Remark

From Debruijn to co-Debruijn using Category Theory

Everybody's Got To Be Somewhere[1]

Marius Weidner

Chair of Programming Languages, University of Freiburg
weidner@cs.uni-freiburg.de

Abstract. We explore the connection between De Bruijn indices used for variable representation in an intrinsically scoped syntax and the usage of co-De Bruijn representation. To understand the duality between De Bruijn and co-De Bruijn, we will look into the categorical concepts underpinning scopes, binders, and intrinsically scoped syntaxes in general. Lastly, we will address challenges encountered in expressing these concepts and investigate the practical application of co-De Bruijn to specific syntaxes within theorem proving contexts.

Table of Contents

1 Scopes and Binders Categorically

1.1 The Category of Scopes

Definition 1. *A category consists of a set of objects and a set of morphisms. Morphisms are indexed by two objects: source and target. Given a Category $\mathcal{C}$, we denote its objects as $|\mathcal{C}|$, and its morphisms from specified source to target as $\mathcal{C}(S,T)$, where $S, T \in |\mathcal{C}|$.*

Definition 2. *For every set X, we define the category of scopes as Δ_+^X with objects $\bar{x}, \bar{y}, \bar{s} \in X^*$ and morphisms $f, g \in \Delta_+^X(\bar{x}, \bar{y})$ inductively defined by the following inference rules in infix notation[1]:*

$$\frac{}{\varepsilon \sqsubseteq \varepsilon} \; \cdot \qquad \frac{\bar{x} \sqsubseteq \bar{y}}{\bar{x}x \sqsubseteq \bar{y}x} \; 1 \qquad \frac{\bar{x} \sqsubseteq \bar{y}}{\bar{x} \sqsubseteq \bar{y}y} \; 0$$

Example 1. Let $X = \top$, where $\top$ is the set with a single element k. In the category $\Delta_+^\top$, objects correspond to *numbers* $n, p, q \in \top^* = \mathbb{N}$. In this scenario, objects symbolize scopes with n variables, all of the same *kind* k. Morphisms map from a smaller scope p to a larger scope q by selecting p variables in q and mapping all variables from p to them.

$$3 \xrightarrow{\;10101\;} 5$$

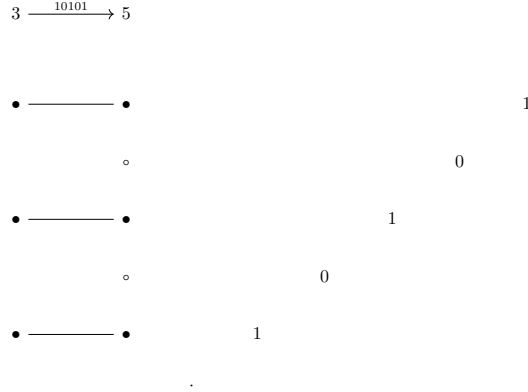|		
•——————•		1
∘		0
•——————•		1
∘	0	
•——————•	1	

.

Fig. 1. Embedding a scope with 3 variables into a scope with 5.

Example 2. Let $X = \mathbb{B}$, where $\mathbb{B}$ is the set with two elements k_1 and k_2. In the category $\Delta_+^\mathbb{B}$, objects correspond to *bit vectors* $\bar{b}, \bar{b}_1, \bar{b}_2 \in \mathbb{B}^*$. Here, objects represent scopes with $\|b\|$ variables of either kind k_1 or k_2. Morphisms map from a smaller scope $\bar{b}_1$ to a larger scope $\bar{b}_2$ by inserting kinds into $\bar{b}_1$ while preserving the order of the variables in $\bar{b}_1$.

---

[1] In the construction of morphisms using inference rules, we omit the axiom rule · most of the time.

Remark 1. Morphisms in Δ_+^X can be represented by *bit vectors* $\bar{b} \in \{0,1\}^*$ with one bit per variable of the target scope telling whether it has been mapped to or skipped by the source scope. We use this representation from time to time.

