On the Properties of B-terms

Keisuke Nakano[1]

1 **The University of Electro-Communications**
1-5-1 Chofugaoka, Japan
`ksk@cs.uec.ac.jp`

Abstract

B-terms are built from the B combinator alone defined by $B \equiv \lambda f.\lambda g.\lambda x.f\ (g\ x)$, which is well-known as a function composition operator. This paper introduces an interesting property of B-terms, that is, whether repetitive right applications of a B-term circulates or not. The decision problem of the property is investigated through a canonical representation of B-terms and a sound and complete equational axiomatization. Several remaining problems related to the property are also discussed.

1998 ACM Subject Classification F.4.1 Mathematical Logic

Keywords and phrases Combinatory logic, B combinator, Equational axiomatization

Digital Object Identifier 10.4230/LIPIcs.xxx.yyy.p

1 Introduction

The 'bluebird' combinator $B = \lambda f.\lambda g.\lambda x.f\ (g\ x)$ is well-known [8] as a bracketing combinator or composition operator, which has a principal type $(\alpha \to \beta) \to (\gamma \to \alpha) \to \gamma \to \beta$. A function $B\ f\ g$ (also written as $f \circ g$) synthesized from two functions f and g takes a single argument to apply g and returns the result of an application f to the output of g.

In the general case where g takes n arguments to pass the output to f, the synthesized function is defined by $\lambda x_1. \cdots \lambda x_n.f\ (g\ x_1\ \ldots\ x_n)$. Interestingly, the function can be expressed by $B^n\ f\ g$ where e^n is an n-fold composition of function e such that $e^0 = \lambda x.x$ and $e^{n+1} = B\ e^n\ e$ for $n \geq 0$. We call *n-argument composition* for the generalized composition represented by B^n.

Now we consider the 2-argument composition expressed by B^2. From the definition, we have $B^2 = B\ B\ B$ where function application is considered left-associative, that is, $f\ g\ h = (f\ g)\ h$. Thus B^2 is defined by an expression in which all applications nest to the left, never to the right. We call such an expression *flat* [7]. In particular we write $X_{(n)}$ for a flat expression involving only a combinator X, which is defined by

$$X_{(1)} = X \qquad\qquad X_{(n+1)} = X_{(n)}\ X \quad (n \geq 1).$$

For instance, we write $B^2 = B_{(3)}$. Okasaki [7] shows that there is no universal combinator X that can represent any combinator by $X_{(n)}$ with some n.

Let us consider the 3-argument composition expressed by B^3. Does n exist such that

Conference title on which this volume is based on.
Editors: Billy Editor and Bill Editors; pp. 1–14
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

$$\begin{tikzcd}[column sep=small]
B_{(1)} \arrow[r, dash] & B_{(2)} \arrow[r, dash] & B_{(3)} \arrow[r, dash] & B_{(4)} \arrow[r, dash] & B_{(5)} \arrow[r, dash] & \begin{array}{l} B_{(6)} \\ (= B_{(10)} = B_{(14)} = \dots) \end{array} \arrow[r, dash] \arrow[d, dash] & \begin{array}{l} B_{(7)} \\ (= B_{(11)} = B_{(15)} = \dots) \end{array} \arrow[d, dash] \\
& & & & & \begin{array}{l} B_{(9)} \\ (= B_{(13)} = B_{(17)} = \dots) \end{array} \arrow[r, dash] & \begin{array}{l} B_{(8)} \\ (= B_{(12)} = B_{(16)} = \dots) \end{array}
\end{tikzcd}$$

Figure 1 ρ-property of the B combinator

$B^3 = B_{(n)}$? The answer is yes. Using $\underline{f}\ (\underline{\underline{g}}\ x) = B\ f\ g\ x$, we have

$$\begin{aligned}
B^3 &= B\ B^2\ B \\
&= \underline{B}\ (\underline{\underline{B\ B}}\ B)\ B \\
&= \underline{B\ B}\ (\underline{\underline{B}}\ B)\ B\ B \\
&= \underline{B}\ (\underline{\underline{B}}\ B)\ B\ B\ B\ B \\
&= B\ B\ B\ B\ B\ B\ B\ B.
\end{aligned}$$

Hence, $B^3 = B_{(8)}$.

How about the 4-argument composition expressed by B^4? Indeed, there is no integer n such that $B^4 = B_{(n)}$ with respect to $\beta\eta$-equality. It can be proved by *ρ-property* of combinator B, that is introduced in this paper. We say that a combinator X has ρ-property if there exists two distinct integer i and j such that $X_{(i)} = X_{(j)}$. If such a pair i, j is found, we have $X_{(i+k)} = X_{(j+k)}$ for any $k \geq 0$ (a la *finite monogenic semigroup* [3]). In the case of B, we can check $B_{(6)} = B_{(10)} = \lambda x.\lambda y.\lambda z.\lambda w.\lambda v.\ x\ (y\ z)\ (w\ v)$ hence $B_{(i)} = B_{(i+4)}$ for $i \geq 6$. Fig. 1 shows a computation graph of $B_{(k)}$. The ρ-property is named after the shape of the graph. The ρ-property implies that the set $\{\ B_{(k)} \mid k \geq 1\ \}$ is finite. Since none of the terms in the set is equal to B^4 up to the $\beta\eta$-equivalence of the corresponding λ-terms, we conclude that there is no integer n such that $B^4 = B_{(n)}$.

