Size-based termination of higher-order rewriting

FRÉDÉRIC BLANQUI
INRIA, ENS / Université Paris-Saclay,
LSV, 61 avenue du Président Wilson, 94235 Cachan Cedex, France
(e-mail: frederic.blanqui@inria.fr)

Abstract
We provide a general and modular criterion for the termination of simply typed \(\lambda\)-calculus extended with function symbols defined by user-defined rewrite rules. Following a work of Hughes, Pareto and Sabry for functions defined with a fixpoint operator and pattern matching, several criteria use typing rules for bounding the height of arguments in function calls. In this paper, we extend this approach to rewriting-based function definitions and more general user-defined notions of size.

1 Introduction
In this paper, we are interested in the termination of Church’s simply typed \(\lambda\)-calculus (Church, 1940) extended with function symbols defined by user-defined rewrite rules (Dershowitz & Jouannaud, 1990; TeReSe, 2003) like the ones of Figure 1. Our results could be used to check the termination of typed functional programs (e.g. in OCaml (2017) or Haskell (2017)), rewriting-based programs (e.g. in Maude (2015)) or function definitions in proof assistants (e.g. (Coq, 2017), Agda (2017), Dedukti (2018)). By termination, we mean the strong normalization property, that is, the absence of infinite rewrite sequences \(t_0 \rightarrow t_1 \rightarrow \ldots\). The mere existence of a normal form is a weaker property called weak normalization. Termination is an important property in program verification.

The rewrite system of Figure 1 defines the subtraction and division functions on the sort \(\mathbb{N}\) of natural numbers in unary notation, i.e. with the constructors \(0 : \mathbb{N}\) for zero and \(s : \mathbb{N} \Rightarrow \mathbb{N}\) for the successor function. A way to prove the termination of this system is to show that, in two successive functions calls, arguments are strictly decreasing wrt some well-founded order. A natural order, based on the inductive nature of \(\mathbb{N}\), is to compare the height of terms. More precisely, let the size of a terminating term \(t\) of sort \(\mathbb{N}\) be the number of \(s\) symbols at the top of the normal form of \(t\) (this rewrite system is weakly orthogonal and thus confluent van Oostrom (1994)). While the termination of \(\text{sub}\) (i.e. the absence of infinite reductions starting from a term of the form \(\text{sub} \ t \ u\) with \(t\) and \(u\) in normal form) is not very difficult to establish (the size of the first argument is strictly decreasing in recursive calls), proving the termination of \(\text{div}\) requires the observation that \(\text{sub}\)
\[\begin{align*}
\text{sub } x \ 0 & \rightarrow x \\
\text{sub } 0 \ y & \rightarrow 0 \\
\text{sub } (s \ x) \ (s \ y) & \rightarrow \text{sub } x \ y \\
\text{div } 0 \ (s \ y) & \rightarrow 0 \\
\text{div } (s \ x) \ (s \ y) & \rightarrow s \ (\text{div } (\text{sub } x \ y) \ (s \ y))
\end{align*}\]

Fig. 1. Rewrite system defining subtraction and division on natural numbers.

is not size increasing, that is, the size of \((\text{sub } t \ u)\) is less than or equal to the size of \(t\).

The idea of sized types, introduced by Hughes et al. (1996) for fixpoint-based function definitions, is to consider an abstract interpretation of this notion of size into an algebra of symbolic size expressions, and turn the usual typing rules of simply typed \(\lambda\)-calculus into deduction rules on the size of terms. This allows one to automatically deduce some information on the size of terms, and thus prove termination by checking that, for instance, the size of some given argument decreases in every recursive call. Hence, termination is reduced to checking typing and abstract size decreasingness.

In our example, this amounts to saying the second rule of \(\text{div}\) does not jeopardize termination since, assuming that \(x\) is instantiated by a term \(t\) of abstract size \(\alpha\), and \(y\) is instantiated by a term \(u\) of abstract size \(\beta\), then \(\text{div } (s \ t) \ (s \ u)\) terminates because its first argument is of size \(\alpha + 1\), while in the recursive call \(\text{div } (\text{sub } t \ u) \ (s \ u)\), the first argument has a size smaller than or equal to \(\alpha\).

The goal of this work is to automate this kind of inductive reasoning, and check the information given by the user (here, the fact that \(\text{sub}\) is not size-increasing). However, when considering type constructors taking functions as arguments (e.g. Sellink’s model of \(\mu\)CRL (Sellink, 1993), Howard’s constructive ordinals in Example 5), the size of a term is generally not a finite natural number but a transfinite ordinal number. However, abstract size expressions can also handle transfinite sizes.

Before explaining our contributions and detailing the outline of the paper, we give hereafter a short survey on the use of ordinals for proving termination since this is at the heart of our work though, in the end, we provide an ordinal-free termination criterion.

### 1.1 Ordinal-based termination

A natural (and trivially complete) method for proving the termination of a relation \(\rightarrow\) consists in considering a well-founded domain \((\mathbb{D}, <_{\mathbb{D}})\), e.g. some ordinal \((h, <_{h})\), assigning a “size” \(\|t\| \in \mathbb{D}\) to every term \(t\), and checking that every rewrite step (including \(\beta\)-reduction) makes the “size” strictly decrease: \(\|t\| >_{\mathbb{D}} \|u\|\) whenever \(t \rightarrow u\).

In theory, it is enough to take \(\mathbb{D} = \omega\) (the first infinite ordinal) when the rewrite relation is finitely branching. However, after Gödel’s incompleteness theorem (Gödel, 1931), defining \(\|\cdot\|\) and proving that \(\|t\| >_{\mathbb{D}} \|u\|\) whenever \(t \rightarrow u\), may require the use
of much bigger ordinals. For instance, the termination of cut elimination in Peano arithmetic requires induction up to the ordinal $\varepsilon_0 = \omega^{\omega}$, but Peano arithmetic cannot prove the well-foundedness of $\varepsilon_0$ itself (Gentzen, 1935). Yet, there is a function $\| \|$ from the terms of Gödel’s system T (Gödel, 1958) (which extends Peano arithmetic) to $\omega$ such that $\|t\| >_D \|u\|$ whenever $t \to u$ (Weiermann, 1998).

An equivalent approach is finding a well-founded relation containing $\to$. For instance, Dershowitz’s recursive path ordering (RPO) (Dershowitz, 1979b, 1982) or its extension to the higher order case by Jouannaud and Rubio (Jouannaud & Rubio, 1999, 2007; Blanqui et al., 2015). But, in this paper, we will focus on the explicit use of size functions. For a connection between RPO and ordinals, see, for instance, Dershowitz & Okada (1988).

Early examples of this approach are given by Ackermann’s proof of termination of second-order primitive recursive arithmetic functions using $\eta = \omega^{\omega^{\omega}}$ (Ackermann, 1925), Gentzen’s proof of termination of cut elimination in Peano arithmetic using $\eta = \varepsilon_0$ (Gentzen, 1935; Howard, 1970; Wilken & Weiermann, 2012), Turing’s proof of weak normalization of Church’s simply typed $\lambda$-calculus (Turing, 1942) and Howard’s proof of termination of his system V (an extension of Gödel’s system T with an inductive type for representing ordinals) using Bachmann’s ordinal (Howard, 1972). This approach developed into a whole area of research for measuring the logical strength of axiomatic theories, involving ever growing ordinals, that can hardly be automated. See for instance Rathjen (2006) for some recent survey. Instead, Monin & Simonot (2001) developed an algorithm for trying to find size assignments in $\eta = \omega^{\omega}$.

But, up to now, there has been no ordinal analysis for powerful theories like second-order arithmetic: the termination of cut elimination in such theories is based on another approach introduced by Girard (1972) and Girard et al. (1988), which consists in interpreting types by so-called computability predicates and typing by the membership relation.

In the first-order case, i.e. when there is no rule with abstraction or applied variables, size decreasingness can be slightly relaxed by conducting a finer analysis of the possible sequences of function calls. This led to the notions of dependency pair in the theory of first-order rewrite systems (Arts, 1996; Arts & Giesl, 2000; Hirokawa & Middeldorp, 2005; Giesl et al., 2006), and size-change principle for first-order functional programs (Lee et al., 2001). These two notions are thoroughly compared in Thiemann & Giesl (2005). In both cases, it is sufficient to define a measure on the class of terms which are arguments of a function call only. Various extensions to the higher order case have been developed (Sakai et al., 2001; Wahlstedt, 2007; Jones & Bohr, 2008; Kusakari et al., 2009; Kop, 2011), but no general unifying theory yet.

The present paper is not concerned with this problem but with defining a practical notion of size for simply typed $\lambda$-terms inhabiting inductively defined types.

Note by the way that the derivational complexity of a rewrite system, i.e. the function mapping every term $t$ to the maximum number of successive rewrite steps one can do from $t$ (Hofbauer & Lautemann, 1989), does not seem to be related, at
least in a simple way, to the ordinal necessary to prove its termination: there are rewrite systems whose termination can be proved by induction up to $\omega$ only and yet have huge derivational complexities (Moser, 2017), unless perhaps one bounds the growth rate of the size of terms (measured here as the number of symbols) (Schmitz, 2014). The notion of runtime complexity, i.e. the function mapping every $n \in \mathbb{N}$ to the maximum number of successive rewrite steps one can do from a term whose subterms are in normal form and whose size is smaller than $n$, seems to provide a better (Turing related) complexity model (Avanzini & Moser, 2010).

1.2 Model-based termination

Manna & Ness (1970) proposed to interpret every term whose free variables are $x_1, \ldots, x_n$ by a function from $\mathbb{E}^n$ to $\mathbb{E}$, where $(\mathbb{E}, <_E)$ is a well-founded domain. That is, $\mathcal{I}$ is the set of all the functions from some power of $\mathbb{E}$ to $\mathbb{E}$ and $<_\mathcal{I}$ is the pointwise extension of $<_E$, i.e. $f : \mathbb{E}^n \to \mathbb{E} <_\mathcal{I} g : \mathbb{E}^n \to \mathbb{E}$ if, for all $x_1, \ldots, x_n \in \mathbb{E}$, $f(x_1, \ldots, x_n) <_E g(x_1, \ldots, x_n)$.

In the first-order case, this can be done in a structured way by interpreting every function symbol $f$ of arity $n$ by a function $f_\mathcal{I} : \mathbb{E}^n \to \mathbb{E}$ and every term by composing the interpretations of its symbols, e.g. $\|f(g(x))\|$ is the function mapping $x$ to $f_\mathcal{I}(g_\mathcal{I}(x))$. If moreover these interpretation functions are monotone in each argument, then checking that rewriting is size decreasing can be reduced to checking that every rule is size decreasing.

A natural domain for $(\mathbb{E}, <_E)$ is of course $(\mathbb{N}, <_\mathbb{N})$. In this case, both monotony and size decreasingness can be reduced to absolute positivity. Indeed, $f(x_1, \ldots, x_p) > g(x_1, \ldots, x_q)$ is equivalent to $f(x_1, \ldots, x_p) - g(x_1, \ldots, x_q) - 1 \geq 0$ and monotony is equivalent to checking that, for all $i$, $f(\ldots, x_i+1, \ldots) - f(\ldots, x_i, \ldots) - 1 \geq 0$. By restricting the class of functions, e.g. to polynomials of bounded degree, one can develop heuristics for trying to automatically find monotone polynomial interpretation functions making rules size decrease (Cherifa & Lescanne, 1987; Lucas, 2005; Contejean et al., 2005; Fuhs et al., 2007). Unfortunately, polynomial absolute positivity is undecidable on $\mathbb{N}$ since it is equivalent to the solvability of Diophantine equations (Proposition 6.2.11 in TeReSe (2003)), which is undecidable (Matiyasevich, 1970, 1993). Yet, these tools get useful results in practice by restricting degrees and coefficients to small values, e.g. 2.

A similar approach can be developed for dense sets like $\mathbb{Q}^+$ or $\mathbb{R}^+$ by ordering them with the (not well-founded!) usual orderings on $\mathbb{Q}^+$ and $\mathbb{R}^+$ if one assumes moreover that the functions $f_\mathcal{I}$ are strictly extensive (i.e. $f_\mathcal{I}(x_1, \ldots, x_n) > x_i$ for all $i$) (Dershowitz, 1979a), or with the well-founded relation $<_\delta$ where, for some fixed $\delta > 0$, $x <_\delta y$ if $x + \delta \leq y$ (Lucas, 2005; Fuhs et al., 2008). In the case of $\mathbb{R}^+$, polynomial absolute positivity is decidable but of exponential complexity (Tarski, 1948; Collins, 1975). Useful heuristics have however been studied (Hong & Jakuš, 1998).

These approaches have also been successfully extended to linear functions on domains like $\mathbb{E} = \mathbb{B}^n$ (vectors of dimension $n$) or $\mathbb{E} = \mathbb{B}^{n \times n}$ (square matrices of
dimension \( n \) (Endrullis et al., 2008; Courtieu et al., 2010), where \( \mathbb{B} \) is a well-founded domain.

Instead of polynomial functions, Cichoń considered the class of Hardy functions (Hardy, 1904) indexed by ordinals smaller than \( \varepsilon_0 \) (Cichoń & Touzet, 1996). The properties of Hardy functions (composition is addition of indices, etc.) can be used to reduce the search of appropriate Hardy functions to solving inequalities on ordinals.

Manna and Ness’ approach has also been extended to the higher order case. Gandy (1980b) remarks that terms of the \( \lambda I \)-calculus (i.e. when, in every abstraction \( \lambda x t \), \( x \) freely occurs at least once in \( t \)) can be interpreted in the set of hereditary strictly monotone functions on some well-founded set \( (\mathbb{E},<_{\mathbb{E}}) \), that is, a closed term of base type \( \mathbb{B} \) is interpreted in the set \( \llbracket \mathbb{B} \rrbracket = \mathbb{E} \), a closed term of type \( T \Rightarrow U \) is interpreted by a monotone function from \( \llbracket T \rrbracket \) to \( \llbracket U \rrbracket \). and \( f : \llbracket T \Rightarrow U \rrbracket <_{\llbracket T \Rightarrow U \rrbracket} \ g : \llbracket T \Rightarrow U \rrbracket \) if, for all \( x \in \llbracket T \rrbracket \), \( f(x) <_{\llbracket U \rrbracket} g(x) \) (note that, in contrast with the first-order case, \( x \) itself may be a function). Then, by taking \( \mathbb{E} = \mathbb{N} \) and extending the \( \lambda \)-calculus with constants \( 0 : o, s : o \Rightarrow o \) and \( + : o \Rightarrow o \Rightarrow o \) for each base type \( o \), he defines a size function that makes \( \beta \)-reduction size decrease and provide an upper bound to the number of rewrite steps. An exact upper bound was later computed by de Vrijer (1987).

Gandy’s approach was later extended by van de Pol (1993, 1996) and Kahrs (1995) to arbitrary higher order rewriting à la Nipkow (1991) and Mayr & Nipkow (1998), that is, to rewriting on terms in \( \beta \)-normal \( \eta \)-long form with higher order pattern matching (Miller, 1991). But this approach has been implemented only recently (Fuhs & Kop, 2012).

Interestingly, van de Pol also showed that, in the simply typed \( \lambda \)-calculus, Gandy’s approach can be seen as a refinement of Girard’s proof of termination based on computability predicates (Van de Pol, 1995, 1996).

Finally, a general categorical framework has been developed by Hamana (2006), that is complete wrt the termination of binding term rewrite systems, a formalism based on Fiore, Plotkin and Turi’s binding algebra (Fiore et al., 1999) and close to a typed version of Klop’s combinatory reduction systems (Klop et al., 1993).

To the best of our knowledge, nobody seems to have studied the relations between Howard’s approach based on ordinals (Howard, 1970; Wilken & Weiermann, 2012) and Gandy’s approach based on interpretations (Gandy, 1980b; De Vrijer, 1987; Van de Pol, 1996).

Note also that the existence of a quasi interpretation, i.e. \( \llbracket t \rrbracket \geq_{\mathbb{D}} \llbracket u \rrbracket \) whenever \( t \rightarrow u \), not only may give useful information on the complexity of a rewrite system (Bonfante et al., 2011) but, sometimes, may also simplify the search of a termination proof. Indeed, Zantema (1995) proved that the termination of a first-order rewrite system \( \mathcal{R} \) is equivalent to the termination of \( \text{lab}(\mathcal{R}) \cup \geq_{\mathbb{D}} \), where \( \text{lab}(\mathcal{R}) \) are all the variants of \( \mathcal{R} \) obtained by annotating function symbols by the interpretation of their arguments, a transformation called semantic labeling. Although usually infinite, the obtained labeled system may be simpler to prove terminating, and some heuristics have been developed to use this technique in automated termination tools (Middeldorp et al., 1996; Koprowski & Zantema, 2006; Sternagel

https://doi.org/10.1017/S0956796818000072 Published online by Cambridge University Press
1.3 Termination based on typing with size annotations

Finally, there is another approach based on the semantics of inductive types, that has been developed for functions defined with a fixpoint combinator and pattern matching (Burstall et al., 1980).

The semantics of an inductive type $B$, $\sem{B}$, is usually defined, following Hessenberg (1909) theorem, Knaster & Tarski (1928) theorem or Tarski (1955) theorem, as the smallest fixpoint of a monotone function $\mathcal{H}^B$ on some complete lattice. Moreover, following Kuratowski (1922) and Cousot & Cousot (1979), such a fixpoint can be reached by transfinite iteration of $\mathcal{H}^B$ from the smallest element of the lattice $\bot$. Hence, every element $t \in \sem{B}$ can be given as size the smallest ordinal $\alpha$ such that $t \in \mathcal{S}_\alpha^B$, where $\mathcal{S}_\alpha^B$ is the set obtained after $\alpha$ transfinite iterations of $\mathcal{H}^B$ from $\bot$.

In particular, terms of a first-order data type like the type of Peano integers, lists, binary trees, ... always have a size smaller than $\omega$.

Mendler used this notion of size to prove the termination of an extension of Gödel’s system T (Gödel, 1958) and Howard’s system V (Howard, 1972) to functionals defined by recursion on higher order inductive types, i.e. types with constructors taking functions as arguments (Mendler, 1987, 1991), in which case the size of a term can be bigger than $\omega$.

In Hughes et al. (1996) and Pareto (2000), Hughes, Pareto and Sabry proposed to internalize this notion of size by extending the type system with, for each data type $B$, new type constants $B_0, B_1, \ldots, B_\infty = B$ for typing the terms of type $B$ of size smaller than or equal to 0, 1, $\ldots$, $\infty$, respectively, and the subtyping relation induced by the fact that a term of size at most $a$ is also of size at most $b$ whenever $a \leq_N b$ or $b = \infty$. More generally, to provide some information on how a function behaves wrt sizes, they consider as size annotations not only 0, 1, $\ldots$, $\infty$ but any first-order term built from the function symbols 0 for zero, $s$ for successor and $+$ for addition, and arbitrary size variables $\alpha, \beta, \ldots$, that is the language of Presburger arithmetic (Presburger, 1929). So, for instance, the usual list constructor $\text{cons}$ gets the type $\text{N} \Rightarrow \text{L}_\alpha \Rightarrow \text{L}_{\alpha\alpha}$, and the usual $\text{map}$ function on lists can be typed by $(\text{N} \Rightarrow \text{N}) \Rightarrow \text{L}_\alpha \Rightarrow \text{L}_\alpha$, where $\alpha$ is a free size variable that can be instantiated by any size expression in a way similar to type instantiation in ML-like programming languages (Milner, 1978).

Hughes, Pareto and Sabry do not actually prove the termination of their calculus but provide a domain-theoretic model (Scott, 1972). However, following Plotkin (1977), a closed term of first-order data type terminates iff its interpretation is not $\bot$. The first termination proof for arbitrary terms seems to have been given by Amadio & Coupet-Grimal (1997, 1998), who independently developed a system similar to the one of Hughes, Pareto and Sabry, inspired by Giménez’s work on the use of typing annotations for termination and productivity (Giménez, 1996). Giménez (1998) himself later proposed a similar system but provided no termination proof. Note that Plotkin’s result was later extended to higher order types and

https://doi.org/10.1017/S0956796818000072 Published online by Cambridge University Press

Size annotations are an abstraction of the semantic notion of size that one can use to prove properties on the actual size of terms like termination (size decreasingness) or the fact that a function is not size increasing (e.g. \texttt{map}), which can in turn be used in a termination proof (Walther, 1988; Giesl, 1997). Following Cousot (1997), it could certainly be described as an actual abstract interpretation.

Hence, termination can be reduced to checking that a term has some given type in the system with size-annotated type constants and subtyping induced by the ordering on size annotations, the usual typing rules being indeed valid deduction rules wrt the size of terms (e.g. if \( t : N_a \Rightarrow N_b \) and \( u : N_a \), then \( tu : N_b \)).

But, in such a system, a term can have infinitely many different types because of size instantiation or because of subtyping. As already mentioned, size instantiation is similar to type instantiation in Hindley–Milner’s type system (Hindley, 1969; Milner, 1978), where the set of types of a term has a smallest element wrt the instantiation ordering if it is not empty (Huet, 1976). In this case, there is a complete type-checking algorithm for \((t, T)\) which consists of checking that \( T \) is an instance of the smallest type of \( t \) (Hindley, 1969). Unfortunately, with subtyping, there is no smallest type wrt the instantiation ordering (e.g. \( \lambda xx \) has type \( \forall x \Rightarrow x \) for all \( x \), and type \( B \Rightarrow C \) if \( B < C \), but \( B \Rightarrow C \) is not an instance of \( \forall x \Rightarrow x \)), or subtyping composed with instantiation (e.g. \( \lambda f\lambda x(f(x)) \) has type \( \forall x \Rightarrow (\forall x \Rightarrow \forall x) \) for all \( x \), and type \( (B \Rightarrow C) \Rightarrow (B \Rightarrow C) \) if \( B < C \), but no instance of \( (\forall x \Rightarrow \forall x) \Rightarrow (\forall x \Rightarrow \forall x) \) is a subtype of \( (B \Rightarrow C) \Rightarrow (B \Rightarrow C) \)) (Fuh & Mishra, 1990). To recover a notion of smallest type and completeness, all the works we know on type inference with subtyping extend the notion of type to include subtyping constraints.

We will not follow this approach though. One reason is that we consider Church-style \( \lambda \)-terms (i.e. with type-annotated abstractions) instead of Curry-style \( \lambda \)-terms and, in this case, as we will prove it, there is a smallest type wrt to subtyping composed with instantiation when size expressions are only built from variables, the successor symbol and an arbitrary number of constants (the “successor” size algebra). Note moreover that, although structural (function types and base types are incomparable), subtyping is not well-founded in this case since, for instance, \( N_x \Rightarrow N > N_{sz} \Rightarrow N > \ldots \) However, if we disregard how size annotations are related to the semantics of inductive types, our work has important connections with more general extensions of Hindley–Milner’s type system with subtypes (Mitchell, 1984; Fuh & Mishra, 1990; Pottier, 2001), indexed types (Zenger, 1997), DML(C) (Xi, 2002), HM(X) (Sulzmann, 2000) or generalized algebraic data types (Cheney, 2003; Xi et al., 2003), which are all a restricted form of dependent types (De Bruijn, 1970; Martin-Löf, 1975).

Hughes, Pareto and Sabry’s approach was later extended to higher order data types (Barthe et al., 2004), polymorphic types (Abel, 2004; Barthe et al., 2005; Abel, 2006, 2008), rewriting-based function definitions in the calculus of constructions (Blanqui, 2004, 2005a), conditional rewriting (Blanqui & Riba, 2006), product types (Barthe et al., 2008), and fixpoint-based function definitions in the calculus of constructions (Barthe et al., 2006; Grégoire & Sacchini, 2010; Sacchini, 2011).
It should be noted that, in contrast with the ordinal-based approach, not all terms are given a size, but only those of base type. Moreover, although ordinals are used to define the size of terms, no ordinal is actually used in the termination criterion since one considers an abstraction of them. Indeed, when comparing two terms, one does not need to actually know their size: it is enough to differentiate between their size. Hence, transfinite computations can be reduced to finite ones.

Finally, Roux and the author proved in Blanqui & Roux (2009) that size annotations provide a quasi-model, and thus can be used in a semantic labeling. Terms whose type is annotated by \( \infty \) (unknown size) are interpreted by using a technique introduced by Hirokawa & Middeldorp (2006). Interestingly, semantic labeling allows one to deal with function definitions using matching on defined symbols, like in a rule for associativity (e.g. \((x + y) + z \rightarrow x + (y + z)\)), while termination criteria based on types with size annotations are restricted to matching on constructor symbols.

Current implementations of termination checkers based on typing with size annotations include ATS (Xi, 2003; AT, 2018), MiniAgda (Abel, 2010; MiniAgda, 2014), Agda (Agda, 2017), cicminus (Sacchini, 2011; Cicminus, 2015) or HOT (HOT, 2012). Most of these tools assume given the annotated types of function symbols (e.g. to know whether the size of a function is bounded by the size of one of its arguments). Heuristics for inferring the annotations of function symbols have been proposed in Telford & Turner (2000) and Chin & Khoo (2001). They are both based on abstract interpretation techniques (Cousot, 1996).

### 1.4 Contributions

1. The first contribution of the present paper is to give a rigorous and detailed account, for the simply typed \( \lambda \)-calculus, of the approach and results sketched in Blanqui (2004, 2005a), hence providing the first complete account of the extension of Hughes, Pareto and Sabry’s approach to rewriting-based function definitions (Dershowitz & Jouannaud, 1990; TeReSe, 2003).

2. In all the works on size-annotated types, the size algebra is fixed. In those considering first-order data types only, the size algebra is usually the language of Presburger arithmetic, the first-order theory of which is decidable (Presburger, 1929; Fischer & Rabin, 1974). In those considering higher order data types, the successor symbol \( s \) is usually the only symbol allowed, except in Barthe et al. (2008) which allows addition too. Yet, there are various examples showing that, within a richer size algebra, more functions can be proved terminating since types are more precise.

The second contribution of the present paper is to provide a type-checking algorithm for a general formulation of Hughes, Pareto and Sabry’s calculus parameterized, for size annotations, by a quasi-ordered first-order term algebra \((A, \leq_A)\) interpreted in ordinals. In particular, we prove that this algorithm is complete whenever size function symbols are monotone, the existential fragment of \((A, \leq_A)\) is decidable and every satisfiable set of size constraints admits a smallest solution.
3. In all the previous works, the notion of size is also fixed: the size of \( t \) is the height of the set-theoretical tree representation of the normal form of \( t \) (an abstraction being represented as an infinite set of trees).

The third contribution of the paper is to enable users to define their own notion of size by annotating the types of constructors. These annotations generate a stratification of the interpretation of inductive types. We prove that one can build such a stratification in the domain of Girard's computability predicates (Girard, 1972; Girard et al., 1988) when annotations form monotone and extensive functions.

4. The fourth contribution is the proof that, in the successor algebra, the satisfiability of a finite set of constraints is decidable in polynomial time, and every satisfiable finite set of constraints has a smallest solution computable in polynomial time too.

In contrast with Blanqui (2004, 2005a), the present paper

- includes a short survey on the use of ordinals in termination proofs;
- develops a stratification-based notion of size for inhabitants of inductive types;
- introduces the notion of constructor size function;
- shows how to define a stratification from constructor size functions that are monotone and strictly extensive on recursive arguments;
- proves the existence and polynomial complexity of the computation of a smallest solution for a solvable set of constraints in the successor algebra, using max-plus algebra techniques instead of pure linear algebra techniques.

### 1.5 Organization of the paper

In Section 2, we recall the definitions of types, terms and rewriting, and the interpretation of types as computability predicates. In Section 3, we introduce the notions of stratification, size and constructor size functions, and prove properties on the size of computable terms. In Section 4, we present the termination criterion. The main ingredient of the termination criterion is a type system with subtyping, parameterized by a quasi-ordered first-order term algebra for abstract size expressions. It also requires that annotations of arguments are minimal in some sense. In Section 5, we provide a sufficient syntactic condition for the minimality property to be satisfied when the size is defined as the height. In Section 6, we provide various examples of the expressive power of our termination criterion. In Section 7, we provide a complete algorithm for checking subject-reduction and size decreasingness under some general assumptions on the size algebra. In Section 8, we show how subtyping problems can be reduced to ordering problems in the size algebra. Finally, in Section 9, we prove that the simplest possible algebra, the successor algebra, satisfies the required conditions for the type-checking algorithm to be complete.
2 Types, terms and computability

In this section, we define the set of terms that we consider (Church’s simply typed \(\lambda\)-calculus with constants (Church, 1940)), the operational semantics (the combination of \(\beta\)-reduction and user-defined rewrite rules (Dershowitz & Jouannaud, 1990; TeReSe, 2003)) and the notion of computability used to prove termination.

Given a set \(E\), we denote by \(E^*\) the set of words or sequences over \(E\) (i.e. the free monoid containing \(E\)), the empty word by \(\varepsilon\), the concatenation of words by juxtaposition, the length of a word \(w\) by \(|w|\). We also use \(\vec{e}\) to denote a (possibly empty) sequence \(e_1, \ldots, e_{|\vec{e}|}\) of elements of \(E\).

Given a partial function \(f : A \rightarrow B\), \(a \in A\) and \(b \in B\), let \([a : b, f]\) be the function mapping \(a\) to \(b\) and every \(x \in \text{dom}(f) - \{a\}\) to \(f(x)\).

We recall that, if \(X\) is a bounded set of ordinals, i.e. when there is \(b\) such that \(x \leq b\) for all \(x \in X\), then the least upper bound of \(X\), written \(\text{sup} X\), exists. In particular, \(\text{sup} \emptyset = 0\).

### 2.1 Types

Following Church, we assume given a non-empty countable set \(\mathbb{S}\) of sorts \(B, C, \ldots\) and define the set \(\mathbb{T}\) of (simple) types as follows:

- Sorts are types.
- If \(T\) and \(U\) are types, then \(T \Rightarrow U\) is a type.

Implication associates to the right. So, \(T \Rightarrow U \Rightarrow V\) is the same as \(T \Rightarrow (U \Rightarrow V)\). Moreover, \(\hat{T} \Rightarrow U\) is the same as \(T_1 \Rightarrow T_2 \Rightarrow \ldots \Rightarrow T_n \Rightarrow U\) where \(n = |\hat{T}|\).

The arity of a type \(T\), \(\text{ar}(T)\), is defined as follows: \(\text{ar}(B) = 0\) and \(\text{ar}(T \Rightarrow U) = 1 + \text{ar}(U)\).

### 2.2 Terms

Given disjoint countable sets \(\mathbb{V}, \mathbb{C}\) and \(\mathbb{F}\), for variables, constructors and function symbols, respectively, we define the set of pre-terms as follows:

- Variables, constructors and function symbols are pre-terms.
- If \(x\) is a variable, \(T\) a type and \(u\) a pre-term, then \(\lambda x^T u\) is a pre-term.
- If \(t\) and \(u\) are pre-terms, then \(tu\) is a pre-term.

Application associates to the left. So, \(tuv\) is the same as \((tu)v\). Moreover, \(t\vec{u}\) is the same as \((\cdots((tu_1)u_2)\ldots u_{n-1})u_n\) where \(n = |\vec{u}|\).

As usual, the set of terms \(\mathbb{L}\) is obtained by quotienting pre-terms by \(\alpha\)-equivalence, i.e. renaming of bound variables, assuming that \(\mathbb{V}\) is infinite (Curry & Feys, 1958).

As usual, positions in a tree (type or term) are denoted by words on positive integers. Word concatenation is denoted by juxtaposition and the empty word by \(\varepsilon\).

Given a tree \(t\) and a position \(p\) in \(t\), let \(t|_p\) be the subtree of \(t\) at position \(p\), and \(\text{Pos}(u, t)\) be the set of positions \(p\) in \(t\) such that \(t|_p = u\).
A substitution \( \theta \) is a map from variables to terms whose domain \( \text{dom}(\theta) = \{ x \in V \mid \theta(x) \neq x \} \) is finite. In the following, any finite map \( \theta \) from variables to terms is implicitly extended into the substitution \( \theta \cup \{(x, x) \mid x \notin \text{dom}(\theta)\} \). Let \( \text{FV}(\theta) = \bigcup\{\text{FV}(\theta(x)) \mid x \in \text{dom}(\theta)\} \). The application of a substitution \( \theta \) to a term \( t \) is written \( t\theta \). We have \( x\theta = \theta(x), (tu)\theta = (t\theta)(u\theta) \) and \( (\lambda xT)\theta = \lambda xT(u\theta) \) if \( x \notin \text{dom}(\theta) \cup \text{FV}(\theta) \), which can always be achieved by \( \alpha \)-equivalence.

### 2.3 Typing

We assume given a map \( \Theta \) assigning a type to every symbol \( s \in C \cup F \), and will sometimes write \( s: T \) instead of \( (s,T) \in \Theta \) or \( \Theta(s) = T \).

A **typing environment** is a finite map \( \Gamma \) from variables to types. The usual deduction rules assigning a type to a term in a typing environment are recalled in Figure 2. As mentioned at the beginning of the section, \( [x : U, \Gamma] \) is the function mapping \( x \) to \( U \) and every \( x \in \text{dom}(\Gamma) \setminus \{x\} \) to \( \Gamma(x) \).

Given a symbol \( s \), let \( r^s = \text{ar}(\Theta(s)) \) be the maximum number of terms that can be applied to. For all \( s \), there are types \( T_1, \ldots, T_n \) and a sort \( B \) such that \( \Theta(s) = T_1 \Rightarrow \cdots \Rightarrow T_n \Rightarrow B \).

Given \( B \in S \), let \( C^B = \{(c, \tilde{t}, \tilde{T}) \mid c \in C, c : \tilde{T} \Rightarrow B, |\tilde{t}| = |\tilde{T}|\} \) be the set of tuples \( (c, \tilde{t}, \tilde{T}) \) such that \( c \) is maximally applied in \( c\tilde{t} \) and \( \tilde{T} \) are the types declared for the arguments of \( c \) (but \( t_i \) does not need to be of type \( T_i \)).

\[
\begin{array}{c}
(s,T) \in \Theta \cup \Gamma \\
\Gamma \vdash s : T \\
\Gamma \vdash t : U \Rightarrow V \\
\Gamma \vdash u : U \\
\Gamma \vdash tu : V \\
[x : U, \Gamma] \vdash v : V \\
\Gamma \vdash \lambda x^U v : U \Rightarrow V
\end{array}
\]

Fig. 2. Typing rules.

### 2.4 Rewriting

Given a relation on terms \( R \), let \( R(t) = \{ t' \in L \mid tRt' \} \) be the set of immediate reducts of a term \( t \), \( R^\ast \) be the reflexive and transitive closure of \( R \) and \( R^{-1} \) be its inverse \( (xR^{-1}y \text{ if } yRx) \). \( R \) is **finely branching** if, for all \( t \), \( R(t) \) is finite. It is **monotone** (or congruent, stable by context, compatible with the structure of terms) if \( tuRtu', utRu' \) and \( \lambda x^U tR\lambda x^U t' \) whenever \( tRt' \). It is **stable** (by substitution) if \( t\theta Rt' \theta \) whenever \( tRt' \).

Given two relations \( R \) and \( S \), let \( RS \) (or \( R \circ S \)) be their composition \( (tRSv \text{ if there is } u \text{ such that } tRu \text{ and } uSv) \). A relation \( R \) is **locally confluent** if \( R^{-1}R \subseteq R^\ast(R^{-1})^\ast \), and **confluent** if \( (R^{-1})^\ast R^\ast \subseteq R^\ast(R^{-1})^\ast \).