Definition 3. *We define the infix operation of composition $f;g$ for two morphisms $f \in \Delta_+^X(\bar{x}, \bar{y})$ and $g \in \Delta_+^X(\bar{y}, \bar{z})$, resulting in a morphism $h \in \Delta_+^X(\bar{x}, \bar{z})$, inductively on the inference rules of morphisms:*

$$\begin{aligned}
\cdot \;\;\; ; \;\; \cdot \;\;\; &= \;\; \cdot \\
f1 \;\; ; \;\; g1 \;\; &= \;\; (f;g)1 \\
f0 \;\; ; \;\; g1 \;\; &= \;\; (f;g)0 \\
f \;\;\; ; \;\; g0 \;\; &= \;\; (f;g)0
\end{aligned}$$

Remark 2. The operation $f;g$ reads 'f than g' and applies f first and the result to g, in contrast to $f \circ g$ which reads 'f after g'. The operation $f;g$ reads 'f than g' and executes the morphism f first, and then applies g to the result. This contrasts with the composition $f \circ g$, which reads 'f after g', indicating the application of g followed by f.

Example 3. Let's reconsider the category $\Delta_+^\top$. Objects are represented by numbers, and morphisms can be encoded as bit vectors.
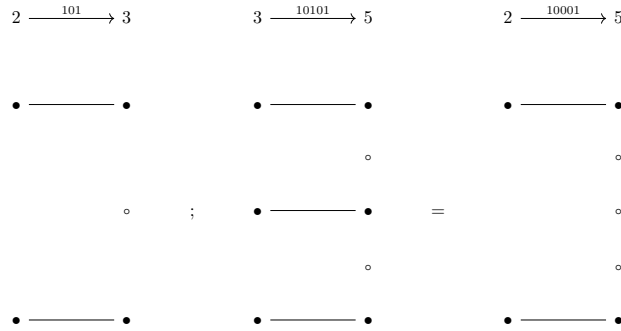

Fig. 2. Composition of two embeddings in the category $\Delta_+^\top$

Corollary 1. *Composition of morphisms in Δ_+^X is associative, i.e., $f;(g;h) = (f;g);h$ for $f \in \Delta_+^X(\bar{x}, \bar{y})$, $g \in \Delta_+^X(\bar{y}, \bar{z})$, and $h \in \Delta_+^X(\bar{x}, \bar{s})$. The proof follows straightforward induction.*

Corollary 2. *For every object $\bar{x}$ in the category Δ_+^X, there exists an identity morphism $id_{\bar{x}} : \bar{x} \sqsubseteq \bar{x}$. The identity morphism can be constructed for every $\bar{x}$ using the axiom inference rule followed by $\|\bar{x}\|$ times applying the 1 rule. Additionally, the equalities $id_{\bar{x}}; f = f$ and $g; id_{\bar{x}} = g$ hold for all $\bar{x}$ and morphisms $f \in \Delta_+^X(\bar{x}, \bar{y})$ and $g \in \Delta_+^X(\bar{y}, \bar{x})$, respectively. This can straightforwardly be proven through induction.*

Proposition 1. *The category of scopes is a well-defined category. The composition operation $f; g$ is associative. Identity morphisms exist within Δ_+^X and act neutral in composition.*

1.2 Intrinsically Scoped De Bruijn Syntax

Definition 4. *The indexed set of intrinsically scoped lambda calculus terms is defined as $\mathbb{N} \to Tm$, where the index corresponds to the number of variables $n \in |\Delta_+^\top| = \mathbb{N}$ present in the term. Terms within this set can be constructed using following inference rules[2]:*

$$\frac{1 \sqsubseteq n}{Tm \; n} \; \# \quad \frac{Tm \; n \quad Tm \; n}{Tm \; n} \; \$ \quad \frac{Tm \; (n+1)}{Tm \; n} \; \lambda$$

Remark 3. The λ rule binds a new variable in its body, incrementing n by one within that body. In contrast, the variable rule $\#$ points to a variable using a morphism in the scope category, precisely selecting one of the n bound variables. The referencing of a variable can be represented by a bit vector of length n with precisely one 1. This representation in turn is equal to a number $m \in \mathbb{N}$ with $m \le n$, indicating the position of the sole 1 in the bit vector.