This paper discusses the ρ-property of combinatory terms, particularly in $\mathbf{CL}(B)$, called *B-terms*, that are built from B alone. Interestingly, $B\ B$ enjoys the ρ property with $(B\ B)_{(52)} = (B\ B)_{(32)}$ and so does $B\ (B\ B)$ with $(B\ (B\ B))_{(294)} = (B\ (B\ B))_{(258)}$ as reported by the author [4, 5, 6]. Several combinators other than B-terms can be found enjoy the ρ-property, for example, $K = \lambda x.\lambda y.x$ and $C = \lambda x.\lambda y.\lambda z.\ x\ z\ y$ because of $K_{(3)} = K_{(1)}$ and $C_{(4)} = C_{(3)}$. They are not so interesting in the sense that the cycle starts immediately and its size is very small, comparing with B-terms like $B\ B$ and $B\ (B\ B)$. As we will see later, $B\ (B\ (B\ (B\ (B\ (B\ B)))))(\equiv B^6\ B)$ has the ρ-property with the cycle of the size more than 3×10^{11} which starts after more than 2×10^{12} repetitive right application. This is why the ρ-property of B-terms is intensively discussed in the present paper. From his observation, the author conjectured [5] that "B-term X has the ρ-property if and only if X is equivalent to $B^n B$ with some n".

The contributions of the paper are two-fold. One is to give a characterization of $\mathbf{CL}(B)$ and another is to provide a necessary condition for the anti-ρ-property of B-terms. In the former, a canonical representation of B-terms is introduced and sound and complete equational axiomatization for $\mathbf{CL}(B)$ is established. In the latter, the only-if-part of the conjecture above is illustrated. The if-part still remains open, though.

$\rho(B^0B) = (6, 4)$	$\rho(B^4B) = (191206, 431453)$
$\rho(B^1B) = (32, 20)$	$\rho(B^5B) = (766241307, 234444571)$
$\rho(B^2B) = (258, 36)$	$\rho(B^6B) = (2641033883877, 339020201163)$
$\rho(B^3B) = (4240, 5796)$	

Figure 2 ρ-properties of B-terms in a particular form

2 ρ-property of terms

The ρ-property of combinator X is that $X_{(i)} = X_{(j)}$ holds for some $i > j \geq 0$. We adopt $\beta\eta$-equivalence of corresponding λ-terms for the equivalence of combinatory terms in this paper. Furthermore, for simplicity, we only deal with the case where $X_{(n)}$ is normalizable for all n.

Let us write $\rho(X) = (i, j)$ when a combinator X has the ρ-property due to $X_{(i)} = X_{(i+j)}$ with minimum positive integers i and j. For example, we can write $\rho(B) = (6, 4)$, $\rho(C) = (3, 1)$, $\rho(K) = (1, 2)$ and $\rho(I) = (1, 1)$. Besides them, several combinators introduced in the Smullyan's book [8] has ρ-properties:

$$
\begin{aligned}
&\rho(D) = (32, 20) && \text{where } D = \lambda x.\lambda y.\lambda z.\lambda w.x\ y\ (z\ w)\\
&\rho(F) = (3, 1) && \text{where } F = \lambda x.\lambda y.\lambda z.z\ y\ x\\
&\rho(R) = (3, 1) && \text{where } R = \lambda x.\lambda y.\lambda z.y\ z\ x\\
&\rho(T) = (2, 1) && \text{where } T = \lambda x.\lambda y.y\ x\\
&\rho(V) = (3, 1) && \text{where } V = \lambda x.\lambda y.\lambda z.z\ x\ y.
\end{aligned}
$$

Except the B and $D(= B\ B)$ combinator, the property is 'trivial' in the sense that loop starts early and the size of cycle is very small.

No other combinators in the book have the ρ-property. The reasons are divided into two: (A) all terms in the sequence generated by repetitive right application differ; (B) a certain number of repetitive right application generates non-normalizable terms. The combinators $S = \lambda x.\lambda y.\lambda z.x\ z\ (y\ z)$, and $O = \lambda x.\lambda y.y\ (x\ y)$ do not have the ρ-property for reason (A), which is illustrated by

$$
\begin{aligned}
S_{(2n+1)} &= \lambda x.\lambda y.\underbrace{x\ y\ (x\ y\ (\ldots (x\ y}_{n}\ (\lambda z.x\ z\ (y\ z)))\ldots)), \text{ and}\\
O_{(n+1)} &= \lambda x.\underbrace{x\ (x\ (\ldots (x}_{n}\ (\lambda y.y\ (x\ y)).
\end{aligned}
$$

The combinators $L = \lambda x.\lambda y.x\ (y\ y)$, $M = \lambda x.x\ x$, $U = \lambda x.\lambda y.y\ (x\ x\ y)$ and $W = \lambda x.\lambda y.x\ y\ y$ do not have the ρ-property for reason (B). We can check them by that $L_{(4)}$, $M_{(2)}$, $U_{(2)}$ and $W_{(4)}$ are not normalizable.