The relation of \( \beta \)-**rewriting** \( \rightarrow_\beta \) is the smallest monotone relation containing all the pairs \( ((\lambda x^U t)u, t'(x,u)) \).

A **rewrite rule** is a pair of terms \( (l, r) \), written \( l \rightarrow r \), such that there are \( f \in F \), \( \tilde{l}, \Delta \) and \( T \) such that \( l = \tilde{l}, \text{FV}(r) \subseteq \text{FV}(l), \Delta \vdash l : T \) and, (SR) for all \( \Gamma \) and \( U \), \( \Gamma \vdash r : U \) whenever \( \Gamma \vdash l : U \).

Given a set \( \mathcal{R} \) of rewrite rules, let \( \rightarrow_{\mathcal{R}} \) denote the smallest monotone and stable relation containing \( \mathcal{R} \). The condition (SR) implies that \( \rightarrow_{\mathcal{R}} \) preserves typing: if
Γ ⊬ t : U and t → R u, then Γ ⊬ u : U (subject-reduction property). Note that it is satisfied if, for instance, l contains no abstraction and no subterm of the form x t (Barbanera et al., 1997).

All over the paper, we assume given a set R of rewrite rules and let SN be the set of terms strongly normalizing wrt

\[ \rightarrow = \rightarrow_\beta \cup \rightarrow_R \]

We will assume that → is finitely branching, which is in particular the case if R is finite.

Given B and t, let \( \mathcal{C}_B^C(t) = \{ (c, \vec{t}, \vec{T}) \in \mathcal{C}_B^C | t \rightarrow^* c \vec{t} \} \) be the set of triples \((c, \vec{t}, \vec{T})\) such that \( t \) reduces to \( c \vec{t} \), \( c \) is maximally applied in \( c \vec{t} \), and \( \vec{T} \) are the types of the arguments of \( c \).

Given a relation \( R \), let \( \vec{x} R_{\text{prod}} \vec{y} \) if \( |\vec{x}| = |\vec{y}| \) and there is \( i \) such that \( x_i R y_i \) and, for all \( j \neq i \), \( x_j = y_j \). Given \( n \) relations \( R_1, \ldots, R_n \), let \( \vec{x} (R_1, \ldots, R_n)_{\text{lex}} \vec{y} \) if \( |\vec{x}| \geq n, |\vec{y}| \geq n \) and there is \( i \) such that \( x_i R_y y_i \) and, for all \( j < i \), \( x_j = y_j \). \( R_{\text{prod}} \) and \( (R_1, \ldots, R_n)_{\text{lex}} \) are well-founded whenever \( R, R_1, \ldots, R_n \) so are.

### 2.5 Computability

Following Tait (1967), Girard (1972), Girard et al. (1988), Mendler (1987), Okada (1989), Brezu-Tannen & Gallier (1989), Jouannaud & Okada (1991) and Blanqui et al. (2002), ... termination of a rewrite relation on simply typed \( \lambda \)-terms can be obtained by interpreting types by computability predicates and checking that function symbols are computable, that is, map computable terms to computable terms.

However, to handle matching on constructors taking functions as arguments (or matching on function symbols), one needs to modify Girard’s definition of computability. In the following, we recall the definition that we will use some of its basic properties, and refer the reader to Blanqui (2016) and Riba (2009) for more details on the theory of computability predicates with rewriting.

**Definition 1 (Computability predicates)** A term \( t \) is neutral if it is of the form \( x \vec{v} \), \((\lambda x)t \vec{w}\) or \( f \vec{t} \) with \( |\vec{t}| \geq \sup\{ |\vec{l}| \mid \exists r, \vec{l} \rightarrow r \in R \} \).\(^1\) A computability predicate is a set of terms \( \mathcal{P} \) satisfying the following properties:

- \( \mathcal{P} \subseteq \text{SN} \).
- \( \rightarrow (\mathcal{P}) \subseteq \mathcal{P} \).
- If \( t \) is neutral and \( \rightarrow (t) \subseteq \mathcal{P} \), then \( t \in \mathcal{P} \).

Let \( \mathcal{P} \) be the set of all the computability predicates. An element of a computability predicate is said to be computable.

\(^1\) The supremum exists since, by assumption, for all \( f \vec{l} \rightarrow r \in R \), \( \vec{l} \) is typable and thus \( |\vec{l}| \leq t' \).

https://doi.org/10.1017/S0956796818000072 Published online by Cambridge University Press
In our definition of neutral terms, not every redex is neutral as it is the case in Girard’s definition. However, the following key property is preserved: application preserves neutrality, that is, if \( t \) is neutral, then \( tu \) is neutral too. This definition also works with polymorphic and dependent types. It only excludes infinite rewrite systems where the number of arguments to which a function symbol is applied is unbounded (at the top of rule left-hand sides only, not in every term).

Computability predicates enjoy the following properties:

- The set \( \mathcal{V} \) of variables is included in every computability predicate.
- Given a computability predicate \( \mathcal{S} \), \((\lambda x^\mathcal{U}v)u \in \mathcal{S} \) iff \( v\{x,u\} \in \mathcal{S} \) and \( u \in \mathcal{S}_{\text{SN}} \).
- \( \mathcal{I} \) is a complete lattice wrt inclusion.

The greatest lower bound of a set \( \mathcal{Q} \subseteq \mathcal{I} \) is \( \bigcap \mathcal{Q} \) if \( \mathcal{Q} \neq \emptyset \), and \( \mathcal{S}_{\text{SN}} \) (the greatest element of \( \mathcal{I} \)) otherwise. Note however that the lowest upper bound of \( \mathcal{Q} \), written \( \text{lub}(\mathcal{Q}) \), is not necessarily the union. For instance, with the non-confluent system \( \mathcal{R} = \{ f \to a, f \to b \} \), \( \mathcal{I}(\{a\}) \cup \mathcal{I}(\{b\}) \) is not a computability predicate since it does not contain \( f \). There are a number of cases where the union of two computability predicates is known to be a computability predicate, but this is for a different notion of neutral term:

- Riba (2007, 2008) prove that the set of computability predicates is stable by union if \( \mathcal{R} \) is an orthogonal constructor rewrite system.
- Werner (1994) (Lemma 4.14 p. 96) proves that the set of computability predicates is stable by well-ordered union.

Luckily, Werner’s proof does not depend on the definition of neutral terms:

**Lemma 1** If \( \rightarrow \) is finitely branching and \( \mathcal{Q} \) is a non-empty set of computability predicates well-ordered wrt inclusion, then \( \bigcup \mathcal{Q} \) is a computability predicate.

**Proof**

- Let \( t \in \bigcup \mathcal{Q} \). Then, there is \( \mathcal{S} \in \mathcal{Q} \) such that \( t \in \mathcal{S} \). Since \( \mathcal{S} \subseteq \mathcal{S}_{\text{SN}} \), we have \( t \in \mathcal{S}_{\text{SN}} \).
- Let \( t \in \bigcup \mathcal{Q} \) and \( u \) such that \( t \rightarrow u \). Then, there is \( \mathcal{S} \in \mathcal{Q} \) such that \( t \in \mathcal{S} \). Since \( t \rightarrow (\mathcal{S}) \subseteq \mathcal{S} \), we have \( u \in \mathcal{S} \) and thus \( u \in \mathcal{Q} \).
- Let \( t \) be a neutral term such that \( t \rightarrow (t) \subseteq \bigcup \mathcal{Q} \). If \( t \rightarrow (t) = \emptyset \), then \( t \) belongs to every element of \( \mathcal{Q} \). Therefore, \( t \in \bigcup \mathcal{Q} \). Otherwise, since \( \rightarrow \) is finitely branching, we have \( t \rightarrow (t) = \{t_1,\ldots,t_n\} \) with \( n \geq 1 \). For every \( i \in \{1,\ldots,n\} \), there is \( \mathcal{S}_i \in \mathcal{Q} \) such that \( t_i \in \mathcal{S}_i \). Since \( \mathcal{Q} \) is well-ordered wrt inclusion, there is \( k \in \{1,\ldots,n\} \) such that \( \mathcal{S}_k \) is the biggest element of \( \{\mathcal{S}_1,\ldots,\mathcal{S}_n\} \) wrt inclusion. Hence, \( t \rightarrow (t) \subseteq \mathcal{S}_k \) and \( t \in \mathcal{S}_k \). Therefore, \( t \in \bigcup \mathcal{Q} \).

The interpretation of arrow types is defined as usual, in order to ensure the termination of \( \beta \)-reduction:

**Definition 2 (Interpretation of arrow types)** A (partial) interpretation of sorts, that is, a (partial) function \( \mathbb{I} : \mathcal{S} \to \wp(\mathcal{L}) \) (powerset of \( \mathcal{L} \)), is extended into a (partial) interpretation of types \( \mathbb{I} : \mathcal{T} \to \wp(\mathcal{L}) \) as follows:
\textbullet\hspace{1em} \tilde{\Pi}(B) = \Pi(B).

\textbullet\hspace{1em} \tilde{\Pi}(U \Rightarrow V) = \tilde{\Pi}(U) \Rightarrow \tilde{\Pi}(V), \text{ where } \forall u \in \mathcal{U}, \forall t \in \mathcal{V}, \{t \in \mathcal{L} | \forall u \in \mathcal{U}, tu \in \mathcal{V}\}.

Note that \(\tilde{\Pi}(T)\) is defined whenever \(\Pi\) is defined on every sort occurring in \(T\), and \(\tilde{\Pi}(T) = \tilde{\Pi}(T)\) whenever \(\Pi\) and \(\Pi\) are defined and equal on every sort occurring in \(T\).

Note also that \(\forall u \Rightarrow \forall v\) is a computability predicate whenever \(\forall u\) and \(\forall v\) so are. Hence, \(\tilde{\Pi}(T)\) is a computability predicate whenever \(\Pi(B)\) so is for every sort \(B\) occurring in \(T\).

For interpreting sorts, one could take the computability predicate \(SN\). But this interpretation does not allow one to prove the computability of functions defined by induction on types with constructors taking functions as arguments.

Moreover, a computable term may have non-computable subterms. Consider, for instance, \(c : (B \Rightarrow C) \Rightarrow B, f : B \Rightarrow (B \Rightarrow C), \mathcal{R} = \{f(c \ x) \rightarrow x\} \) and \(t = \lambda x^B f xx\). Then, assuming that \(\Pi(B) = SN\), we have \((c \ t) \in \Pi(B)\), but \(t \notin \Pi(B) \Rightarrow \Pi(C)\) since \(t(c \ t) \notin SN\) because \(t(c \ t) \rightarrow_\mathcal{R} f(c \ t)(c \ t) \rightarrow_\mathcal{R} t(c \ t)\). It is however possible to enforce that a direct subterm of type \(T\) of a computable term of sort \(B\) is computable if \(B\) occurs in \(T\) at positive positions only (Mendler, 1987):

**Definition 3 (Positive and negative positions in a type)** The subsets of positive \((s = +)\) and negative \((s = -)\) positions in a type \(T\), \(\text{Pos}^s(T)\), are defined as follows:

\textbullet\hspace{1em} \text{Pos}^s(B) = \{c | c = +\},

\textbullet\hspace{1em} \text{Pos}^s(U \Rightarrow V) = \{p \in \text{Pos}^{-s}(U) \cup \{2p \mid p \in \text{Pos}^s(V)\},

where \(-+ = +\) and \(-+ = -\).

Note that the sets of positive and negative positions of a type are disjoint. However, in a type, a sort can have both positive and negative occurrences. For instance, \(\text{Pos}^+(B, B \Rightarrow B) = \{2\}\) and \(\text{Pos}^-(B, B \Rightarrow B) = \{1\}\).

**Definition 4 (Accessible arguments)** We assume given a well-founded ordering on sorts \(<_S\). The \(i\)th argument of a constructor \(c : T \Rightarrow B\) is

\textbullet\hspace{1em} recursive if \(\text{Pos}(B, T_i) \neq \emptyset\),

\textbullet\hspace{1em} accessible if \(T_i\) is positive wrt. \(B\), that is,

\textbullet\hspace{1em} every sort occurring in \(T_i\) is smaller than or equal to \(B\):

\textbullet\hspace{1em} for all \(C\), \(\text{Pos}(C, T_i) = \emptyset\) or \(C \leq_S B\), where \(\leq_S\) is the reflexive closure of \(<_S\);

\textbullet\hspace{1em} \(B\) occurs only positively in \(T_i\): \(\text{Pos}(B, T_i) \subseteq \text{Pos}^+(T_i)\).

In the following, we will assume wlog\(^2\) that there are \(0 \leq p^c \leq q^c\) such that

\textbullet\hspace{1em} the arguments \(1\) to \(p^c\) are accessible and recursive,

\textbullet\hspace{1em} the arguments \(p^c + 1\) to \(q^c\) are accessible and not recursive:

\(\Theta(c) = T_1 \Rightarrow \ldots \Rightarrow T_{p^c} \Rightarrow T_{p^c+1} \Rightarrow \ldots \Rightarrow T_{q^c} \Rightarrow T_{q^c+1} \Rightarrow \ldots \Rightarrow T_{q^c} \Rightarrow B\)

\text{rec. acc. args} \hspace{1em} \text{non-rec. acc. args} \hspace{1em} \text{non-acc. args}

\(^2\) Arguments can be permuted if needed.
For instance, for the sort \( \mathbb{N} \) of natural numbers with the constructors \( 0 : \mathbb{N} \) and \( \mathsf{s} : \mathbb{N} \to \mathbb{N} \) (successor) (Peano, 1889), we can take \( p^0 = q^0 = 0 \) and \( p^\mathsf{s} = q^\mathsf{s} = 1 \) since \( \mathbb{N} \) occurs only positively in \( \mathbb{N} \). Similarly, for the sort \( \mathbb{O} \) of Howard’s constructive ordinals with the constructors \( \mathsf{zero} : \mathbb{O} \), \( \mathsf{succ} : \mathbb{O} \to \mathbb{O} \) (successor) and \( \mathsf{lim} : (\mathbb{N} \to \mathbb{O}) \to \mathbb{O} \) (limit) (Howard, 1972), we can take \( p^{\mathsf{zero}} = q^{\mathsf{zero}} = 0 \), \( p^{\mathsf{succ}} = q^{\mathsf{succ}} = 1 \) since \( \mathbb{O} \) occurs only positively in \( \mathbb{O} \), and \( p^{\mathsf{lim}} = q^{\mathsf{lim}} = 1 \) since \( \mathbb{O} \) occurs only positively in \( \mathbb{N} \to \mathbb{O} \) if one takes \( \mathbb{N} \prec \mathbb{O} \). Now, for the sort \( \mathbb{L} \) of lists of natural numbers with the constructors \( \mathsf{nil} : \mathbb{L} \) and \( \mathsf{cons} : \mathbb{N} \to \mathbb{L} \Rightarrow \mathbb{L} \), we can take \( p^{\mathsf{cons}} = 1 \) and \( q^{\mathsf{cons}} = 2 \) if one takes \( \mathbb{N} \prec \mathbb{L} \).

Non-accessible arguments are usually forbidden by requiring all the arguments to be positive, or even strictly positive\(^3\) as it is the case in the Coq proof assistant (Coquand & Paulin-Mohring, 1988). Here, we do not forbid non-positive arguments and do not require arguments to be strictly positive. Hence, one can have a sort \( \mathbb{D} \) with the constructors \( \mathsf{app} : \mathbb{D} \Rightarrow \mathbb{D} \Rightarrow \mathbb{D} \) and \( \mathsf{lam} : (\mathbb{D} \Rightarrow \mathbb{D}) \Rightarrow \mathbb{D} \), for which we must have \( p^{\mathsf{lam}} = q^{\mathsf{lam}} = 0 \) since the first argument of \( \mathsf{lam} \) is not positive. However, the termination conditions will enforce that, although one can use in a rule left-hand side (\( \mathsf{lam} \ x \)) as a pattern, \( x \) cannot be used in the corresponding rule right-hand side: in a rule, constructors with non-positive arguments can be pattern matched in the left-hand side, but only their positive arguments can be used by themselves in the right-hand side.

For the sake of simplicity, we consider an ordering instead of a quasi-ordering, although a quasi-ordering might \textit{a priori} be necessary for dealing with mutually defined inductive types (e.g. the types of trees and forests with the constructors \( \mathsf{F}, \mathsf{add} : \mathbb{F} \Rightarrow \mathbb{T} \Rightarrow \mathbb{F} \) and \( \mathsf{node} : \mathbb{F} \Rightarrow \mathbb{T} \)). The results described in this paper can however still be applied if one identifies mutually defined inductive types, because a term typable with mutually defined inductive types is \textit{a fortiori} typable in the type system where they are identified. This abstraction is correct but not necessarily complete since more terms get typable when two types are identified (e.g. \( \mathsf{add} \) is typable if \( \mathbb{T} = \mathbb{F} \)).

Since \( \prec \mathbb{S} \) is well-founded, we can define an interpretation \( \mathbb{I} \) for every sort by well-founded induction on it as follows. Let \( \mathbb{B} \) be a sort and assume that \( \mathbb{I} \) is defined for every sort smaller than \( \mathbb{B} \). Then, let \( \mathbb{I}(\mathbb{B}) \) be the least fixpoint of the monotone function \( \mathbb{H}^\mathbb{B} \) on the complete lattice \( \wp(\mathbb{L}) \) such that

\[
\mathbb{H}^\mathbb{B}(\mathcal{A}) = \{ t \in \mathbb{S}N \mid \forall (\mathbf{c}, \mathbf{T}) \in \mathcal{C}^\mathbb{B}_{\mathcal{A}}(t), \forall k \in \{1, \ldots, q^\mathbf{c}\}, t_k \in \mathbb{B} : \mathcal{A}, \mathbb{I}(T_k) \}
\]

where \( \mathbb{B} : \mathcal{A}, \mathbb{I} \) is introduced in Definition 2.

That such a least fixpoint exists follows from Knaster and Tarski’s fixpoint theorem (Knaster & Tarski, 1928; Tarski, 1955) and the following fact:

**Proposition 1** (Blanqui, 2005b) Let \( \mathbb{B} \) be a sort, \( \mathbb{I} \) be an interpretation for every sort smaller than \( \mathbb{B} \), and \( \mathbb{T} \) be a type positive wrt \( \mathbb{B} \). Then, \( \mathbb{B} : \mathcal{A}, \mathbb{I}(\mathbb{T}) \) is monotone wrt \( \mathcal{A} \).

\(^3\) The \( i \)th argument of \( \mathbf{c} \) is \textit{strictly positive} if \( \text{Pos}(\mathbb{B}, T_i) = \emptyset \), or \( T_i = \widehat{U} \Rightarrow \mathbb{B} \) and \( \text{Pos}(\mathbb{B}, \mathbb{U}) = \emptyset \).
Moreover, one can easily check that $\mathbb{P}^0(\mathcal{X})$ is a computability predicate whenever $\mathcal{X}$ so is. Hence, for every type $T$, $\mathbb{P}(T)$ is a computability predicate.

In the following, for the sake of simplicity, we will not mention $\mathbb{P}$ anymore and simply write $t \in T$ instead of $t \in \mathbb{P}(T)$, and $t \in [B : \mathcal{X}]T$ instead of $t \in [B : \mathcal{X}, \mathbb{P}](T)$.

### 3 Size of computable terms

In this section, we study a general way of attributing an ordinal size to computable terms of base type by defining, for each sort, a stratification of computable terms of this sort using a size function for each constructor, and assuming that $\to$ is finitely branching.

By Hartogs (1915) theorem, there is an ordinal the elements of which cannot be injected into $\wp(L)$, where $L$ is the set of terms (note that this theorem does not require the axiom of choice). Let $\eta$ be the smallest such ordinal. Since $V$ is countably infinite and $C$ and $F$ are countable, $\eta$ is the successor cardinal of $|\wp(L)| = 2^\omega$ (Hrbacek & Jech, 1999).

#### 3.1 Stratifications

**Definition 5 (Stratification of a sort)** Given a family $(S_a)_{a < \eta}$ of computability predicates, let $S_\eta = \text{lub}(\{S_a \mid a < \eta\})$.

A stratification of a computability predicate $S$ is a monotone sequence of computability predicates $(S_a)_{a < \eta}$ included in $S$ and converging to $S$, that is, such that $S_\eta = S$.

A stratification of a type $T$ is a stratification of $\mathbb{P}(T)$.

Given a stratification $S$, the size of an element $t \in S_\eta$, written $o_S(t)$, is the smallest ordinal $a < \eta$ such that $t \in S_a$.

A stratification is continuous if, for all limit ordinals $0 < a < \eta$, $S_a = \text{lub}(\{S_b \mid b < a\})$.

Because $\to$ is finitely branching, we immediately remark:

**Lemma 2** For all continuous stratifications $S$ and limit ordinal $0 < a \leq \eta$, we have $S_a = \bigcup(\{S_b \mid b < a\})$.

**Proof**

By definition, $S$ is monotone. So, for all $a \leq \eta$, $\{S_b \mid b < a\}$ is well-ordered wrt inclusion. Since $\to$ is finitely branching, the conclusion follows from Lemma 1.

We now prove some properties of $o_S(t)$:

**Lemma 3** Let $S$ be a stratification and $t \in S_\eta$.

- If $t \to t'$, then $t' \in S_\eta$ and $o_S(t) \geq o_S(t')$.
- If $S$ is continuous, then either $o_S(t) = 0$ or $o_S(t) = b + 1$ for some ordinal $b$.
Lemma 4

If \( \text{Cousot, 1979} \):

Assume that \( o_\mathcal{F}(t) \) is a limit ordinal \( a > 0 \). Since \( \mathcal{F} \) is continuous, we have \( \mathcal{F}_a = \bigcup \{ \mathcal{F}_b \mid b < a \} \). Therefore, \( t \in \mathcal{F}_b \) for some \( b < a \). Contradiction.

By Proposition 1, \([B : \mathcal{F}](T)\) is monotone wrt \( \mathcal{F} \) whenever \( T \) is positive wrt \( B \). Hence, any stratification \( \mathcal{F} \) of \( B \) provides a way to define a stratification of \( T \):

Definition 6 (Stratification of a positive type) Given a stratification \( \mathcal{F} \) of a sort \( B \) and a type \( T \) positive wrt \( B \), let \([B : \mathcal{F}](T)\) denote the stratification \( \mathcal{F} \) of \( T \) obtained by taking \( \mathcal{F}_a = [B : \mathcal{F}_a](T) \).

Note that \([B : \mathcal{F}](T)\) is not continuous in general (see Example 1 below).

Lemma 4 If \( \mathcal{F} \) is a stratification of \( B \), \( v \in \hat{U} \Rightarrow B \) and \( \text{Pos}(B, \hat{U}) \neq \emptyset \), then
\[
o_\{B : \mathcal{F}\}([U = B]v) = \sup \{ o_\mathcal{F}(v\hat{u}) \mid \hat{u} \in \hat{U} \}.
\]

Proof
Let \( a = o_{[B : \mathcal{F}]}([U = B]v) \) and \( b = \sup \{ o_\mathcal{F}(v\hat{u}) \mid \hat{u} \in \hat{U} \} \). By definition of \( a \), we have \( v \in \hat{U} \Rightarrow \mathcal{F}_a \). So, for all \( \hat{u} \in \hat{U} \), \( v\hat{u} \in \mathcal{F}_a \) and \( o_\mathcal{F}(v\hat{u}) \leq a \). Thus, \( b \leq a \). We now prove that \( a \leq b \). To this end, it suffices to prove that \( v \in \hat{U} \Rightarrow \mathcal{F}_b \). Let \( \hat{u} \in \hat{U} \). By definition of \( b \), \( o_\mathcal{F}(v\hat{u}) \leq b \). So, \( v\hat{u} \in \mathcal{F}_b \).

A continuous stratification of a sort \( B \) can be obtained by the transfinite iteration of \( H^B \) from the smallest computability predicate \( \bot \) (Kuratowski, 1922; Cousot & Cousot, 1979):

- \( D_0^B = \bot \).
- \( D_{\alpha+1}^B = H^B(D_\alpha^B) \).
- \( D_\alpha^B = \text{lub}(\{D_b^B \mid b < \alpha\}) \), if \( \alpha \) is an infinite limit ordinal.

The fact that \( D^B \) is monotone follows from the facts that \( D_0^B \subseteq D_1^B \) and \( H^B \) is monotone (Cousot & Cousot, 1979). Now, by definition of \( h \), \( D^B \) is not injective. Therefore, there are \( c < d < h \) such that \( D_c^B = D_d^B \). Since \( D^B \) is monotone, \( D_c^B = D_{c+1}^B = D_0^B = D_h^B = B \) (Rubin & Rubin, 1963).

We call this stratification the default stratification. It is the one used in all the previous works on sized types, except in Abel (2012) where, after Sprenger & Dam (2003), Abel uses a stratification having better properties, namely \( F_a^B = \text{lub}(\{H^B(D_b^B) \mid b < \alpha\}) \).

The size wrt the default stratification of a term \( t \) is the set-theoretical height of the tree representation of \( t \) when abstractions are interpreted as set-theoretical functions. If no constructor of \( B \) has accessible functional arguments and \( \rightarrow \) is finitely branching, then every element of \( B \) has a size smaller than \( \omega \). Hence, when considering first-order data types only (e.g. natural numbers, lists and binary trees) and a finitely branching rewrite relation \( \rightarrow \), one can in fact take \( h = \omega \).

On the other hand, when one wants to consider constructors with accessible functional arguments, then one can get terms of size bigger than \( \omega \):
Example 1 Take the sort $O$ of Howard’s constructive ordinals mentioned in the previous section and let $\text{inj} : N \Rightarrow O$ be the usual injection from $N$ to $O$ defined by the rules $\text{inj} 0 \rightarrow \text{zero}$ and $\text{inj} (s \, x) \rightarrow \text{succ} (\text{inj} \, x)$. Let us prove that $o_{\mathcal{O}} (\lim \, \text{inj}) = \omega + 1$. By definition, $o_{\mathcal{O}} (\lim \, \text{inj})$ is the smallest ordinal $\alpha$ such that $\lim \, \text{inj} \in \mathcal{D}_{\alpha}$. By definition of $\mathcal{D}$, $\alpha = o_{\mathcal{D}} (\text{inj}) + 1$ where $\mathcal{D} = \{ O : \mathcal{D}_{O} \} [N \Rightarrow O]$. By Lemma 4, $o_{\mathcal{D}} (\text{inj}) = \sup \{ o_{\mathcal{O}} (\text{inj} \, t) \mid t \in N \}$. Now, a term of the form $(\text{inj} \, u)$ can only reduce to a term of the form $(\text{inj} \, u)$, zero or $(\text{succ} \, u)$. Hence, $o_{\mathcal{O}} (\text{inj} \, t) < \omega$. Finally, one can easily prove that, for all $n < \omega$, $o_{\mathcal{O}} (\text{inj} (s^n 0)) = n + 1$. Therefore, $o_{\mathcal{D}} (\text{inj}) = \omega$ and $o_{\mathcal{O}} (\lim \, \text{inj}) = \omega + 1$. Moreover, $\mathcal{D}$ is not continuous since $\text{inj} \in \mathcal{D}_\omega - \bigcup \{ \mathcal{D}_n \mid n < \omega \}$.

One can also get terms of size bigger than $\omega$ by considering infinitely branching and non-confluent rewrite relations with $\mathcal{R} = \{ f \rightarrow s^i 0 \mid i \in N \}$, one gets $o_{\mathcal{O}} (f) = \omega + 1$.

### 3.2 Stratifications based on constructor size functions

We now introduce a general way of defining a stratification:

**Definition 7 (Constructor size function)** A size function for $c : T \Rightarrow B$ is given by

- a function $\Sigma^c : \mathfrak{H}_T^c \rightarrow \mathfrak{H}$ for computing the size of a term of the form $ct$ from the sizes of its accessible arguments;
- for every non-recursive accessible argument $k \in \{ p^c + 1, \ldots, q^c \}$, a sort $B_k^c <_S B$ occurring in $T_k$, only positively, and with respect to which we will measure the size of the $k$th argument of $c$ (in the following, we let $B_k^c = B$ if $k \in \{ 1, \ldots, p^c \}$).

In practice, there is usually no choice for $B_k^c$. For having a choice, the order of $T_k$ must be greater than or equal to 2. For instance, if $T_k = (C \Rightarrow D) \Rightarrow E$, then one can choose between $C$ and $E$ if both are different from $D$.

On the other hand, there are many possible choices for $\Sigma^c$. For instance, consider the type $T$ of labeled binary trees with the constructors $\text{leaf} : B \Rightarrow T$ and $\text{node} : T \Rightarrow T \Rightarrow B \Rightarrow T$, where $B <_S B$ is a sort for labels. We can take $p^\text{leaf} = 0$, $q^\text{leaf} = 1$, $p^\text{node} = 2$, $q^\text{node} = 3$, $\Sigma^\text{leaf} (a) = 0$ and $\Sigma^\text{node} (a, b, c) = a + b + 1$, so that the size of a tree is not its height as in the default stratification but the number of its nodes.

Interestingly, $\Sigma^c$ may depend on all accessible arguments, including the non-recursive ones. For instance, one can measure the size of a pair of natural numbers by the sum of their sizes: given a type $P$ for pairs of natural numbers with the constructor $\text{pair} : N \Rightarrow N \Rightarrow P$, one can take $p^\text{pair} = 0$, $q^\text{pair} = 2$, $B_1^\text{pair} = B_2^\text{pair} = N$ and $\Sigma^\text{pair} (a, b) = a + b$.

Finally, $\Sigma^c$ can be defined by combining of size of recursive and non-recursive arguments. For instance, the size of a list of natural numbers can be defined as the sum of the sizes of its components. With this notion of size, a list with only one big element can be greater than a list with many small elements.

**Definition 8 (Stratification defined by size functions)** Assume that $\rightarrow$ is finitely branching. Given a size function $\Sigma^c$ for every constructor $c$, we define a continuous
stratification \( \mathcal{S}^B \) for every sort \( B \) by induction on \( >_S \) as follows, where, given \((c, \bar{i}, \bar{T}) \in \mathfrak{C}_{<S}^B(t)\), \( o_{\mathcal{S}^B}(\bar{i}) \) denotes the sequence \( o_{\mathcal{S}^B}(t_1), \ldots, o_{\mathcal{S}^B}(t_n) \) with \( n = q^c \) and \( \mathcal{S}^B_{c,k} = [B^c : \mathcal{S}^B_{a}] T_k \), that is, \( o_{\mathcal{S}^B}(t_k) \) is the size of \( t_k \) in \( T_k \) wrt \( B^c_k \) (which is \( B \) if \( k \in \{1, \ldots, p^c\} \)):

- \( \mathcal{S}^B_0 \) is the set of terms \( t \in SN \) such that, for all \((c, \bar{i}, \bar{T}) \in \mathfrak{C}_{<S}^B(t)\):
  - \( p^c = 0 \) (i.e. \( c \) has no recursive argument),
  - \( \forall k \in \{p^c + 1, \ldots, q^c\}, \ t_k \in T_k \),
  - \( \Sigma^c(o_{\mathcal{S}^B}(\bar{i})) \leq 0 \).
- \( \mathcal{S}^B_{a+1} \) is the set of terms \( t \in SN \) such that, for all \((c, \bar{i}, \bar{T}) \in \mathfrak{C}_{<S}^B(t)\):
  - \( \forall k \in \{1, \ldots, p^c\}, \ t_k \in [B : \mathcal{S}^B] T_k \),
  - \( \forall k \in \{p^c + 1, \ldots, q^c\}, \ t_k \in T_k \),
  - \( \Sigma^c(o_{\mathcal{S}^B}(\bar{i})) \leq a + 1 \).
- \( \mathcal{S}^B_a = \text{lub}\{\mathcal{S}^B_b | b < a\} \), if \( a \) is an infinite limit ordinal.

Note that \( \mathcal{S} \) is well-defined because

- in the case of \( \mathcal{S}^B_0 \):
  - \( p^c = 0 \) and thus, for all \( k \in \{1, \ldots, q^c\}, \ o_{\mathcal{S}^B}(t_k) = o_{[B^c : \mathcal{S}^B]}(t_k) \) is well-defined
    since \( t_k \in T_k \) and \( B^c_k <_S B \).
- in the case of \( \mathcal{S}^B_{a+1} \):
  - \( \forall k \leq p^c, \ o_{\mathcal{S}^B}(t_k) = o_{[B : \mathcal{S}^B]}(t_k) \) is well-defined and \( \leq a \) since \( t_k \in [B : \mathcal{S}^B] T_k \);
  - \( \forall k \in \{p^c + 1, \ldots, q^c\}, \ o_{\mathcal{S}^B}(t_k) \) is well-defined since \( t_k \in T_k \) and \( B^c_k <_S B \).

The definition of \( \mathcal{S}^B \) is similar to the definition of the default stratification except that the size functions \( \Sigma^c \) are used to enforce lower bounds on the size of terms. Hence, if one takes for every \( \Sigma^c \) the constant function equal to 0, then one almost gets the default stratification. To get the default stratification, one has to slightly change the definition of \( \mathcal{S}^B \) by taking \( \mathcal{S}^B_0 = \bot \). The current definition has the advantage that both variables and nullary constructors whose size function is 0 have size 0. Hence, if one takes \( \Sigma_0 = \Sigma_S(a) = 0 \), then \( o_{\mathcal{S}^B}(s^i \cdot x) = o_{\mathcal{S}^B}(s^i \cdot 0) = i \) while, in the default stratification, \( o_{\mathcal{S}^B}(s^i \cdot x) = i \) and \( o_{\mathcal{S}^B}(s^i \cdot 0) = i + 1 \) (nullary constructors do not belong to \( \bot \)).

We now check that \( \mathcal{S}^B \) is indeed a stratification of \( B \).

**Lemma 5** For every sort \( B \) and ordinal \( a < h \), \( \mathcal{S}^B_a \subseteq B \).

**Proof**

We proceed by induction on \( _S \) and \( a \).

- Let \( t \in \mathcal{S}^B_0 \). Then, \( t \in SN \). Let now \((c, \bar{i}, \bar{T}) \in \mathfrak{C}_{<S}^B(t) \) and \( k \in \{1, \ldots, q^c\} \). Then, \( p^c = 0 \) and \( t_k \in T_k \). Hence, \( t \in B \) since \( B = H^B(B) \).
- Let \( t \in \mathcal{S}^B_{a+1} \). Then, \( t \in SN \). Let now \((c, \bar{i}, \bar{T}) \in \mathfrak{C}_{<S}^B(t) \) and \( k \in \{1, \ldots, q^c\} \). If \( k \leq p^c \), then \( t_k \in [B : \mathcal{S}^B] T_k \). By induction hypothesis, \( \mathcal{S}^B_{a} \subseteq B \). Since \( B \) occurs only positively in \( T_k \), Proposition 1 gives \([B : \mathcal{S}^B] T_k \subseteq [B : B] T_k = T_k \). Therefore, \( t_k \in T_k \). Now, if \( k \in \{p^c + 1, \ldots, q^c\} \), then \( t_k \in T_k \) too. Therefore, \( t \in B \) since \( B = H^B(B) \).
Let $a$ be an infinite limit ordinal. Then, $\mathcal{S}^B_a = \text{lub}\{\mathcal{S}^B_b \mid b < a\}$. For every $b < a$, by induction hypothesis, $\mathcal{S}^B_b \subseteq B$. Therefore, $\mathcal{S}^B_a \subseteq B$.

Lemma 6 For every sort $B$ and ordinal $a < h$, $\mathcal{S}^B_a$ is a computability predicate.

