^+,u) + 1.$$

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

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

$$(D_1^n s \cdot u^n)0 \simeq_\beta \left(\frac{\partial^n \star}{\partial x^n}\cdot u^n\right)[0/x] = \begin{cases} 0 & \text{if } n \neq 0 \\ \star & \text{if } n = 0 = L(s,u) \end{cases}$$

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

$$(D_1^n s \cdot u^n)0 \simeq_\beta (s)0 \simeq_\beta (0)\bar{t} = 0.$$

Otherwise if $n\geqslant 1$ we have

$$(D_1^n s \cdot u^n)0 \simeq_\beta \left(\frac{\partial^n (x)\bar{t}}{\partial x^n}\cdot u^n\right)[0/x] \simeq_\beta n\left(\frac{\partial^{n-1} (u)\bar{t}}{\partial x^{n-1}}\cdot u^{n-1}\right)[0/x]$$

by Lemma 38. So we have

$$(D_1^n s \cdot u^n)0 \simeq_\beta n(D_1^{n-1} s^+ \cdot u^{n-1})0.$$

But by inductive hypothesis

$$(D_1^{n-1} s^+ \cdot u^{n-1})0 \simeq_\beta \begin{cases} (n-1)!\star & \text{if } n-1 = L(s^+,u) = L(s,u)-1, \\ 0 & \text{otherwise} \end{cases}$$

and the result is proved. □

The number $L(s,u)$ counts the substitutions of the successive head variables of s in the *linear head reduction* of $(s)u$ ([7]). The head variable of s is the only occurrence of variable in $(s)u$ which may be considered as linear. So $L(s,u)$ 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.

Going now in the opposite direction, we want to conclude this section by computing the Taylor expansion of the outermost application in the well known nonsolvable term $(\delta)\delta$, where $\delta=\lambda x\,(x)x$. This expansion is

$$\sum_{n=0}^{\infty} \frac{1}{n!}(D_1^n \delta \cdot \delta^n)0.$$

Let $t_n=(D_1^n\delta\cdot\delta^n)0$. Up to reduction, we have $t_n=((\partial^n(x)x/\partial x^n)\cdot\delta^n)[0/x]$. By Lemma 38 (with $u=\delta$ and $\bar{t}=x$), we get (again, up to conversion) $t_{n+1}=(n+1)t_n$ for each n.

But clearly $t_0 = 0$ and hence $t_n = 0$ for each n. Hence the Taylor expansion of $(\delta)\delta$ is 0. This reflects of course the fact that this term is unsolvable, and this observation enforces the idea that the Taylor expansion provides an approach to term approximations similar in spirit to the Böhm tree approach, the role of Ω being played by 0.

Acknowledgements

We would like to thank especially one of the referees of this paper, who made many insightful comments on an earlier version of this work and derived us to clarify several delicate syntactical aspects of the differential lambda-calculus.

Appendix A. Short survey of the semantics

In [10], the first author introduced a semantics of linear logic based on Köthe spaces which are locally convex topological vector spaces of a quite particular kind, in some sense similar to coherence spaces. We present here shortly a simplified version of this semantics, based on the notion of *finiteness spaces*, a "discrete" analogue of Köthe spaces. This model will be presented more thoroughly in [11].

Given a set I and a subset $\mathscr{F}$ of $\mathscr{P}(I)$, let us denote by $\mathscr{F}^\perp$ the set of all subsets of I which have a *finite* intersection with *all* the elements of $\mathscr{F}$. A finiteness space is a pair $X = (|X|, \mathsf{F}(X))$, where $|X|$ is a set (the *web* of X) and $\mathsf{F}(X)$ is a subset of $\mathscr{P}(|X|)$ which satisfies $\mathsf{F}(X) = \mathsf{F}(X)^{\perp\perp}$ and whose elements are called the *finitary* subsets of $|X|$. Given a finiteness space X, we define $R\langle X\rangle$ as the subset of all $x \in R^{|X|}$ such that the set $|x| = \{a \in |X| \mid x_a \neq 0\}$ belongs to $\mathsf{F}(X)$. The set $R\langle X\rangle$ has clearly a module structure (all operations being defined pointwise), since the union of two finitary sets is still finitary.[6]