The definition of the ρ-property is naturally extended from single combinators to terms built from combinators. The author found by computation that several B-terms, built from the B combinator alone, has nontrivial ρ-properties as shown in Fig. 2. The implementation is based on two techniques: one is the Floyd's cycle-finding algorithm [2] (also called the tortoise and the hare algorithm) to detect the ρ-property with a constant memory usage; another is a canonical representation of B-terms, which will be introduced in the next section, to perform right application and equivalence check efficiently. Even by such efficient algorithms, it took more than 1 month to find the ρ-property of $B^6\ B$.

3 Characterization of B-terms

We call B-terms for combinatory terms constructed only the B combinator $\lambda x.\lambda y.\lambda z.x\ (y\ z)$, and let $\mathbf{CL}(B)$ denote the set of all B-terms (or corresponding λ-terms). If e_1 and e_2, e are B-terms, then $e_1 \circ e_2$ and e^n with non-negative integer n are also B-terms from the definition of $\circ$ and $-^n$.

Certain of B-terms are equivalent in a sense of the $\beta\eta$-equality over corresponding λ-terms. For instance, $B\ B\ (B\ B)$ and $B\ (B\ B)\ B\ B$ are β-equivalent which is easy to find from $B\ x\ y\ z = x\ (y\ z)$. For another (non-trivial) instance, $B\ B\ (B\ B)$ and $B\ (B\ (B\ B))\ B$ are equivalent which is illustrated by the fact that two are equivalent to $\lambda x.\lambda y.\lambda z.\lambda w.\lambda v.x\ (y\ z)\ (w\ v)$. These kinds of equality make it hard to investigate the ρ-property of B-terms.

To solve the annoying issue, a canonical representation of B-terms is introduced in this section. We will finally find that for any B-term e there exists a unique finite non-empty weakly-decreasing sequence of non-negative integers $n_1 \geq n_2 \geq \cdots \geq n_k$ such that e is equivalent to $(B^{n_1}\ B) \circ (B^{n_2}\ B) \cdots \circ (B^{n_k}\ B)$. It will be proved after introducing several equation laws on B-terms.

3.1 Three equation laws on B-terms

First three equation laws on B-terms are presented. As shown later, they are enough to show any equation of B terms.

▶ **Lemma 1.** *Let e_1, e_2, and e_3 be combinatory terms. Then the following equations hold:*

$$B\ e_1\ e_2\ e_3 = e_1\ (e_2\ e_3) \tag{1}$$

$$B\ (B\ e_1\ e_2) = B\ (B\ e_1)\ (B\ e_2) \tag{2}$$

$$B\ B\ (B\ e_1) = B\ (B\ (B\ e_1))\ B \tag{3}$$

Proof. Equation (1) is immediate from the definition of B. Equation (2) is shown by

$$\begin{aligned}
B\ (B\ e_1\ e_2) &= \lambda x.\lambda y.\ B\ e_1\ e_2\ (x\ y)\\
&= \lambda x.\lambda y.\ e_1\ (e_2\ (x\ y))\\
B\ (B\ e_1)\ (B\ e_2) &= \lambda x.B\ e_1\ (B\ e_2\ x)\\
&= \lambda x.\lambda y.\ e_1\ (B\ e_2\ x\ y)\\
&= \lambda x.\lambda y.\ e_1\ (e_2\ (x\ y))
\end{aligned}$$

where the α-renaming is implicitly used. Similarly, equation (3) is shown by

$$\begin{aligned}
B\ B\ (B\ e_1) &= \lambda x.\ B\ (B\ e_1\ x)\\
&= \lambda x.\lambda y.\lambda z.\ B\ e_1\ x(y\ z)\\
&= \lambda x.\lambda y.\lambda z.\ e_1\ (x\ (y\ z))\\
B\ (B\ (B\ e_1))\ B &= \lambda x.\ B\ (B\ e_1)\ (B\ x)\\
&= \lambda x.\lambda y.\ B\ e_1\ (B\ x\ y)\\
&= \lambda x.\lambda y.\lambda z.\ e_1\ (B\ x\ y\ z)\\
&= \lambda x.\lambda y.\lambda z.\ e_1\ (x\ (y\ z)).
\end{aligned}$$

◀

Equation (2) has been employed by Statman [9] to show that no $B\omega$-term can be a fixed-point combinator where $\omega = \lambda x.x\ x$. This equation exposes an interesting feature of the B combinator. Write the equation as

$$B\ (e_1 \circ e_2) = (B\ e_1) \circ (B\ e_2) \tag{4}$$

by replacing every B combinator with $\circ$ infix operator if it has exactly two arguments. The equation is an distributive law of B over $\circ$, which will be used to obtain the canonical representation of B-terms. Equation (3) is also used for the same purpose as the form of

$$B \circ (B\ e_1) = (B\ (B\ e_1)) \circ B. \tag{5}$$

One may expect a natural equation

$$B\ (B\ e_1\ e_2)\ e_3 = B\ e_1\ (B\ e_2\ e_3) \tag{6}$$

which represents associativity of function composition, i.e., $(e_1 \circ e_2) \circ e_3 = e_1 \circ (e_2 \circ e_3)$. This is shown with equations (2) and (1) in Lemma 1 by