Proof
We proceed by induction on $<_S$ and $a$. If $a$ is an infinite limit ordinal, then $\mathcal{S}^B_a$ is a computability predicate by definition of lub since, by induction hypothesis, for all $b < a$, $\mathcal{S}^B_a$ is a computability predicate.

We are left with the cases of 0 and successor ordinals. Given a predicate $P$ on triples $(c, t, \bar{T})$, let $S^B(P) = \{t \in \text{SN} \mid \mathcal{C}^B_{\leq}(c, t) \subseteq P\}$. We have $\mathcal{S}^B_0 = \text{SN}^B(P_0)$ for some predicate $P_0$, and $\mathcal{S}^B_{a+1} = \text{SN}^B(P_{a+1})$ for some predicate $P_{a+1}$. However, for all predicates $P$, $\text{SN}^B(P)$ is a computability predicate:

- $\text{SN}^B(P) \subseteq \text{SN}^B$ by definition.
- If $t \in \text{SN}^B(P)$ and $t \rightarrow t'$, then $t' \in \text{SN}^B(P)$, since $t' \in \text{SN}^B$ and $\mathcal{C}^B_{\leq}(c, t') \subseteq \mathcal{C}^B_{\leq}(c, t)$.
- Assume now that $t$ is neutral and $t \rightarrow (t) \subseteq \text{SN}^B(P)$. Then, $t \in \text{SN}^B$. Assume, moreover, that $(c, t, \bar{T}) \in \mathcal{C}^B_{\leq}(t)$. Since $t$ is neutral, there is $t'$ such that $t \rightarrow t'$ and $(c, t, \bar{T}) \in \mathcal{C}^B_{\leq}(t')$. Therefore, $(c, t, \bar{T}) \in P$ and $t \in \text{SN}^B(P)$.

Lemma 7 For every sort $B$, $\mathcal{S}^B$ is monotone.

Proof
We prove that, for all $(a, b, c)$, if $b \leq c \leq a$, then (1) $\mathcal{S}^B_b \subseteq \mathcal{S}^B_c$, hence $\mathcal{S}^B_a |_{a}$ is monotone, (2) $\mathcal{S}^B_c \subseteq \mathcal{S}^B_a$, and (3) $\mathcal{S}^B_a \subseteq \mathcal{S}^B_{a+1}$, by induction on $a$. There are three cases:

- $a = 0$. Then, $b = c = 0$ and (1) and (2) hold trivially. We now prove (3). Let $t \in \mathcal{S}^B_0$. We prove that $t \in \mathcal{S}^B_1$:

  - $t \in \text{SN}$ since, by definition, $\mathcal{S}^B_0 \subseteq \text{SN}$.

  Let now $(c, \bar{T}) \in \mathcal{C}^B_{\leq}(t)$.

  - We have to prove that, for all $k \in \{1, \ldots, p^c\}$, $t_k \in [B : \mathcal{S}^B_0]T_k$. Since $t \in \mathcal{S}^B_0$, we have $p^c = 0$. Therefore, the property holds since there is no $k \in \{1, \ldots, p^c\}$.

  - We have to prove that, for all $k \in \{p^c + 1, \ldots, q^c\}$, $t_k \in T_k$. This holds since $t \in \mathcal{S}^B_0$.

  - Finally, we have to prove that $\Sigma^c(o_{\mathcal{S}^B}(\bar{t})) \leq 1$. This holds since $t \in \mathcal{S}^B_0$ and thus $\Sigma(\mathcal{C}(o_{\mathcal{S}^B}(\bar{t}))) \leq 0$.

- $a = a' + 1$.

  1. If $c \leq a'$, then (1) holds by induction hypothesis (1) on $(a', b, c)$. Otherwise, $c = a' + 1$. If $b = c$, then (1) holds trivially. Otherwise, $b \leq a'$ and (1) holds by induction hypothesis (1) and (3) on $(a', b, a')$, and transitivity of $\leq$.

  2. If $c \leq a'$, then (2) holds by induction hypothesis (2) and (3) on $(a', b, c)$, and transitivity of $\leq$. Otherwise, (2) holds trivially.
3. Let \( t \in \mathcal{S}^{B}_{a+1} \). We prove that \( t \in \mathcal{S}^{B}_{a+2} \):

- \( t \in SN \) since, by definition, \( \mathcal{S}^{B}_{a+1} \subseteq SN \).

Let now \((c, i, \bar{T}) \in \mathcal{C}^{B}_{\kappa}(t)\) and \( k \in \{1, \ldots, \bar{q}^{c}\} \).

- Assume that \( k \leq p^{c} \). Since \( t \in \mathcal{S}^{B}_{a+1} \), we have \( t_{k} \in [B : \mathcal{S}^{B}_{a}]T_{k} \). Therefore, \( t_{k} \in [B : \mathcal{S}^{B}_{a+1}]T_{k} \) since \( B \) occurs only positively in \( T_{k} \) and \( \mathcal{S}^{B}_{a} \subseteq \mathcal{S}^{B}_{a+1} \) by induction hypothesis (3) on \((a', a', a')\).

- Assume that \( k > p^{c} \). Then, \( t_{k} \in T_{k} \) since \( t \in \mathcal{S}^{B}_{a+1} \).

Since \( t \in \mathcal{S}^{B}_{a+1} \), we have \( \Sigma^{c}(o_{\mathcal{S}^{B}}(i)) \leq a' + 1 \). Therefore, \( \Sigma^{c}(o_{\mathcal{S}^{B}}(i)) \leq a' + 2 \).

- \( a \) is an infinite limit ordinal. Then, \( \mathcal{S}^{B}_{a} = \text{lub}\{\mathcal{S}^{B}_{b} \mid b < a\} \).

1. If \( c < a \), then (1) follows by induction hypothesis (1) on \((c, b, c)\). Otherwise, \( c = a \). If \( b = c \), then (1) holds trivially. Otherwise, \( b < c \) and (1) holds by definition of lub.

2. (2) holds by definition of lub.

3. Let \( t \in \mathcal{S}^{B}_{a} \). We have to prove that \( t \in \mathcal{S}^{B}_{a+1} \).

After (1), \( \mathcal{S}^{B}_{a} \mid a \) is monotone. Therefore, by Lemma 1, \( \mathcal{S}^{B}_{a} = \bigcup_{b \ (< a)} \{\mathcal{S}^{B}_{b} \mid b < a\} \) and \( t \in \mathcal{S}^{B}_{b} \) for some \( b < a \). Now, since \( a \) is a limit ordinal, \( b + 1 < a \). Therefore, by induction hypothesis (2) on \((b + 1, b, b)\), \( \mathcal{S}^{B}_{b} \subseteq \mathcal{S}^{B}_{b+1} \) and \( t \in \mathcal{S}^{B}_{b+1} \). We now prove that \( t \in \mathcal{S}^{B}_{a+1} \):

- \( t \in SN \) since \( \mathcal{S}^{B}_{a} \) is a computability predicate.

Let now \((c, i, \bar{T}) \in \mathcal{C}^{B}_{\kappa}(t)\) and \( k \in \{1, \ldots, \bar{q}^{c}\} \).

- Assume that \( k \leq p^{c} \). Since \( t \in \mathcal{S}^{B}_{b+1} \), we have \( t_{k} \in [B : \mathcal{S}^{B}_{b}]T_{k} \). Therefore, \( t_{k} \in [B : \mathcal{S}^{B}_{a+1}]T_{k} \) since \( B \) occurs only positively in \( T_{k} \) and \( \mathcal{S}^{B}_{b} \subseteq \mathcal{S}^{B}_{a+1} \).

- Assume that \( k > p^{c} \). Then, \( t_{k} \in T_{k} \) since \( t \in \mathcal{S}^{B}_{b+1} \).

Since \( t \in \mathcal{S}^{B}_{b+1} \), we have \( \Sigma^{c}(o_{\mathcal{S}^{B}}(i)) \leq b + 1 \). Therefore, \( \Sigma^{c}(o_{\mathcal{S}^{B}}(i)) \leq a + 1 \).

\( \square \)

**Lemma 8** For every sort \( B \), \( \mathcal{S}^{B}_{h} = B \).

**Proof**

By Lemma 5, \( \mathcal{S}^{B}_{a} \subseteq B \). Now, since \( B = \mathcal{S}^{B}_{h} \), where \( \mathcal{S}^{B} \) is the default stratification, it suffices to prove that, for all \( a, \mathcal{S}^{B}_{a} \subseteq \mathcal{S}^{B}_{h} \), that is, for all \( a \), there is \( b < h \) such that \( \mathcal{S}^{B}_{a} \subseteq \mathcal{S}^{B}_{b} \). We proceed by induction on \(<_{S} \) and \( a \).

- \( \mathcal{S}^{B}_{0} = \perp \subseteq \mathcal{S}^{B}_{0} \).

- Let \( a \) be an infinite limit ordinal smaller than \( h \). By induction hypothesis, for all \( b < a \), \( \mathcal{S}^{B}_{b} \subseteq \mathcal{S}^{B}_{h} \). Therefore, \( \mathcal{S}^{B}_{a} = \text{lub}\{\mathcal{S}^{B}_{b} \mid b < a\} \subseteq \mathcal{S}^{B}_{h} \).

- Let now \( a + 1 < h \). By induction hypothesis, \( \mathcal{S}^{B}_{a} \subseteq \mathcal{S}^{B}_{h} \).

Since \( h \) is a successor cardinal, it is regular, that is, it is equal to its cofinality. And since it is uncountable, it is \( \omega \)-complete, that is, every countable subset of \( h \) has a least upper bound in \( h \).

Let \( c = \text{sup}(X) \) where \( X = \{o_{\mathcal{S}^{B}}(t) \mid t \in \mathcal{S}^{B}_{a}\} \). Since \( \left| X \right| \leq \left| \mathcal{S}^{B}_{a} \right| \leq \left| \mathcal{I} \right| \leq \omega \), we have \( c < h \) and \( \mathcal{S}^{B}_{a} \subseteq \mathcal{S}^{B}_{c} \).
Let now \( \delta = \text{sup}(X \cup Y) \) where \( Y \) is the set of the ordinals \( \Sigma^c(o_{\mathcal{B}}(\vec{t})) \) such that there are \( t \in \mathcal{D}_0^\mathcal{B} \) and \( (c, \vec{t}, \vec{T}) \in \mathcal{C}_{c-a}^\mathcal{B}(t) \). Since \( |Y| \leq \omega \) (\( \mathcal{D}_0^\mathcal{B} \subseteq \text{SN} \) and \( \rightarrow \) is finitely branching), we have \( \text{sup}(Y) < \delta \) and thus \( \delta < \eta \). Since \( \eta \) is a limit ordinal, \( \delta + 1 < \eta \).

We now prove that \( \mathcal{D}_{\alpha+1}^\mathcal{B} \subseteq \mathcal{D}_\delta^\mathcal{B} \). Let \( t \in \mathcal{D}_{\alpha+1}^\mathcal{B} \). Then, \( t \in \text{SN} \). Let now \( (c, \vec{t}, \vec{T}) \in \mathcal{C}_{c-a}^\mathcal{B}(t) \) and \( k \in \{1, \ldots, \eta^2\} \). If \( k > \eta^2 \), then \( t_k \in T_k \). Otherwise, \( t_k \in [B : \mathcal{D}_a^\mathcal{B}] T_k \). Since \( \mathcal{B} \) occurs only positively in \( T_k \), we have \( [B : \mathcal{D}_a^\mathcal{B}] T_k \subseteq [B : \mathcal{D}_a^\mathcal{B}] T_k \). Since \( c \leq \delta \) and \( \mathcal{D}^\mathcal{B} \) is monotone by Lemma 7, we have \( [B : \mathcal{D}_a^\mathcal{B}] T_k \subseteq [B : \mathcal{D}_a^\mathcal{B}] T_k \). Finally, \( \Sigma^c(o_{\mathcal{B}}(\vec{t})) \leq \delta \). Therefore, \( t \in \mathcal{D}_\delta^\mathcal{B} \).

This ends the proof that \( \mathcal{D}^\mathcal{B} \) is a stratification of \( \mathcal{B} \). We now see some of its properties:

**Lemma 9**
\[
\begin{align*}
\bullet & \quad t \in \mathcal{D}_0^\mathcal{B} \text{ iff } t \in B \text{ and, for all } (c, \vec{t}, \vec{T}) \in \mathcal{C}_{c-a}^\mathcal{B}(t), \Sigma^c(o_{\mathcal{B}}(\vec{t})) = \eta^2 = 0. \\
\bullet & \quad t \in \mathcal{D}_{\alpha+1}^\mathcal{B} \text{ iff } t \in B \text{ and, for all } (c, \vec{t}, \vec{T}) \in \mathcal{C}_{c-a}^\mathcal{B}(t), \Sigma^c(o_{\mathcal{B}}(\vec{t})) \leq \alpha + 1 \text{ and, for all } k \in \{1, \ldots, \eta^2\}, \ o_{\mathcal{B}}(t_k) \leq \alpha. 
\end{align*}
\]

**Proof**
- Immediate.
- Assume that \( t \in \mathcal{D}_{\alpha+1}^\mathcal{B} \). Then, \( t \in B \). Assume moreover that \( (c, \vec{t}, \vec{T}) \in \mathcal{C}_{c-a}^\mathcal{B}(t) \). Then, \( \Sigma^c(o_{\mathcal{B}}(\vec{t})) \leq \alpha + 1 \) and, for all \( k \in \{1, \ldots, \eta^2\}, t_k \in [B : \mathcal{D}_a^\mathcal{B}] T_k = \mathcal{D}_a^\mathcal{B} \). Hence, \( o_{\mathcal{B}}(t_k) \leq \alpha \). Conversely, if \( o_{\mathcal{B}}(t_k) \leq \alpha \), then \( t_k \in [B : \mathcal{D}_a^\mathcal{B}] T_k \).

**Lemma 10** If \( (c, \vec{t}, \vec{T}) \in \mathcal{C}^\mathcal{B} \) and \( c \vec{t} \in B \), then
\[
\begin{align*}
\bullet & \quad o_{\mathcal{B}}(c \vec{t}) \geq \Sigma^c(o_{\mathcal{B}}(\vec{t})), \\
\bullet & \quad o_{\mathcal{B}}(c \vec{t}) > o_{\mathcal{B}}(t_k) \text{ for all } k \in \{1, \ldots, \eta^2\}. 
\end{align*}
\]

**Proof**
Let \( a = o_{\mathcal{B}}(c \vec{t}) \). Since \( \mathcal{D}^\mathcal{B} \) is continuous, by Lemma 3, either \( a = 0 \) or \( a = b + 1 \) for some \( b \).

- If \( a = 0 \), then \( c \vec{t} \in \mathcal{D}_0^\mathcal{B} \) and \( \Sigma^c(o_{\mathcal{B}}(\vec{t})) \leq \alpha \) by definition of \( \mathcal{D}_0^\mathcal{B} \). Otherwise, \( c \vec{t} \in \mathcal{D}_{b+1}^\mathcal{B} \) and \( \Sigma^c(o_{\mathcal{B}}(\vec{t})) \leq \alpha \) by definition of \( \mathcal{D}_{b+1}^\mathcal{B} \).
- If \( a = 0 \), then \( c \vec{t} \in \mathcal{D}_0^\mathcal{B} \) and \( \eta^2 = 0 \). So, there is no \( k \in \{1, \ldots, \eta^2\} \). Otherwise, \( c \vec{t} \in \mathcal{D}_{b+1}^\mathcal{B} \) and \( t_k \in [B : \mathcal{D}_b^\mathcal{B}] T_k \). Thus, \( o_{\mathcal{B}}(t_k) \leq b < a \).

**Lemma 11** If \( t \in B \), then \( o_{\mathcal{B}}(t) = \delta \sup(R \cup S \cup T) \), where
\[
\begin{align*}
\bullet & \quad \delta a = a + 1 \text{ if } a \text{ is an infinite limit ordinal, and } \delta a = a \text{ otherwise}; \\
\bullet & \quad R = \{ o_{\mathcal{B}}(t') | t \rightarrow t' \}; \\
\bullet & \quad S = \{ o_{\mathcal{B}}(t_k) + 1 | (c, \vec{t}, \vec{T}) \in \mathcal{C}_B, t = c \vec{t}, k \in \{1, \ldots, \eta^2\} \}; \\
\bullet & \quad T = \{ \Sigma^c(o_{\mathcal{B}}(\vec{t})) | (c, \vec{t}, \vec{T}) \in \mathcal{C}_B, t = c \vec{t} \}.
\end{align*}
\]
Let $a = \sup(R \cup S \cup T)$ and $b = o_{\prod}(t)$.

We first prove that $b \geq \delta a$. Let $t'$ such that $t \to t'$. Then, $b \geq o_{\prod}(t')$ by Lemma 3.

Assume now that $(c, \hat{t}, \hat{T}) \in \mathbf{C}^{B}$ and $t = c \hat{t}$. By Lemma 10, $b \geq \Sigma^c(o_{\prod}(\hat{t}))$ and, for all $k \in \{1, \ldots, p^c\}$, $b > o_{\prod}(t_k)$. Therefore, $b \geq a$.

Since $\mathcal{I}^{B}$ is continuous, $b$ cannot be an infinite limit ordinal. So, if $a$ is an infinite limit ordinal, then $b > a$ and $b \geq a + 1 = \delta a$. Otherwise, $\delta a = a$ and $b \geq \delta a$.

Now, to have $b \leq \delta a$, we prove that $t \in \mathcal{I}^{B}_{\delta a}$ using Lemma 9:

• Case $\delta a = 0$. Then, $a = 0$. Let $(c, \hat{i}, \hat{T}) \in \mathbf{C}^{\mathbf{i}}(t)$.
  
  — Case $t = c \hat{i}$. Then, $S = \emptyset$, $p^c = 0$ and $\Sigma^c(o_{\prod}(\hat{i})) = 0$. Therefore, $t \in \mathcal{I}^{B}_{0}$.
  
  — Case $t \to t' \to^* c \hat{i}$. Then, $o_{\prod}(t') = 0$. So, $p^c = 0$, $\Sigma^c(o_{\prod}(\hat{i})) = 0$ and $t \in \mathcal{I}^{B}_{0}$.

• Case $\delta a = a' + 1$. Let $(c, \hat{i}, \hat{T}) \in \mathbf{C}^{\mathbf{i}}(t)$.
  
  — Case $t = c \hat{i}$. First, $\Sigma^c(o_{\prod}(\hat{i})) \leq \sup(T) \leq a \leq \delta a = a' + 1$. Second, if $k \in \{1, \ldots, p^c\}$, then $o_{\prod}(t_k) < o_{\prod}(t_k) + 1 \leq \sup(S) \leq a \leq \delta a$. Therefore, $o_{\prod}(t_k) \leq a' + 1$ and $t \in \mathcal{I}^{\mathbf{i}}_{\delta a}$.
  
  — Case $t \to t' \to^* c \hat{i}$. First, $\Sigma^c(o_{\prod}(\hat{i})) \leq a' + 1$ since, by Lemma 10, $\Sigma^c(o_{\prod}(\hat{i})) \leq o_{\prod}(c \hat{i})$ and, by Lemma 3, $o_{\prod}(c \hat{i}) \leq o_{\prod}(t') \leq \sup(S) \leq a \leq \delta a$. Second, if $k \in \{1, \ldots, p^c\}$, then $o_{\prod}(t_k) \leq a' + 1$ since, by Lemma 10, $o_{\prod}(t_k) < o_{\prod}(c \hat{i})$ and, by Lemma 3, $o_{\prod}(c \hat{i}) \leq o_{\prod}(t') \leq \sup(R) \leq a \leq \delta a = a' + 1$. So, $t \in \mathcal{I}^{\mathbf{i}}_{\delta a}$.

Note that taking $\Sigma^c(\tilde{a}) \leq \sup\{a_k + 1 \mid k \in \{1, \ldots, p^c\}\}$ gives the same notion of size as taking $\Sigma^c(\tilde{a}) = 0$. On the other hand, if $\Sigma^c(\tilde{a}) \geq \sup\{a_k + 1 \mid k \in \{1, \ldots, p^c\}\}$, then $\Sigma^c$ gives the size of irreducible terms of the form $c \hat{i}$:

**Corollary 1** Assume that $\Sigma^c$ is strictly extensive wrt recursive arguments (i.e. $a_k < \Sigma^c(\tilde{a})$ if $k \in \{1, \ldots, p^c\}$) and $\Sigma^c(\tilde{a})$ is never an infinite limit ordinal. Then, for all $(c, \hat{i}, \hat{T}) \in \mathbf{C}^{D}$ such that $c \hat{i} \in \mathbf{B}$ and $c \hat{i}$ is irreducible, we have $o_{\prod}(c \hat{i}) = \Sigma^c(o_{\prod}(\hat{i}))$.

**Proof**

Since $c \hat{i}$ is irreducible, $R = \emptyset$. Let $a = \Sigma^c(o_{\prod}(\hat{i}))$. Since $a > o_{\prod}(t_k)$ whenever $k \in \{1, \ldots, p^c\}$, $o_{\prod}(c \hat{i}) = \delta a$. Since $a$ is not an infinite limit, $\delta a = a$. 

**Corollary 2** Assume that $\Sigma^c$ is monotone wrt every argument, strictly extensive wrt recursive arguments and never returns an infinite limit ordinal. Then, for all $(c, \hat{i}, \hat{T}) \in \mathbf{C}^{D}$ with $c \hat{i} \in \mathbf{B}$, we have $o_{\prod}(c \hat{i}) = \Sigma^c(o_{\prod}(\hat{i}))$.

**Proof**

We proceed by induction on $\hat{i}$ with $\to_{\prod}$ as well-founded relation. Assume that $c \hat{i} \to u$. Then, there are $\hat{u}$ such that $u = c \hat{u}$ and $\hat{i} \to_{\prod} \hat{u}$. Hence, $o_{\prod}(\hat{u}) \leq_{\prod} o_{\prod}(\hat{i})$ and, by induction hypothesis, $o_{\prod}(c \hat{u}) = \Sigma^c(o_{\prod}(\hat{u}))$. So, $o_{\prod}(c \hat{i}) \leq \Sigma^c(o_{\prod}(\hat{i}))$ since $\Sigma^c$ is monotone. Therefore, $o_{\prod}(c \hat{i}) = \Sigma^c(o_{\prod}(\hat{i}))$. 


Finally, we are going to prove that, if $\rightarrow$ is locally confluent, hence confluent on strongly normalizing terms (Newman, 1942), then the size of a term is equal to the size of its normal form when its type is a strictly positive sort:

**Definition 9 (Strictly positive sorts)** A sort $B$ is strictly positive if, for every constructor $c : \hat{T} \Rightarrow B$ and argument $k \in \{1, \ldots, q^c\}$, $T_k$ is positive wrt $B$ and either $T_k$ is a strictly positive sort\(^4\) $C <_S B$ or $T_k$ is of the form $\vec{U} \Rightarrow B$ with $\text{Pos}(B, \vec{U}) = \emptyset$.

Examples of strictly positive sorts are Peano numbers and Howard constructive ordinals.

**Lemma 12** Assume that $\rightarrow$ is locally confluent and, for every constructor $c$, $\Sigma^c$ is monotone wrt every argument, strictly extensive wrt recursive arguments and never returns an infinite ordinal. Then, for every strictly positive sort $B$ and term $t \in B$, $o_{\mathcal{R}}(t) = o_{\mathcal{R}}(t\downarrow)$, where $t\downarrow$ is the normal form of $t$.

**Proof**

First note that $o_{\mathcal{R}}(t\downarrow) \leq o_{\mathcal{R}}(t)$ since $t \rightarrow^* t\downarrow$. We now prove that, for all strictly positive $B$, for all $t \in B$, $o_{\mathcal{R}}(t) \leq o_{\mathcal{R}}(t\downarrow)$, hence that $o_{\mathcal{R}}(t) = o_{\mathcal{R}}(t\downarrow)$, by induction on $(B, o_{\mathcal{R}}(t), t)$ with $(<_S, <, \leftarrow)$ as well-founded relation. By Lemma 11, $o_{\mathcal{R}}(t) = \delta \sup(R \cup S \cup T)$. Since $\Sigma^c$ is strictly extensive, $o_{\mathcal{R}}(t) = \delta \sup(R \cup T)$. Since $o_{\mathcal{R}}(t\downarrow)$ cannot be an infinite ordinal, it is sufficient to prove that $\sup(R \cup T) \leq o_{\mathcal{R}}(t\downarrow)$.

Assume that $t \rightarrow u$. Then, $o_{\mathcal{R}}(u) \leq o_{\mathcal{R}}(t)$. Hence, by induction hypothesis on the second or third component, $o_{\mathcal{R}}(u) \leq o_{\mathcal{R}}(u\downarrow) = o_{\mathcal{R}}(t\downarrow)$.

Assume now that $(c, \vec{t}, \vec{T}) \in \mathcal{C}^B$ and $t = \vec{c}\vec{t}$. By Corollary 2, $o_{\mathcal{R}}(t) = \Sigma^c(o_{\mathcal{R}}(\vec{t}))$ and $o_{\mathcal{R}}(t\downarrow) = \Sigma^c(o_{\mathcal{R}}(\vec{t}\downarrow))$. Since $\Sigma^c$ is monotone, it suffices to prove that, for all $k \in \{1, \ldots, q^c\}$, $o_{\mathcal{R}^c}(t_k) \leq o_{\mathcal{R}^c}(t_k\downarrow)$. Since $B$ is strictly positive, there are two cases:

- $T_k$ is a strictly positive sort $C <_S B$. Then, by induction hypothesis on the first component, $o_{\mathcal{R}^c}(t_k) = o_{\mathcal{R}^c}(t_k) \leq o_{\mathcal{R}^c}(t_k\downarrow) = o_{\mathcal{R}^c}(t_k\downarrow)$.
- There is $\vec{U}$ such that $T_k = \vec{U} \Rightarrow B$ and $\text{Pos}(B, \vec{U}) = \emptyset$. Then, by Lemma 4, $o_{\mathcal{R}^c}(t_k) = \sup\{o_{\mathcal{R}}(t_k\vec{u}) \mid \vec{u} \in \vec{U}\}$ and $o_{\mathcal{R}^c}(t_k\downarrow) = \sup\{o_{\mathcal{R}}(t_k\downarrow\vec{u}) \mid \vec{u} \in \vec{U}\}$. Let $\vec{u} \in \vec{U}$. Since $o_{\mathcal{R}}(t_k\downarrow\vec{u}) \leq o_{\mathcal{R}}(t_k\vec{u}) < o_{\mathcal{R}}(t)$, by induction hypothesis on the second component, $o_{\mathcal{R}}(t_k\vec{u}) = o_{\mathcal{R}}(t_k\downarrow\vec{u})$. Therefore, $o_{\mathcal{R}^c}(t_k) = o_{\mathcal{R}^c}(t_k\downarrow)$.

We end this section by introducing the reflexive and transitive closure of the notion of accessible argument (Definition 4) and prove some properties about it. In order to keep track of the sort with respect to which the size is measured, we consider a relation on triples $(t, T, B)$ made of a term $t$, its type $T$ and the sort $B$ used to measure the size of $t$ in $[B : \mathcal{R}^B]T$.

\(^4\) This is a restriction wrt the definition given in Coquand & Paulin-Mohring (1988), where $T_k$ can be any type where $B$ does not occur.
Definition 10 (Accessible subterm) We say that \((u, U, C)\) is accessible in \((t, T, B)\), written \((u, U, C) \preceq_a (t, T, B)\), if \((u, U, C) = (t, T, B)\) or there are \((c, \vec{t}, \vec{T}) \in C^B\) and \(k \in \{1, \ldots, q^c\}\) such that \(t = c \vec{t}, T = B\) and \((u, U, C) \preceq_a (t_k, T_k, B_k^c)\), where \(B_k^c = B\) if \(k \leq p^c\), and \(B_k^c\) is given by the size function of \(c\) if \(k > p^c\) (see Definition 7).

For example,
- \((x, N, N)\) is accessible in \((s x, N, N)\) if \(s : N \Rightarrow N\);
- \((f, N \Rightarrow O, O)\) is accessible in \((\lim f, O, O)\) if \(\lim : (N \Rightarrow O) \Rightarrow O\);
- \((x, N, N)\) is accessible in \((\text{pair}(s x), P, P)\) if \(\text{pair} : (N \Rightarrow N) \Rightarrow P\) and \(s : N \Rightarrow N\);
- \((x, B \Rightarrow C, B)\) is not accessible in \((c x, B, B)\) if \(c : (B \Rightarrow C) \Rightarrow B\), because \(B\) occurs negatively in \(B \Rightarrow C\) and thus \(q^c = 0\).

Note that \(\preceq_a\) is stable by substitution, and that \(C\) occurs only positively in \(U\) whenever \((u, U, C) \preceq_a (t, T, B)\), where \(\preceq_a\) is the strict part of \(\preceq_a\).

Lemma 13 If \((u, U, C) \preceq_a (t, T, B)\) and \(t \in T\), then \(u \in U\).

Proof
We proceed by induction on \(\preceq_a\). If \((u, U, C) = (t, T, B)\), this is immediate. Otherwise, there are \((c, \vec{t}, \vec{T}) \in C^B\) and \(k \in \{1, \ldots, q^c\}\) such that \(t = c \vec{t}, T = B\) and \((u, U, C) \preceq_a (t_k, T_k, B_k^c)\). By definition of \(\tilde{\Pi}(B)\), we have \(t_k \in T_k\). So, by induction hypothesis, \(u \in U\).

4 Termination criterion

In this section, we describe a termination criterion that capitalizes on the fact that some terms can be assigned an ordinal size. The idea is simple: if for every rewrite step \(fl \to r\) and every function call \(gm\) in \(r\), the size of \(m\) is strictly smaller than the size of \(l\), then there cannot be any infinite reduction.

The idea, dating back to Hughes et al. (1996), consists of introducing symbolic expressions representing ordinals and logical rules for deducing information about the size of terms, namely, that it is bounded by some expression. Hence, termination is reduced to checking the decreasingness of symbolic size expressions.

Following these authors, we replace every sort \(B\) by a pair \((B, a)\), written \(B_a\), where \(a\) is a symbolic expression from an algebra interpretable in ordinals, so that a term is of size-annotated type \(B_a\) if it is of type \(B\) and of size \textit{smaller than or equal to} the interpretation of \(a\). The typing rules of Figure 2 are then easily turned into valid deduction rules on size annotations. Moreover, the monotony of stratifications naturally induces a notion of subtyping on size-annotated types: a term of type \(B_a\) is also of type \(B_b\) if \(a \leq b\).

4.1 Size-annotated types

In the previously mentioned works, only two particular algebras have been considered so far. First, the successor algebra (Definition 12). Second, when \(h\) is restricted to \(\omega\) (e.g. when inductive types are restricted to first-order data types), the
algebra of Presburger arithmetic generated from the symbols 0, s and + interpreted
by zero, the successor function and the addition on natural numbers respectively,
the first-order theory of which is decidable (Presburger, 1929).

Other algebras are however interesting as we shall see in some examples. For
instance, the max-successor algebra, that is, the successor algebra extended by a max
operator, and the max-plus algebra, that is, the algebra generated by the symbols 0,
1, + and max.

So, in the following, we consider an arbitrary size algebra and prove general results
under some conditions on it. Then, in Section 9, we prove that these conditions are
in particular satisfied by the successor algebra.

**Definition 11 (Size algebra)** A size algebra is given by

- a first-order term algebra \( A \) built from a set \( V \) of size variables \( \alpha, \beta, ... \) and a
  set \( F \) of size function symbols \( f, g, ... \) of fixed arity, disjoint from \( V \);
- a quasi-order \( \leq_A \) on \( A \) stable by substitution: \( a \varphi \leq_A b \varphi \) whenever \( a \leq_A b \) and
  \( \varphi : V \rightarrow A \);
- a strict order \( <_A \subseteq \leq_A \) also stable: \( a \varphi <_A b \varphi \) whenever \( a <_A b \) and
  \( \varphi : V \rightarrow A \);
- for each size function symbol \( f \in F \) of arity \( n \), a function \( f_h : h^n \rightarrow h \) so that,
  for every valuation \( \mu : V \rightarrow h \), \( a\mu \leq b\mu \) (\( a\mu < b\mu \) resp.) whenever \( a <_A b \) (\( a <_A b \)
  resp.) where, as usual, \( \alpha\mu = \mu(\alpha) \) and \( (fa_1...a_n)\mu = f_h(a_1\mu,...,a_n\mu) \).

A size algebra is **monotone** if every size function symbol is monotone wrt \( \leq_A \)
in every argument, that is, \( f\hat{a} \leq_A f\hat{b} \) whenever \( \hat{a} \leq_A \hat{b} \). Given a size substitution \( \varphi \) and
a set \( V \) of variables, let \( \varphi|_V = \{ (\alpha, \alpha\varphi) | \alpha \in V \} \).

Let \( a \leq_{ext} b \) (\( a <_{ext} b \) resp.) iff, for all \( \mu \), \( a\mu \leq b\mu \) (\( a\mu < b\mu \) resp.). Note that
\( (\leq_{ext}, <_{ext}) \) satisfies the above conditions and, for every pair of relations (\( \leq_A, <_A \))
satisfying the above conditions, we have \( \leq_A \subseteq \leq_{ext} \) and \( <_A \subseteq <_{ext} \). So, one can
always take \( \leq_{ext} \) (\( <_{ext} \) resp.) for \( \leq_A \) (\( <_A \) resp.).

As remarked in Giesl et al. (2002), the strict part of a stable quasi-order \( \leq_A \), that
is \( \leq_A = \leq_A - >_A \), is not necessarily stable. On the other hand, its stable-strict part
\( <_A \) is stable, where \( a <_A b \) iff, for all closed substitution \( \varphi \), \( a\varphi \leq_A b\varphi \).

The simplest size algebra is:

**Definition 12 (Successor algebra)** The **successor** size algebra is obtained by taking

- \( F = C \cup \{ s \} \) where \( C \) is an infinite set of constants and \( s \) a unary symbol
  interpreted by the successor function\(^5\);
- \( <_A \) is the smallest strict ordering on \( A \) such that, for all \( a, a <_A s a \);
- \( \leq_A \) is the reflexive closure of \( <_A \).

Although this algebra may seem overly simple, it is already sufficient to overtake
the Coq termination checker (see Section 6 for various examples using it). We will
study the properties of this algebra in Section 9.

\(^5\) \( h \) is closed by successor since it is a limit ordinal.
Definition 13 (Size-annotated types) The set $\mathbb{T}_\Lambda$ of annotated types is defined as follows:

- If $T$ is a type, then $T \in \mathbb{T}_\Lambda$.
- If $B$ is a sort and $a$ a size expression, then $B_a \in \mathbb{T}_\Lambda$.
- If $U$ and $V$ belong to $\mathbb{T}_\Lambda$, then $U \Rightarrow V \in \mathbb{T}_\Lambda$.

Let $\text{Var}(T)$ be the set of size variables occurring in $T$.

Given an annotated type $T$, let $|T|$ be the type obtained by removing every annotation.

Given a sort $B$, a size expression $a$ and a type $T$, let $\text{Annot}(T, B, a)$ be the annotated type obtained by annotating in $T$ every occurrence of $B$ by $a$.