Example 4. We examine the $\mathbb{K}$ and $\mathbb{S}$ combinators expressed in lambda calculus, comparing their representations with variables as names and intrinsic De Bruijn notation:

$\mathbb{K} = \lambda x.\lambda y.x \qquad\qquad = \lambda \; \lambda \; \#1$
$\mathbb{S} = \lambda x.\lambda y.\lambda z.x \; z \; (y \; z) = \lambda \; \lambda \; \lambda \; \#2 \; \#0 \; (\#1 \; \#0)$

Note that, for example, a De Bruijn term like $\lambda \#1$ could not even be defined using the given inference rules.

2 From De Bruijn to co-De Bruijn

2.1 The Slice Category of Scopes

Definition 5. *If $\mathcal{C}$ is a category and $T, O \in |\mathcal{C}|$, the slice category $\mathcal{C}/O$ has pairs (T, f) as objects, where $f \in \mathcal{C}(T, O)$. A morphism in $(\mathcal{C}/O)((T, f), (S, g))$ is represented by $h \in \mathcal{C}(T, S)$ such that $f = h; g$.*

Definition 6. *We refer to the slice category $\Delta_+^X \backslash \bar{s}$ as the category of subscopes, where objects (subscopes of $\bar{s}$) are denoted by $\bar{b} \in |\Delta_+^X|$ and morphisms (embedding of subscopes given some $\bar{s}$) are represented by $h \in [\Delta_+^X](\bar{b}_1, \bar{b}_2)$.*

Remark 4. Objects $\bar{b}$ in $\Delta_+^X \backslash \bar{s}$ are *bit vectors* $\bar{b} \in \{0, 1\}^*$ with one bit per variable of scope $\bar{s}$, telling whether it has been selected.

---

[2] When constructing the $\$$ rule, we usually omit the dollar sign and, instead, express it by using a space between the two subterms.

Example 5. The morphism 0111 in $\Delta_+^T \setminus 5$ from 01110 to 11110, denoted as $(3, 01110)$ to $(4, 11110)$, embeds a scope with 3 variables into one with 4 by inserting an additional unused variable at the very beginning. Since we sliced with object 5, both 01110 and 11110 are subscopes of scope 4, and the morphism 0111 effectively embeds one subscope into the other.
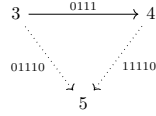

Fig. 3. In the diagram, objects in the slice category $\Delta_+^\top \setminus 5$ are represented by dotted arrows along with their source, while normal arrows are actual morphisms.
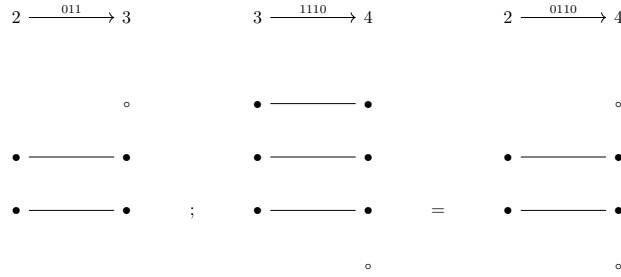
Example 6. We can also compose embeddings of subscopes:

Fig. 4. In the diagram, objects in the slice category $\Delta_+^\top \setminus 4$ at the top are represented solely by their numbers. The corresponding bit vector can be inferred from the dots and lines below.

Definition 7.

2.2   A Monad over Sets Indexed by Scopes

2.3   The Notion of Relevant Pairs

2.4   Intrinsically Scoped co-De Bruijn Syntax

References

[1]   Conor McBride. "Everybody's Got To Be Somewhere". In: *Electronic Proceedings in Theoretical Computer Science* 275 (July 2018), pp. 53–69. ISSN: 2075-2180. DOI: 10.4204/eptcs.275.6. URL: http://dx.doi.org/10.4204/EPTCS.275.6.