$$B\ (B\ e_1\ e_2)\ e_3 = B\ (B\ e_1)\ (B\ e_2)\ e_3 = B\ e_1\ (B\ e_2\ e_3).$$

3.2 Canonical representation of B-terms

We only consider B-terms in which every B has at most two arguments. One can easily reduce the number of arguments of every B to less than 3 by repeatedly rewriting the occurrence of $B\ e_1\ e_2\ e_3$ into $e_1\ (e_2\ e_3)$ as long as such a pattern exists. This repetition terminates because each rewriting decreases the number of B. Thereby B-terms treated in this paper range over e defined by the syntax:

$$e ::= B \ \mid\ B\ e\ \mid\ e \circ e$$

where the $\circ$ infix operator is used for B that has exactly two arguments. The advantage of $\circ$ is to employ its associativity implicitly, avoiding to use equation (6) everywhere. Parentheses are needed due to the left associativity of the space (for application) and its precedence over $\circ$, e.g., $B\ (B\ B)$ and $B\ (B \circ B)$.

According to the syntactic definition above, the B combinators and the $\circ$ operators occur in arbitrary position. A histogram representation puts a restriction to the syntax so that no B combinator occurs outside of the $\circ$ operator, that is, the syntax is given as

$$\begin{aligned} e &::= e_B \mid\ e \circ e \\ e_B &::= B \mid B\ e_B \end{aligned}$$

The histogram representation is obtained by replacing $\underbrace{B\ (\dots (B}_{n}\ \ B) \dots)$ by $B^n\ B$.

▶ **Definition 2.** A *histogram representation* is given as the form of

$$(B^{n_1}\ B) \circ (B^{n_2}\ B) \cdots \circ (B^{n_k}\ B)$$

where $k > 0$ and $n_1, \dots, n_k \geq 0$ are integers. In particular, a histogram representation is called *decreasing* when $n_1 \geq n_2 \geq \cdots \geq n_k$. The *length* of a histogram representation h is defined by adding 1 to the number of $\circ$ in h. The numbers $n_1, n_2, \dots, n_k$ are called *height*.

In the rest of this subsection, we prove that for any B-term e there exists a unique and decreasing histogram representation equivalent to e. First, we show the existence of at least one histogram representation equivalent to e.

▶ **Lemma 3.** For any B-term e, there exists a histogram representation equivalent to e.

Proof. Let e be a B-term. We prove the statement by induction on structure of e. In the case of $e \equiv B$, the term itself is a histogram representation. In the case of $e \equiv B\ e_1$, assume that e_1 has equivalent histogram representation $(B^{n_1}\ B) \circ (B^{n_2}\ B) \cdots \circ (B^{n_k}\ B)$. Repeatedly applying equation (4) to $B\ e_1$, we obtain the histogram representation as $B\ e_1 = (B^{n_1+1}\ B) \circ (B^{n_2+1}\ B) \cdots \circ (B^{n_k+1}\ B)$. In the case of $e \equiv e_1 \circ e_2$, assume that e_1 and e_2 have equivalent histogram representations h_1 and h_2, respectively. Then the histogram representation equivalent to $e_1 \circ e_2$ is given by $h_1 \circ h_2$. ◀

▶ Remark. Histogram representations of length 1 are called *monomial* while the others are called *polynomial* in [9], where these terminologies are defined for $\mathbf{CL}(B, M)$ with $M \equiv \lambda x.\ x\ x$, though. Corollary 0.1 in the paper is a statement similar to the above.

Next we show that for any histogram representation h there exists a decreasing histogram representation equivalent to h. A key equation of the proof is

$$(B^m\ B) \circ (B^n\ B) = (B^{n+1}\ B) \circ (B^m\ B) \quad \text{when} \quad m < n, \tag{7}$$

which is shown by

$$\begin{aligned}
(B^m\ B) \circ (B^n\ B) &= B^m\ (B \circ (B^{n-m}\ B)) \\
&= B^m\ (B \circ (B(B^{n-m-1}\ B))) \\
&= B^m\ ((B(B(B^{n-m-1}\ B))) \circ B) \\
&= (B^{n+1}\ B) \circ (B^m\ B)
\end{aligned}$$

using equations (4) and (5).

▶ **Lemma 4.** For any histogram representation h, there exists a decreasing histogram representation h' equivalent to h such that

- The length of h and h' are equal, and
- The smallest heights of h and h' are equal.

Proof. Let h be a histogram representation. We prove the statement by induction on length of h. When the length is 1, that is, h includes no $\circ$ operator, h itself is decreasing and the statement holds. When the length of h is k, assume that any histogram representation of length $k-1$ has equivalent decreasing histogram representations. From the hypothesis, taking h_1 such that $h \equiv h_1 \circ (B^n\ B)$, there exists a decreasing histogram representation $h'_1 \equiv (B^{n_1}\ B) \circ (B^{n_2}\ B) \cdots \circ (B^{n_{k-1}}\ B)$ equivalent to h_1. If $n_{k-1} \geq n$, then $h' \equiv h'_1 \circ (B^n\ B)$ is decreasing and equivalent to h. Since the smallest heights of h and h' are n, the statement holds. If $n_{k-1} < n$, h is equivalent to

$$\begin{aligned}
&(B^{n_1}\ B) \circ (B^{n_2}\ B) \cdots \circ (B^{n_{k-2}}\ B) \circ (B^{n_{k-1}}\ B) \circ (B^n\ B) \\
&= (B^{n_1}\ B) \circ (B^{n_2}\ B) \cdots \circ (B^{n_{k-2}}\ B) \circ (B^{n+1}\ B) \circ (B^{n_{k-1}}\ B)
\end{aligned}$$

due to equation (7). Putting the last term as $h_2 \circ (B^{n_{k-1}}\ B)$, the length of h_2 is $k-1$ and the smallest height of h_2 is greater than or equal to n_{k-1}. From the induction hypothesis, h_2 has an equivalent decreasing histogram representation h'_2 of length $k-1$ and the smallest height of h'_2 greater than or equal to n_{k-1}. Thereby we obtain a decreasing histogram representation $h'_2 \circ (B^{n_{k-1}}\ B)$ equivalent to h and the statement holds. ◀