The main purpose of these definitions is that, given $x \in R\langle X\rangle$ and $x' \in R\langle X^\perp\rangle$, it is possible to define $\langle x, x'\rangle \in R$ as $\sum_{a\in|X|} x_a x'_a$ since this sum is finite[7] by definition (of course, $X^\perp$ is defined as $(|X|, \mathsf{F}(X)^\perp)$). In this way, the pair $(R\langle X\rangle, R\langle X^\perp\rangle)$ carries a well-behaved duality (each module can be seen as the topological dual of the other for a suitable *linear* topology in the sense of [16], but this needs not be explained here). A morphism from X to Y (finiteness spaces) is a linear function from $R\langle X\rangle$ to $R\langle Y\rangle$ which is continuous for the topologies mentioned above. But these morphisms admit a more concrete matricial characterization as we shall see.

We can use finiteness spaces for interpreting all the formulae of propositional linear logic and we briefly survey now the corresponding space constructions and their main properties. Let X and Y be finiteness spaces.

[6] The point of the condition $\mathsf{F}(X) = \mathsf{F}(X)^{\perp\perp}$ is that, for checking that $u \subseteq |X|$ is finitary in X, one has only to check that u has a finite intersection with all the elements of $\mathsf{F}(X)^\perp$; therefore if u and v are finitary in X, so is $u \cup v$.

[7] More precisely, it has only finitely many nonzero terms.

- The direct sum $X \oplus Y$ and the direct product $X \,\&\, Y$ of X and Y are the same space whose web is $|X| + |Y|$ (disjoint union) and where a subset of this disjoint union is finitary if each of its restrictions to $|X|$ and $|Y|$ is finitary (in X and Y, respectively). Infinite direct sums and products can be defined as well but do not coincide anymore.
- The tensor product of X and Y is the space $X \otimes Y$ whose web is $|X| \times |Y|$ and where a subset of that web is finitary if its two projections are finitary (in X and Y respectively). It can be checked that the subset $\mathsf{F}(X \otimes Y)$ of $\mathscr{P}(|X| \times |Y|)$ defined in this way satisfies indeed that $\mathsf{F}(X \otimes Y)^{\perp\perp} = \mathsf{F}(X \otimes Y)$.
- The linear function space $X \multimap Y$ is defined as $(X \otimes Y^{\perp})^{\perp}$. An element of $R\langle X \multimap Y\rangle$ should be seen as a matrix indexed over $|X| \times |Y|$, with coefficients in R. Given $A \in R\langle X \multimap Y\rangle$ and $B \in R\langle Y \multimap Z\rangle$, the product of matrices $BA \in R^{|X| \times |Z|}$ is given by $(BA)_{a,c} = \sum_{b \in |Y|} A_{a,b} B_{b,c}$. Due to the fact that $|A| \subseteq \mathsf{F}(X \multimap Y)$ and $|B| \subseteq \mathsf{F}(Y \multimap Z)$, this sum indeed is always finite and it is not hard to check that $BA \in R\langle X \multimap Z\rangle$. The identity matrix $I \in R\langle X \multimap X\rangle$ is defined as usual by $I_{a,b} = \delta_{a,b}$ and is the neutral element for matrix composition. In this way we have defined a category of finiteness spaces, where a morphism from X to Y is an element of $R\langle X \multimap Y\rangle$, called *linear category* in the sequel.

If $A \in R\langle X \multimap Y\rangle$ and $x \in R\langle X\rangle$, we can define $A \cdot x \in R\langle Y\rangle$ by $(A \cdot x)_b = \sum_{a \in |X|} A_{a,b} x_a$, thus allowing one to see any element of $R\langle X \multimap Y\rangle$ as a (linear and continuous[8]) function from $R\langle X\rangle$ to $R\langle Y\rangle$. This map from matrices to linear and continuous functions turns out to be a bijection.

