A Structural Approach to the Stretching Lemma of Simply-Typed Lambda-Calculus

Ralph Matthes*

Institut de Recherche en Informatique de Toulouse (IRIT),
C.N.R.S. and University of Toulouse, France

This work is about an analysis of an advanced aspect of the inhabitation problem in simply-typed λ-calculus by way of a λ-calculus notation for proof search in minimal implicational logic, introduced in joint work with José Espírito Santo and Luís Pinto [1] (see also the revised and extended version [2]). By proofs in minimal implicational logic we understand η-long β-normal λ-terms that are well-typed according to the rules of simply-typed λ-calculus. One has to treat the general case of terms with open assumptions: a sequent σ is of the form $\Gamma \Rightarrow A$ with a finite set Γ of declarations $x_i : A_i$, where the x_i are variables of λ-calculus. This fits with the typing relation of λ-calculus but, viewed from the logical side, presents the particularity of named hypotheses. The total discharge convention that plays a role in the paper by Takahashi et al [5] goes into the opposite direction and considers λ-terms where there is only one term variable per type. In the joint work of the author, cited above, no total discharge convention is needed for obtaining a finitary description of the whole solution space $\mathcal{S}(\sigma)$ for a given sequent σ. The solution space $\mathcal{S}(\sigma)$ itself is a *coinductive* expression formed from the grammar of β-normal forms of λ-calculus and an operator for finite sums expressing proof alternatives. Its potential infinity reflects the *a priori* unlimited depth of proof search and serves the specification of proof search problems. For simply-typed λ-calculus, the subformula property allows to describe the solution spaces finitely. This may be seen as a coinductive extension of work done already by Ben-Yelles in his 1979 PhD thesis with a very concrete λ-calculus approach and by Takahashi et al [5] by using formal grammar theory (but the latter need the total discharge convention for reaching finiteness). The announced λ-calculus notation for proof search is thus [1, 2]:

$$
\begin{array}{llll}
\text{(terms)} & N & ::= & \lambda x^A.N \mid \mathsf{gfp}\ X^\sigma.E_1 + \cdots + E_n \mid X^\sigma \\
\text{(elimination alternatives)} & E & ::= & xN_1 \ldots N_k
\end{array}
$$

where X is assumed to range over a countably infinite set of *fixpoint variables*, and the sequents σ are supposed to be atomic, i. e., with atomic conclusion. A fixpoint variable may occur with different sequents in a term, but only *well-bound* terms are generated when building a finitary representation of $\mathcal{S}(\sigma)$, and only well-bound terms are given a semantics. In essence, a term is well-bound if the fixed-point operator gfp with X^σ only binds free occurrences of $X^{\sigma'}$ where the σ' are inessential extensions of σ in the sense that they have the same conclusion and maybe more bindings, but only with types/formulas that already have a binding in σ. The main result is that there is a term $\mathcal{F}(\sigma)$ without free fixpoint variables (called *closed* term) whose semantics is (modulo a notion of bisimilarity that considers the sums of alternatives as sets) $\mathcal{S}(\sigma)$ [1].

In subsequent work, based on this framework, the same authors studied (among other questions) a decision algorithm for the problem "given a type A, is there only a finite number of β-normal η-long inhabitants of type A in simply-typed λ-calculus?" [3] (with only a few more references, unfortunately missing out the work by Zaionc, e.g., with David, and with comments on PSPACE-completeness). While in that work, the only quantitative information

*This work was financed by the project *Climt*, ANR-11-BS02-016, of the French Agence Nationale de la Recherche.

that is studied concerns the number of such inhabitants (which would often be smaller under the total discharge convention), the present work refines the analysis of the decision procedure with quantitative information on the depth (height ignoring λ-abstractions, as defined in Hindley's book [4], but a similar notion where variables are assigned value 1 instead of 0 is called height—of Böhm trees—by Takahashi et al [5]).

A measure δ on the terms involving fixpoint variables is defined analogously to depth and height, by ignoring λ-abstractions and by maximizing over sums of elimination alternatives and over arguments (even as in the height definition), but an application is reduced to 0 if any argument has measure 0. And a fixpoint variable counts ∞. One can show that for any term T, $\delta(T) = 0$ iff a decidable predicate of T holds that, for closed terms, is known to characterize the absence of inhabitants [3]. Another result of that previous work characterizes by a predicate FF the closed terms that have only finitely many inhabitants (to be more precise, the characterizations, applied to terms of the form $\mathcal{F}(\sigma)$, have this form, but they need to specify the general situation of not necessarily closed T). The latter result is refined in the present work by bounding the depth of all inhabitants to strictly below $\delta(\mathcal{F}(\sigma))$ in case $\mathsf{FF}(\mathcal{F}(\sigma))$—again, a statement for all T had to be found. This result is always informative since—as can be shown—for all closed T, $\mathsf{FF}(T)$ implies $\delta(T) < \infty$.

Ben-Yelles proved the stretching lemma (see Hindley's book [4, 8D2]) saying that if there is an η-long β-normal inhabitant of type A of depth at least the number n of atoms occurring in A, then there are infinitely many such inhabitants. Turning it around, if there are only finitely many inhabitants, then all inhabitants have depth strictly less than n. By the result above, an alternative proof of the stretching lemma would be obtained by showing that $\delta(\mathcal{F}(\sigma)) \leq n$, with σ the sequent that asks for A in the empty context, under the proviso $\mathsf{FF}(\mathcal{F}(\sigma))$.

Currently, only the case $n = 1$—called the monatomic case [4]—has been solved by the author in this manner. Apart from the trivial cases of an isolated atom (no inhabitant) or an implication without any nesting ($\delta = 1$), at least one composite hypothesis appears. In this case, $\mathsf{FF}(\mathcal{F}(\sigma))$ implies $\delta(\mathcal{F}(\sigma)) = 0$. The proof needs to be more informative in refuting the condition in case any of the appearing hypotheses is the atom alone. And it needs to be more general in speaking also about terms appearing in the construction process for $\mathcal{F}(\sigma)$; it then goes by induction on the term structure (with provisos since the statement is even false for fixpoint variables alone).

The advantage of the proposed method (as is the case already of the cited work it is refining) is that most arguments are by recursion on structure, in particular on the expressions of the λ-calculus for proof search. Hopefully, the full stretching lemma will be obtained following the same path.

References

[1] José Espírito Santo, Ralph Matthes, and Luís Pinto. A coinductive approach to proof search. In David Baelde and Arnaud Carayol, editors, *Proceedings of FICS 2013*, volume 126 of *EPTCS*, pages 28–43, 2013. http://dx.doi.org/10.4204/EPTCS.126.3.

[2] José Espírito Santo, Ralph Matthes, and Luís Pinto. A calculus for a coinductive analysis of proof search. http://arxiv.org/abs/1602.04382, February 2016.

[3] José Espírito Santo, Ralph Matthes, and Luís Pinto. Inhabitation in simply-typed lambda-calculus through a lambda-calculus for proof search. http://arxiv.org/abs/1604.02086, April 2016.

[4] J. Roger Hindley. *Basic Simple Type Theory*, volume 42 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, 1997.

[5] Masako Takahashi, Yohji Akama, and Sachio Hirokawa. Normal proofs and their grammar. *Inf. Comput.*, 125(2):144–153, 1996.