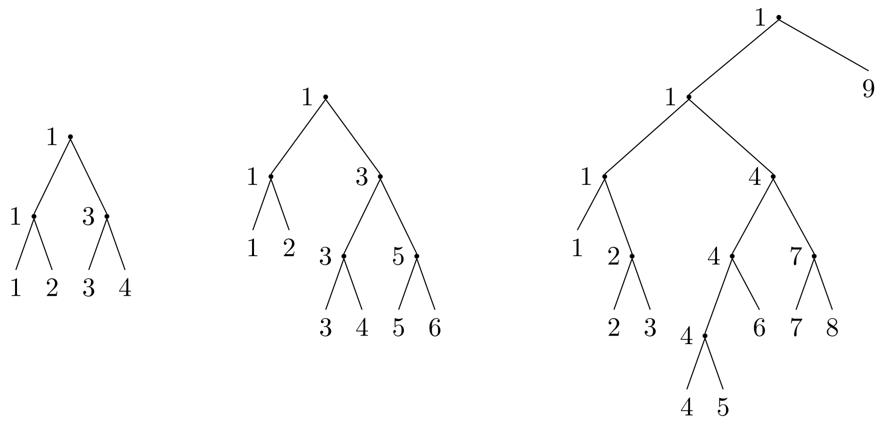

Figure 3 Labeled binary trees obtained from λ-terms e_1, e_2 and e_3 by the labeling procedure

Every B-term has an equivalent decreasing histogram representation as shown so far. To conclude this subsection, we show the uniqueness of decreasing histogram representations for B-terms, that is, for any B-term e, there is no two distinct decreasing histogram representation equivalent to e.

The proof is based on the idea that B-terms correspond to unlabeled binary trees. Since every λ-term in $\mathbf{CL}(B)$ is linear and ordered, that is, every argument of a function is used exactly once in the order it was applied, the corresponding λ-term must be in the form of $\lambda x_1.\lambda x_2.\ldots.\lambda x_k.\ e$ where e is build by putting parentheses to appropriate positions in the sequence $x_1\ x_2\ \ldots\ x_k$. Whereas the body e can be regarded as an unlabeled binary tree consisting of application and variables, a λ-term in $\mathbf{CL}(B)$ is constructed for any unlabeled binary tree by putting a variable to each leaf in the order of $x_1, x_2, \ldots$ and enclosing it with k-fold lambda abstraction $\lambda x_1.\lambda x_2.\ldots.\lambda x_k.[\ ]$ where k is a number of leaves of the binary tree.

A *labeling* procedure for a given unlabeled binary tree puts a label to nodes and leaves of the tree in the following steps. First each leaf of a given tree is labeled by 1,2,. . . , in left-to-right order. Then each binary node of the tree is labeled by the same label as its leftmost descendant leaf. Figure 3 shows examples of labeled binary trees obtained by the labeling procedure: each of which corresponding to λ-terms in $\mathbf{CL}(B)$,

$$e_1 \equiv \lambda x_1.\lambda x_2.\lambda x_3.\lambda x_4.\ x_1\ x_2\ (x_3\ x_4),$$
$$e_2 \equiv \lambda x_1.\lambda x_2.\lambda x_3.\lambda x_4.\lambda x_5.\lambda x_6.\ x_1\ x_2\ (x_3\ x_4\ (x_5\ x_6)), \text{ and}$$
$$e_3 \equiv \lambda x_1.\lambda x_2.\lambda x_3.\lambda x_4.\lambda x_5.\lambda x_6.\lambda x_7.\lambda x_8.\lambda x_9.\ x_1\ (x_2\ x_3)\ (x_4\ x_5\ x_6\ (x_7\ x_8))\ x_9.$$

From any distinct λ-terms in $\mathbf{CL}(B)$, the labeling procedure constructs different labeled binary trees.