The definition of positive ($s = +$) and negative ($s = -$) positions in a type (Definition 3) is extended to annotated types as follows:

- $\text{Pos}^+(B_b) = \{ip \mid p \in \text{Pos}^+(b)\}$.
- $\text{Pos}^+(\alpha) = \{\varepsilon \mid s = +\}$.
- $\text{Pos}^+(f) = \{\varepsilon \mid s = +\}$ if $f$ is of arity 0.
- $\text{Pos}^+(f b_1 \ldots b_n) = \{ip \mid i \in \text{Mon}^+(f), p \in \text{Pos}^+(b_i)\} \cup \{ip \mid i \in \text{Mon}^-(f), p \in \text{Pos}^+(b_i)\}$ if $f$ is of arity $n > 0$,

where $\text{Mon}^+(f)$ (resp. $\text{Mon}^-(f)$) is the set of arguments in which $f$ is monotone (anti-monotone resp.) wrt $\preceq_\Lambda$.

In order to combine terms with annotated and unannotated types, we extend $\Lambda$ by a greatest element $\infty$ and identify $B_{\infty}$ with $B$:

Definition 14 (Top-extension of a size algebra) The top-extension of a size algebra $\Lambda$ is a set $\overline{\Lambda} = \Lambda \cup \{\infty\}$ with $\infty \not\in \Lambda$. Given $B \in \mathbb{S}$, let $B_{\infty} = B$ (we identify $B_{\infty}$ and $B$). Given size expressions $a, b \in \overline{\Lambda}$, let $a \leq_{\overline{\Lambda}} b$ if $a \leq_\Lambda b$ or $b = \infty$. Given $\varphi : V \rightarrow \overline{\Lambda}$, let $aq = \infty$ if $a$ contains a variable $x$ such that $\varphi(x) = \infty$, and $a\varphi$ be the usual substitution otherwise. Terms distinct from $\infty$ are called finite.

We now propose to users a syntactic way to specify their own notions of size through the annotation of constructor types. We assume that every constructor type is annotated in a way that allows us to define a size function, hence a stratification for every sort, and thus an interpretation of every annotated type in computability predicates. To this end, we use notations similar to the ones of Definition 7:

Definition 15 (Annotated types of constructors) We assume that every $c \in C$ with $\Theta(c) = T_1 \Rightarrow \ldots \Rightarrow T_{c^e} \Rightarrow B$ is equipped with an annotated type $\overline{\Theta}(c) = \overline{T}_1 \Rightarrow \ldots \Rightarrow \overline{T}_{c^e} \Rightarrow B_{c^e}$ with

- for all $i \in \{1, \ldots, q^c\}$, $\overline{T}_i = \text{Annot}(T_i, B^c_i, x^{c^e}_i)$;
- for all $i \in \{q^c + 1, \ldots, r^c\}$, $\overline{T}_i = T_i$;
- $x^{c^e}_1, \ldots, x^{c^e}_{r^c} \in V$;
- $x^{c^e}_{r^c + 1}, \ldots, x^{c^e}_{q^c} \in V \cup \{\infty\}$;
- the variables of $\{x^{c^e}_1, \ldots, x^{c^e}_{q^c}\}$ are either pairwise equal or pairwise distinct.

https://doi.org/10.1017/50956796818000072 Published online by Cambridge University Press
for all \( i \in \{1, \ldots, p^c\} \), \( B^c_i = B \);
• for all \( i \in \{p^c + 1, \ldots, q^c\} \) with \( \alpha^c_i \in V \), \( B^c_i \) occurs in \( T_i \);
• for all \( i \in \{p^c + 1, \ldots, q^c\} \) with \( \alpha^c_i \in V \), \( \text{Pos}(B^c_i, T_i) \subseteq \text{Pos}^+(T_i) \);
• \( \sigma^c \in \mathbb{K} \);
• for all \( i \in \{1, \ldots, q^c\} \) with \( \alpha^c_i \in V \), \( \text{Pos}(B^c_i, T_i) \subseteq \text{Pos}^+(T_i) \);
• \( \sigma^c \in \mathbb{A} \);
• for all \( i \in \{1, \ldots, q^c\} \), \( \text{Pos}(\alpha^c_i, \sigma^c) \subseteq \text{Pos}^+(\sigma^c) \) (\( \sigma^c \) is monotone wrt every \( \alpha^c_i \));
• for all \( i \in \{1, \ldots, p^c\} \), \( \alpha^c_i <_A \sigma^c \) (\( \sigma^c \) is strictly extensive wrt recursive arguments).

The semantics of these annotations is given by the next definition. The intuition is that the size of a term of the form \( c \vec{t} \) will be given by the interpretation in ordinals of \( \sigma^c \) with each \( \alpha^c_i \), the abstract size of the \( i \)th argument of \( c \), interpreted by the actual size of \( t_i \) in \( [B^c_i : B^c]T_i \).

We now extend the interpretation of types in computability predicates to annotated types, by defining a size function \( \Sigma^c \) for each constructor \( c \):

**Definition 16 (Interpretation of size-annotated types)** First, for each constructor \( c \) with \( \Theta(c) \) as in Definition 15, we define a size function \( \Sigma^c \) (see Definition 7) as follows:

\[
\Sigma^c(\vec{a}) = \begin{cases} 
0 & \text{if } \sigma^c = \infty \\
\sigma^c v \text{ otherwise where } v(x) = \begin{cases} 
 a_i & \text{if } x = \alpha^c_i \text{ and all the } \alpha^c_i \in V \text{ are distinct} \\
 \sup \{a_i \mid i \in \{1, \ldots, q^c\}, \alpha^c_i \in V \} & \text{otherwise}
\end{cases}
\end{cases}
\]

Then, given a valuation \( \mu : V \to h \), we interpret annotated types as follows:

• \( B^c \mu = B \).
• \( B^c a \mu = \mathcal{S}_{\mu}^c \) if \( a \in A \), where \( \mathcal{S} \) is the stratification defined by \( \Sigma \) (see Definition 8).
• \( (U \Rightarrow V) \mu = U \mu \Rightarrow V \mu \).

Note that \( \Sigma^c \) is monotone wrt every argument and strictly extensive wrt recursive arguments since \( \sigma^c \) so is.

Note also that, by definition of sup, if \( x \) is distinct from every \( \alpha^c_i \), then \( v(x) = 0 \).

In the successor algebra, a constructor \( c \) can always be annotated as in Definition 15 above by taking:

**Example 2 (Canonical annotations in the successor algebra)** The canonical type of a constructor \( c \) in the successor algebra is obtained by taking

• \( \alpha^c_1 = \ldots = \alpha^c_{p^c} \),
• \( \alpha^c_{p^c+1} = \ldots = \alpha^c_{q^c} = \infty \),
• \( \sigma^c \in V \) if \( p^c = 0 \),
• \( \sigma^c = s \alpha^c_1 \) otherwise.

In this case, we get \( \Sigma^c(a_1, \ldots, a_{q^c}) = \sup\{a_1 + 1, \ldots, a_{p^c} + 1\} \), that is, the size is the constructor height, the size of a constant being 0.

For the constructors of the sort \( O \) of Howard’s constructive ordinals, we get:

• \( \text{zero} : O, \sigma^{\text{zero}} = x \) and \( \Sigma^{\text{zero}} = 0 \);
Size-based termination

4.2 Termination conditions

An important ingredient of the termination criterion is the way the sizes of function arguments are compared. In frameworks where functions are defined by fixpoint and case analysis, exactly one argument must decrease at a time. Here, we allow the comparison of various arguments simultaneously, possibly through some interpretation functions \( \zeta \).

Since not every term can be assigned a notion of size, and since two function calls can have different numbers of arguments, we first need to specify what arguments have to be taken into account and how their sizes are compared:

Definition 17 (Order on function calls) We assume given

- a well-founded quasi-ordering \( \leq_F \) on \( F \) (precedence) that we extend into a well-founded quasi-ordering on \( V \cup C \cup F \) by taking \( s <_F f \) whenever \( s \in V \cup C \) and \( f \in F \);
- for every \( f : T_1 \Rightarrow \ldots \Rightarrow T_r \Rightarrow B \):
  - a number \( q^f \) such that, for all \( i \in \{1, \ldots, q^f\} \), \( T_i \) is a sort \( B^f_i \) (the first \( q^f \) arguments of \( f \) are the arguments that will be taken into account for proving termination);
  - an annotated type \( \Theta(f) = T_1 \Rightarrow \ldots \Rightarrow T_r \Rightarrow B_{\sigma^f} \) such that
    - for all \( i \in \{1, \ldots, q^f\} \), \( T_i = \text{Annot}(B^f_i, B^f_i, a^f_i) \);
    - for all \( i \in \{q^f + 1, \ldots, r^f\} \), \( T_i = T_i \);
    - \( \bar{a} \) are distinct variables;
    - \( \sigma^f \in \bar{a} \);
    - \( \text{Var}(\sigma^f) \subseteq \{\bar{a}\} \);
  - for each \( X \in \{A, h\} \), a set \( D^f_X \), a quasi-order \( \leq^f_X \) on \( D^f_X \), a well-founded relation \( <^f_X \subseteq \leq^f_X \) and a map \( \zeta_X : X^{q^f} \rightarrow D^f_X \) such that
    - \( (D^f_X, \leq^f_X, <^f_X) = (D^g_X, \leq^g_X, <^g_X) \) whenever \( f \approx_F g \);
    - \( \bar{a} \mu <^f_X \bar{b} \nu \) whenever \( \bar{a} <^g_X \bar{b} \) and \( \mu : V \rightarrow h \);
\[\tilde{a} \preceq^{\tilde{g}_{\tilde{f}}} \tilde{c} \quad \text{whenever} \quad \tilde{a} \preceq^{\tilde{g}_{\tilde{f}}} \tilde{b} \quad \text{and} \quad \tilde{b} \preceq^{\tilde{g}_{\tilde{f}}} \tilde{c}, \quad \text{that is,} \quad <^{\tilde{g}_{\tilde{f}}} \preceq^{\tilde{g}_{\tilde{f}}} \subseteq <^{\tilde{g}_{\tilde{f}}};
\]
\[\tilde{a} \preceq^{\tilde{g}_{\tilde{f}}} \tilde{c} \quad \text{whenever} \quad \tilde{a} \preceq^{\tilde{g}_{\tilde{f}}} \tilde{b} \quad \text{and} \quad \tilde{b} \preceq^{\tilde{g}_{\tilde{f}}} \tilde{c}, \quad \text{that is,} \quad <^{\tilde{g}_{\tilde{f}}} \circ <^{\tilde{g}_{\tilde{f}}} \subseteq <^{\tilde{g}_{\tilde{f}}};
\]
where \(\langle x_1, \ldots, x_{q_{\tilde{f}}} \rangle <^{\tilde{g}_{\tilde{f}}} (y_1, \ldots, y_{q_{\tilde{f}}})\) iff \(g \cong f\) and \(\zeta^{\tilde{g}_{\tilde{f}}} \zeta^{\tilde{g}_{\tilde{f}}} (x_1, \ldots, x_{q_{\tilde{f}}}) <^{\tilde{g}_{\tilde{f}}} (y_1, \ldots, y_{q_{\tilde{f}}}).\)

The condition \(<^{\tilde{g}_{\tilde{f}}} \circ <^{\tilde{g}_{\tilde{f}}} \subseteq <^{\tilde{g}_{\tilde{f}}}\) is used in Theorem 1 (in the case (app-decr)). On the other hand, the condition \(<^{\tilde{g}_{\tilde{f}}} \circ <^{\tilde{g}_{\tilde{f}}} \subseteq <^{\tilde{g}_{\tilde{f}}}\) is only used in Lemma 22.

Note that, because \(<^{\tilde{g}_{\tilde{f}}} \subseteq <^{\tilde{g}_{\tilde{f}}}\) is only defined on terms of \(A\), if \(\tilde{a} <^{\tilde{g}_{\tilde{f}}} \tilde{b}\) and \(\tilde{b} <^{\tilde{g}_{\tilde{f}}} \tilde{c}\), then \(\tilde{a}\) must be in \(A\) too since, by definition, \(a \preceq_b b\) if \(a \preceq b\) or \(b = \infty\).

In the following, we may drop the exponent \(f\) when there is no ambiguity.

For the sake of simplicity, we assume that termination arguments come first. This is not a real restriction since arguments can always be permuted if needed.

For \(\zeta^{\tilde{f}_{\tilde{X}}}\), one can often take the identity (assuming that \(q^{\tilde{f}_{\tilde{X}}} = q^{\tilde{g}}\) whenever \(f \cong g\)). In Example 7, we use a different function. When \(\zeta^{\tilde{f}_{\tilde{X}}}\) is the identity, one can for instance take for \(\preceq^{\tilde{f}_{\tilde{X}}} \preceq^{\tilde{f}_{\tilde{F}}}\) the lexicographic or multiset extension (Dershowitz

\[\frac{(h, \tilde{h} \Rightarrow \tilde{V}) \in \Gamma \cup \{\tilde{h} \} \quad h <^{\tilde{g}} f \lor (h \cong f \land |\tilde{h}| \geq q^{\tilde{h}})}{\psi : \{\tilde{h}\} \rightarrow \bar{A} \quad (h, \psi) <^A (f, \varphi) \quad \forall \tilde{V} \Gamma \vdash_{\psi} T_{\tilde{V}} : V_{\psi}} \]  
(app-decr)

\[\frac{\Gamma, x : U \vdash^f_{\varphi} w : V}{\lambda x : U \vdash^f_{\varphi} \lambda x : U \Rightarrow V} \quad \frac{\Gamma, x : U \vdash^f_{\varphi} w : V}{\Gamma, x : U \vdash T \quad U \leq V} \quad \frac{\Gamma, x : U \vdash T \quad U \leq V \quad \Gamma \vdash^f_{\varphi} t : U \quad U' \leq V \quad V \leq V'}{\Gamma \vdash^f_{\varphi} t : V} \quad \frac{T \leq U \quad U \leq V}{T \leq V} \]  
(lam)  
(sub)  
(prod)  
(trans)  
(size)  
(refl)  

Fig. 3. Computability closure of \((f, \varphi)\).

Fig. 4. Subtyping rules.
& Manna, 1979) of $\leq_A$ ($\leq$ resp.), or some combination thereof, for which one can easily prove the compatibility of $\leq_A^i$ ($\leq_B^i$ resp.) with $\leq_A^\infty$ ($\leq_B^\infty$ resp.). Indeed, we have $(\leq_A^\infty)_{\text{prod}} \circ (\leq_A^\infty)_{\text{lex}} \subseteq (\leq_A^\infty)_{\text{lex}}$ since $\leq_A^\infty \circ \leq_A \subseteq \leq_A^\infty$, and $\leq_A^\infty \circ \leq_{\text{prod}} \subseteq \leq_{\text{lex}}$.

We can now state our general termination theorem. In Section 6, we will provide many examples of rewrite systems whose termination is implied by it.

**Theorem 1 (Termination criterion)** Assume that constructor types are annotated as in Definition 15. By Definitions 16 and 8, this provides us with a size function $\Sigma$ and a stratification $\mathcal{S}$. Assume moreover that $\rightarrow_{\mathcal{S}}$ is finitely branching and no $\sigma^c$ can be interpreted by an infinite limit ordinal.

Then, the relation $\rightarrow = \rightarrow_{\mathcal{B}} \cup \rightarrow_{\mathcal{S}}$ terminates on the set $\mathcal{T}$ of well-typed terms if, for each rule $l \rightarrow r \in \mathcal{R} \subseteq \mathcal{T}^2$, $l$ is of the form $t\bar{l}$, the type of $t$ is annotated as in Definition 17, $|\bar{l}| \geq q^i$ and there are

- a typing environment $\Gamma : \text{FV}(r) \rightarrow \mathcal{T}_A$ with, for every $(x, U) \in \Gamma$, an integer $k_x$ such that $x$ occurs in $l_x$, a sort $B^x$ occurring only positively in $|U|$ and a size variable $x^\phi$ such that $U = \text{Annot}(|U|, B^x, x^\phi)$, indicating how to measure the size of $x^\phi$;
- finite symbolic size upper bounds $\phi : \{\bar{x}^i\} \rightarrow A$ for $l_1, \ldots, l_q$;

such that

- **Monotony.** For all $i \in \{1, \ldots, q^i\}$, $\text{Pos}(x^i, \sigma^i) \subseteq \text{Pos}^+(\sigma^i)$;
- **Accessibility.** For every $(x, U) \in \Gamma$, one of the following holds:
  - $x = l_x$ and $U = T_{k_x} \phi$,
  - $T_{k_x}$ is a sort and $(x, |U|, B^x) \leq_A (l_x, T_{k_x}, T_{k_x})$;
- **Minimality.** For all substitutions $\theta$ with $l \theta \in \widehat{T}$, there exists a valuation $v$ such that
  - for all $(x, U) \in \Gamma$, $o_{B^x}^{\mathcal{S}^\theta} |U|(x \theta) \leq x^\phi v$,
  - for all $i \in \{1, \ldots, q^i\}$, $\bar{x}^i \phi v = o_{\mathcal{S}^\theta}(l \theta)$;
- **Subject-reduction and decreasingness.**
  $\Gamma \vdash_{\phi} r : T_{|l|+1} \Rightarrow \ldots \Rightarrow T_i \Rightarrow B_{\sigma^i} \phi$, where $\vdash_{\phi}$ is defined in Figures 3 and 4.

**Proof**

**Computability of constructors.** We first prove that, for all $(c, \mu, \bar{l})$ with $\Theta(c) = T_1 \Rightarrow \ldots \Rightarrow T_1 \Rightarrow B$, $\overline{\Theta(c)} = \overline{T_1} \Rightarrow \ldots \Rightarrow \overline{T_1} \Rightarrow B_{\sigma}$ as in Definition 15 (we drop the $c$'s in exponents), $|\bar{l}| = r$ and $(\forall i) t_i \in \overline{T_i} \mu$, we have $c \bar{l} \in B_{\sigma} \mu$. First, we have $c \bar{l} \in \text{SN}$ since $\bar{l} \in \text{SN}$ and there is no rule of the form $c \bar{l} \rightarrow r$. Second, by Proposition 1, for every $i \in \{1, \ldots, q\}$, we have $\overline{T_i} \mu \leq T_i$ since $\overline{T_i} = \text{Annot}(T_i, B_i, z_i)$ and $\text{Pos}(B_i, t_i) \leq \text{Pos}^+(T_i)$. Therefore, $c \bar{l} \in \text{SN}$. Now, if $\sigma = \infty$, then we are done. Otherwise, we are left to prove that $o_{\mathcal{S}^\theta}(c \bar{l}) \leq \sigma \mu$. By Corollary 2, $o_{\mathcal{S}^\theta}(c \bar{l}) = \Sigma(o_{\mathcal{S}^\theta}(\bar{l}))$. By definition, $\Sigma(o_{\mathcal{S}^\theta}(\bar{l})) = \sigma v$ where $v$ is defined in Definition 16. Since $\sigma$ is monotone and

---

6 Note that, if we do not care about the size of $x$, or if no sort occurs only positively in $U$, then we can always take for $B^x$ any sort not occurring in $U$.

7 Lemma 17 provides a syntactic condition for checking minimality in the successor algebra.
Var(\sigma) \subseteq \{\bar{z}\}$, it suffices to prove that, for all $i$, such that $z_i \in \mathcal{V}$, $z_i\nu \leq z_i\mu$. If all the $z_i \in \mathcal{V}$ are distinct, then $z_i\nu = \sigma_{\nu}(t_i) \leq z_i\mu$ since $t_i \in \overrightarrow{T_i}\mu$ and $z_i$ occurs only positively in $\overrightarrow{T_i}$. Otherwise, all the $z_i \in \mathcal{V}$ are equal. If there is no $z_i \in \mathcal{V}$, then the property holds trivially. Otherwise, all the $z_i \in \mathcal{V}$ are equal to some variable $x$ and $z_i\nu = \sup\{\sigma_{\nu}(t_i) \mid z_i = x\} \leq z\mu$ since, for all $i$ such that $z_i = z$, $t_i \in \overrightarrow{T_i}\mu$ and $z$ occurs only positively in $\overrightarrow{T_i}$.

**Computability of function symbols.** We now prove that, for all $((f,\mu), \bar{t})$ with $\Theta(f) = T_1 \Rightarrow \ldots \Rightarrow T_r \Rightarrow \mathbb{B}$ and $\overrightarrow{\Theta}(f) = T_1 \Rightarrow \ldots \Rightarrow T_r \Rightarrow \mathbb{B}_\sigma$ as in Definition 17 (we drop the $f$’s in exponents), $\bar{[\bar{t}]} = r$ and $(\forall i) t_i \in \overrightarrow{T_i}\mu$, we have $\bar{t} \bar{i} \in \mathbb{B}_\sigma\mu$, by induction on $((f,\mu), \bar{t})$ with $(<_{b}, \leftarrow_{\text{prod}})_{\text{lex}}$ as well-founded relation (1). Since $\bar{t} \bar{i}$ is neutral, it suffices to prove that, for all $u$ such that $\bar{t} \bar{i} \rightarrow u$, we have $u \in \mathbb{B}_\sigma\mu$. There are two cases:

a. $u = f \bar{u}$ and $\bar{i} \rightarrow_{\text{prod}} \bar{u}$. Since computability is preserved by reduction, $(\forall i) u_i \in \overrightarrow{T_i}\mu$. Therefore, by induction hypothesis (1), $u \in \mathbb{B}_\sigma\mu$.

b. $\bar{i} = \bar{l} \bar{u}$, $\bar{l} \rightarrow r \in \mathcal{R}$ and $u = r \bar{u} \bar{u}$. Let $\nu$ be the valuation given by minimality. For all $i \leq q$, $z_i\phi\nu = \sigma_{\nu}(l_i\theta)$. Since $l_i \in \overrightarrow{T_i}\mu$ and $T_i = B_{l_i\sigma\nu}$, we have $\phi\nu \leq \mu$ (*).

i. **Correctness of the computability closure.** We prove that, for all $(\Gamma, t, T, \theta)$, if $\Gamma \vdash^t \theta : T$ and $\chi\theta \in U\psi\nu$ when $(\chi, U) \in \Gamma$, then $t\theta \in TV$, by induction on $\vdash^t_\phi$ (2).

- (app-decr) By induction hypothesis (2), $\nu_i\theta \in V_{\psi}\nu$. There are three cases:

  - $h \in \mathcal{V}$. Then, $h\theta\bar{\nu} \in V_{\psi}\nu$ since $\psi = \emptyset$ and $h\theta \in (\bar{V} \Rightarrow V)\nu$ by assumption.

  - $h \in \mathcal{C}$ and $V = \bar{U} \Rightarrow C_\sigma$. For all $\bar{u} \in \bar{U}\psi\nu$, we have $h\theta\bar{\nu} \in C_\sigma\psi\nu$ by computability of constructors. Therefore, by Definition 2, $h\theta\bar{\nu} \in V_{\psi}\nu$.

  - $h \in \mathcal{F}$ and $V = \bar{U} \Rightarrow C_\sigma$. Since $(h, \psi) \lessdot_{\emptyset} (f, \phi)$ and $\bar{a}\nu \lessdot_{\emptyset} \bar{b}\nu$ whenever $\bar{a} \lessdot_{\emptyset} \bar{b}$, we have $(h, \psi\nu) \lessdot_{\emptyset} (f, \phi\nu)$. Since $\phi\nu \leq \mu$ and $\lessdot_{\emptyset} \leq_{\text{prod}} \leq_{\emptyset} \lessdot_{\emptyset}$, we have $(h, \psi\nu) \lessdot_{\emptyset} (f, \mu\nu)$. Hence, for all $\bar{u} \in \bar{U}\psi\nu$, we have $h\theta\bar{\nu} \in C_\sigma\psi\nu$ by induction hypothesis (1). Therefore, by Definition 2, $h\theta\bar{\nu} \in V_{\psi}\nu$.

- (lam) Wlog. we can assume that $x \notin \text{dom}(\theta) \cup \text{FV}(\theta)$. We have $(\lambda x U. w)(\theta) = \lambda x U. (w\theta) \in U\psi\nu \Rightarrow V\nu$ because, for all $u \in U\nu$, $(w\theta)(\{x, u\}) \in V\nu$ (cf. remarks after Definition 1) since $(w\theta)(\{x, u\}) = w\theta'$ where $\theta' = \theta \cup \{(x, u)\}$ and $\theta' \in \bar{V}\nu$ by induction hypothesis (2).

- (sub) We prove that $U\nu \subseteq V\nu$ whenever $U \subseteq V$ by induction on $\subseteq$ (3):

  - (size) If $b = \infty$, then $B_{a\nu} \subseteq \mathbb{B}$ by definition. Otherwise, $a\nu \leq b\nu$ and $B_{a\nu} = \mathcal{F}_{a\nu} \subseteq \mathcal{F}_{b\nu} = B_{b\nu} \nu$ since $\mathcal{F}_{b\nu}$ is monotone.

  - (prod) Let $t \in U\nu \rightarrow V\nu$ and $u' \in U'\nu$. By induction hypothesis (3), $U'\nu \subseteq U\nu$. Hence, $u \in U\nu$ and $tu \in V\nu$. By induction hypothesis (3), $V\nu \subseteq V'\nu$. Thus, $tu' \in V'\nu$.

- (refl) Immediate.

- (trans) By induction hypothesis (3) and transitivity of $\subseteq$. 

https://doi.org/10.1017/50956796818000072 Published online by Cambridge University Press
ii. **Computability of the matching substitution:** For all \((x, U) \in \Gamma\), \(x\theta \in Uv\). By assumption, there is \(k\) such that \(x\) occurs in \(l_k\), and \(l_k\theta \in \overline{T_k\mu}\). After the accessibility condition, there are two cases:

- \(x = l_k\) and \(U = \overline{T_k\phi}\). If \(k > q\), then \(\overline{T_k} = T_k\) and \(\overline{T_k\mu} = Uv\). Therefore, \(x\theta \in Uv\) since \(l_k\theta \in \overline{T_k\mu}\). If \(k \leq q\), then \(\overline{T_k} = B_{z_k}\) for some sort \(B\). By minimality, \(z_k\phi = o_{\phi}(l_k\theta)\). Therefore, \(x\theta \in Uv\) since \(U = B_{z_k}\phi\).

- \(T_k\) is a sort and \((x, |U|, B^x) \leq_{\alpha} (l_k, T_k, T_k)\). By Lemma 13, \(x\theta \in |U|\) since, by assumption, \(l_k\theta \in T_k\). By assumption, \(U = \text{Annot}(|U|, B^x, x^v)\) and Pos\((B^x, |U|) \subseteq \text{Pos}^+(|U|)\). By minimality, \(o_{B^x, U}(|U|)(x\theta) \leq x^v\). Therefore, \(x\theta \in Uv\).

iii. We can now end the proof that \(u \in B_{\sigma\mu}\). Since \(\Gamma \vdash u : V\phi\) with \(V = \overline{T_{|u|+1}} \Rightarrow \ldots \Rightarrow \overline{T_r} \Rightarrow B_{\sigma}\), and \(x\theta \in Uv\) whenever \((x, U) \in \Gamma\) by (ii), we have \(r\theta \in V\phi\) by (i). Hence, \(u = r\theta \overline{u} \in B_{\sigma\phi\nu}v\). We now prove that \(B_{\sigma\phi\nu}v \subseteq B_{\sigma\mu}\).

If \(\sigma = \infty\), then \(B_{\sigma\phi\nu}v = B_{\sigma\mu}\). Otherwise, we have \(\phi \sigma \neq \infty\) since \(\phi : \{\overline{2}\} \rightarrow \Lambda\). Moreover, we have seen in (ii) that, for all \(k \leq q\), \(\overline{T_k} = B_{z_k}\) for some sort \(B\) and \(z_k\phi = o_{\phi}(l_k\theta)\). Since \(l_k\theta \in \overline{T_k\mu}\), \(z_k\phi \leq z_k\mu\). Now, by monotony, Pos\((z_k, \sigma) \subseteq \text{Pos}^+(\sigma)\). Therefore, by Proposition 1, \(B_{\sigma\phi\nu}v \subseteq B_{\sigma\mu}\).

**Computability of well-typed terms.** Now, it is easy to prove that every well-typed term is computable by proceeding as for the correctness of the computability closure: if \(\Gamma \vdash t : T\) and \(x\theta \in U\) whenever \((x, U) \in \Gamma\), then \(t\theta \in T\). We just detail the case of a function symbol \(f\) with \(\Theta(f) = T_1 \Rightarrow \ldots \Rightarrow T_r \Rightarrow B\) and \(\overline{\Theta(f)} = \overline{T_1} \Rightarrow \ldots \Rightarrow \overline{T_r} \Rightarrow B_{\sigma}\).

After Definition 2, \(f \in \Theta(f)\) iff, for all \(\overline{t} \in \overline{T}\) such that \(|\overline{t}| = r\), \(f \overline{t} \in B\). By assumption, for all \(i \in \{1, \ldots, q\}\), \(T_i\) is a sort \(B_i\). Let \(\mu\) be the valuation mapping, for every \(i \in \{1, \ldots, q\}\), \(z_i\) to the smallest ordinal \(h_i < h\) such that \(\overline{S}_{B_i} = B_i\). Then, \(t_i \in B_{\overline{B_i}\mu}\) and, by computability of function symbols, \(f \overline{t} \in B_{\sigma\mu} \subseteq B\). Finally, we conclude by noting that the identity substitution is computable (cf. remark after Definition 1).

It is worth remarking that this criterion is modular since the above conditions are for each rule. Hence, if both \(\mathcal{R}_1\) and \(\mathcal{R}_2\) satisfy the criterion with the same parameters, then \(\mathcal{R}_1 \cup \mathcal{R}_2\) satisfies the criterion too.

We now discuss each condition in turn.

**Accessibility.** The accessibility is easy to check. As explained in Section 2.5, not every subterm of a computable term is computable. The definition of computability ensures that all accessible subterms so are (Lemma 13). The accessibility condition ensures that each free variable \(x\) of the right-hand side is accessible in some \(l_i\). Hence, every instance of \(x\) is computable if the arguments of \(f\) so are. Now, when \(x\) is accessible in a termination argument \((k^x \leq q^f)\), there must be a sort \(B^x\) with respect to which the size of the instances of \(x\) are measured. Since \(x\) can be instantiated by terms of any size, the type of \(x\) should be of the form \(\text{Annot}(|U|, B^x, x^v)\), that is, every occurrence of \(B^x\) should be annotated by some size variable \(x^v\), and no other sort should be annotated.

**Subject-reduction and decreasingness.** This condition enforces two properties at once. First, the right-hand side has the same type as the left-hand side. This subject-
reduction property is required since the interpretation of a type has to be stable by reduction. So, there should be no rule $f \vdash r$ such that the size of $r$ is strictly bigger than the size of $f \hat{l}$. Second, by (app-decr), in every function call $h \bar{t}$, the symbolic upper bounds $\psi$ of the actual sizes of the termination arguments of $h$ are strictly smaller than those of $f \hat{l}$ given by $\varphi$.

In (app-decr), $\psi$ is any size substitution of the size variables of $\bar{V}$. This rule works like the rule for type instantiation in Hindley–Milner type system (Hindley, 1969; Milner, 1978) except that, here, $\psi$ is not a type substitution but a size substitution. Hence, if $s$ is declared of type $N_x \Rightarrow N_{az}$, then by (app-decr), $\vdash^f s : N_a \Rightarrow N_{sa}$ for any size expression $a$. This means that, in annotated types, size variables are implicitly universally quantified.

The rule (app-decr) is a compact formulation that subsumes in a single rule the usual rules of simply typed $\lambda$-calculus for variables ($\Gamma \vdash^f x : T$ if $(x, T) \in \Gamma$), constructors and function symbols ($\Gamma \vdash^f c : T\psi$ if $(c, T) \in \bar{\Theta}$ and $\psi$ is any size substitution), and application ($\Gamma \vdash^f tu : V$ if $\Gamma \vdash^f t : U \Rightarrow V$ and $\Gamma \vdash^f u : U$), with the following restrictions on application and function symbols. First, the head of an application cannot be an abstraction: $\vdash^f$ only accepts terms in $\beta$-normal form since rule right-hand sides usually so are. Second, if an application is headed by a function symbol $g$, then $g <_f f$ (note that $h <_f f$ whenever $h \in \mathbf{V} \cup \mathbb{C}$), or we have $g \simeq f$, $g$ applied to at least $q^g$ arguments, and the sizes of the arguments of $g$, represented by $\psi$, are smaller than $\varphi$ in $<_\lambda$.

Hence, in (app-decr), $h$ is either a variable of $\Gamma$, in which case $\bar{V} \Rightarrow V$ is the type of $h$ declared in $\Gamma$, or a constructor or function symbol, in which case $\bar{V} \Rightarrow V$ is the annotated type of $h$ declared in $\bar{\Theta}$. In addition, if $h$ is a variable, a constructor symbol or a function symbol strictly smaller than $f$, then $h$ can be applied to any number of arguments compatible with its type. On the other hand, if $h$ is a function symbol equivalent to $f$, then it must be applied to at least $q^h$ arguments, and the abstract sizes of these arguments, given by the size substitution $\psi$, must be strictly smaller than $\varphi$ in $<_\lambda$.

In the examples below, we will however use (var), (cons) and (prec) to denote the rule (app-decr) when $h$ is variable, a constructor or a function symbol smaller than $f$, respectively.

Note that the typability of $r$ may require two variables $x$ and $y$ to have the same size over-approximation, that is, to have $\alpha^x = \alpha^y$. This will always be the case in the successor algebra when $x$ and $y$ are two recursive arguments of a constructor because, in this algebra, the types of constructor arguments are annotated by the same variable. For instance, if $x$ and $y$ are the first two arguments of node : $T_x \Rightarrow T_x \Rightarrow B \Rightarrow T_{az}$, we must have $\alpha^x = \alpha^y$.

Note also that the termination conditions do not require $l$ itself to be typable in $\vdash^f$. Hence, for instance, assuming that $B$ has two constructors $c : B_x \Rightarrow B_{az}$ and $b : B_x \Rightarrow B_x \Rightarrow B_{az}$, we can handle the rule $f(b x_1 (c x_2)) \rightarrow f x_2$ by taking $\Gamma = [x_2 : B_{x_2}]$ and $\alpha^f_1 = sz^{x_2}$. On the other hand, we cannot handle the rule $f(b x_1 (c x_2)) \rightarrow f(b x_1 x_2)$. Indeed, in this case, we can have $o_{\varphi = \bar{\psi}}(b x_1 \theta x_2 \theta) = o_{\varphi = \bar{\psi}}(b x_1 \theta (c x_2 \theta))$ if $o(x_2 \theta) < o(x_1 \theta)$: the height is not a decreasing measure in this case.
The relation $\vdash^f_\phi$ is similar to the notion of computability closure introduced in Blanqui et al. (2002) and Blanqui (2016) except that, when comparing function arguments, it uses the sizes given by the type system instead of the structure of terms. As already mentioned in the introduction, using the size information instead of the structure of terms relates our termination technique to well-founded monotone algebras (Manna & Ness, 1970; Van de Pol, 1996; Hamana, 2006), semantic labeling (Zantema, 1995; Hamana, 2007) or the notion of size-change principle (Lee et al., 2001; Hyvernat, 2014). Now, as remarked in (Blanqui, 2006a) and (Kusakari & Sakai, 2007), the notion of computability closure itself has strong connections with the notion of dependency pair (Arts & Giesl, 2000). It is also a tool for defining and strengthening the higher order RPO (Blanqui, 2006b; Jouannaud & Rubio, 2007; Blanqui et al., 2015). Finally, some relations between these notions have been formally established: size-change principle and dependency pairs (Thiemann & Giesl, 2005), semantic labeling and RPO (Kamin & Lévy, 1980), dependency pairs and RPO (Dershowitz, 2013), and size-based termination and semantic labeling (Blanqui & Roux, 2009).