This linear category is a model of multiplicative-additive linear logic, that is, a $\star$-autonomous category with finite sums and products (see [3]). In particular, all the operations we have defined are functorial. Of course, we need more for getting a model of full linear logic, or of simply typed lambda-calculus. We have to define an exponential.

Given a finiteness space X, we define $!X$ as follows: $|!X|$ is the set of all finite multi-sets of elements of $|X|$, and if U is a collection of such multi-sets, we decide that it is finitary (that is, belongs to $\mathsf{F}(!X)$) if the union of the supports[9] of the elements of U is finitary in X. It turns out that this collection $\mathsf{F}(!X)$ of subsets of $|!X|$ satisfies our basic requirement, namely $\mathsf{F}(!X)^{\perp\perp} = \mathsf{F}(!X)$.

This operation on objects can be turned into an endofunctor on the linear category described above by defining its action on matrices: if $A \in R\langle X \multimap Y\rangle$, it is possible to define $!A \in R\langle !X \multimap !Y\rangle$. We do not describe this operation here, we just give its fundamental property (which completely characterizes it). Given $x \in R\langle X\rangle$, we can define $x^! \in R\langle !X\rangle$ by $(x^!)_m = x^m = \prod_{a \in |X|} x_a^{m(a)}$ (since m is finite, this product is well defined; by $m(a)$, we denote the number of occurrences of a in the multi-set m). Then we have $(A \cdot x)^! = !A \cdot x^!$. It turns out that this endofunctor has all the structure required for interpreting the "bang" modality of linear logic (basically: it is a comonad and there is a natural isomorphism between $!(X \,\&\, Y)$ and $!X \otimes !Y$).

[8] With respect to the linear topologies we mentioned above on $R\langle X\rangle$ and $R\langle Y\rangle$.

[9] The support of a multi-set m is the set of all the elements which occur at least once in m.

Given $\varphi \in R\langle !X \multimap Y\rangle$ and $x \in R\langle X\rangle$, we can set

$$\varphi(x) = \varphi \cdot x^! = \sum_{b\in|Y|}\left(\sum_{m\in|!X|}\varphi_{m,b}x^m\right)e_b \in R\langle Y\rangle,$$

where $(e_b)_{b\in|Y|}$ is the "canonical basis" of $R\langle Y\rangle$ defined by $(e_b)_{b'} = \delta_{b,b'}$ so that the elements of $!X \multimap Y$ can be considered as power series from $R\langle X\rangle$ to $R\langle Y\rangle$. In view of all the structure presented so far, it is a standard fact in the semantics of linear logic that the category whose objects are the finiteness spaces and where a morphism from X to Y is an element of $R\langle !X \multimap Y\rangle$, and equipped with a notion of composition defined in terms of the comonad structure of the "!" functor, [10] is a cartesian closed category, that is, a model of simply typed lambda-calculus: the Kleisli category of the comonad "!". See for instance [4].

We finish this short presentation by a word about differentiation. Due to the fact that finite sums and products coincide, we can build a canonical linear morphism M from $!X \otimes !X$ to $!X$ (we apply the ! functor to the co-diagonal of the sum $X \oplus X$ which is an element of $R\langle X \oplus X \multimap X\rangle = R\langle X \,\&\, X \multimap X\rangle$; seen as a linear map from $R\langle X\rangle \times R\langle X\rangle$ to $R\langle X\rangle$, this morphism is just *addition*). By pre-composing a power series $\varphi \in R\langle !X \multimap Y\rangle$ with M, we obtain an element ψ of $R\langle !X \otimes !X \multimap Y\rangle$, that is, a two-parameter power series, which is characterized by $\psi(x, y) = \varphi(x + y)$. We have on the other hand a linear morphism ∂^0 from X to $!X$ which actually embeds [11] X into $!X$; this morphism is the matrix given by $\partial^0_{a,m} = \delta_{[a],m}$. If φ is as above, by pre-composing ∂^0 with φ in the linear category, we get a linear morphism from X to Y which is easily seen to be the derivative of φ (considered as a power series) at 0. Now it should be clear how to use M and ∂^0 to compute the derivative of φ at any point x of $R\langle X\rangle$, and not only at 0: translate φ using M and then use ∂^0 for derivating the obtained power series $y \mapsto \varphi(x + y)$ at 0.