We define a function $\mathcal{L}$ which characterizes λ-terms in $\mathbf{CL}(B)$ using the labeling procedure. The $\mathcal{L}$ function takes λ-terms in $\mathbf{CL}(B)$ and returns a bag (also called multiset) that consists of only node labels of the binary tree obtained from the lambda term with the labeling procedure, provided that label 1's are excluded. For example, we have

$$\mathcal{L}(e_1) = \{3\},$$
$$\mathcal{L}(e_2) = \{3, 3, 5\}, \text{ and}$$
$$\mathcal{L}(e_3) = \{2, 4, 4, 4, 7\}$$

for the λ-terms above. Note that by excluding the label 1's it may happen to be $\mathcal{L}(e) = \mathcal{L}(e')$ for two distinct λ-terms e and e' in $\mathbf{CL}(B)$. However, e and e' must be η-equivalent due to the definition of the labeling procedure. This is one of reasons why the definition of $\mathcal{L}$ excludes the label 1's. Another reason is to introduce an important fact on the correspondence between $\mathcal{L}(e)$ and its decreasing histogram representation equivalent to e, that is,

If $e \in \mathbf{CL}(B)$ is equivalent to $(B^{n_1}\ B) \circ (B^{n_2}\ B) \circ \cdots \circ (B^{n_k}\ B)$ with $n_1 \geq \cdots \geq n_k$, then $\mathcal{L}(e) = \{n_1+2, n_2+2, \ldots, n_k+2\}$ holds, $\cdots(\bigstar)$

which is shown in the proof of the following lemma that illustrates the uniqueness of decreasing histogram representation. Thereby, we could find

$$\begin{aligned}
e_1 &= B^1\ B,\\
e_2 &= (B^3\ B) \circ (B^1\ B) \circ (B^1\ B), \text{ and}\\
e_3 &= (B^5\ B) \circ (B^2\ B) \circ (B^2\ B) \circ (B^2\ B) \circ (B^0\ B)
\end{aligned}$$

where the last equation holds up to η-equivalence.

▶ **Lemma 5.** B-terms e_1 and e_2 are not equivalent if they have distinct decreasing histogram representations.

Proof. It suffices to show the statement ($\bigstar$) because B-terms e_1 and e_2 are not equivalent if $\mathcal{L}(e_1)$ and $\mathcal{L}(e_2)$ differs. We prove the statement that for any decreasing histogram representation $h = (B^{n_1}\ B) \circ (B^{n_2}\ B) \circ \cdots \circ (B^{n_k}\ B)$ and any λ-term e equivalent to h, $\mathcal{L}(e) = \{n_1+2, n_2+2, \ldots, n_k+2\}$ holds, by induction on length of h.

When $h = B^n\ B$, the corresponding λ-term is found to form

$$e \equiv \lambda x_1.\lambda x_2.\ldots.\lambda x_{n+1}.\lambda x_{n+2}.\lambda x_{n+3}.\ x_1\ x_2\ \ldots\ x_{n+1}\ (x_{n+2}\ x_{n+3})$$

by induction on n. Since we have $\mathcal{L}(e) = \{n+2\}$, the statement holds.

When $h = (B^{n_1}\ B) \circ \cdots \circ (B^{n_{k-1}}\ B) \circ (B^n\ B)$ with $k \geq 2$ and $n_1 \geq \ldots n_{k-1} \geq n$, suppose that a λ-term e is equivalent to $(B^{n_1}\ B) \circ \cdots \circ (B^{n_{k-1}}\ B)$. From the induction hypothesis, $\mathcal{L}(e) = \{n_1+2, \ldots, n_{k-1}+2\}$ holds. Therefore, e must have the form of $\lambda x_1.\ldots.\lambda x_l.\ x_1\ \ldots\ x_{n_{k-1}+1}\ (x_{n_{k-1}+2} \ldots) \ldots$ in which none of $x_2, \ldots, x_{n_{k-1}+1}$ occurs in function position. From $B^n\ B = \lambda y_1.\ldots.\lambda y_{n+3}.\ y_1\ y_2\ \ldots\ y_{n+1}\ (y_{n+2}\ y_{n+3})$ and $n_{k-1} \geq n$, we compute a λ-term e' corresponding to h by

$$\begin{aligned}
h &= \lambda x.e\ (B^n\ B\ x)\\
&= \lambda x.(\lambda x_1.\ldots.\lambda x_l.\ x_1\ \ldots\ x_{n_{k-1}+1}\ (x_{n_{k-1}+2} \ldots) \ldots)\\
&\qquad\qquad (\lambda y_2.\ldots.\lambda y_{n+3}.\ x\ y_2\ \ldots\ y_{n+1}\ (y_{n+2}\ y_{n+3})\\
&= \lambda x.\lambda x_2.\ldots.\lambda x_l.\ (\lambda y_2.\ldots.\lambda y_{n+3}.\ x\ y_2\ \ldots\ y_{n+1}\ (y_{n+2}\ y_{n+3}))\\
&\qquad\qquad\qquad\qquad x_2\ \ldots\ x_{n_{k-1}+1}\ (x_{n_{k-1}+2} \ldots) \ldots\\
&= \lambda x.\lambda x_2.\ldots.\lambda x_l.\ (\lambda y_{n+1}.\lambda y_{n+2}.\lambda y_{n+3}.\ x\ x_2\ \ldots\ x_n\ y_{n+1}\ (y_{n+2}\ y_{n+3}))\\
&\qquad\qquad\qquad\qquad x_{n+1}\ \ldots\ x_{n_{k-1}+1}\ (x_{n_{k-1}+2} \ldots) \ldots\\
&= \begin{cases}
\lambda x.\lambda x_2.\ldots.\lambda x_l.\ x\ x_2\ \ldots\ x_n\ x_{n+1}\ (x_{n+2} \ldots\ldots) & (n = n_{k-1})\\
\lambda x.\lambda x_2.\ldots.\lambda x_l.\ x\ x_2\ \ldots\ x_n\ x_{n+1}\ (x_{n+2}\ (x_{n_{k-1}+2} \ldots)) \ldots & (n+1 = n_{k-1})\\
\lambda x.\lambda x_2.\ldots.\lambda x_l.\ x\ x_2\ \ldots\ x_n\ x_{n+1}\ (x_{n+2}\ (x_{n+3} \ldots)) \ldots & (n+2 \leq n_{k-1}).
\end{cases}
\end{aligned}$$