The decidability of $\vdash^f_\phi$ will be studied in Section 7 and following.

**Monotony.** The monotony condition is easy to check. It requires the size of terms generated by $f$ to be monotone wrt the sizes of its termination arguments. It can always be satisfied by taking $\sigma^f = \infty$. It is also satisfied if $\Lambda$ is monotone. This condition also appears in Abel (2004) and Barthe et al. (2004). It is necessary because, in the rule (app-decr), $\psi$ is not necessarily minimal: it may be set to a strict upper bound by using the rule (sub) beforehand. This could lead to invalid deductions wrt sizes. Take, for instance, the subtraction on natural numbers defined by the rules of Figure 1 and assume that $\text{sub} : N_x \Rightarrow N_\beta \Rightarrow N_{x-\beta}$ in the size algebra with $\leq_A = \leq_{\text{ext}}$ and 0, $s$ and $-$ interpreted by 0, successor and minus, respectively. Then, given $f : N_x \Rightarrow N$ with $\text{sub} \ll f$, the rule $f (s x) \rightarrow f (\text{sub} (s x))$ satisfies the other conditions. Indeed, take $\Gamma = [x : N_x]$ and $\phi = \{(x, s x)\}$. By (var), $\vdash^f_\phi x : N_x$.

By (cons), $\vdash^f_\phi s x : N_{s x}$. By (sub), $\vdash^f_\phi x : N_{a x}$. By (prec), $\vdash^f_\phi \text{sub} (s x) x : N_{s x} - s x$. By (sub), $\Gamma \vdash^f_\phi \text{sub} (s x) x : N_0$ (while $\sigma_{\phi N}(\text{sub} (s x)) > 0$!). Therefore, $\Gamma \vdash^f_\phi f (\text{sub} (s x)) : N$ since 0 $<_A s x$, but the system does not terminate since $f (s x) \rightarrow f (\text{sub} (s x)) \rightarrow f (s x)$.

**Minimality.** Since $\phi$ provides symbolic upper bounds only, this does not suffice for getting termination. We also need $\phi$ to be minimal. Indeed, consider the rule $f x \rightarrow f x$ with $f : N_x \Rightarrow N$ and $\Gamma = [x : N_x]$. By taking $z \phi = s x$, one has $\Gamma \vdash^f_\phi f x : N$ since $\Gamma \vdash^f_\phi x : N_x$ and $x <_A s x$.

In Theorem 1, minimality is expressed in the most general way by using the interpretation of annotated types in computability predicates. With some acquaintance, it is not too difficult to check this condition by hand on simple systems as shown in Example 3. In fact, we think that it is always possible to find a minimal $\phi$ when the type of every constructor $c$ is annotated in the max-successor algebra (extension of the successor algebra with a max operator) in the canonical way, that is, by taking $\sigma^c \in \mathbb{V}$ if $p^c = 0$ and $\sigma^c = s (\text{max} \alpha_1 \ldots \alpha_n)$ with distinct variables for $\alpha_1, \ldots, \alpha_n$ otherwise. However, in this paper, we want to focus on the
successor algebra and, in this case, minimality may not be satisfiable whatever \( \varphi \) is. This is due to the fact that, in the successor algebra, one often needs to approximate the sizes of two distinct term variables by the same size variable. Indeed, in the successor algebra, there is no function symbol of arity \( \geq 2 \). Hence, the annotated type of a binary constructor can only be of the form \( B_2 \Rightarrow B_2 \Rightarrow B_\sigma \) with the same size variable \( x \) for both arguments.

In the following section, we study in more details the size of constructor terms when the size is defined as the height like it is the case with the canonical annotations of constructor types in the successor and max-successor algebras. Then, we give a syntactic condition for minimality to be satisfied in the successor algebra.

### 5 Minimality property when the size is defined as the height

In this section, we provide sufficient conditions for the minimality condition of Theorem 1 to be satisfied when the notion of size is the height and the size of constants is 0, that is, when, for every constructor \( c \), we have

\[
\Sigma^c(a_1, \ldots, a_q^c) = \sup \{a_1 + 1, \ldots, a_p^c + 1\}.
\]

After Definition 16, this can be achieved in the successor algebra by taking the canonical annotation for constructor types (cf. Example 2).

To check the minimality condition, we need to know how the size of a term of the form \( t \theta \) depends on the sizes of the subterms \( x \theta \) where \( x \) is a variable of \( t \). To this end, we introduce a number of definitions to express what are the subterms that contribute to the size of a term and how they contribute to it:

**Definition 18 (Recursive subterms)** Let \( \mathcal{D} \) be the set of triples \( (u, U, k) \) made of a term \( u \), a type \( U \) and a number \( k \in \mathbb{N} \). Given a sort \( B \) and \( (u, U, k) \in \mathcal{D} \), let

\[
\text{Sub}_B(u, U, k) = \begin{cases} 
\{(u_i, U_i, k + 1) \mid i \in \{1, \ldots, p^c\}\} & \text{if there is } (c, \vec{u}, \vec{U}) \in \mathcal{C}^B \text{ such that } u = c \vec{u} \text{ and } U = B \\
\emptyset & \text{otherwise}
\end{cases}
\]

Then, let \( \rightarrow_B \) be the relation on finite sets of triples such that \( S \rightarrow_B S' \) if there is \( d \in S \) such that \( \text{Sub}_B(d) \neq \emptyset \) and \( S' = (S \setminus \{d\}) \cup \text{Sub}_B(d) \) (we replace \( d \) by \( \text{Sub}_B(d) \)). We say that a set \( S \subseteq \mathcal{D} \) is a set of \( B \)-recursive subterms of a term \( t \) if \( \{(t, B, 0)\} \rightarrow_B^* S \).

For instance, if \( a : B, c : B \Rightarrow B, p^c = 1, b : B \Rightarrow B \Rightarrow B, p^b = 2 \) and \( t = b(c(ca))x \), then \( \{(t, B, 0)\} \rightarrow_B \{(c(ca), B, 1), (x, B, 1)\} \rightarrow_B \{(ca, B, 2), (x, B, 1)\} \rightarrow_B \{(a, B, 3), (x, B, 1)\} \).

**Lemma 14** If \( S \) is a set of \( B \)-recursive subterms of \( t \in \mathcal{B} \), then

\[
o_{\varphi \theta}(t) = \sup \{o_{[\varphi \theta]}(u) + k \mid (u, U, k) \in S\}.
\]
Proof
Let \( M(S) = \{ o_{[B,:\tau]}U(u) + k \mid (u, U, k) \in S \} \). The lemma trivially holds for \( S = \{(t, B, 0)\} \). Hence, if suffices to check that, if it holds for \( S \) and \( S \rightarrow_B S' \), then it holds for \( S' \) too. So, assume that there is \((c \tilde{u}, B, k) \in S \) such that \( \text{Sub}_B(c \tilde{u}, B, k) \neq \emptyset \). Then, \( M(S') = (M(S) - \{ o_{[B,:\tau]}U(u) + k_1 \} \cup \{ o_{[B,:\tau]}U(u_i) + k + 1 \mid i \in I \} \) where \( I = \{1, \ldots, p^\circ\} \). But, by Corollary 2, \( o_{[B,:\tau]}B(c \tilde{u}) = o_{[B,:\tau]}(c \tilde{u}) = \Sigma^\circ(o_{[B,:\tau]}(u_1), \ldots, o_{[B,:\tau]}(u_p)) = \sup\{ o_{[B,:\tau]}(u_i) + 1 \mid i \in I \} = \sup\{ o_{[B,:\tau]}U(u_i) + 1 \mid i \in I \} \). Therefore, \( \sup M(S) = \sup M(S') \).

\( \square \)

Lemma 15 If \( S \) is a set of \( B \)-recursive subterms of \( t \) and \( \theta \) is a substitution, then \( S\theta = \{(u\theta, U, k) \mid (u, U, k) \in S \} \) is a set of \( B \)-recursive subterms of \( t\theta \).

Proof
The lemma holds for \( S = \{(t, B, 0)\} \). Hence, if suffices to check that, if it holds for \( S \) and \( S \rightarrow_B S' \), then it holds for \( S' \) too. But \( \text{Sub}_B(u\theta, U, k) = \text{Sub}_B(u, U, k)\theta \).

Note that \( \rightarrow_B \) terminates (it acts on finite sets and replaces a term by smaller subterms) and is confluent (it is orthogonal). Hence, every finite set has a \( \rightarrow_B \)-normal form.

Definition 19 (Simple terms) Given a sort \( B \) and a term \( t \), let \( \text{Sub}_B(t) \) be the \( \rightarrow_B \)-normal form of \( \{ (t, B, 0) \} \). A term \( t \) of sort \( B \) is simple if, for all \((u, U, k) \in \text{Sub}_B(t)\), either \( u \in \mathcal{V} \) or there is \((c, \tilde{u}, \tilde{U}) \in \mathcal{C}^B \) such that \( u = c \tilde{u}, U = B \) and \( p^\circ = 0 \) (\( c \) has no recursive argument).

Lemma 16 If \( t \) is a simple term of sort \( B \) and \( t\theta \in B \), then

\[
o_{[B,:\tau]}(t\theta) = \sup\{ d_B(t) \cup \{ o_{[B,:\tau]}V(x\theta) + d_B^\tau(t) \mid (x, V) \in \text{Var}_B(t) \} \}
\]

where

- \( \text{Var}_B(t) = \{(x, U) \mid \exists k, (x, U, k) \in \text{Sub}_B(t)\} \),
- \( d_B^x(t) = \sup\{ k \mid \exists U, (x, U, k) \in \text{Sub}_B(t)\} \),
- \( d_B(t) = \sup\{ k \mid \exists u, U, (u, U, k) \in \text{Sub}_B(t)\} \).

Proof
By Lemma 15, \( \text{Sub}_B(t) \theta \) is a set of \( B \)-recursive subterms of \( t\theta \). Hence, by Lemma 14, \( o_{[B,:\tau]}(t\theta) = \sup\{ o_{[B,:\tau]}U(u) + k \mid (u, U, k) \in \text{Sub}_B(t)\theta \} = \sup\{ o_{[B,:\tau]}U(u\theta) + k \mid (u, U, k) \in \text{Sub}_B(t)\} \). Let \((x, V) \in \text{Var}_B(t)\). Since \( t \) is well-typed, for all \((x, V') \in \text{Var}_B(t)\), we have \( V' = V \). Hence, \( \sup\{ o_{[B,:\tau]}U(u\theta) + k \mid (u, U, k) \in \text{Sub}_B(t), u = x \} = o_{[B,:\tau]}V(x\theta) + d_B^\tau(t) \). Let now \((u, U, k) \in \text{Sub}_B(t)\) with \( u \notin \mathcal{V} \). Since \( t \) is simple, there is \((c, \tilde{u}, \tilde{U}) \in \mathcal{C}^B \) such that \( u = c \tilde{u}, U = B \) and \( p^\circ = 0 \). By Corollary 2, \( o_{[B,:\tau]}U(c \tilde{u}\theta) = o_{[B,:\tau]}(c \tilde{u}) = \Sigma^\circ(o_{[B,:\tau]}(u_1\theta), \ldots, o_{[B,:\tau]}(u_p\theta)) = \sup\{ o_{[B,:\tau]}(u_i\theta) + 1 \mid i \in \{1, \ldots, p^\circ\} \} = 0 \). Therefore, \( \sup\{ o_{[B,:\tau]}U(u\theta) + k \mid (u, U, k) \in \text{Sub}_B(t), u \notin \mathcal{V} \} = d_B(t) \) and \( o_{[B,:\tau]}(t\theta) = \sup\{ d_B(t) \cup \{ o_{[B,:\tau]}V(x\theta) + d_B^\tau(t) \mid (x, V) \in \text{Var}_B(t) \} \} \).

To carry on with the previous example, \( t = b(c(ca))x \) is simple and we have \( o_{[B,:\tau]}(t\theta) = \sup\{ o_{[B,:\tau]}(c(ca)) + 1, o_{[B,:\tau]}(x\theta) + 1 \} = \sup\{ 3, o_{[B,:\tau]}(x\theta) + 1 \} = \sup\{ d_B(t), o_{[B,:\tau]}(x\theta) + d_B^\tau(t) \} \).

https://doi.org/10.1017/50956796818000072 Published online by Cambridge University Press
Assume now that we are under the conditions of Theorem 1 for some rule \( t \overset{ f }{\rightarrow} r \in \mathcal{R} \), typing environment \( \Gamma = [ x_1 : U_1, \ldots, x_n : U_n ] \) and substitution \( \phi : \{ \bar{z} \} \rightarrow \mathcal{A} \). In particular, 
\[
\bar{\Theta}(f) = \mathcal{B}_{1x_1} \Rightarrow \ldots \Rightarrow \mathcal{B}_{q_aq} \Rightarrow \mathcal{T}_{q+1} \Rightarrow \ldots \Rightarrow \mathcal{T}_r \Rightarrow \mathcal{B}_\sigma
\]
with \( \bar{z} \) distinct variables, \( \sigma \in \mathcal{K} \) and \( \text{Var}(\sigma) \subseteq \{ \bar{z} \} \).

Assume, moreover that, for all \( j \in \{ 1, \ldots, q \} \), \( l_j \) is a simple term of sort \( \mathcal{B}_j \) and there are \( n_j \in \mathbb{N} \) and \( \gamma_j \in \mathbb{V} \) such that \( x_j \phi = s^{n_j}\gamma_j \).

Then, after Lemma 16, the minimality property is equivalent to the following purely numerical problem on ordinals: for all \( a_1, \ldots, a_n \) (for the sizes of \( x_1 \theta, \ldots, x_n \theta \), respectively), there are \( b_1, \ldots, b_n \) (for \( x^{\gamma_1} \nu, \ldots, x^{\gamma_q} \nu \), respectively) and \( c_1, \ldots, c_q \) (for \( \gamma_1 \nu, \ldots, \gamma_q \nu \), respectively), such that

1. \((\forall j)(\forall k)\ b_j = b_k\text{ if }x_j^\gamma = x_k^\gamma,\)
2. \((\forall j)(\forall k)\ c_j = c_k\text{ if }\gamma_j = \gamma_k,\)
3. \((\forall j)(\forall k)\ b_j = b_k\text{ if }x_j^\gamma = \gamma_k,\)
4. \((\forall j)\ a_j \leq b_j,\)
5. \((\forall j)\ c_j + n_j = \text{sup}\{(d_{B_j}(l_j)) \cup \{a_m + d_{B_j}^\mathbb{N}(l_j) \mid x_m \in \text{dom}(\text{Sub}_{B_j}(l_j))\}\}.\)

The first three constraints are coherence conditions for \( \nu \) to be well-defined. The last two correspond to the first and second conditions of the minimality property, respectively.

One of the problems for these inequations to be satisfied is when two arguments of \( f \), say \( l_1 \) and \( l_2 \), share some variable but have distinct sets of variables, or when shared variables occur at different depths. Take, for instance, \( l_1 = x_1 \) and \( l_2 = b(c(x_1))(c(x_2)) \) with constructors annotated in the canonical way in the successor algebra, that is, \( c : \mathcal{B}_z \Rightarrow \mathcal{B}_{ax} \) and \( b : \mathcal{B}_z \Rightarrow \mathcal{B}_x \Rightarrow \mathcal{B}_{ax} \). Then, for having \( b \eta_{x_1} \eta_{x_2} \) in the right-hand side, we need to take \( x_1^\gamma = x_2^\gamma = \gamma_1 = \gamma_2 \). In this case, the minimality condition says that, for all \( a_1, a_2 \), there is a \( b \) such that \( a_1 \leq b, a_2 \leq b, a_1 = b + n_1 \) and \( \text{sup}\{a_1 + 2, a_2 + 2\} = b + n_2 \), which is not possible. Take now \( l_1 = b(c(x_1))(c(x_2)) \) and \( l_2 = b(c(c(x_1)))(c(x_2)) \). Again, for having \( b \eta_{x_1} \eta_{x_2} \) in the right-hand side, we need to take \( x_1^\gamma = x_2^\gamma = \gamma_1 = \gamma_2 \). In this case, the minimality condition says that, for all \( a_1, a_2 \), there is a \( b \) such that \( a_1 \leq b, a_2 \leq b, \text{sup}\{a_1 + 2, a_2 + 2\} = b + n_1 \) and \( \text{sup}\{a_1 + 3, a_2 + 2\} = b + n_2 \), which is not possible either.

We now give sufficient conditions for the above set of inequations to be satisfied:

**Lemma 17** Under the conditions of Theorem 1, assume that constructor types are annotated in the canonical way in the successor algebra (cf. Example 2). Then, the minimality property is satisfied if, for all \( j \in \{ 1, \ldots, q \} \):

a. \( l_j \) is a simple term of sort \( \mathcal{B}_j \);
b. \( n_j \in \mathbb{N} \) and \( \gamma_j \in \mathbb{V} \) such that \( x_j \phi = s^{n_j}\gamma_j \);
c. \( n_j \leq \text{inf}(\text{range}(D_j)) \);
d. for all \( k \in \{ 1, \ldots, q \} \), if \( \gamma_j = \gamma_k \), then \( n_j = n_k, \ d_{\mathcal{B}_j}(l_j) = \ d_{\mathcal{B}_k}(l_k) \) and \( D_j = D_k \);
e. for all \( x \in \text{dom}(\Gamma) \), if \( \gamma_j = x^\gamma \) then \( x \in \text{dom}(D_j) \);

where \( D_j = \{ (x, d_{\mathcal{B}_j}(l_j)) \mid x \in \text{dom}(\text{Sub}_{\mathcal{B}_j}(l_j)) \} \), \( \text{Sub}_{\mathcal{B}}(l) \) is introduced in Definition 19, \( d_{\mathcal{B}} \) and \( d_{\mathcal{B}}^\mathbb{N} \) are defined in Lemma 16.

https://doi.org/10.1017/50956796818000072 Published online by Cambridge University Press
Proof
Let \( c_i = \sup(\{d_B(l_i)\} \cup \{a_p + d \mid (x_p, d) \in D_i\}) - n_i \). It is well-defined since \( n_i \leq \inf(\range(D_i)) \leq d_B(l_i) \). Now, let \( b_i = c_m \) if \( x^j = y^m \) for some \( m \), and \( b_i = \sup\{a_p \mid x^y = x^j \} \) otherwise. It is well-defined since, if \( \gamma_j = \gamma_k \), then \( c_j = c_k \) because \( n_j = n_k \), \( d_B(l_j) = d_B(l_k) \) and \( D_j = D_k \). We now prove that the five numerical constraints equivalent to minimality are satisfied:

1. Assume that \( x^y = x^m \). If \( x^y = y^m \), then \( b_j = c_m = b_k \).
   Otherwise, \( b_j = \sup\{a_m \mid x^y = x^j \} = b_k \).
2. Assume that \( \gamma_j = \gamma_k \). Then, \( c_j = c_k \).
3. Assume that \( x^y = \gamma_k \). Then, \( b_j = c_k \).
4. For all \( j, a_j \leq b_j \). Indeed, if \( x^y = y^m \), then \( b_j = c_m \geq a_j \) since \( x_j \in \dom(D_m) \).
   Otherwise, \( b_j = \sup\{a_p \mid x^y = x^j \} \geq a_j \).
5. For all \( j, c_j + n_j = \sup(\{d_B(l_j)\} \cup \{a_m + d_B(l_j) \mid x_m \in \dom(Sub_B(l_j))\}) \) by definition of \( c_j \).

For instance, with the last rule of Figure 1, \( \div(s \, x) \, (s \, y) \rightarrow s \, (\div(sub \, x \, y) \, (s \, y)) \), if we take \( \div : \mathbb{N} \Rightarrow \mathbb{N} \Rightarrow \mathbb{N}, \Gamma = [x : \mathbb{N}, y : \mathbb{N}], x^x = x, x^y = y \) and \( \varphi = \{(x, s \, x)\} \), we have \( n_1 = 1, \gamma_1 = x = x^x \) and \( D_1 = \{(x, 1)\} \). So, the conditions above are satisfied.

On the contrary, if \( l_1 = c \, x_1, \, l_2 = b \, (c \, x_1) \, (c \, x_2), \, x^{x_1} = x^{x_2} = \gamma_1 = \gamma_2, \, n_1 = 1 \) and \( n_2 = 2 \), then (d) is not satisfied because \( \gamma_1 = \gamma_2 \) but \( D_1 = \{(x_1, 1)\} \) and \( D_2 = \{(x_1, 2), (x_2, 2)\} \).

6 Examples

In this section, we show various examples whose termination can be established by using Theorem 1. In proofs of \( \mathcal{L}_\varphi^e \) judgments, \((\text{var}), (\text{cons})\) and \((\text{prec})\) will refer to the specialization of \((\text{app-decr})\) to variables, constructors and function symbols smaller than \( f \), respectively.

We will use the following sorts and constructors with \( N <_S L \) and \( N <_S O \):

- **B**: the sort of booleans with the constructors \( \text{true} : B \) and \( \text{false} : B \).
- **N**: the sort of natural numbers with the constructors \( 0 : N \) and \( s : N \Rightarrow N \).
- **O**: the sort of Howard's constructive ordinals with the constructors \( \text{zero} : O \), \( \text{succ} : O \Rightarrow O \) and \( \text{lim} : (N \Rightarrow O) \Rightarrow O \).
- **L**: the sort of lists with the constructors \( \text{nil} : L \) and \( \text{cons} : L \Rightarrow N \Rightarrow L \).

Unless stated otherwise, we always use the successor algebra (Definition 12) and, for constructor types, the canonical annotations (Example 2).

**Example 3 (Division)** Consider the function symbols \( \text{sub} \) (subtraction) and \( \div \) (division) both of type \( N \Rightarrow N \Rightarrow N \) defined by the rules of Figure 1.
We take \( \text{sub} \preceq \text{div} \). For annotated types, we take, for each \( f \in \{ \text{sub}, \text{div} \} \), 
\( \overline{\Theta}(f) = N_x \Rightarrow N \Rightarrow N_z \), \( q^f = 1 \), \( \mathcal{B}_1^f = N \), \( x_1^f = x \), which expresses the fact that these functions are not size-increasing. And, for \( \zeta_{X}^{\text{sub}} \) and \( \zeta_{X}^{\text{div}} \), we take the identity.

We now detail the conditions of Theorem 1 for each rule in turn (monotony is trivial).

1. sub \( x \ 0 \to x \). Take \( \Gamma = [x : N_x] \), \( k^x = 1 \), \( B^x = N \), \( x^x = x \) and \( \phi = \{(x, x)\} \). Then, \( N_x = \text{Annot}(N, B^x, x^x) \), and:
   - Accessibility. \( x \) is accessible since \( x = I_{k^x} \) and \( N_x = N_{x, \phi} \).
   - Minimality. One could use Lemma 17. We give a direct proof instead. Let \( \theta \) be such that \( x \theta \in N \). We have to prove that there exists \( v \) such that \( o_{\text{gen}}(x \theta) \leq x^x v \) and \( x \phi v = o_{\text{gen}}(x \theta) \). It suffices to take \( v(x) = o_{\text{gen}}(x \theta) \).
   - Subject-reduction. By (var), \( \models_{\phi} \text{sub} \ x : N_x = N_{x, \phi} \).

2. sub \( 0 \ y \to 0 \). Take \( \Gamma = \phi = \emptyset \). Then,
   - Minimality. Let \( \theta \) be such that \( y \theta \in N \). We have to prove that there exists \( v \) such that \( x \phi v = o_{\text{gen}}(0) \). It suffices to take \( v(x) = o_{\text{gen}}(0) \).
   - Subject-reduction. By (cons), \( \models_{\phi} \text{sub} \ 0 : N_x = N_{x, \phi} \).

3. sub \( (s \ x) \ (s \ y) \to \text{sub} \ x \ y \). Take \( \Gamma = [x : N_x, y : N_y] \), \( k^x = 1 \), \( B^x = N \), \( x^x = x \), \( k^y = 2 \), \( B^y = N \), \( x^y = y \), \( \phi = \{(x, s, x)\} \). Then, \( N_x = \text{Annot}(N, B^x, x^x) \), \( N_y = \text{Annot}(N, B^y, x^y) \) and:
   - Accessibility. \( x \) is accessible since \((x, N, N) \sqsubseteq_{\alpha} (I_{k^x}, N, N) \). \( y \) is accessible since \((y, N, N) \sqsubseteq_{\alpha} (I_{k^y}, N, N) \).
   - Minimality. Let \( \theta \) be such that \( s \ x \theta \in N \) and \( s \ y \theta \in N \). We have to prove that there exists \( v \) such that \( o_{\text{gen}}(s x \theta) \leq x^x v \), \( o_{\text{gen}}(s y \theta) \leq x^y v \) and \( x \phi v = o_{\text{gen}}(s x \theta) = o_{\text{gen}}(s x \theta) + 1 \). It suffices to take \( v(x) = o_{\text{gen}}(s x \theta) \) and \( v(y) = o_{\text{gen}}(s y \theta) \).
   - Subject-reduction. Let \( \models = \models_{\phi} \text{sub} \). By (var), \( \models x : N_x \) and \( \models y : N_y \). By (app-decr), \( \models \text{sub} \ x \ y : N_x \) since \( x <_{\alpha} s \ x \). Therefore, by (sub), \( \models \text{sub} \ x \ y : N_{ax} = N_{x, \phi} \).

4. div \( 0 \ (s \ y) \to 0 \). Like for rule (2).

5. div \( (s \ x) \ (s \ y) \to s \ (\text{div} \ (\text{sub} \ x \ y) \ (s \ y)) \). Take \( \Gamma = [x : N_x, y : N_y] \), \( k^x = 1 \), \( B^x = N \), \( x^x = x \), \( k^y = 2 \), \( B^y = N \), \( x^y = y \) and \( \phi = \{(x, s, x)\} \). Then, \( N_x = \text{Annot}(N, B^x, x^x) \), \( N_y = \text{Annot}(N, B^y, x^y) \) and:
   - Accessibility and minimality like for rule (3).
   - Subject-reduction. Let \( \models = \models_{\phi} \text{div} \). By (var), \( \Gamma \models x : N_x \) and \( \Gamma \models y : N_y \). By (prec), \( \Gamma \models \text{sub} \ x \ y : N_x \). By (cons), \( \Gamma \models s \ y : N_{ay} \). By (app-decr), \( \Gamma \models \text{div} \ (\text{sub} \ x \ y) \ (s \ y) : N_x \) since \( x <_{\alpha} s \ x \). Finally, by (cons), \( \Gamma \models s \ (\text{div} \ (\text{sub} \ x \ y) \ (s \ y)) : N_{ax} = N_{x, \phi} \).

Example 4 (Map and filter) Consider the function symbols map : \( L \Rightarrow (N \Rightarrow N) \Rightarrow L \)\(^9\) if \( L \Rightarrow L \Rightarrow B \Rightarrow L \) and filter : \( L \Rightarrow (N \Rightarrow B) \Rightarrow L \) defined by the rules:

\(^9\) We permuted the usual order of the arguments of map so that its type conforms to Definition 17.
map nil \( f \) \( \rightarrow \) nil
map (cons \( l \) \( x \)) \( f \) \( \rightarrow \) cons (map \( l \) \( f \)) (\( f \) \( x \))
if \( x \) \( y \) true \( \rightarrow \) \( x \)
if \( x \) \( y \) false \( \rightarrow \) \( y \)
filter nil \( f \) \( \rightarrow \) nil
filter (cons \( l \) \( x \)) \( f \) \( \rightarrow \) if (cons (filter \( l \) \( f \)) \( x \)) (filter \( l \) \( f \)) (\( f \) \( x \))

For annotated types, we could take in the successor algebra, map : \( L_2 \) \( \Rightarrow \) (\( N \Rightarrow \) \( N \)) \( \Rightarrow \) \( L_2 \), q\text{map} = 1, \( B^1_1\text{map} = L \), \( x^1_\text{map} = x \), if : \( L_2 \) \( \Rightarrow \) \( L_2 \) \( \Rightarrow \) \( B \) \( \Rightarrow \) \( L_2 \), q\text{if} = 2, \( B^f_1 = B^f_2 = L \), \( x^f_1 = x^f_2 = x \), filter : \( L_2 \) \( \Rightarrow \) (\( N \Rightarrow \) \( B \)) \( \Rightarrow \) \( L_2 \) and q\text{filter} = 1, expressing the fact that these functions are not size-increasing.

Unfortunately, the annotated type of if does not satisfy the conditions of Definition 17 because \( x^f_1 = x^f_2 \) (the variables \( x^f_i \) should be distinct). There are however two solutions to get around this problem:

1. Annotate if in the max-successor algebra by taking if : \( L_2 \) \( \Rightarrow \) \( L_\beta \) \( \Rightarrow \) \( B \) \( \Rightarrow \) \( L_\text{max} \\beta \).
2. Introduce a new type \( C \geq_S \) \( L \) with constructor cond : \( L_2 \) \( \Rightarrow \) \( L_2 \) \( \Rightarrow \) \( B \) \( \Rightarrow \) \( C_2 \), a new function symbol newif : \( C_2 \) \( \Rightarrow \) \( L_2 \) with q\text{newif} = 1, and define newif and filter by the following rules instead:

\[
\text{newif (cond } x \ y \ \text{true) } \rightarrow x
\]
\[
\text{newif (cond } x \ y \ \text{false) } \rightarrow y
\]
\[
\text{filter nil } f \rightarrow \text{nil}
\]
\[
\text{filter (cons } l \ x \text{) } f \rightarrow \text{newif (cond (cons (filter } l \ f \text{) } x \text{)) (filter } l \ f \text{) (} f \ x \text{)}
\]

One can easily check the conditions on annotated types and the monotony condition.

For the other conditions, we only detail the case of the last rule of filter by taking \( \Gamma = [f : N \Rightarrow B, x : N, l : L_l], \varphi = \{(x, s l)\}, k^f_l = 2, \) any sort distinct from \( N \) \( \) and \( B \) \( \) for \( B^l \) (we do not care about the size of \( f \)), \( x^l = f, k^x_l = 1, \) any sort distinct from \( N \) \( \) for \( B^x \) (we do not care about the size of \( x \)), \( x^x = x, k^l_l = 1, B^l = L, x^f = l, \) newif \( <_F \) filter, cond \( <_F \) filter and the identity for \( \zeta^\text{filter} \).

One can easily check the accessibility and minimality conditions.

We now check subject-reduction. Let \( \vdash = \vdash_{\phi_{filter}} \). By (var), \( \Gamma \vdash x : N \) and \( \Gamma \vdash l : L_l \). By (var), \( \Gamma \vdash f \ x : B \). By (app-decr), \( \Gamma \vdash \text{filter } l \ f : L_l \) since \( l \prec_A s l \). By (cons), \( \Gamma \vdash \text{cons (filter } l \ f \text{) } x : L_{s l} \). By (sub), \( \Gamma \vdash \text{filter } l \ f : L_{s l} \) since \( l \leq_F s l \). By (cons), \( \Gamma \vdash \text{cond (cons (filter } l \ f \text{) } x \text{) (filter } l \ f \text{) (} f \ x \text{)} : L_{s l} \). Therefore, by (prec), \( \Gamma \vdash \text{newif (cond (cons (filter } l \ f \text{) } x \text{)) (filter } l \ f \text{) (} f \ x \text{)} : L_{s l} = L_{x \varphi} \).

**Example 5 (Gödel’ system T and Howard’ system V)** Consider the recursor on natural numbers \( \text{rec}_N^T : N \Rightarrow T \Rightarrow (N \Rightarrow T \Rightarrow T) \Rightarrow T \) from Gödel’ system T (Gödel, 1958), and the recursor on ordinals \( \text{rec}_O^T : O \Rightarrow T \Rightarrow (O \Rightarrow T \Rightarrow T) \Rightarrow ((N \Rightarrow O) \Rightarrow (N \Rightarrow T) \Rightarrow T) \Rightarrow T \) from Howard’ system V (Howard, 1972) defined by the following rules:
The functions $\text{pair}$, $\text{fst}$, $\text{snd}$, $\text{le}$, and $\text{qs}$ are defined by the rules:

- $\text{pair} \, l \, m \rightarrow l$
- $\text{pair} \, l \, m \rightarrow m$
- $\text{pair} \, (\text{cons} \, l \, x) \, m \rightarrow \text{pair} \, (\text{cons} \, (\text{cons} \, p_1 \, x) \, p_2) \, (\text{pair} \, p_1 \, (\text{cons} \, p_2 \, x)) \, (\text{le} \, x \, y)$
- $\text{qs} \, \text{nil} \, m \rightarrow m$
- $\text{qs} \, \text{cons} \, l \, x \, m \rightarrow \text{qs} \, p_1 \, (\text{cons} \, (\text{qs} \, p_2 \, m) \, x)$
- $\text{qs} \, \text{nil} \, l \rightarrow \text{qs} \, \text{nil} \, l$

For the annotated types of constructors, we take the canonical annotations except for $\text{pair} : L_x \Rightarrow L_x \Rightarrow P_x$ and $\text{cond} : P_x \Rightarrow P_x \Rightarrow B \Rightarrow C_x$. Hence, a term of type $P_x$ is a pair of lists of length smaller than or equal to $x$.

Now, for function symbols, we take $\text{fst}, \text{snd} : P_x \Rightarrow L_x$, $\text{le} : N_x \Rightarrow N_x \Rightarrow B$, $\text{qs} : L_x \Rightarrow L_x \Rightarrow L_x \Rightarrow L_x$ and $\text{qsort} : L_x \Rightarrow L_x$. We now detail the subject-reduction condition for the last rule of $\text{qs}$ by taking $\Gamma = [x : N_x,l : L_l,m : L_l]$, $\varphi = \{(x,s l)\}$, $\text{fst}, \text{snd}, \text{pivot} <_{\text{qs}} \text{qs}$ and the identity for $\zeta^{\text{qs}}$. Let $\vdash = \vdash^{\text{qs}}$. By (var), $\Gamma \vdash x : N$, $\Gamma \vdash l : L_l$ and $\Gamma \vdash m : L$. By (prec), $\Gamma \vdash p : P_l$, $\Gamma \vdash p_1 : L_l$ and $\Gamma \vdash p_2 : L_l$. Since $l <_{\text{qs}} s l$, by (app-decr), $\Gamma \vdash \text{qs} \, p_2 \, m : L$. 

Example 6 (Quicksort) Let $P$ be the sort of pairs of lists with the constructor pair : $L \Rightarrow L \Rightarrow P$, and $C$ be the sort with the constructor cond : $P \Rightarrow P \Rightarrow B \Rightarrow C$. Then, let the functions $\text{fst}, \text{snd} : P \Rightarrow L$, $\text{le} : N \Rightarrow N \Rightarrow B$, if : $C \Rightarrow P$, pivot : $L \Rightarrow N \Rightarrow P$, $\text{qs} : L \Rightarrow L \Rightarrow L$ and $\text{qsort} : L \Rightarrow L$ be defined by the rules:

\[
\begin{align*}
\text{fst} \, (\text{pair} \, l \, m) & \rightarrow l & \text{le} \, 0 \, 0 & \rightarrow \text{true} \\
\text{snd} \, (\text{pair} \, l \, m) & \rightarrow m & \text{le} \, (s \, x) \, 0 & \rightarrow \text{false} \\
\text{if} \, (\text{cond} \, \text{true} \, p \, q) & \rightarrow p & \text{le} \, (s \, x) \, (s \, y) & \rightarrow \text{le} \, x \, y \\
\text{if} \, (\text{cond} \, \text{false} \, p \, q) & \rightarrow q
\end{align*}
\]
By (cons), \( \Gamma \vdash \text{cons} (\text{qs } p_2 \ m) \ x : L \). Finally, since \( l <_A s \ l \), by (app-decr) again, \( \Gamma \vdash \text{qs } p_1 (\text{cons} (\text{qs } p_2 \ m) \ x) : L \).