Cartesian closedness of the Kleisli category and these operations M and ∂^0 are the basic ingredients for interpreting the differential lambda-calculus in this category. A type will be interpreted as a finiteness space (implication on types being interpreted by the operation $(X, Y) \mapsto !X \multimap Y$ on spaces), and a closed term of type A will be interpreted by an element of $R\langle X\rangle$ where X is the space interpreting A.

References

[1] T. Altenkirch, N. Ghani, C. McBride, Derivatives of containers, in: M. Hofmann (Ed.), Proc. Sixth Typed Lambda Calculi and Applications Conf, Lecture Notes in Computer Science, Vol. 2701, Springer, Berlin, 2003, pp. 16–30.

[2] H. Barendregt, The lambda calculus, in: Studies in Logic and the Foundations of Mathematics, Vol. 103, North-Holland, Amsterdam, 1984.

[3] M. Barr, ∗-autonomous categories, in: Lecture Notes in Mathematics, Vol. 752, Springer, Berlin, 1979.

[10] This notion of composition coincides of course with the standard composition of the power series associated with morphisms.

[11] In the sense that there is a morphism d from $!X$ to X such that $\mathrm{d} \circ \partial^0 = \mathrm{Id}$; this latter morphism d corresponds to the dereliction rule of linear logic and is part of the comonad structure of the ! functor.

[4] G. Bierman, What is a categorical model of intuitionistic linear logic? in: M. Dezani-Ciancaglini, G.D. Plotkin (Eds.), Proc. Second Typed Lambda-Calculi and Applications Conf, Lecture Notes in Computer Science, Vol. 902, Springer, Berlin, 1995, pp. 73–93.

[5] G. Boudol, The lambda calculus with multiplicities, Tech. Report 2025, INRIA Sophia-Antipolis, 1993.

[6] G. Boudol, P.-L. Curien, C. Lavatelli, A semantics for lambda calculi with resource, Mathematical Structures in Computer Science 9 (4) (1999) 437–482.

[7] V. Danos, H. Herbelin, L. Regnier, Games semantics and abstract machines, in: Proc. Eleventh Symp. on Logic in Computer Science, IEEE Computer Society Press, Silver Spring, MD, 1996.

[8] V. Danos, L. Regnier, Head linear reduction, 2003, submitted for publication.

[9] N.G. De Bruijn, Generalizing automath by means of a lambda-typed lambda calculus, in: D.W. Kueker, E.G.K. Lopez-Escobar, C.H. Smith (Eds.), Mathematical Logic and Theoretical Computer Science, Lecture Notes in Pure and Applied Mathematics, Marcel Dekker, New York, 1987, pp. 71–92. (Reprinted in: Selected papers on Automath, Studies in Logic, vol. 133, pp. 313–337, North-Holland, Amsterdam, 1994).

[10] T. Ehrhard, On Köthe sequence spaces and linear logic, Mathematical Structures in Computer Science 12 (2002) 579–623.

[11] T. Ehrhard, Finiteness spaces, Preliminary version accessible from http://iml.univ-mrs.fr/~ehrhard/, 2003.

[12] J.-Y. Girard, Linear logic, Theoretical Computer Science 50 (1987) 1–102.

[13] J.-Y. Girard, Coherent banach spaces: a continuous denotational semantics, Theoretical Computer Science 227 (1999) 275–297.

[14] A. Kriegl, P.W. Michor, The convenient setting of global analysis, in: Mathematical Surveys and Monographs, vol. 53, American Mathematical Society, Providence, RJ, 1997.

[15] J.-L. Krivine, Lambda-Calculus, Types and Models, Ellis Horwood Series in Computers and Their Applications, Ellis Horwood, Chichester, UK, 1993 (Translation by René Cori from French 1990 edition Masson).

[16] S. Lefschetz, Algebraic topology, in: American mathematical society colloquium publications, Vol. 27, American Mathematical Society, Providence, RI, 1942.

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