Renaming x into x_1, we can find that the 'offset' of variables are preserved, hence, all of $\mathcal{L}(e)$ are in $\mathcal{L}(e')$. Since only x_{n+2} is added to $\mathcal{L}(e')$ in any case, the statement holds. ◀

From Lemmas 3, 4 and 5, we can conclude the uniqueness of decreasing histogram representation shown in the following theorem.

▶ **Theorem 6.** Every B-term e has a unique decreasing histogram representation.

3.3 Equational axiomatization of B-terms

Using the canonical representation of B-terms, it is not difficult to give equational axiomatization to $\mathbf{CL}(B)$. Let us put three equations in Lemma 1 as axioms. Then it forms a sound and complete equational system for $\mathbf{CL}(B)$, where the soundness means that all equations over B-terms derived from these axioms must be correct equations and the completeness means that all equations over B-terms can be derived from these axioms. Although this axiomatization does not directly contribute to discussion on the ρ-property of B-terms, it is worthwhile to mention it as one of characterization of $\mathbf{CL}(B)$.

The soundness is immediate from Lemma 1. The proof of the completeness is given as follows. Let e_1 and e_2 be equivalent B-terms. From Lemma 5, they have the same decreasing histogram representation. Note that even if the histogram representation is the same, they can differ as B-terms because of the $\circ$ operator. There are many possibilities of replacing $\circ$ with the B combinator because of the associativity of $\circ$. It does not matter, however, since the associativity (6) can be proved by equations (2) and (1).

▶ **Theorem 7.** Three equations

$$\begin{aligned} B\ e_1\ e_2\ e_3 &= e_1\ (e_2\ e_3) \\ B\ (B\ e_1\ e_2) &= B\ (B\ e_1)\ (B\ e_2) \\ B\ B\ (B\ e_1) &= B\ (B\ (B\ e_1))\ B \end{aligned}$$

gives a sound and complete equational system for $\mathbf{CL}(B)$.

4 Proving (anti-)ρ-property of B-terms

The ρ-property of a term can be investigated by sharp observation on its right application. In this section, we first observe application of one B-terms to another in terms of decreasing histogram representation. Then we prove that no B term which has histogram representation of length more than 1 has the ρ-property. It implies that a B-term must be equivalent to $B^n\ B$ with some n if it has the ρ-property.

4.1 Application on decreasing histogram representation

For two given decreasing histogram representations h_1 and h_2, a decreasing histogram representation h equivalent to $h_1\ h_2$ can be derived systematically as introduced later. The method is based on the following lemma about application of one B-term to another B-term.

▶ **Lemma 8.** Let e_1 and e_2 be B-terms. Then,

$$e_1 \circ (B\ e_2) = B\ (e_1\ e_2) \circ B^k$$

holds with some $k \geq 0$.

Proof. Let h_1 be a decreasing histogram representation equivalent to e_1. We prove the statement by case analysis on the maximum height in h_1. When the maximum height is 0, we can take $k' \geq 1$ such that $h_1 \equiv \underbrace{B \circ \cdots \circ B}_{k'} = B^{k'}$. Then,

$$e_1 \circ (B\ e_2) = \underbrace{B \circ \cdots \circ B}_{k'} \circ (B\ e_2) = (B^{k'+1}\ e_2) \circ \underbrace{B \circ \cdots \circ B}_{k'} = B\ (e_1\ e_2) \circ B^{k'}$$

where equation (5) is used k' times in the second equation. Therefore the statement holds by with $k = k'$. When the maximum height is greater than 0, we can take a decreasing histogram representation h' and $k' \geq 0$ such that $h_1 = (B\ h') \circ \underbrace{B \circ \cdots \circ B}_{k'} = (B\ h') \circ B^{k'}$ due to equation (4). Then,

$$\begin{aligned} e_1 \circ (B\ e_2) &= (B\ h') \circ \underbrace{B \circ \cdots \circ B}_{k'} \circ (B\ e_2) \\ &= (B\ h') \circ (B^{k'+1}\ e_2) \circ \underbrace{B \circ \cdots \circ B}_{k'} \\ &= B\ (h' \circ (B^{k'}\ e_2)) \circ B^{k'} \end{aligned}$$

Since $e_1\ e_2 = h_1\ e_2 = B\ h'\ (B^{k'}\ e_2) = h' \circ (B^{k'}\ e_2)$, the statement holds with $k = k'$. ◀

This lemma indicates that, from two decreasing histogram representations h_1 and h_2, a decreasing histogram representation h equivalent to $h_1\ h_2$ can be obtained by the following steps:

1. Build h'_2 by raising each height of h_2 by 1, i.e., when $h_2 \equiv (B^{n_1}\ B) \circ \cdots \circ (B^{n_l}\ B)$, $h'_2 \equiv (B^{n_1+1}\ B) \circ \cdots \circ (B^{n_l+1}\ B)$.
2. Find a decreasing histogram representation h_{12} corresponding to $h_1 \circ h'_2$;
3. Obtain h by lowering each height of h_{12} after eliminating the trailing 0-height units, i.e., when $h_{12} \equiv (B^{n_1}\ B) \circ \cdots \circ (B^{n_l}\ B) \circ (B^0\ B) \circ \cdots \circ (B^0\ B)$ with $n_1 \geq \cdots \geq n_l > 0$, $h \equiv (B^{n_1-1}\ B) \circ \cdots \circ (B^{n_l-1}\ B)$.