We proved the termination of this system. However, we cannot express that qsort is not size-increasing, that is, take \( \text{qsort : } L_x \Rightarrow L_x \). To do so, we need a more precise type system with existential quantifiers and constraints on size variables where pivot can be given the type:

\[
(\forall x)L_x \Rightarrow N \Rightarrow (3\beta)(3\gamma)(x = \beta + \gamma) L_\beta \times L_\gamma \quad \text{(Blanqui \& Riba, 2006)}.
\]

We now give an example using interpretation functions \( \zeta^L_x \) different from the identity:

\textbf{Example 7 (Reverse)} List reversal can be defined as follows (Huet \& Hullot, 1982):

\[
\begin{align*}
\text{last } \text{nil} \ x & \rightarrow \ x \\
\text{last } (\text{cons } l \ y) \ x & \rightarrow \ \text{last } l \ y \\
\text{revremlast } \text{nil} \ x & \rightarrow \ \text{nil} \\
\text{revremlast } (\text{cons } l \ y) \ x & \rightarrow \ \text{rev } (\text{cons } (\text{rev } (\text{revremlast } l \ y)) \ x) \\
\text{rev } \text{nil} & \rightarrow \ \text{nil} \\
\text{rev } (\text{cons } l \ x) & \rightarrow \ \text{cons } (\text{revremlast } l \ x) \ (\text{last } l \ x)
\end{align*}
\]

where \( \text{rev : } L \Rightarrow L \), \( \text{revremlast : } L \Rightarrow N \Rightarrow L \) and \( \text{last : } L \Rightarrow N \Rightarrow N \).

Since we have a first-order data type, we can assume that \( h = \omega \). Let \( A \) be the size algebra with the constant \( 1 \) interpreted by \( 1 \) and the binary function symbol \( + \) interpreted by the addition. Let \( \leq_A \) and \( <_A \) be \( \leq_{\text{ext}} \) and \( <_{\text{ext}} \), respectively (cf. remark after Definition 11).

Consider the fourth rule. Take \( \Gamma : L_x \Rightarrow N \Rightarrow L_{x+1}, \text{rev : } L_x \Rightarrow L_x, \text{revremlast : } L_x \Rightarrow N \Rightarrow L_x, \Gamma = [ x : N, y : N, l : L_l ] \) and \( \varphi = \{ (x, l + 1) \} \). One can easily check monotony, accessibility and minimality. We now check subject-reduction. Let \( \vdash = \vdash_{\text{revremlast}}^\varphi \). For comparing termination arguments, take \( \text{rev } \approx_{\text{revremlast}} \), \( \zeta^\text{rev}(a) = 2a \) (formally \( a + a \)) and \( \zeta^\text{revremlast}(a) = 2a + 1 \). By (var), \( \Gamma \vdash x : N, \Gamma \vdash y : N \) and \( \Gamma \vdash l : L_l \). By (app-decr), \( \Gamma \vdash \text{revremlast } l \ y : N_l \) since \( \zeta^\text{revremlast}(l) = 2l + 1 <_A \zeta^\text{revremlast}(l + 1) = 2(l + 1) + 1 = 2l + 3 \). By (app-decr), \( \Gamma \vdash \text{rev } (\text{revremlast } l \ y) : N_l \) since \( \zeta^\text{rev}(l) = 2l < 2l + 3 \). By (cons), \( \Gamma \vdash \text{cons } (\text{rev } (\text{revremlast } l \ y)) \ x : L_{l+1} \). Finally, by (app-decr), we get \( \Gamma \vdash r : L_{l+1} \), where \( r = \text{rev } (\text{cons } (\text{rev } (\text{revremlast } l \ y)) \ x) \), since \( \zeta^\text{rev}(l + 1) = 2l + 2 < 2l + 3 \).

We end this series of examples with one using non-standard constructor size annotations:

\textbf{Example 8 (Normalization of conditionals)} Let \( C \) be the sort of conditional expressions with the constructors \( \text{at : } C \) and \( \text{if : } C^3 \Rightarrow C \). Following Boyer \& Moore (1979), one can define a normalization function \( \text{nm : } C \Rightarrow C \) as follows:

\[
\begin{align*}
\text{nm at} & \rightarrow \ \text{at} \\
\text{nm } (\text{if } y z) & \rightarrow \ \text{if } (\text{nm } y) \ (\text{nm } z) \\
\text{nm } (\text{if } u v w) \ y z) & \rightarrow \ \text{nm } (\text{if } u \ (\text{nm } (\text{if } v w) \ y z)) \ (\text{nm } (\text{if } w y) \ z))
\end{align*}
\]
In Paulson (1986) is given a measure on terms due to Shostak that is decreasing in recursive calls. Hence, we can prove the termination of \( \text{nm} \) by using the following annotated types: \( \alpha : \text{C}_z \), if \( \alpha : \text{C}_x \Rightarrow \text{C}_y \Rightarrow \text{C}_{(x+1)(y+z+3)} \) and \( \text{nm} : \text{C}_x \Rightarrow \text{C}_x \). One can easily check the monotony condition.

Now, for the third rule, let \( \Gamma = \{u : \text{C}_{uvw} \Rightarrow \text{C}_{uwz} \Rightarrow \text{C}_{uyz} \} \) where \( u = ((u+1)(v+w+3)+1)(y+z+3) = uwy+uwz+uvw+uwz+3uw+3uw+3uy+3uz+vy+wv+vw+wz+9u+3v+3w+4y+4z+12, \) \( \gamma^m \) be the identity, and \( t = t^m \).

One can easily check monotony, accessibility and minimality. We now check subject-reduction. By (cons), \( \Gamma \vdash t^m \) if \( vwyz : \text{C}_{(w+1)(y+z+3)} \) and \( \Gamma \vdash vwyz : \text{C}_{(w+1)(y+z+3)} \). By (app-decr), \( \Gamma \vdash \text{nm} (vwyz) : \text{C}_{(w+1)(y+z+3)} \) since \( v+1)(y+z+3) = vwy+vwz+vy+zu+3u<_{\alpha} a \) and \( \Gamma \vdash \text{nm} (vwyz) : \text{C}_{(w+1)(y+z+3)} \) since \( w+1)(y+z+3) = wy+wz+y+z+3<_{\alpha} a \) and \( \Gamma \vdash \text{nm} (vwyz) : \text{C}_{(w+1)(y+z+3)} \) since \( (w+1)(y+z+3) = wy+wz+y+z+3<_{\alpha} a \). Finally, by (app-decr), \( \Gamma \vdash \text{nm} (u \text{nm} (vwyz)) \) if \( \text{nm} (vwyz) \) : \( \text{C}_h \) where \( b = (u+1)((v+1)(y+z+3)+(w+1)(y+z+3)+3) \) since \( b = wwy+wyz+wuy+wuz+2uy+2uz+vy+vz+wv+wz+9u+3v+3w+4y+4z+9<_{\alpha} a \). So, by (sub), \( \Gamma \vdash \text{nm} (u \text{nm} (vwyz)) \)

\( \text{nm} (vwyz) : \text{C}_{u} \).

7 Decidability of \( \vdash_{\phi} \)

In this section, we provide an algorithm for deciding the relation \( \vdash_{\phi} \) used in Theorem 1 and defined in Figures 3 and 4, under general conditions on the size algebra \( \alpha \). We will prove in Section 9 that these conditions are satisfied by the successor algebra.

The differences between \( \vdash_{\phi} \) and the usual typing relation for simply typed \( \lambda \)-calculus are the following. First, the set of typable symbols is restricted to those smaller than or equivalent to \( f \). Second, the application of \( t \) to \( u \) is restricted to the terms \( t \) whose head is not an abstraction. Moreover, when the head of \( f \) is a symbol equivalent to \( f \), the number of arguments must be bigger than \( q^f \) and the size of the arguments must be decreasing.

If we remove the decreasingness condition, we get the relation \( \vdash \) defined by the same rules as those of \( \vdash_{\phi} \) except (app-decr) replaced by

\[
\begin{align*}
(h, \hat{V} \Rightarrow V) &\in \Gamma \cup \overline{\Theta} \quad h <_{\text{F}} f \vee (h \simeq_{\text{F}} f \wedge |\hat{V}| > q^h) \\
(\psi : \{\hat{x}^h\} \rightarrow k) &\in \forall (\forall i) \Gamma \vdash \text{i} w_i : V_i \psi \\
\Gamma &\vdash \text{h} \hat{w} : V \psi
\end{align*}
\]

that is (app-decr) without the decreasingness condition \( (h, \psi) <_{\alpha} (t, \phi) \). Hence, deciding \( \Gamma \vdash_{\phi} t : T \) can be reduced to finding a derivation of \( \Gamma \vdash t : T \) where, at each (app) node, the decreasingness condition is satisfied.

In Section 7.1, we provide an algorithm for deciding \( \vdash \). Then, in Section 7.2, we show how to use this algorithm to decide \( \vdash_{\phi} \).

7.1 Decidability of \( \vdash \)

First note that, in a given typing environment \( \Gamma \), a typable term \( t \) may have several and even infinitely many types for two reasons. First, in (app), the size variables
of function types can be instantiated arbitrarily. Second, subtyping is generally not bounded. For instance, in the successor algebra, $\mathbb{N}_x \leq \mathbb{N}_{2x} \leq \ldots$

The relation $\vdash_{\ell}$ differs from Curry and Feys’ typing relation with functional characters or type schemes (a type with type variables) (Curry & Feys, 1958) in two points. First, our type schemes are not built from type variables but from size variables. Second, we have a subtyping relation. We will however see that some techniques developed for Curry and Feys’ type system or, more generally, Milner’s type system (Milner, 1978) and its extensions, can be adapted to our framework.

The decidability of type-checking in Curry and Feys’ system has been proved by Hindley (1969). Hindley’s algorithm is based on the fact the set of types of a typable term has a smallest element wrt some ordering $\sqsubseteq$. Hence, to decide whether $\Gamma \vdash t : T$, the algorithm proceeds in two steps. First, it computes the smallest type of $t$, say $U$, and then checks whether $U \sqsubseteq T$.

In Curry and Feys’ system, $\sqsubseteq$ is the instantiation ordering: a type-scheme $U$ is an instance of a type-scheme $T$, or $T$ is more general than $U$, written $T \sqsubseteq U$, if $T\theta = U$ for some substitution $\theta$. Huet (1976) proved that every non-empty set of terms has a greatest lower bound wrt $\sqsubseteq$. So, in particular, $\{T \mid \Gamma \vdash t : T\}$ has a greatest lower bound if $t$ is typable in $\Gamma$.

For computing the most general type, Hindley uses an algorithm based on unification (Herbrand, 1930; Robinson, 1965). Unifying two terms $T$ and $U$ consists in solving the equation $T = U$, that is, in finding a substitution $\theta$ such that $T\theta = U\theta$. Huet (1976) proved that solving $T = U$ is equivalent to finding an $\sqsubseteq$-upper bound to both $T$ and $U$. He also showed that every non-empty bounded set of terms has a least upper bound wrt $\sqsubseteq$. Hence, every solvable unification problem has a most general solution.

Hindley’s work was later extended in many directions by considering richer types, more complex constructions or by improving the algorithm computing the most general type-scheme. One of the most advanced generalizations seems to be Sulzmann’s HM($X$) system (Sulzmann, 2001), where the type variables of a type scheme are required to satisfy a formula of an abstract constraint system $X$. For his system, Sulzmann provides a generic constrained-type inference algorithm assuming a procedure for solving constraints in $X$. It would be interesting to study whether our framework can fit in this general setting. However, in this paper, we will simply follow Hindley’s approach.

But, since we also have subtyping, we define $\sqsubseteq$ as follows:

**Definition 20 (More general type)** We say that an annotated type $T$ is more general than another annotated type $U$, written $T \sqsubseteq U$, if there is a substitution $\theta$ such that $T\theta$ is a subtype of $U$, i.e. $T\theta \leq U$.

One can easily check that $\sqsubseteq$ is a quasi-ordering.

**Definition 21 (Subtyping problem)** A subtyping problem $P$ is either $\perp$ or a finite set of subtyping constraints, a subtyping constraint being a pair of types $(T, U)$ written $T \leq^? U$. It has a solution $\varphi : V \rightarrow \mathbb{K}$ if $P \neq \perp$, $\text{dom}(\varphi) \subseteq \text{Var}(P)$ and, for all
Fig. 5. Type inference algorithm.

\[
\frac{\Gamma, x : U \vdash f \; V}{\Gamma \vdash \lambda x^U \; V} \quad \text{(inf-lam)}
\]

\[
(h, \tilde{V} \Rightarrow V) \in \Gamma \cup \Theta \quad h <_T f \vee (h <_T f \wedge |\tilde{V}| \geq q^h)
\]

\[
(\forall i)\Gamma \vdash U_i \quad \rho_1, \ldots, \rho_n \text{ permutations on } V (n = |\tilde{V}| = |\tilde{w}|)
\]

\[
(\forall i)\text{Var}(U_i\rho_i) \cap \text{Var}(\tilde{V} \Rightarrow V) = \emptyset
\]

\[
(\forall i)(\forall j)i \neq j \Rightarrow \text{Var}(U_i\rho_i) \cap \text{Var}(U_j\rho_j) = \emptyset
\]

\[
\eta = \text{mgs} \left( \left\{ U_1\rho_1 \leq \eta V_1, \ldots, U_n\rho_n \leq \eta V_n \right\} \right)
\]

\[
\frac{\Gamma \vdash f \; V}{\Gamma \vdash h \tilde{w} \Rightarrow V \eta} \quad \text{(inf-app)}
\]

Again, one can easily check that the ordering \( \sqsubseteq \) on substitutions is a quasi-ordering.

In order to compute the most general type of a term, we make the following assumptions:

- Every solvable subtyping problem \( P \) has a most general solution \( \text{mgs}(P) \).
- There is an algorithm for deciding whether a subtyping problem is solvable and, if so, computing its most general solution.

We will see in Section 9 that these assumptions are satisfied when types are annotated in the successor algebra. On the other hand, they are not generally satisfied in an algebra with addition.

Now, following Hindley (1969), the computation of the most general type is defined by the rules of Figure 5 where \( \Gamma \vdash f \; U \) means that, in the typing environment \( \Gamma \), the most general type of \( f \) is \( U \). In the case of an application \( h \tilde{w} \), the algorithm proceeds as follows:

1. Check that \( h \) is declared. Let \( T \) be its declared type.
2. Check that \( h \) can take \( n = |\tilde{w}| \) arguments, i.e. \( T \) is of the form \( \tilde{V} \Rightarrow V \) with \( |\tilde{V}| = n \).
3. If \( h \) is a function symbol equivalent to \( f \), check that \( |\tilde{V}| \geq q^h \).
4. Try to infer the types of every \( w_i \).
5. If this succeeds with \( U_i \) for the type of \( w_i \), then rename the variables of every \( U_i \) using a permutation \( \rho_i \), so that, for all \( i \), \( U_i\rho_i \) has no variable in common with \( T \) and, for all \( i \neq j \), \( U_i\rho_i \) and \( U_j\rho_j \) have no variable in common.
6. Finally, try to compute the most general solution \( \eta \) of the problem \( \{ U_1\rho_1 \leq \eta V_1, \ldots, U_n\rho_n \leq \eta V_n \} \) and return \( V_\eta \) in case of success.
Example 9  To carry on with Example 3, let \( r = s \ (\text{div} \ (\text{sub} \ x \ y) \ (s \ y)) \) be the right-hand side of the last rule of Figure 1. We would like to infer the type of \( \Gamma = [x : N_x, y : N] \) when \( s : N_y \Rightarrow N_{az}, \text{sub} : N_x \Rightarrow N \Rightarrow N_x \) and \( \text{div} : N_x \Rightarrow N \Rightarrow N_x \). Let \( \Gamma \vdash t \rightarrow \text{div} \) and assume wlog that \( x \) is a constant of the successor algebra.

By (inf-app), we get (1) \( \Gamma \vdash x \rightarrow N_x \) and (2) \( \Gamma \vdash y \rightarrow N \). From (1) and (2), by (inf-app), we get (3) \( \Gamma \vdash x \, y \rightarrow N_x \) since, as we shall see in Section 9.2, \( \text{mgs}\{N_x \leq^\# N_x, N \leq^\# N\} = \{(x, x)\} \). From (2), by (inf-app), we get (4) \( \Gamma \vdash s \, y \rightarrow N \) since \( \text{mgs}\{N_x \leq^\# N_x, N \leq^\# N\} = \{(x, x)\} \). From (3) and (4), we get (5) \( \Gamma \vdash \text{div} \ (x \, y) \ (s \, y) \rightarrow N_x \) since \( \text{mgs}\{N_x \leq^\# N_x, N \leq^\# N\} = \{(x, x)\} \).

We now prove that this algorithm is correct and complete wrt \( \Gamma \vdash t \rightarrow \) when the size algebra is monotone and the algorithm is applied to an environment \( \Gamma \) having no size variables. To extend in the next section this result to \( \Gamma \vdash t \rightarrow \), we need to make derivations explicit:

**Definition 22 (Derivation)** Derivations of \( \Gamma \vdash t : T \) are defined as follows:

- If \( (h, \hat{V} \Rightarrow V) \in \Gamma \cup \Theta \) and, for all \( i, \pi_i \) is a derivation of \( \Gamma \vdash w_i : V_i \psi_i \), written \( \pi_i \triangleright \Gamma \vdash w_i : V_i \psi_i \), then \( a(\Gamma, h \, \hat{w}, \psi_i, \hat{\pi}) \) is the derivation of \( \Gamma \vdash h \, \hat{w} : V \psi \) whose last rule is (app).
- If \( \pi \triangleright \Gamma, x : U \vdash v : V \), then \( l(\pi) \) is the derivation of \( \Gamma \vdash \lambda x \uparrow U v : U \Rightarrow V \) whose last rule is (lam).
- If \( \pi \triangleright \Gamma \vdash t : U \) and \( U \leq V \), then \( s(\pi, V) \) is the derivation of \( \Gamma \vdash t : V \) whose last rule is (sub).

Similarly, derivations of \( \Gamma \vdash t \rightarrow T \) are defined as follows:

- If \( (h, \hat{V} \Rightarrow V) \in \Gamma \cup \Theta, \hat{\rho} \) are permutations satisfying the conditions of (inf-app) and, for all \( i, \pi_i \triangleright \Gamma \vdash w_i \rightarrow U_i \), then \( i(\Gamma, h \, \hat{w}, \hat{\rho}, \hat{\pi}) \) is the derivation of \( \Gamma \vdash h \, \hat{w} \rightarrow V \eta \) whose last rule is (inf-app).
- If \( \pi \triangleright \Gamma, x : U \vdash v \rightarrow V \), then \( l(\pi) \) is the derivation of \( \Gamma \vdash \lambda x \uparrow U v \rightarrow U \Rightarrow V \) whose last rule is (inf-lam).

Given a derivation \( \pi \) for \( \pi \triangleright \Gamma \vdash t : T \), we write \( \pi \triangleright \Gamma \vdash t : T \rightarrow \) if, at every (app) node in \( \pi \), the decreasingness condition of (app-decr) is satisfied, that is, if \( \Gamma \vdash t : T \rightarrow \).

Note that \( \Gamma \vdash t \rightarrow \) has at most one derivation.

**Lemma 18** If \( \pi \triangleright \Gamma \vdash t : T \), then, for every size substitution \( \varphi \), \( \pi \varphi \triangleright \Gamma \varphi \vdash t : T \varphi \).

*Proof* Straightforward induction using the fact that \( \leq_A \) and thus \( \leq_A^\# \) and \( \leq \) are stable by substitution. In the case of (app), by induction hypothesis, we have \( \Gamma \varphi \vdash t_i : V_i \psi'_i \) with \( \psi'_i = \psi_i \varphi \). Therefore, by (app), \( \Gamma \varphi \vdash h \, \hat{w} : V \psi' = (V \psi) \varphi \).

**Lemma 19 (Correctness wrt. \( \vdash \))** If \( \pi \triangleright \Gamma \vdash t \rightarrow T \) and \( \text{Var}(\Gamma) = \emptyset \), then there is \( |\pi| \) such that \( |\pi| \triangleright \Gamma \vdash t : T \). In particular, for (inf-app), \( |i(\Gamma, h \, \hat{w}, \hat{\rho}, \hat{\pi})| = a(\Gamma, h \, \hat{w}, \eta, \hat{v}) \), where \( v = s(|\pi|, \rho, \eta, V \eta) \).
Proof
By induction on $\Gamma \vdash^f t \uparrow T$. We only detail the case (inf-app). By induction hypothesis, $\Gamma \vdash^f w_i : U_i$. By Lemma 18, $\Gamma \vdash^f w_i : U_i \rho \eta$ since $\text{Var}(\Gamma) = \emptyset$. Since $U_i \rho \eta \leq V \eta$, by (sub), $\Gamma \vdash^f w_i : V \eta$. Therefore, by (app), $\Gamma \vdash^f h \hat{w} : V \eta$.

Lemma 20 If $\phi \leq^\infty_\$ \psi$ and, for all $\alpha$, $\text{Pos}(\alpha, T) \subseteq \text{Pos}^+(T)$, then $T \phi \leq T \psi$.

Proof
We say that $T \in \mathbb{T}_A \cup \lambda$ is $\delta \in \{+, -\}$ if, for all $\alpha$, $\text{Pos}(\alpha, T) \subseteq \text{Pos}^\delta(T)$. We first prove that $(\ast)$ if $a$ is $\delta$ then $a \phi (\leq^\infty_\$ \delta a \psi$ where, for any relation $R$, $R^+ = R$ and $R^- = R^\ast$. We proceed by induction on $a$:

- $a$ is a variable $\alpha$. Then, $\delta = +$ and $\alpha \phi \leq^\infty_\$ \alpha \psi$ since $\phi \leq^\infty_\$ \psi$.
- $a = f a_1 \ldots a_n$. Let $i \in \{1, \ldots, n\}$. If $i \in \text{Mon}^f(\$)$ (cf. Definition 13), then $a_i$ is $\delta \epsilon$ and, by induction hypothesis, $a_i \phi (\leq^\infty_\$ a_i \psi$. If $i \notin \text{Mon}^+(\$) \cup \text{Mon}^-(\$)$ then $a_i$ contains no variable and $a_i \phi = a_i \psi$. Therefore, by monotony of $\$ in every $i \in \text{Mon}^+(\$)$, anti-monotony of $\$ in every $i \in \text{Mon}^-(\$)$ and transitivity, we get $a \phi (\leq^\infty_\$ a \psi$.

We now prove that, if $T$ is $\delta$, then $T \phi \leq^\delta T \psi$, by induction on $T$.

- $T = U \Rightarrow V$. Then, $U$ is $-\delta$ and $V$ is $\delta$. So, by induction hypothesis, $U \phi \leq^\delta U \psi$ and $V \phi \leq^\delta V \psi$. Therefore, by (prod), $(U \Rightarrow V) \phi \leq^\delta (U \Rightarrow V) \psi$.
- $T = B_a$. Then, $a$ is $\delta$ and, by $(\ast)$, $a \phi \leq^\infty_\$ a \psi$. Therefore, by (size), $T \phi \leq T \psi$.

Lemma 21 (Completeness wrt $\vdash^f$) In monotone algebras, if $\pi \triangleright \Gamma \vdash^f t : T$ and $\text{Var}(\Gamma) = \emptyset$, then there are $U$ and $\pi \uparrow$ such that $\pi \uparrow \triangleright \Gamma \vdash^f t \uparrow U$ and $U \subseteq T$. In particular, $a(\Gamma, h \hat{w}, \hat{\rho}, \hat{\pi} \uparrow) = i(\Gamma, h \hat{w}, \hat{\rho}, \hat{\pi} \uparrow)$ where $\hat{\rho}$ are permutations satisfying the conditions of rule (inf-app).

Proof
We proceed by induction on $\Gamma \vdash^f t : T$. We only detail the case (app) when $h \in \mathbb{C} \cup \mathbb{F}$. By induction hypothesis, $\Gamma \vdash^f w_i \uparrow U_i$ and there is $\chi_i$ such that $U_i \chi_i \leq V \psi$. Wlog. we can assume that $\text{dom}(\chi_i) \subseteq \text{Var}(U_i)$. Let now $\rho_1, \ldots, \rho_n$ be permutations satisfying the conditions of (inf-app), and $\zeta = \{(z, z \psi) \mid z \in \text{Var}(\hat{V} \Rightarrow V)\} \cup \{(z, z \rho_i^{-1} \chi_i) \mid z \in \text{Var}(U_i \rho_i), 1 \leq i \leq n\}$. Then, for all $i$, $U_i \rho_i \chi_i = U_i \chi_i \leq V \psi = V_i \zeta$. Therefore, $P = \{U_1 \rho_1 \leq^2 V_1, \ldots, U_n \rho_n \leq^2 V_n\}$ is solvable, $\eta = \text{mgs}(P)$ exists and there is $\chi$ such that $\eta \chi \leq^\infty_\$ \zeta$. Hence, by (inf-app), $\Gamma \vdash^f h \hat{w} \uparrow \eta \psi$. By the monotony condition, variables occur only positively in $V$. Therefore, by Lemma 20, $V \eta \chi \leq V \zeta = V \psi$. Hence, $V \eta \subseteq V \psi$.

7.2 Decidability of $\vdash^f_\$\phi$
We now prove that, when the size algebra is monotone, for checking $\Gamma \vdash^f_\$ t : T$, it is sufficient to check whether there are $U$ and $\chi$ such that $\Gamma \vdash^f t \uparrow U$, $U \chi \leq T$ and also that, if one denotes by $v$ the (unique) derivation of $\Gamma \vdash^f t \uparrow U$, then
\( |\nu|\chi \triangleright \Gamma \vdash_{\varphi} t : U\chi \), that is, at every (app) node in \( |\nu|\chi \), the decreasingness condition is satisfied.

**Lemma 22** In monotone algebras, if \( \pi \triangleright \Gamma \vdash f : T, \pi \xi \triangleright \Gamma \vdash_{\varphi} t : T \xi \) and \( \zeta \leq_{h}^{\infty} \zeta' \), then \( \pi \chi \triangleright \Gamma \vdash_{\varphi} t : T \xi' \).

**Proof** By induction on \( \pi \triangleright \Gamma \vdash f : T \), we only detail the case (app) when \( h \simeq_{\eta} f \). We have \( (h, \dot{V} : V) \in \Theta, \pi i \triangleright \Gamma \vdash_{\varphi} w_i : V_i \psi, \pi \xi \triangleright \Gamma \vdash f : V_i \gamma \) and \( \tilde{\alpha} \psi \zeta <_{h} \tilde{\beta} \psi \zeta' \). By induction hypothesis, \( \pi \chi \triangleright \Gamma \vdash_{\varphi} w_i : V_i \psi \). Since \( \zeta \leq_{h}^{\infty} \zeta' \) and the size algebra is monotone, we have \( \psi \zeta \leq_{h}^{\infty} \psi \zeta' \). Therefore, \( \pi \chi \triangleright \Gamma \vdash_{\varphi} w_i : V_i \psi \). By Lemma 21, \( \pi \chi \triangleright \Gamma \vdash_{\varphi} t : U\chi \).

**Lemma 23 (Completeness wrt \( \vdash_{\varphi} \))** Let \( A \) be a monotone algebra. Assume that \( \pi \triangleright \Gamma \vdash_{\varphi} t : T \) and \( \text{Var}(\Gamma) = \emptyset \). By lemma 21, there are \( U \) and \( \chi \) such that \( \pi \uparrow \Gamma \vdash_{\varphi} t \uparrow U \) and \( U\chi \leq_{h} T \). Then, \( \pi \uparrow |\chi| \triangleright \Gamma \vdash_{\varphi} t : U\chi \).

**Proof** We prove that if \( \pi \triangleright \Gamma \vdash_{\varphi} t : T \), \( \text{Var}(\Gamma) = \emptyset \), \( \pi \uparrow \Gamma \vdash_{\varphi} t \uparrow U \) and \( U\chi \leq_{h} T \), then \( s(|\pi|\chi, T) \triangleright \Gamma \vdash_{\varphi} t : T \), by induction on \( \pi \triangleright \Gamma \vdash_{\varphi} t : T \). We only detail the case (app-decr) when \( \eta \) is given by the rule (inf-app). We have \( \pi = a(\Gamma, h \dot{w}, \psi, \tilde{\eta}) \), \( \pi \uparrow = i(\Gamma, h \dot{w}, \tilde{\rho}, \tilde{\eta} \uparrow) \), \( |\pi \uparrow| = a(\Gamma, h \dot{w}, \eta, \tilde{\nu}) \) where \( v_i = s(|\pi_i|\chi_i, V_i \eta) \), \( |\pi \uparrow| \chi = a(\Gamma, h \dot{w}, \tilde{\eta}\chi, \tilde{\nu}\chi) \) and, for all \( i \), \( \pi i \triangleright \Gamma \vdash_{\varphi} w_i : V_i \psi, \pi i \uparrow \Gamma \vdash_{\varphi} w_i \uparrow U_i \) and \( U_i \chi_i \leq_{h} V_i \psi \) for some \( \chi_i \). By induction hypothesis, \( s(|\pi_i|\chi_i, V_i \psi) \triangleright \Gamma \vdash_{\varphi} w_i : V_i \psi \). In particular, \( |\pi_i|\chi_i \triangleright \Gamma \vdash_{\varphi} w_i : U_i \chi_i \), that is, \( |\pi_i|\rho_i \chi_i \triangleright \Gamma \vdash_{\varphi} w_i : U_i \rho_i \chi_i \), where \( \chi_i \) is defined in the proof of Lemma 21. Since \( \kappa \chi \leq_{h} \zeta', \pi \chi \leq_{h} \zeta' \), by Lemma 22, we get \( |\pi_i|\rho_i \eta \chi \triangleright \Gamma \vdash_{\varphi} w_i : U_i \rho_i \eta \chi \). Hence, \( v_i \chi \triangleright \Gamma \vdash_{\varphi} w_i : V_i \eta \chi \). Moreover, since \( \tilde{\alpha}_h \eta \chi \leq_{h} \zeta' \tilde{\alpha}_h \zeta = \tilde{\alpha}_h \psi \) and \( \tilde{\alpha}_h \psi \leq_{h} \tilde{\beta}_h \phi \), we get \( \tilde{\alpha}_h \eta \chi \leq_{h} \tilde{\beta}_h \phi \) by assumption on \( \leq_{h} \). Therefore, \( s(|\pi|\chi, T) \triangleright \Gamma \vdash_{\varphi} t : T \).

The previous lemmas assume that there are no size variables in \( \Gamma \). So, to use these lemmas, we need to be able to replace size variables by constants (aka eigenvariables). Under this assumption, we can conclude:

**Theorem 2 (Decidability of \( \vdash_{\varphi} \))** Assume that \( A \) is an algebra such that

- \( A \) is monotone;
- \( A \) contains an infinite set of constants \( C \) such that, if \( a \leq_{h} b \) (\( \leq_{h} \)) resp. then, for all \( c \in C \) and \( e \in A \), \( a \delta <_{h} b \delta \) (\( \delta \)) resp., where \( \delta \) replaces every \( c \) by \( e \);
- \( \leq_{h} \) is decidable and, for all \( g \simeq_{\eta} f \), \( <_{h} \) is decidable;
- the satisfiability of a subtyping problem is decidable;
- every satisfiable problem \( P \) has a most general solution \( \text{mgs}(P) \) that is computable.

Given \( \Gamma, t \) and \( T \), one can decide whether \( \Gamma \vdash_{\varphi} t : T \) by using the algorithm of Figure 6.
Proof

- Correctness. Assume that the algorithm succeeds. Then, \( \Gamma \vdash^f \gamma \vdash t : U \chi \) and \( U \chi \leq T \gamma \). By (sub), \( \Gamma \vdash^f \gamma \vdash t : T \gamma \). Then, by applying \( \delta = \gamma^{-1} \), we get \( \Gamma \vdash^f \gamma \vdash t : T \).

- Completeness. Assume that the algorithm fails in step 1 or 2 then, by Lemma 21, \( t \) is not typable in \( \Gamma \gamma \). Therefore, it is not typable in \( \Gamma \) either. Finally, if the algorithm fails in step 3 then, by Lemma 23, there is no derivation of \( \Gamma \gamma \vdash^f \gamma \vdash t : T \gamma \). Therefore, there is no derivation of \( \Gamma \vdash^f \gamma \vdash t : T \) either.

\[\Box\]

1. Check whether there is \( U \) such that \( \Gamma \gamma \vdash^f \gamma \vdash t \uparrow U \), where \( \gamma \) is an injection from the set of size variables of \( \Gamma, T \) and \( \varphi \) to the set of constants \( C \) not occurring in \( \Gamma, T \) and \( \varphi \). If it fails, then \( t \) is not typable in \( \Gamma \).
2. If it succeeds, try to compute \( \chi = \text{mgs}\{U \leq T \gamma \} \). If it fails, \( \Gamma \vdash^f \gamma \vdash t : T \) does not hold.
3. If it succeeds, then check whether \( |\varphi| \chi > \Gamma \gamma \vdash^f \gamma \vdash U \chi \) where \( \chi \) is the unique derivation of \( \Gamma \gamma \vdash^f \gamma \vdash t \uparrow U \).
   If it succeeds, then \( \Gamma \vdash^f \gamma \vdash t : T \) holds. Otherwise, \( \Gamma \vdash^f \gamma \vdash t : T \) does not hold.

Fig. 6. Algorithm for deciding whether \( \Gamma \vdash^f \gamma \vdash t : T \).

That the successor algebra satisfies the first two conditions follows from Lemma 25.

Example 10 To carry on with Example 9, we now would like to check whether \( \Gamma \vdash r : N_{s \chi} \) where \( |= \vdash^f \chi \) and \( \varphi = \{s x, s x\} \). We have seen that \( \Gamma \vdash^f \chi \vdash r \uparrow N_{s x} \). Hence, \( \chi \) is the identity and we are left to check that, in every (app) node with \( h \cong_{s \chi} \chi \), the decreasingness condition is satisfied. Here, it amounts to check that, in the (app) node for \( (\text{div} (\text{sub} x y) (s y)) \), the size annotation for the type of \( \text{sub} x y \), that is, \( x \), is smaller than \( s \varphi = s x \), which is indeed the case.

8 Reducing subtyping problems to size problems

For the type inference algorithm we just saw, we assumed the existence of an algorithm to compute the most general solution of a subtyping problem. In this section, we show how a subtyping problem can be reduced to solving constraints in \( \mathcal{A} \). As subtyping is not syntax directed, we first prove that it is equivalent to a syntax-directed relation. To this end, we prove that the rules (refl) and (trans) are redundant, that is, they can be eliminated, following a proof technique used by Curien & Ghelli (1992):

Theorem 3 \( T \leq U \) iff \( T \leq_{\mathcal{A}} U \), where \( \leq_{\mathcal{A}} \) is defined by the rules (size) and (prod) only.
Proof
We first prove that (refl) can be eliminated, hence that ≤ = ≤' where ≤' is the
relation defined by (size), (prod) and (trans) only. Indeed, using the reflexivity of
\(\leq_{\pi}^{\infty}\), one can easily prove that \(T \leq_a T\), by induction on \(T\).
We now prove that, in turn, (trans) can be eliminated, hence that ≤ = ≤_a. More
precisely, we prove that, if \(\pi\) is a derivation of \(A \leq B\) of height \(n\), then \(A \leq_a B\), by
induction on \(n\). We proceed by case on the last rule:

(size) Immediate.

(prod) Assume that \(U \Rightarrow V \leq' U' \Rightarrow V'\) ends with (prod). By induction hypothesis,
\(U' \leq_a U\) and \(V \leq_a V'\). Hence, by (prod), \(U \Rightarrow V \leq_a U' \Rightarrow V'\).

(trans) Assume that \(T \leq' U\) and \(U \leq' V\). By induction hypothesis, \(T \leq_a U\) and
\(U \leq_a V\). If \(T \leq_a U\) ends with (size), then \(T = B_a\), \(U = B_b\) and \(a \leq_{\pi}^{\infty} b\).
Therefore, \(U \leq_a V\) ends with (size) too, \(V = B_c\) and \(b \leq_{\pi}^{\infty} c\). Hence, by
transitivity of \(\leq_{\pi}^{\infty}\), \(T \leq_a V\). Similarly, if \(U \leq_a V\) ends with (size), then
\(T \leq_a U\) ends with (size) and \(T \leq_a V\). So, we are left with the case where both \(T \leq_a U\) and \(U \leq_a V\) ends with (prod):

\[
\begin{array}{cccc}
\pi_{11} & \pi_{12} & \pi_{21} & \pi_{22} \\
A' \leq_a A & B \leq_a B' & A'' \leq_a A' & B' \leq_a B'' \\
\hline
A \Rightarrow B \leq_a A' \Rightarrow B' & (\text{prod}) & A' \Rightarrow B' \leq_a A'' \Rightarrow B'' & (\text{prod}) & (\text{trans}) \\
\end{array}
\]

But \(A \Rightarrow B \leq' A'' \Rightarrow B''\) can also be proved as follows:

\[
\begin{array}{cccc}
\pi_{21} & \pi_{11} & \pi_{12} & \pi_{22} \\
A'' \leq_a A' & A' \leq_a A & B \leq_a B' & B' \leq_a B'' \\
\hline
A'' \leq' A & (\text{trans}) & B \leq' B'' & (\text{trans}) & (\text{prod}) \\
\hline
A \Rightarrow B \leq' A'' \Rightarrow B'' \\
\end{array}
\]

The derivation heights of \(A'' \leq' A\) and \(B \leq' B''\) are strictly smaller than the
derivation height of \(A \Rightarrow B \leq' A'' \Rightarrow B''\). Therefore, by induction hypothesis,
\(A'' \leq_a A\) and \(B \leq_a B''\). Hence, by (prod), \(A \Rightarrow B \leq_a A'' \Rightarrow B''\).

□

As a consequence, we can prove that a subtyping problem can be reduced to an
equivalent size problem as follows:

Definition 23 (Size problem) A size constraint is a pair of size expressions \((a, b)\),
written \(a \leq \_+ b\). A size problem \(P\) is either \(\perp\) or a finite set of size constraints. It has
a solution \(\varphi : V \rightarrow \mathbf{A}\) if \(P \neq \perp\), \(\text{dom}(\varphi) \subseteq \text{Var}(P)\) and, for all \(a \leq \_+ b \in P\), \(a\varphi \leq_{\pi} b\varphi\).
A solution \(\varphi\) is finite if \(\varphi : V \rightarrow \mathbf{A}\). Let \(\text{Sol}_{\pi}(P)\) (\(\text{Sol}_{\pi}(P)\) resp.) be the set of the (finite
resp.) solutions of \(P\).

We define the size problem associated to a subtyping problem as follows:

- \(|\emptyset| = \emptyset\),
- \(|P \cup Q| = |P| \cup |Q|\) if \(|P| \neq \perp\) and \(|Q| \neq \perp\),
Lemma 24 \( \text{Sol}(P) = \text{Sol}_\mathcal{P}(|P|) \).

Proof
We proceed by induction on \( P \). We only detail the case where \( P = \{ T \leq^? T' \} \):

- Let \( \varphi \in \text{Sol}(P) \). Then, \( T \varphi \leq_a T' \varphi \). If \( T = B_a \), then \( T' = B_b \) and \( \varphi \in \text{Sol}_\mathcal{P}(\{a \leq^? b\}) = \text{Sol}_\mathcal{P}(|P|) \). Otherwise, \( T = U \Rightarrow V \), \( T' = U' \Rightarrow V' \) and \( \varphi \in \text{Sol}(\{U \leq^? U', V \leq^? V'\}) \). By induction hypothesis, \( \varphi \in \text{Sol}_\mathcal{P}(|U' \leq^? U|) \cap \text{Sol}_\mathcal{P}(|V \leq^? V'|) = \text{Sol}_\mathcal{P}(|P|) \).

- Let \( \varphi \in \text{Sol}_\mathcal{P}(|P|) \). If \( T = B_a \), then \( T' = B_b \) and \( \varphi \in \text{Sol}(P) \). Otherwise, \( T = U \Rightarrow V \), \( T' = U' \Rightarrow V' \), \( \varphi \in \text{Sol}_\mathcal{P}(|U' \leq^? U|) \cap \text{Sol}_\mathcal{P}(|V \leq^? V'|) \). By induction hypothesis, \( \varphi \in \text{Sol}(U' \leq^? U) \cap \text{Sol}(V \leq^? V') = \text{Sol}(P) \).

To go further, we need to make more assumptions on the size algebra.

9 Solving size problems in the successor algebra

We have seen in the previous section that solving a subtyping problem can be reduced to solving inequalities in \( \mathcal{A} \). In this section, we consider a specific size algebra \( \mathcal{A} \), the successor algebra, and prove that, in this algebra, the solvability of a size problem is decidable in polynomial time, and that solvable size problems have a most general solution that can be computed in polynomial time too.

The relations \( \leq_{\mathcal{A}} \) and \( <_{\mathcal{A}} \) of the successor algebra (Definition 12) are equivalently defined by the rules of Figure 7. We start by proving basic properties of \( \leq_{\mathcal{A}} \), the quasi-ordering \( \sqsubseteq \) and its associated equivalence relation \( \equiv \) on size substitutions introduced in Definition 21.

\[
\begin{array}{cccc}
\hline
& a <_\mathcal{A} b & & a <_\mathcal{A} b <_\mathcal{A} c \\
\hline
a \leq_{\mathcal{A}} a & & a \leq_{\mathcal{A}} b & & a <_\mathcal{A} a <_\mathcal{A} c \\
\hline
\end{array}
\]

Fig. 7. Ordering in the successor algebra.

Lemma 25

- \( a \leq_{\mathcal{A}} b \) \( (a <_\mathcal{A} b \text{ resp.}) \) iff there is \( k \geq 0 \) \( (k > 0 \text{ resp.}) \) such that \( b = s^k a \).
- \( sa <_\mathcal{A} sb \) iff \( a <_\mathcal{A} b \).

Proof

- One can easily check \( a \leq_{\mathcal{A}} s^k a \) by induction on \( k \geq 0 \). We have \( a \leq_{\mathcal{A}} a \) by definition. Assume now that \( a \leq_{\mathcal{A}} s^k a \). Since \( s^k a <_\mathcal{A} s^{k+1} a \) holds by definition, we get \( a \leq_{\mathcal{A}} s^{k+1} a \) by transitivity.
Similarly, once can easily check \( a <_\mathbb{A} s^k a \) by induction on \( k \geq 1 \). We have \( a <_\mathbb{A} sa \) by definition. Assume now that \( a <_\mathbb{A} s^k a \). Since \( s^k a <_\mathbb{A} s^{k+1} a \) by definition, we get \( a <_\mathbb{A} s^{k+1} a \) by transitivity.

We now prove that, if \( a <_\mathbb{A} b \), then there is \( b' \) such that \( b = sb' \) and \( a \leq_\mathbb{A} b' \), by induction on the derivation height of \( a <_\mathbb{A} b \). If \( b = sa \), then this is immediate. Otherwise, there is \( c \) such that \( a <_\mathbb{A} c \) and \( c <_\mathbb{A} b \). By induction hypothesis, there is \( b' \) such that \( b = sb' \) and \( c \leq_\mathbb{A} b' \). Therefore, \( a \leq_\mathbb{A} b' \) since \( \leq_\mathbb{A} \) is the reflexive closure of \( <_\mathbb{A} \) and \( <_\mathbb{A} \) is transitive.

We finally prove that there is \( k \geq 0 \) whenever \( a \leq_\mathbb{A} b \), by induction on \( b \). If \( a = b \), this is immediate. If \( a <_\mathbb{A} b \), then there is \( b' \) such that \( b = sb' \) and \( a \leq_\mathbb{A} b' \).

By induction hypothesis, \( b' = s^k a \) for some \( k \geq 0 \). Therefore, \( b = s^{k+1} a \).

- If \( sa <_\mathbb{A} sb \), then \( sb = s^{k+1} sa \) for some \( k \). Therefore, \( b = s^{k+1} a \). Conversely, if \( a <_\mathbb{A} b \), then \( b = s^{k+1} a \) for some \( k \). Therefore, \( sb = s^{k+1} sa \).

\[ \qed \]

It follows that the successor algebra is monotone and also that \( \leq_\mathbb{A} \) and \( \leq^\mathbb{A}_\mathbb{A} \) are orderings, as well as their pointwise extensions to substitutions.

**Definition 24 (Successor and head parts of a substitution)** To a substitution \( \varphi : V \to \overline{\mathbb{A}} \), we associate two unique maps \( \varphi_s : V \to \mathbb{N} \) and \( \varphi_h : V \to V \cup C \cup \{\infty\} \) such that, for all \( x \), \( x\varphi = s^{\varphi_s}x\varphi_h \) with \( \varphi_s = 0 \) if \( \varphi_h = \infty \).

**Lemma 26** \( \varphi \sqsubseteq \psi \) iff there is \( \rho : V \to V \cup C \cup \{\infty\} \) such that \( \varphi \rho \leq^\mathbb{A}_\mathbb{A} \psi \).

**Proof**
The “if” part is immediate. We now prove the “only if” part. Assume that there is \( \theta \) such that \( \varphi \theta \leq^\mathbb{A}_\mathbb{A} \psi \). Let \( \rho = \theta_{\varphi_s}|_{\operatorname{Var}(\varphi_h)} \), where \( \operatorname{Var}(\varphi_h) = \bigcup\{\operatorname{Var}(x\varphi_h) \mid x \in \operatorname{dom}(\varphi_h)\} \).

We now check that \( \varphi \rho \leq^\mathbb{A}_\mathbb{A} \psi \). If \( x\varphi \notin V \), then \( x\varphi = x\varphi \theta \leq^\mathbb{A}_\mathbb{A} x\psi \). Otherwise, \( x\varphi = s^{\varphi_s}x\varphi_h \theta_h \leq^\mathbb{A}_\mathbb{A} s^{\varphi_s}x\varphi_h \theta_h = x\varphi \theta \leq^\mathbb{A}_\mathbb{A} x\psi \).

\[ \qed \]

**Lemma 27** Let \( V \) be a set, and \( V_1 \) and \( V_2 \) be subsets of \( V \). If \( \rho_1 : V_1 \to V_2 \) and \( \rho_2 : V_2 \to V_1 \) are injections, then there is a permutation \( \xi : V \to V \) such that \( \xi |_{V_1} = \rho_1 \).

**Proof**
By Cantor–Bernstein theorem, \( V_1 \) and \( V_2 \) are equipotent. Hence, \( V_1 - V_2 \) and \( V_2 - V_1 \) are equipotent as well. Let \( v \) be any bijection from \( V_2 - V_1 \) to \( V_1 - V_2 \), and \( \xi = \{(x, \rho_1) \mid x \in V_1\} \cup \{(x, v) \mid x \in V_2 - V_1\} \). The function \( \xi \) is a bijection on \( V_1 \cup V_2 \) and \( \xi |_{V_1} = \rho_1 \).

\[ \qed \]

**Lemma 28** \( \varphi_2 \equiv \varphi_1 \) iff \( \varphi_2 = \varphi_1 \xi \) for some permutation \( \xi : V \to V \).

**Proof**
If “if” part is immediate. We now prove the “only if” part. Huet (1976) proved this result when \( \leq^\mathbb{A}_\mathbb{A} \) is the equality. His proof can be adapted to our more general situation since \( x \leq^\mathbb{A}_\mathbb{A} \beta \) iff \( x = \beta \). By assumption and Lemma 26, there are \( \rho_1, \rho_2 : \)}
\( V \rightarrow V \cup C \cup \{ \infty \} \) such that \( \varphi_1 \rho_1 \preceq^\infty \varphi_2 \) and \( \varphi_2 \rho_2 \preceq^\infty \varphi_1 \). By stability, we have \( \varphi_1 \rho_1 \rho_2 \preceq^\infty \varphi_2 \rho_2 \). Hence, by transitivity, \( \varphi_1 \rho_1 \rho_2 \preceq^\infty \varphi_1 \). Similarly, \( \varphi_2 \rho_2 \rho_1 \preceq^\infty \varphi_2 \).

We now prove that \( \rho_1 \) is a function from \( V_1 \) to \( V_2 \), where \( V_1 = \{ \text{Var}(\beta \phi) \mid \beta \in V \} \) and \( V = \text{dom}(\varphi_1) \cup \text{dom}(\varphi_2) \). Let \( x \in V_1 \). Then, there is \( \beta \in V \) such that \( x \in \text{Var}(\beta \phi) \). Hence, \( \beta \phi_1 = s^k x \) for some \( k \in \mathbb{N} \). Since \( \varphi_1 \rho_1 \rho_2 \preceq^\infty \varphi_1 \), we have \( \beta \phi_1 \rho_1 \rho_2 = s^k \rho_1 \rho_2 \preceq^\infty \beta \phi_1 = s^k x \). Therefore, \( x \rho_1 \rho_2 = x \) and \( \rho_1 \) is a function on \( V_1 \).

We now prove that \( \varphi_1 \) is immediate. Otherwise, we proceed by case on \( \text{algebra} \):

**Definition 25 (Successor–iterator algebra)** Let \( \varphi_1 \rho_1 \rho_2 \preceq^\infty \varphi_2 \rho_2 \). Hence, by transitivity, \( \varphi_1 \rho_1 \rho_2 \preceq^\infty \varphi_1 \). Similarly, \( \varphi_2 \rho_2 \rho_1 \preceq^\infty \varphi_2 \).

To check whether a problem is satisfiable, we are going to introduce a terminating rewrite system that will put the problem into some normal form whose satisfiability is easy to establish. To do so, we first need to extend the successor algebra as follows:

**9.1 Satisfiability**

To check whether a problem is satisfiable, we are going to introduce a terminating rewrite system that will put the problem into some normal form whose satisfiability is easy to establish. To do so, we first need to extend the successor algebra as follows:

**Definition 25 (Successor–iterator algebra)** Let \( B \) be the following multi-sorted algebra:

- **Sorts**: \( A \) interpreted by \( h \), and \( N \) interpreted by \( 0 \).
- **Function symbols**: \( 0 : N \rightarrow N \) and \( s : A \rightarrow A \) interpreted by the successor function, \( c : A \) for every \( c \in C \), \( s : N \times A \rightarrow A \), with \( s(a, b) \) written \( s^a b \), interpreted as the iteration of the successor function: \( s^a b = b + a \).
- **Variables**: the variables \( x, y, \ldots \in V \) are of sort \( A \). In addition, we assume given a set \( V_\infty \), disjoint from \( V \cup C \), of variables \( x, y, \ldots \) of sort \( \infty \), and an injection \( x : V \rightarrow V_\infty \).
- **Equality**: \( \preceq_B = \preceq_A \).
- **Congruence**: \( \simeq_B = \simeq_A \cup \simeq_A \) where \( \simeq_A \) is the smallest congruence satisfying the following semantically valid equations on terms of sort \( A \):

\[
\begin{align*}
s^a x & \simeq_A x \\
{s^a x} x & \simeq_A s(s^a x) \\
s^x(sz) & \simeq_A s(s^x z)
\end{align*}
\]
In the top-extension of \( B \), the symbol \( \infty \) is of sort \( A \). Let \( \text{Var}_A(a) \) be the variables of sort \( s \) occurring in \( a \). A problem is constant-free if it contains no constants \( c \in C \).

Note that, in a multi-sorted algebra, substitutions map a variable of sort \( s \) to a term of sort \( s \) (hence a substitution cannot map a variable of sort \( \mathbb{N} \) to \( \infty \)), and constraints are pairs of terms of the same sort. A problem is of sort \( s \) if all its constraints are of sort \( s \).

One can easily check that, when oriented from left to right, the equations defining \( \simeq_A \) form a confluent and terminating rewrite system. Hence, every term has a unique normal form and two equivalent terms have the same normal form. So, wlog, we can always assume that terms are in normal form, in which case \( \simeq_A \) is the equality and \( \leq_B \) is \( \leq_A \).

In the following, we use the letters \( e \) and \( f \) (\( k \) and \( l \) resp.) to denote arbitrary (closed resp.) terms of sort \( \mathbb{N} \). Closed terms of sort \( \mathbb{N} \) are isomorphic to natural numbers. Hence, we identify \( s_\mathbb{N} \ldots s_\mathbb{N} \cdot 0 \) (\( k \) times \( s_\mathbb{N} \)) with \( k \), denote \( s_\mathbb{N} \ldots s_\mathbb{N} \cdot x \) (\( k \) times \( s_\mathbb{N} \)) by \( x + k \), and call a problem of sort \( \mathbb{N} \) an integer problem. However, \( s^k x \) will not denote \( s^{a_\mathbb{N} \ldots s_\mathbb{N} \cdot 0} \cdot x \) (\( k \) times \( s_\mathbb{N} \)) but its normal form \( s \ldots s \cdot x \) (\( k \) times \( s \)).

Given a problem \( P \) in \( \overline{A} \), since \( \text{Sol}_{\overline{B}}(P) \) may contain solutions not expressible in \( \overline{A} \), we consider the following subset instead:

**Definition 26 (\( \mathbb{N} \)-closed solutions)** A term \( a \in \overline{B} \) is \( \mathbb{N} \)-closed if \( \text{Var}_\mathbb{N}(a) = \emptyset \). A solution to a problem \( P \) is \( \mathbb{N} \)-closed if it maps every \( x \in \text{Var}(P) \) to an \( \mathbb{N} \)-closed term. Let \( \text{Sol}_{\overline{B}}^\mathbb{N}(P) \) (\( \text{Sol}_{\overline{B}}^0(P) \) resp.) be the set of all the \( \mathbb{N} \)-closed (finite resp.) solutions of \( P \).

**Lemma 29**

- A term of sort \( A \) belongs to \( \overline{A} \) iff it is \( \mathbb{N} \)-closed.
- Given a problem \( P \) in \( \overline{A} \), \( \text{Sol}_{\overline{A}}(P) = \text{Sol}_{\overline{B}}^\mathbb{N}(P) \).

**Proof**

- This is immediate if \( a = \infty \). Otherwise, \( a = s \ldots s_\mathbb{N} \cdot x \ldots s_\mathbb{N} \cdot b \) with \( b \in V \cup C \). If \( a \in A \), then \( n = 0 \) and \( a \) is \( \mathbb{N} \)-closed. Conversely, if \( a \) is \( \mathbb{N} \)-closed, then \( n = 0 \) and \( a \in A \).
- Immediate consequence of the previous property.

Note that a \( \mathbb{N} \)-closed solution maps every variable of sort \( \mathbb{N} \) to an integer. Hence, for an integer problem \( P \), \( \text{Sol}_{\overline{B}}^\mathbb{N}(P) = \text{Sol}_{\overline{B}}^\mathbb{N}(P) \) (solutions to integer problems are always finite) and, given \( \varphi, \psi \in \text{Sol}_{\overline{B}}^\mathbb{N}(P) \), \( \varphi \sqsubseteq \psi \) iff \( \varphi \leq_\mathbb{N} \psi \), i.e. for all \( x \in \text{Var}(P) \), \( x_\varphi \leq_\mathbb{N} x_\psi \).

Now, to a problem in \( \overline{A} \), we associate a graph as follows:

**Definition 27 (Graph associated to a problem in \( \overline{B} \))** Let \( \mathbb{H} = V \cup V_\mathbb{N} \cup C \cup \{\infty\} \). To a problem \( P \) in \( \overline{B} \), we associate a directed graph \( G(P) \) on \( \mathbb{H} \cup \{\infty\} \) with the following labeled edges:

- \( x \xrightarrow{k+l} y \) for each constraint \( x + k \leq_\mathbb{N} y + l \in P \) with \( x, y \in V_\mathbb{N} \cup \{\infty\} \).
- \( 0 \xrightarrow{0} y \) for each variable \( y \in \text{Var}_\mathbb{N}(P) \).
• $\alpha^k \beta$ for each constraint $s^k \alpha \leq^? s^l \beta \in P$.
• $\alpha \rightarrow \infty$ for each $\alpha \in \text{Var}_A(P)$.
• $\infty \rightarrow 0 \beta$ for each constraint $\infty \leq^? s^l \beta \in P$.
• $c \rightarrow \beta$ for each constraint $s^c \in P$.

The weight of a path $a_1 \rightarrow^{k_1} \ldots \rightarrow^{k_n} a_{n+1}$ is $\sum_{i=1}^{n} k_i \in \mathbb{N}$. A cycle (i.e. when $a_{n+1} = a_1$) is positive if its weight is $> 0$.

Let $\leq_P$ be the smallest quasi-ordering on $H$ (we exclude $\infty$) such that $a \leq_P b$ if there is a path from $a$ to $b$ in $G(P)$.

A triple $(\alpha, c, d)$ such that $c \leq_P \alpha, d \leq_P \alpha$ and $c \neq d$, is called incompatible.

For instance, the graph of the problem $P = \{c \leq^? \alpha, s\alpha \leq^? \beta, \beta \leq^? \alpha\}$ is

If we replace $\alpha$ by $x_\alpha \in \mathbb{V}_N$, $\beta$ by $x_\beta \in \mathbb{V}_N$ and $c$ by 0, we get the integer problem $I(P) = \{0 \leq^? x_\alpha, x_\alpha + 1 \leq^? x_\beta, x_\beta \leq^? x_\alpha\}$ whose graph is

Following Pratt (1977), an integer problem $P$ has an integer solution iff $G(P)$ has no positive cycles, which can be decided in polynomial time “e.g. by forming the max/+ transitive closure of the graph and searching for a self-edge with a positive label.”

In the graph of $I(P)$, the cycle $x_\alpha \rightarrow_1 x_\beta \rightarrow_0 x_\alpha$ has weight 1 and thus is positive. So, $I(P)$ has no integer solution. On the other hand, $P$ can be solved by taking $\alpha = \beta = \infty$.

Next, we introduce a data structure used to transform an arbitrary problem into a problem in normal form using the rules of Figure 8:

**Definition 28 (Configuration)** A term is admissible if it contains at most one variable. A constraint $a \leq^? b$ is admissible if both $a$ and $b$ are admissible.

A configuration $C$ is $\perp$ or a tuple $(C_0, C_1, C_2, C_3, C_4)$ with

• $C_0 \subseteq \mathbb{V}$,
• $C_1 \subseteq \mathbb{V}$,
• $C_2$ is a finite map from $\mathbb{V}$ to $\mathbb{C}$,
• $C_3$ is a set of admissible constraints of sort $\mathbb{N}$,
• $C_4$ is a set of admissible constraints of sort $\mathbb{A}$,
• $C_0, C_1, \text{dom}(C_2)$ and $\text{Var}(C_4)$ are pairwise disjoint.
Let $\text{Sol}_B^0(C) = \text{Sol}_B^0(\pi(C))$ and $\text{Var}(C) = \text{Var}(\pi(C))$, where $\pi(\bot) = \bot$ and $\pi(C_0, \ldots, C_4)$ is the union of

- $\{x \leq^2 \infty \mid x \in C_0\}$,
- $\{\infty \leq^2 x \mid x \in C_1\}$,
- $\{x \leq^2 s^x c \mid (x, c) \in C_2\} \cup \{s^x c \leq^2 x \mid (x, c) \in C_2\}$,
- $C_3 \cup C_4$. 

$C$ is normal if there is no $D$ such that $C \sim D$ where $\sim$ is defined in Figure 8. Finally, given $C$ and $\varphi$, let

- $\sigma_0(C, \varphi) = \{(x, x\varphi) \mid x \in C_0\}$,
- $\sigma_1(C) = \{(x, \infty) \mid x \in C_1\}$,
- $\sigma_2(C, \varphi) = \{(x, s^x \varphi c) \mid (x, c) \in C_2\}$,
- $\sigma_{3, 4}(C, \varphi) = \{(x, x \varphi) \mid x \in \text{Var}(C_3 \cup C_4)\}$,
- $\sigma_{4a}(C, \varphi) = \{(x, x \varphi) \mid x \in \text{Var}_A(C_4)\}$. 

$C_0$ records the variables with no constraints, $C_1$ records the variables that must be set of $\infty$, $C_2$ records the variables that must be set to a value of the form $s^x c$, $C_3$ contains the constraints on integer variables, and $C_4$ contains all the other constraints.

Note that Figure 8 describes an infinite set of rules since $a$ stands for an arbitrary size expression of sort $\mathbb{A}$, $e$ and $f$ for arbitrary size expressions of sort $\mathbb{N}$, $k$ for an arbitrary natural number, $x$ for an arbitrary size variable of sort $\mathbb{A}$, $c$ and $d$ for arbitrary constants, and $P \cup Q$ for an arbitrary set with two disjoint parts, $P$ and $Q$.

$(-\infty)$ removes the constraints of the form $a \leq^2 \infty$ that are always satisfied, and records in $C_0$ variables not occurring elsewhere.

$(\infty x_1)$ detects variables that must be set to $\infty$ because they belong to a positive cycle.

$(\infty x_2)$ detects variables $x$ that must be set to $\infty$ because some constraints imply that it should otherwise be set to a term of the form $s^x c$ and some other constraints that it should be set to a term of the form $s^y d$ with $c \neq d$.

$(\infty c)$ detects an unsatisfiable constraint of the form $\infty \leq^2 s^x c$.

$(cd)$ detects an unsatisfiable constraint of the form $s^x c \leq^2 s^y d$ with $c \neq d$.

$(cc)$ replaces a constraint of the form $s^x c \leq^2 s^y c$ by the integer constraint $e \leq^2 f$.

$(ac)$ replaces a constraint of the form $s^x x \leq^2 s^x c$ by recording in $C_2$ that $x$ must be set to a term of the form $s^x c$, propagating it in other constraints and recording in integer constraints the fact that $x + k \leq^2 e$.

The rule $(\infty x_2)$ is not necessary for deciding the satisfiability of a problem. It is included here because it is useful to compute a most general solution in next section.
Let $S(C) = \{ \sigma_0(C, \varphi) \cup \sigma_1(C) \cup \sigma_2(C, \psi) \cup \sigma_3, \psi(1, C, \psi) \mid \varphi \text{ N-closed, } \psi \in \text{Sol}_\emptyset^0(C \cup C_4) \}$. One can easily check that $S(C) \subseteq \text{Sol}_\emptyset^0(C_4)$. Assume now that $\varphi \in \text{Sol}_\emptyset^0(C)$. Then, $\varphi = \sigma_0(C, \varphi) \cup \sigma_1(C) \cup \sigma_2(C, \psi) \cup \sigma_3, \psi(1, C, \psi)$ where $\varphi = \varphi|_{\text{Var}(C_4)}$. Indeed, if $x \in C_1$, then $\alpha \leq^e x \in \pi(C)$. Hence, $x \varphi = \infty$. Now, if $(x, c) \in C_2$, then $\pi(C)$ contains $x \leq^e x^c$ and $x^c \varphi = x^c \psi$ since $\{x_\alpha \mid x \in \text{dom}(C_2)\} \subseteq \text{Var}(C_3)$. One can easily check that $(0, 0, 0, 0, P)$ is a configuration. The fact that $\text{Sol}_\emptyset^0(P) = \text{Sol}_\emptyset^0(0, 0, 0, 0, P)$ directly follows from the previous property.


4. a. One can easily check that all the conditions defining what is a configuration are preserved by each rule. In particular, $(zc)$ replaces $x$ by $s^x c$, hence every term of $D$ is admissible if every term of $C$ so is.

b. Straightforward.

c. Straightforward.

d. We only detail the following cases:

- Rule $(zc)$. We have $(s^k x) \psi = s^x v c$. So, $x \varphi = \infty$ and $(s^k x) \psi = s^k(x \psi) \leq_A s^e v c$. By Lemma 25, there is $l$ such that $s^e v c = s^l (x \psi)$.
Hence, there is \( m \) such that \( xy = smc \) and \( ev = l + k + m \). Let now \( \varphi = \psi \cup \{ x, z \} \). We have \( \varphi = \psi = smc = (s + c)\varphi \) and \( (x + k)\varphi = m + k \leq l + k + m = ev = e\varphi \). Therefore, \( \varphi \in Sol^0_\beta(D) \) and \( \varphi|_{\text{Var}(C)} = \psi \).

- Rule \((\infty \alpha_1)\). We first prove that, if \( z_1 \xrightarrow{k_1} \ldots \xrightarrow{k_n} z_{n+1} \) is a path in \( G(Q) \), \( \psi \in Sol^0_\beta(Q) \) and \( k = \Sigma^n_{i=1}k_i \geq 0 \) (\( k < 0 \) resp.), then \( s^k z_1 \psi \leq_k s^{-k} z_{n+1} \psi \) (\( z_1 \psi \leq_k s^{-k} z_{n+1} \psi \) resp.) \((*)\), by induction on \( n \). If \( n = 1 \), this is immediate. We now prove it for \( n + 1 \).

  — Case \( k \geq 0 \). By induction hypothesis, \( s^k z_1 \psi \leq_k z_{n+1} \psi \).

    - Case \( k_{n+1} \geq 0 \). Then, \( s^{k_{n+1}} z_{n+1} \psi \leq_k z_{n+1} \psi \).

      - Case \( k + k_{n+1} \geq 0 \). By monotony and transitivity, \( s^{k+k_{n+1}} z_1 \psi \leq_k z_{n+1} \psi \).

      - Case \( k + k_{n+1} < 0 \). Impossible.

    - Case \( k_{n+1} < 0 \). Then, \( z_{n+1} \psi \leq_k s^{-k_{n+1}} z_{n+2} \psi \) and, by transitivity, \( s^k z_1 \psi \leq_k s^{-k_{n+1}} z_{n+2} \psi \).

      - Case \( k + k_{n+1} \geq 0 \). Since \( -k_{n+1} \leq k \), \( s^{k+k_{n+1}} z_1 \psi \leq_k z_{n+2} \psi \).

      - Case \( k + k_{n+1} < 0 \). Since \( k < -k_{n+1} \), \( z_1 \psi \leq_k s^{-k-k_{n+1}} z_{n+2} \psi \).

  — Case \( k < 0 \). Symmetric to previous case.

Assume now that \( Q \) is constant-free and \( G(Q) \) is a positive cycle. If \( G(Q) \) contains \( \infty \), then \( z_1 \psi = \infty \) for all \( z \in \text{Var}(Q) \). Otherwise, \( G(Q) \) is \( z_1 \xrightarrow{k_1} \ldots \xrightarrow{k_n} z_{n+1} \xrightarrow{k_{n+1}} z_1 \). Hence, \( s^k z_1 \psi \leq_k z_{n+1} \psi \) with \( k = \Sigma^n_{i=1}k_i > 0 \). Therefore, \( z_1 \psi = \infty \) and \( z_1 \psi = \infty \) for all \( z \in \text{Var}(Q) \).

- Rule \((\infty \alpha_2)\). We first prove that (a) for any problem \( P \), if \( \beta \leq_P z \) by a path of length \( n \), \( \varphi \in Sol^0_\beta(P) \) and \( \beta \varphi = \infty \), then \( z \varphi = \infty \), by induction on \( n \). If \( n = 0 \), this is immediate. Otherwise, there is \( s^\gamma \beta \leq \gamma s^\delta z \in P \) with \( \gamma \leq_P z \) by a path of length \( n - 1 \). Since \( \varphi \in Sol^0_\beta(P) \) and \( \beta \varphi = \infty \), we have \( \gamma \varphi = \infty \) and, by induction hypothesis, \( z \varphi = \infty \).

  We now prove that (b) if \( \beta \leq_P z \) by a path of length \( n \), \( \varphi \in Sol^0_\beta(P) \) and \( \beta \varphi = s^k \varphi \) for some \( k \), then either \( z \varphi = \infty \) or \( z \varphi = s^l \varphi \) for some \( i \), by induction on \( n \). If \( n = 0 \), this is immediate. Otherwise, there is \( s^\gamma \beta \leq \gamma s^\delta z \in P \) with \( \gamma \leq_P z \) by a path of length \( n - 1 \). Since \( \varphi \in Sol^0_\beta(P) \) and \( \beta \varphi = s^k \varphi \), we have \( s^{\gamma+k} \varphi \leq_k s^{\delta} z \varphi \). If \( \gamma \varphi = \infty \), then, by (a), \( z \varphi = \infty \).

  Otherwise, \( \gamma \varphi = s^l \varphi \) for some \( l \) and, by induction hypothesis, either \( z \varphi = \infty \) or \( z \varphi = s^l \varphi \) for some \( i \).

  Hence, if \((z, c, d)\) is incompatible in \( C \) and \( \varphi \in Sol^0_\beta(C) \), then \( z \varphi = \infty \).

5. Every rule decreases the number of constraints in \( C \) except rule \((\infty \alpha_2)\). In \((\infty \alpha_2)\), this number is unchanged but the number of variables decreases. Since the number of variables in \( C \) never increases, the system terminates.


The properties 4(c) and 4(d) give \( Sol^0_\beta(C) = \{ \varphi|_{\text{Var}(C)} \mid \varphi \in Sol^0_\beta(D) \} \) whenever \( C \leadsto D \).
Definition 29 (Affine problem) A constraint is affine if it is of sort \( \mathbb{N} \), of the form \( s^k \alpha \leq s^l \beta \) or of the form \( s^c \alpha \leq s^l \beta \). A problem is affine if all its constraints are affine.

Lemma 31 In any normal configuration \( C \neq \bot \), \( C_4 \) is an affine problem with no positive cycles and no incompatible triples.

Proof By Lemma 30, every term of sort \( A \) occurring in \( C \) is of the form \( \infty, s^k \alpha \) or \( s^c \alpha \).

Now, \( C_4 \) cannot contain a constraint of the form:

- \( a \leq \infty \) because of rule \((\infty)\),
- \( \infty \leq s^l \beta \) because of rule \((\infty x_1)\),
- \( \infty \leq s^l s^d \) because of rule \((\infty c)\),
- \( s^k \alpha \leq s^l s^d \) because of rule \((\alpha c)\),
- \( s^c \alpha \leq s^l s^d \) because of rules \((cc)\) and \((cd)\).

Therefore, a constraint in \( C_4 \) can only be either of the form \( s^k \alpha \leq s^l \beta \) or of the form \( s^c \alpha \leq s^l \beta \). Moreover, \( G(C_4) \) cannot have positive cycles because of rule \((\infty x_1)\), and cannot have incompatible triples because of rule \((\infty x_2)\). \( \square \)

Since affine problems of sort \( A \) are always satisfiable (by setting their variables to \( \infty \)), we can conclude:

1. Compute a normal form \( C \) of \((\emptyset, \emptyset, \emptyset, \emptyset, P)\) wrt the rules of Figure 8.
2. If \( C = \bot \), then \( P \) is not satisfiable. Otherwise, \( C = (C_0, C_1, C_2, C_3, C_4) \).
3. If \( C_3 \) has a positive cycle, then \( P \) is not satisfiable. Otherwise, \( P \) is satisfiable.

Fig. 9. Algorithm for deciding the satisfiability of a problem \( P \) in the successor algebra.

Theorem 4 (Satisfiability) The satisfiability of a size problem in the successor algebra is decidable in polynomial time wrt the number of symbols by the algorithm of Figure 9.

Proof Let \(|P|\) be the number of symbols in \( P \). Constructing \( G(P) \) requires at most \#Var(\( P \)) + \#P steps, where \#X is the cardinal of \( X \). But \|Var(\( P \))\| \leq 2\#P since there are at most two variables per constraint, and \( 2\#P \leq |P| \) since every constraint is of size 2 at least. Therefore, constructing \( G(P) \) requires at most \( 3|P|/2 \) steps.

Whether there is a positive cycle in a graph is decidable in polynomial time (Pratt, 1977). Whether there is an incompatible triple in a graph can be done in polynomial time too. Hence, whether a rule can be applied is decidable in polynomial time. Now, since \( \sim \) terminates, the algorithm describes a computable function.

We now prove that it is correct and complete. If \( C = \bot \), then, by completeness, \( P \) is unsatisfiable. Otherwise, \( C = (C_0, C_1, C_2, C_3, C_4) \). If \( G(C_3) \) has a positive cycle then, by completeness, \( P \) is unsatisfiable. Otherwise, let \( \phi_3 \in \text{Sol}_B(C_3) \). Then, one can easily check that \( \phi = \phi_3 \cup \{(\alpha, \infty) \mid \alpha \in \text{Var}(C_1) \cup \text{Var}(C_4)\} \cup \{(\alpha, s^{\alpha_0} \phi_3 c) \mid (\alpha, c) \in \)
$C_2 \in \text{Sol}_{\text{g}}^0(C)$. Therefore, by correctness, $\varphi|_{\text{Var}(P)} \in \text{Sol}_{\text{g}}^0(P) = \text{Sol}_{\text{g}}^0(P)$ and $P$ is satisfiable.

Finally, to prove that the complexity for computing $C$ is polynomial, it suffices to show that the number of rewrite steps and the size $|C| = |\pi(C)|$ of intermediate configurations $C$ are polynomially bounded by $|P|$. By definition of $\sim$, $\text{Var}(C) \subseteq \text{Var}(P) \cup \{x_\alpha \mid \alpha \in \text{Var}(P)\}$ and $2\#\text{Var}(C) \leq 2\#\text{Var}(P) \leq 2|P|$. So, after the termination proof, the number of rewrite steps is $\leq \#P \times 2|P| \leq |P|^2$.

Let $\|C\|_\infty$ be the maximum size of a constraint in $\pi(C)$. No rule but $(xc)$ can make $\|C\|_\infty$ increase. $\|C\|_\infty$ can be increased by at most 2 for each replacement of a variable $x$ by $s^\varphi c$. However, there cannot be more than two such replacements in a constraint since, after two such replacements, there is no variable of sort $A$ anymore. Therefore, $\|C\|_\infty \leq \|P\|_\infty + 4 \leq |P| + 4$. Now, $\|\pi(C)\| \leq 5\#P \leq |P|/2$ since $\#P_0 + \#P_1 + \#P_2 \leq 2\#\text{Var}(P) \leq 2\#P$ and $\#P_3 + \#P_4 \leq \#P$. Therefore, $|C| \leq \|C\|_\infty \times \#C \leq (|P| + 4) \times |P|/2$. □

Our procedure can be related to the one described in Barthe et al. (2005) where, like many works on type inference, the authors consider constrained types. But they do not bring out the properties of the size algebra and, in particular that, in the successor algebra, satisfiable sets of constraints have a most general solution as we shall see in next section.

**Example 11** Let $P = \{c \leq \alpha, s c \leq \beta, \beta \leq \alpha, d \leq \beta\}$. We have $(0, 0, 0, 0, P)$

$\sim (0, \{x\}, 0, 0, \{c \leq \infty, \infty \leq \beta, \beta \leq \infty, d \leq \beta\}$, by $(\infty \alpha 1)$ since $c \leq_P \alpha$ and $d \leq_P \alpha$;

$\sim (0, \{\alpha, \beta\}, 0, 0, \{c \leq \infty, \infty \leq \beta\}$, by $(\alpha \alpha 1)$ since $\infty \rightarrow_0 \beta \sim_\infty \infty$ is positive;

$\sim \bot$, by $(\infty \alpha)$. ■

**Example 12** Let $P = \{\alpha \leq s c, \beta \leq \alpha\}$. We have $(0, 0, 0, 0, P)$

$\sim (0, \{\alpha, c\}, \{x_\alpha \leq 1\}, \{\beta \leq s^\varphi c\}$, by $(xc)$;

$\sim (0, \{\alpha, c, (\beta, c)\}, \{x_\alpha \leq 1, x_\beta \leq x_\alpha\}$, by $(xc)$ again. This is a normal form and the graph of $\{x_\alpha \leq 1, x_\beta \leq x_\alpha\}$ has no positive cycle, so it is satisfiable (the solutions for $(x_\alpha, x_\beta)$ are $(0, 0), (1, 0)$ and $(1, 1)$). ■

### 9.2 Computing the most general solution

We now turn to the problem of whether, in the successor algebra $A$, a satisfiable problem $P$ has a most general solution and, if so, how to compute it.

Let $\text{mgs}_A(P)$ (resp. $\text{mgs}_f(P)$) be the set of most general (finite resp.) solutions of $P$, and $\text{mgs}_A^0(C)$ (resp. $\text{mgs}_f^0(C)$) be the set of most general (finite resp.) $\infty$-closed solutions of $C$.

We first prove a refinement of Lemma 26 to $\infty$-closed solutions of a configuration:

**Lemma 32** Given $\varphi, \psi \in \text{Sol}_{\text{g}}^0(C)$, $\varphi \subseteq \psi$ iff there is $\rho : V \rightarrow V \cup \{\infty\}$ such that $\text{dom} \rho \subseteq \text{Var}_A(C)$ and, for all $\alpha \in C_0 \cup \text{Var}(C_3) \cup \text{Var}_A(C_4)$, $\alpha \varphi \rho \leq^\infty A \psi$. 

https://doi.org/10.1017/50956796818000072 Published online by Cambridge University Press
Proof

⇒ By Lemma 26, there is \( \rho : \mathcal{V} \cup \mathcal{V}_B \to \mathcal{V} \cup \mathcal{V}_B \cup \{0, \infty\} \) such that \( \varphi \rho \leq_{\text{A}} \psi \).
Since \( \varphi \) and \( \psi \) are \( \mathbb{N} \)-closed, we also have \( \varphi(\rho|_y) \leq_{\text{A}} \psi \). Indeed, if \( \pi \in \mathcal{V}_B \), then \( \pi(\rho|_y) = \pi \varphi = \pi \varphi \leq_{\text{A}} \pi \psi \). Let now \( \pi \notin \text{Var}(C) \). Then, \( \pi(\rho|_y) = \pi \varphi(\rho|_y) \leq_{\text{A}} \pi \psi = \pi \). Therefore, \( \pi(\rho|_y) = \pi \) and \( \text{dom}(\rho|_y) \subseteq \text{Var}_A(C) \).

⇐ After Lemma 26, it is enough to prove that, for all \( \pi \in \mathcal{V} \cup \mathcal{V}_B \), \( \pi \varphi \leq_{\text{A}} \pi \psi \). By assumption, the property holds if \( \pi \in \mathcal{C}_0 \cup \text{Var}(C) \cup \text{Var}_A(C_4) \). If \( \pi \in \mathcal{C}_1 \), then \( \pi \varphi = \pi \psi \) and \( \pi \varphi \leq_{\text{A}} \pi \psi \). If \( (\pi, \pi) \in \mathcal{C}_2 \), then \( \pi \varphi = \pi \psi = \pi \).

We now prove that the most general solutions of a problem \( P \) in \( \mathbb{N} \) can be obtained from the most general \( \mathbb{N} \)-closed solutions of the normal form of \( (0, 0, 0, 0, P) \).

Lemma 33 Assume that \( (0, 0, 0, 0, P) \sim_{\ast} C \).

- Correctness: if \( \varphi \in \text{msg}_{\mathbb{B}}(C) \), then \( \varphi|_{\text{Var}(P)} \in \text{msg}_{\mathbb{B}}(P) \).
- Completeness: if \( \psi \in \text{msg}_{\mathbb{A}}(P) \), then there is \( \varphi \in \text{msg}_{\mathbb{B}}(C) \) such that \( \varphi|_{\text{Var}(P)} = \psi \).

Proof

Note that \( \text{Var}(C) = \text{Var}(P) \cup \text{Var}(C_3) \).

- Let \( \varphi \in \text{msg}_{\mathbb{B}}(C) \). By correctness of \( \sim_{\ast} \), \( \varphi|_{\text{Var}(P)} \in \text{Sol}_{\mathbb{B}}(P) \). Let now \( \psi \in \text{Sol}_{\mathbb{B}}(P) \).

By completeness of \( \sim_{\ast} \), there is \( \varphi' \in \text{Sol}_{\mathbb{B}}(C) \) such that \( \psi = \varphi'|_{\text{Var}(P)} \). Since \( \varphi = \text{msg}(C) \), \( \varphi \subseteq \varphi' \). By Lemma 32, \( \varphi \rho \leq_{\text{A}} \varphi \rho \) for some \( \rho : \mathcal{V} \to \mathcal{V} \cup \{\infty\} \) such that \( \text{dom}(\rho) \subseteq \text{Var}_A(C) = \text{Var}(P) \). Therefore, for all \( \pi \in \mathcal{V} \cup \mathcal{V}_B \), \( \pi \varphi = \pi \varphi \leq_{\text{A}} \pi \psi \).

- Let \( \psi \in \text{msg}_{\mathbb{A}}(P) \). By completeness of \( \sim_{\ast} \), there is \( \varphi \in \text{Sol}_{\mathbb{B}}(C) \) such that \( \psi = \varphi|_{\text{Var}(P)} \).

Assume now that there is \( \varphi' \in \text{Sol}_{\mathbb{B}}(C) \) such that \( \varphi \leq_{\text{A}} \varphi' \). By correctness of \( \sim_{\ast} \), \( \varphi'|_{\text{Var}(P)} \in \text{Sol}_{\mathbb{B}}(P) \). Since \( \psi = \text{msg}(P) \), \( \psi = \varphi'|_{\text{Var}(P)} \subseteq \varphi'|_{\text{Var}(P)} \), that is, there is \( \rho \) such that \( \varphi|_{\text{Var}(P)} \rho \leq_{\text{A}} \varphi'|_{\text{Var}(P)} \). Since \( \varphi \not\subseteq \varphi' \), there is \( \pi \) such that \( \pi \varphi \not\leq_{\text{A}} \pi \varphi' \).

Since \( \varphi'|_{\text{Var}(P)} \rho \leq_{\text{A}} \varphi'|_{\text{Var}(P)} \), \( \pi = \pi \beta \) for some \( \beta \in \text{Var}(P) \). By definition of \( \text{Sol}_{\mathbb{B}}(C) \), there is \( \pi \) such that \( \beta \varphi = \pi \varphi' \).

Hence, \( \pi \varphi' \leq_{\text{A}} \pi \varphi' \) and \( \pi \varphi \not\leq_{\text{A}} \pi \varphi' \). Contradiction.

We now prove that the most general \( \mathbb{N} \)-closed solutions of \( (C_0, C_1, C_2, C_3, C_4) \) can be obtained from the most general \( \mathbb{N} \)-closed solutions of \( C_3 \cup C_4 \).

Lemma 34 Let \( C = (C_0, C_1, C_2, C_3, C_4) \) be a configuration.

- Correctness: if \( \psi \in \text{msg}_{\mathbb{B}}(C_3 \cup C_4) \) and, for all \( \pi \in \text{Var}_A(C_4) \), \( \text{Var}(\pi) \cap C_0 = \emptyset \), then \( \sigma_1(C) \cup \sigma_2(C, \pi) \cup \sigma_3|_A(C, \psi) \in \text{msg}_{\mathbb{B}}(C) \).
- Completeness: if \( \varphi \in \text{msg}_{\mathbb{B}}(C) \), then \( \varphi|_{\text{Var}(C_3 \cup C_4)} \in \text{msg}_{\mathbb{B}}(C_3 \cup C_4) \).

\(^{10}\) Thanks to Lemma 28, this condition can always be satisfied by applying some permutation to \( \psi \).
Proof

- Let \( \psi' = \sigma_1(C) \cup \sigma_2(C, \psi) \cup \sigma_{3,4}(C, \psi) \) and \( \varphi \in \text{Sol}_B^0(C) \). We have \( \varphi_{3,4} = \varphi|_{\text{Var}(C)\cup C_4} \in \text{Sol}_B^0(C_3 \cup C_4) \). Since \( \psi \in \text{mgs}_B^0(C_3 \cup C_4) \), \( \psi \subseteq \varphi_{3,4} \). By applying Lemma 32 on \((0,0,0,C_3, C_4)\), there is \( \rho : V \rightarrow V \cup C \cup \{\infty\} \) such that \( \text{dom}(\rho) \subseteq \text{Var}_A(C_4) \) and, for all \( \alpha \in \text{Var}(C_3) \cup \text{Var}_A(C_4) \), \( \alpha \rho \leq_k \alpha \varphi_{3,4} \). Then, let \( \rho' = \{(x, \alpha \varphi) \mid \alpha \in C_0 \} \cup \{(x, \alpha \rho) \mid x \in \text{Var}(C_4)\} \). We prove that \( \psi' \subseteq \varphi \) by using Lemma 32. We have \( \text{dom}(\rho') \subseteq \text{Var}(C) \) by definition. If \( x \in C_0 \), then \( x \psi' \rho' = x \varphi \) by definition. If \( x \in \text{Var}(C)_3 \), then \( x \psi' \rho' = x \psi' \rho = x \rho \) because \( \psi \) is \( \mathbb{N} \)-closed, and \( x \psi' \rho' \leq_k x \varphi_{3,4} = x \varphi \). If \( x \in \text{Var}(C)_4 \), then \( x \psi' \rho' = x \psi' \rho = x \rho \) since \( \text{Var}(x \psi) \cap C_0 = \emptyset \) by assumption, and \( x \psi' \rho \leq_k x \varphi_{3,4} = x \varphi \).

- We first check that \( \varphi|_{C_0} \) maps variables to variables and is injective. Let \( \alpha \in C_0 \) and \( \varphi' = \varphi|_{\text{Var}(C)\cup \{x\}} \). Then, \( \varphi' \in \text{Sol}_B(C) \) too since, by definition of configuration, \( \alpha \notin \text{Var}(C) \) for every \( i > 0 \). Hence, \( \varphi \subseteq \varphi' \), that is, there is \( \rho \) such that \( \alpha \rho \leq_k \alpha \varphi' = \alpha \). Therefore, \( \alpha \varphi \) is a variable \( \gamma \). Assume now that \( \gamma = \beta \rho \) for some \( \beta \in C_0 \). Then, \( \varphi'' = \varphi|_{\text{Var}(C)\cup \{x, \beta\}} \in \text{Sol}_B(C) \) too. Hence, \( \varphi \subseteq \varphi'' \), that is, there is \( \rho' \) such that \( \gamma \rho' \leq_k \gamma \varphi'' = \alpha \) and \( \gamma \rho' \leq_k \beta \varphi'' = \beta \). Therefore, \( \alpha = \gamma \rho' = \beta \).

So, by taking in Lemma 27 \( V = V, V_1 = C_0, V_2 = \varphi(C_0), \rho_1 = \{(x, \alpha \varphi) \mid x \in C_0\} \) and \( \rho_2 = \{(x, \alpha \rho) \mid x \in C_0\} \) (the inverse of \( \rho_1 \)), there is a permutation \( \xi : V \rightarrow V \) such that \( \xi|_{C_0} = \varphi|_{C_0} \). By Lemma 28, \( \varphi \xi^{-1} \) is a mgs of \( C \) too. So, wlog, we can assume that \( \varphi|_{C_0} \) is the identity.

We now prove that \( \varphi_{3,4} = \varphi|_{\text{Var}(C)\cup C_4} \in \text{mgs}_B^0(C_3 \cup C_4) \). Let \( \psi \in \text{Sol}_B^0(C_3 \cup C_4) \). By Lemma 30 (1), \( \psi' = \sigma_1(C) \cup \sigma_2(C, \psi) \cup \sigma_{3,4}(C, \psi) \in \text{Sol}_B^0(C) \). Hence, \( \psi \subseteq \psi' \).

By Lemma 32, there is \( \rho : V \rightarrow V \cup C \cup \{\infty\} \) such that \( \text{dom}(\rho) \subseteq \text{Var}_A(C_4) \) and, for all \( \alpha \in C_0 \cup \text{Var}(C_3) \cup \text{Var}_A(C_4) \), \( \alpha \rho \leq_k \alpha \psi' \). For all \( \alpha \in \text{Var}(C_3) \cup \text{Var}_A(C_4) \), \( \alpha \varphi_{3,4} \rho = \alpha \rho \leq_k \alpha \psi' = \alpha \psi \). Therefore, by Lemma 32, \( \varphi_{3,4} \subseteq \psi \).

Next, we prove that, for all affine problems \( P \) with no incompatible triples (like \( C_3 \cup C_4 \) in a normal configuration \( C \)), the set of finite \( \mathbb{N} \)-closed solutions of \( P \) is in bijection with the set of finite \( \mathbb{N} \)-closed solutions of:

**Definition 30 (Integer problem associated to an affine problem)** Given an affine problem \( P \), let \( I(P) \) be the integer problem obtained by replacing in \( P \) every constraint \( \leq \) by \( \leq \), each constraint \( \leq \) by \( \leq \), and each constraint \( \leq \) by \( \leq \).

**Lemma 35** If \( P \) is an affine problem with no incompatible triples, then

1. there is a strictly monotone map \( \psi \mapsto \psi \) from \( \text{Sol}_B^0(I(P)) \) to \( \text{Sol}_B^0(P) \);
2. there is a monotone map \( \varphi \mapsto \varphi \) from \( \text{Sol}_B^0(P) \) to \( \text{Sol}_B^0(I(P)) \);
3. for all \( \psi \in \text{Sol}_B^0(I(P)) \), \( \psi = \psi \);
4. for all \( \varphi \in \text{Sol}_B^0(P) \), there is \( \rho : V \rightarrow V \cup C \) such that \( \varphi = \varphi \rho \), hence \( \varphi \subseteq \varphi \);
5. correctness: if \( \psi \in \text{mgs}_B^0(I(P)) \), then \( \psi \in \text{mgs}_B^0(P) \);
6. completeness: if \( \varphi \in \text{mgs}_B^0(P) \), then \( \varphi \in \text{mgs}_B^0(I(P)) \).

https://doi.org/10.1017/50956796818000072 Published online by Cambridge University Press
Proof
Let $\simeq_P$ be the symmetric and transitive closure of $\leq_P$ and $\eta : H/\simeq_P \to H$ be any function such that, for all equivalence classes $X$, $\eta(X) \in X$ ($H$ and $\leq_P$ are introduced in Definition 27). Such a function always exists because equivalence classes are non-empty. Because $P$ has no incompatible triples, an equivalence class modulo $\simeq_P$ cannot contain two different constants. Hence, we can assume that $\eta(X) = c$ iff $c \in X$.

Given $\psi \in \text{Sol}_B^0(I(P))$, let $\hat{\psi} = \{(x, x\psi) \mid x \in \text{Var}_B^0(P)\} \cup \{(x, z^x \psi x^*) \mid z \in \text{Var}_A^0(P)\}$ where $x^* = \eta([x]_P)$ and $[x]_P$ is the equivalence class of $x$ modulo $\simeq_P$.

Given $\varphi \in \text{Sol}_B^0(P)$, let $\hat{\varphi} = \{(x, x\varphi) \mid x \in \text{Var}_B^0(P)\} \cup \{(x, z\varphi_k) \mid z \in \text{Var}_A^0(P)\}$ ($\varphi_s$ is introduced in Definition 24).

1. We first check that $\hat{\psi} \in \text{Sol}_B^0(P)$ whenever $\psi \in \text{Sol}_B^0(P)$, that is, $\hat{\psi}$ satisfies every constraint of $P$. This is immediate for constraints of sort $\eta$. Otherwise, since $P$ is affine, there are two cases. If $s^k x \leq \eta^k s^l \beta \in P$, then $x^* = \beta^*$ and $x_\beta \psi + k \leq \alpha x_\beta \psi + l$. Hence, $(s^k x) \hat{\psi} = s^k x^* + x_\beta \psi + k \leq \alpha s^k x^* + x_\beta \psi + l$. Therefore, $(s^k x) \hat{\psi} = s^k x^* + x_\beta \psi + l$. Also, $\eta(x) \leq \eta(x) \psi + k \leq \alpha \eta(x) \psi + l$. Hence, there is $\psi \mapsto \hat{\psi}$ is injective ($\psi_1 = \psi_2$ whenever $\psi_1 = \psi_2$) and monotone wrt $\leq_A (\hat{\psi} \leq \hat{\psi}$ whenever $\psi \leq \psi$) and thus wrt $\subseteq$. Therefore, $\psi \mapsto \hat{\psi}$ is strictly monotone wrt $\subseteq$.

2. We first check that $\hat{\varphi} \in \text{Sol}_B^0(I(P))$ whenever $\varphi \in \text{Sol}_B^0(P)$. If $s^k x \leq \eta^k s^l \beta \in P$, then $(s^k x) \varphi = s^k x^* + x_\beta \varphi + k \leq \alpha s^k x^* + x_\beta \varphi + l$. Assume now that $s^k x \leq \eta^k s^l \beta \in P$. Then, $(s^k x) \varphi = s^k x^* + x_\beta \varphi + k \leq \alpha s^k x^* + x_\beta \varphi + l$. Also, $\eta(x) \varphi \leq \eta(x) \psi + k \leq \alpha \eta(x) \psi + l$. Hence, there is $\varphi \mapsto \hat{\varphi}$ is monotone. Let $\varphi_1, \varphi_2 \in \text{Sol}_B^0(P)$ such that $\varphi_1 \subseteq \varphi_2$. Hence, there is $\rho : V \to C \cup V$ such that $\varphi \rho \leq \eta \rho$. Therefore, $\varphi \leq \hat{\varphi}$.

3. Immediate.

4. Let $\rho$ be the map from $V$ to $V \cup C$ such that, if $x^* \in V$, then $x^* \rho = x \varphi_h$. The map $\rho$ is well-defined since $\varphi_h$ is invariant by $\simeq_P$ : if $x \simeq_P \beta$, then $x \varphi_h = \beta \varphi_h$.

5. Let $\varphi \in \text{Sol}_B^0(P)$. By 2, $\hat{\varphi} \in \text{Sol}_B^0(I(P))$ and $\psi \subseteq \hat{\varphi}$. By 1, $\hat{\psi} \subseteq \hat{\varphi}$. By 4, $\hat{\psi} \subseteq \varphi$. Therefore, $\psi \subseteq \varphi$.

6. Let $\psi \in \text{Sol}_B^0(I(P))$. By 1, $\psi \subseteq \hat{\varphi}$. By 2, $\hat{\varphi} \subseteq \psi$. By 3, $\psi = \psi$. Therefore, $\hat{\psi} \subseteq \psi$.

Lemma 36 Every satisfiable integer problem has a smallest $\mathbb{N}$-closed solution that can be computed in polynomial time.

Proof
Let $P$ be a satisfiable integer problem whose variables are $x_1, \ldots, x_n$. We first prove that $P$ is equivalent to a problem in the dioid $(\mathbb{Z}_{\text{max}}^n, \oplus, \ominus)$ where $\mathbb{Z}_{\text{max}} = \mathbb{Z} \cup \{\pm \infty\}$, $\ominus = \max$ and $\oplus = +$ both applied component wise (Baccelli et al., 1992).
Wlog, we can assume that $P$ contains no constraints of the form $0 + k \leq 0$ (since $P$ is satisfiable, these constraints are always satisfied and thus can be removed). Hence, $P$ contains only constraints of the form $x_i + k \leq x_j$, $0 + k \leq x_j$ or $x_i + k \leq 0$, that is, in the syntax of $(\mathbb{Z}_\text{max}, \otimes, \otimes)$, $k \otimes x_i \leq x_j$, $k \leq x_j$ or $x_i \leq -k$.

Given a problem $P$, let $a_{ij} = \text{sup}\{k \in \mathbb{Z}_\text{max} \mid x_i + k \leq x_j \in P\}$, $b_{ij} = \text{sup}\{0\} \cup \{k \in \mathbb{Z}_\text{max} \mid 0 + k \leq x_i \in P\}$ (we add 0 because solutions must be non-negative) and $c_{ij} = \text{inf}\{-k \in \mathbb{Z}_\text{max} \mid x_i + k \leq 0 \in P\}$ with, as usual, $\text{sup}0 = -\infty$ and $\text{inf}0 = +\infty$. Note that, in $b$ and $c$, every column is the same ($b_{ij}$ and $c_{ij}$ do not depend on $j$).

We now prove that, if $\psi \in \text{Sol}_{\psi}(P)$, then there is $x \in \mathbb{Z}^{n \times n}_{\text{max}}$ such that $(a \otimes x) \oplus b \leq x \leq c$ and, for all $j$, $x_{ij} = x_i \psi$ (the columns of $x$ are equal). For all $i$ and $j$, the set of inequations $\{k \otimes x_i \leq x_j \mid x_i + k \leq x_j \in P\}$ is equivalent to $a_{ji} \otimes x_i \leq x_j$ since $k \leq a_{ji}$ and $-\infty \otimes x_i = -\infty \leq x_j$. Hence, $\{a_{ji} \otimes x_i \leq x_j \mid i \in \{1, \ldots, n\}\}$ is equivalent to $\bigoplus_{i=1}^{n} a_{ji} \otimes x_i \leq x_j$. By taking $x_{il} = x_i$ for all $l$, we therefore get $(a \otimes x)_{il} \leq x_{jl}$.

We now prove that, if $\psi \in \text{Sol}_{\psi}(P)$, then there is $x \in \mathbb{Z}^{n \times n}_{\text{max}}$ such that $(a \otimes x) \oplus b \leq x \leq c$, then $\psi_l \in \text{Sol}_{\psi}(P)$ where $\psi_l$ is the substitution such that $x_i \psi_l = x_{il}$ (lth column of $x$).

By Theorem 4.75 in Baccelli et al. (1992), $(a \otimes x) \oplus b \leq x$ has $a^* \otimes b$ as smallest solution, where $a^* = \bigoplus_{k \in \mathbb{N}} a^k$, $a_k^{k+1} = a^k \otimes a$ and $a^0$ is the matrix with 0 on the diagonal and $-\infty$ everywhere else. Since $P$ is satisfiable, $G(P)$ has no positive cycles. Hence, $a^* = \bigoplus_{k=0}^{n} a^k$ (Cuninghame-Green, 1979) (Theorem 3.20 in Baccelli et al. (1992)). Therefore, $\psi \in \text{Sol}_{\psi}(P)$ iff $a^* \otimes b \leq c$, and the smallest solution of $P$ is the function $\psi'$ such that $x_i \psi' = \bigoplus_{k=1}^{n} a^k_{ik} \otimes b_{k1}$, which can be computed in polynomial time.

Therefore, we can now conclude:

1. Apply the algorithm of Figure 9.
2. Compute the most general $N$-closed solution $\psi$ of $C_3 \cup I(C_4)$ using Lemma 36.
3. Compute $\psi$ defined in Lemma 35.
4. Return $\sigma_1(C) \cup \sigma_2(C, \psi) \cup \sigma_4a(C, \psi)$.

Fig. 10. Algorithm computing a most general solution in the successor algebra.

**Theorem 5** In the successor algebra, any satisfiable size problem has a most general solution that can be computed in polynomial time following the algorithm of Figure 10.

**Proof**

Correctness. By Lemma 36, $\psi \in \text{mgs}_B(\text{Sol}_{\psi}(I(C_3 \cup C_4)))$. By Lemma 35 (5), $\psi \in \text{mgs}_B(C_3 \cup C_4) \subseteq \text{mgs}_B(C_3 \cup C_4)$. By Lemma 34 (1), $\varphi = \sigma_1(C) \cup \sigma_2(C, \psi) \cup \sigma_3\varphi(C, \psi) \in \text{mgs}_B(C)$. By Lemma 33 (1), $\varphi|_{\text{Van}(P)} = \sigma_1(C) \cup \sigma_2(C, \psi) \cup \sigma_4a(C, \psi) \in \text{mgs}_B(P)$.

Complexity. After Theorem 4, $C_3 \cup C_4$ is of polynomial size wrt the size of $P$. The computation of $I(C_3 \cup C_4)$ is linear. After Lemma 36, the computation of $\psi$ is
polynomial. After Lemma 35 (1), the computation of $\psi$ is polynomial. Therefore, the algorithm of Figure 10 is polynomial.

**Example 13** In Example 12, we have seen that the normal form of $(0, 0, 0, 0, P)$ where $P = \{ x \leq^* x_c, \beta \leq^* x \}$ is $(0, 0, \{(x, c), (\beta, c)\}, C_3, \emptyset)$ with $C_3 = \{ x_c \leq^* 1, x_{\beta} \leq^* x_x \}$. Following Lemma 36, by taking $x_1 = x_x$ and $x_2 = x_{\beta}$, the corresponding max-linear system is $(a \otimes x) \oplus b \leq x \leq c$ where $a_{11} = \sup \{ k \mid x_x + k \leq x_x \in C_3 \} = \sup \emptyset = -\infty$, $a_{12} = \sup \{ x_{\beta} + k \leq x_x \in C_3 \} = \sup \{ 0 \} = 0$, $a_{21} = \sup \{ k \mid x_x + k \leq x_{\beta} \in C_3 \} = \sup \emptyset = -\infty$, $b_1 = \sup \{ \emptyset \cup \{ k \mid k \leq x_x \in C_3 \} \} = \sup \{ 0 \} = 0$, $b_2 = \sup \{ \{ 0 \} \} = 0$, $c_1 = \inf \{ k \mid x_x \leq k \in C_3 \} = \inf \{ 1 \} = 1$ and $c_2 = \inf \{ k \mid x_{\beta} \leq k \in C_3 \} = \inf \emptyset = +\infty$. To summarize, we have $a = \left( \begin{array}{c} -\infty \\ -\infty \\ 0 \\ 0 \end{array} \right)$, $b = \left( \begin{array}{c} 0 \\ 0 \\ 0 \\ 0 \end{array} \right)$ and $c = \left( \begin{array}{c} 1 \\ 1 \\ 1 \end{array} \right)$.

One can easily check that, if $x = \left( \begin{array}{c} x_x \\ x_\psi \end{array} \right)$, then $(a \otimes x) \oplus b = \left( \begin{array}{c} x_\psi \oplus 0 \\ 0 \end{array} \right)$, hence that $(a \otimes x) \oplus b \leq x \leq c$ is equivalent to $x_x \oplus 0 \leq x_x \leq 1$ and $0 \leq x_{\beta} \leq +\infty$, which is $C_3$. Now, $a^0 = \left( \begin{array}{c} 0 \\ 0 \\ -\infty \\ -\infty \end{array} \right)$ and $a^2 = \left( \begin{array}{c} -\infty \\ -\infty \\ -\infty \\ -\infty \end{array} \right)$. Hence, $a^* = a^0 \oplus a = \left( \begin{array}{c} 0 \\ 0 \\ 0 \\ 0 \end{array} \right)$ and $a^* \oplus b = \left( \begin{array}{c} 0 \\ 0 \end{array} \right)$. So, the smallest solution of $C_3$ is $\psi = \{(x_x, 0), (x_{\beta}, 0)\}$ and the smallest solution of $P$ is $\{ \sigma_C(C) \cup \sigma_c(C, \psi) \cup \sigma_{4k}(C, \psi) = \{(x, c), (\beta, c)\} \}$.

**10 Conclusion**

We have presented a general and modular termination criterion for the combination of $\beta$-reduction and user-defined rewrite rules, based on the use of type-checking with size-annotated types approximating a semantic notion of size defined by the annotations given to constructor symbols. This extends to rewriting-based function definitions and more general notions of size, an approach initiated by Hughes, Pareto and Sabry for function definitions based on a fixpoint combinator and case analysis (Hughes et al., 1996).

First, we have shown that these termination conditions can be reduced to solving problems in the quasi-ordered algebra used for size annotations. Then, we have shown that the successor algebra (successor symbol with arbitrary constants) enjoys nice properties: decidability of the satisfiability of sets of inequalities (in polynomial time), and existence and computability of a most general solution for satisfiable problems (in polynomial time too). As a consequence, we have a complete algorithm for checking the termination conditions in the successor algebra.

We have implemented a simple heuristic that turns this termination criterion into a fully automated termination prover for higher order rewriting called HOT (2012), which tries to detect size-preserving functions and, following Abel & Altenkirch (2002), to find a lexicographic ordering on arguments. Combined with other (non-)termination techniques (Jouannaud & Okada, 1991; Blanqui, 2000; Blanqui et al., 2002), HOT won the 2012 international competition of termination provers (Termination competition, 2017) for higher order rewriting against THOR (2014) and Wanda (2015). It could be improved by replacing the lexicographic ordering by the size-change principle (Lee et al., 2001; Hyvernat, 2014), and using abstract
interpretation techniques for annotating function symbols (Telford & Turner, 2000; Chin & Khoo, 2001). A more complete (and perhaps more efficient) implementation would be obtained by encoding constraints into a SAT problem and send it to state-of-art SAT solvers (Fuhs et al., 2007; Ben-Amram & Codish, 2008; Codish et al., 2011).

A natural following is to study other size algebras like the max-successor algebra (i.e. the successor algebra extended with a \texttt{max} operator), the plus algebra (i.e. the successor algebra extended with addition) or their combination, the max-plus algebra. Indeed, the richer the size algebra is, the more precise the typing of function symbols is, and the more functions can be proved terminating.

Following Blanqui & Riba (2006), it is also possible to consider full Presburger arithmetic (Presburger, 1929) and handle conditional rewrite rules, by extending the system with explicit quantifiers and constraints on size variables, in the spirit of HM(X) (Sulzmann, 2001). Simplification of constraints is then an important issue in practice (Pottier, 2001).

We have presented this criterion in Church’ simply typed \(\lambda\)-terms but, following Blanqui (2005b), it should be possible to extend it to richer type systems with polymorphic and dependent types. Similarly, we considered matching modulo \(\alpha\)-congruence only but, following Blanqui (2016), it should be possible to extend it to rewriting modulo some equational theory and to rewriting on \(\beta\)-normal forms with matching modulo \(\beta\eta\) as used in Klop’s combinatory reduction systems (Klop et al., 1993) or Nipkow’s higher order rewrite systems (Mayr & Nipkow, 1998).

Another interesting extension would be to consider size-annotated types in the computability path ordering Blanqui et al. (2015), following Kamin and Lévy’s extension of Dershowitz’ RPO (Dershowitz, 1979b; Kamin & Lévy, 1980) and Borralleras and Rubio’s extension of Jouannaud and Okada’s higher order RPO (Jouannaud & Rubio, 1999; Borralleras & Rubio, 2001).

Acknowledgments

I would like to thank Christophe Raffalli for a short but useful discussion on max-plus algebra, and Nachum Dershowitz, Jean-Pierre Jouannaud and Sylvain Schmitz for their comments on the introduction and the conclusion. I also want to thank very much the anonymous referees for their very careful reading and the numerous remarks and suggestions they made. This greatly helped me to improve the article.

References


Size-based termination