In the first step, a decreasing histogram representation h'_2 equivalent to $B\ h_2$ is obtained. The second step yields a decreasing histogram representation h_{12} for $h_1 \circ h'_2 = h_1 \circ (B\ h_2)$. Since h_1 and h_2 are decreasing, it is easy to find h_{12} by repetitive application of equation (5) for each unit of h'_2, a la insertion operation in insertion sort. In the final step, a histogram representation h that satisfies $(B\ h) \circ B^k = h_{12}$ with some k is obtained. From Lemma 8 and the uniqueness of decreasing histogram representation, h is equivalent to $h_1\ h_2$.

▶ **Example 9.** Let h_1 and h_2 be decreasing histogram representations given as

$h_1 = (B^7\ B) \circ (B^4\ B) \circ (B^3\ B) \circ (B^0\ B)$, and
$h_2 = (B^3\ B) \circ (B^0\ B)$.

Then a decreasing histogram representation h equivalent to $h_1\ h_2$ is obtained by three steps:

1. Raise each height of h_2 to get $h'_2 = (B^4\ B) \circ (B^1\ B)$.
2. Find a decreasing histogram representation by

$$\begin{aligned} h_1 \circ h'_2 &= (B^7\ B) \circ (B^4\ B) \circ (B^3\ B) \circ (B^0\ B) \circ (B^4\ B) \circ (B^1\ B) \\ &= (B^7\ B) \circ (B^4\ B) \circ (B^3\ B) \circ (B^5\ B) \circ (B^0\ B) \circ (B^1\ B) \\ &= (B^7\ B) \circ (B^4\ B) \circ (B^6\ B) \circ (B^3\ B) \circ (B^0\ B) \circ (B^1\ B) \\ &= (B^7\ B) \circ (B^7\ B) \circ (B^4\ B) \circ (B^3\ B) \circ (B^0\ B) \circ (B^1\ B) \\ &= (B^7\ B) \circ (B^7\ B) \circ (B^4\ B) \circ (B^3\ B) \circ (B^2\ B) \circ (B^0\ B) \end{aligned}$$

where equation (5) is applied in each.

3. By lowering each height after removing trailing $(B^0\ B)$'s,

$$h \equiv (B^6\ B) \circ (B^6\ B) \circ (B^3\ B) \circ (B^2\ B) \circ (B^1\ B)$$

is obtained.

4.2 Proving of anti-ρ-property

To illustrate the anti-ρ-property (not enjoy the ρ-property) for a given B-term, we introduce a measure for histogram representations which always increases by right application with a certain term. First we define a depth as a part of the measure.

▶ **Definition 10.** The *depth* $\llbracket h \rrbracket$ of a histogram representation h is defined by

$$\begin{aligned}\llbracket B^k\ B \rrbracket &= k \\ \llbracket (B^k\ B) \circ h' \rrbracket &= \max\{k,\ 1 + \llbracket h' \rrbracket\}\end{aligned}$$

From the definition, it is easy to see several properties of the depth.

▶ **Lemma 11.** Let h_1 and h_2 be histogram representations and l the length of h_1. Then $\llbracket h_1 \circ h_2 \rrbracket = \max\{\llbracket h_1 \rrbracket,\ l + \llbracket h_2 \rrbracket\}$ holds.

▶ **Lemma 12.** Let h_1 and h_2 be (possibly not decreasing) histogram representations. If h_1 and h_2 are equivalent, then $\llbracket h_1 \rrbracket = \llbracket h_2 \rrbracket$ holds.

Proof. Suppose that histogram representations h_1 and h_2 are equivalent. From Lemma 5, they have the same decreasing histogram representation. From Lemma 4, it suffices to show that $\llbracket h \circ (B^m\ B) \circ (B^n\ B) \rrbracket = \llbracket h \circ (B^{n+1}\ B) \circ (B^m\ B) \rrbracket$ holds for any histogram representation h and non-negative integers $n > m$. Since we obtain

$$\begin{aligned}\llbracket h \circ (B^m\ B) \circ (B^n\ B) \rrbracket &= \max\{\llbracket h \rrbracket, m+1, n+2\} = \max\{\llbracket h \rrbracket, n+2\} \\ \llbracket h \circ (B^{n+1}\ B) \circ (B^m\ B) \rrbracket &= \max\{\llbracket h \rrbracket, n+2, m+2\} = \max\{\llbracket h \rrbracket, n+2\}\end{aligned}$$

by $n > m$, the statements holds. ◀

Let X be a B-term. For a positive integer n, we write L_n, D_n, and Z_n for a length, depth and number of zero unit $(B^0\ B)$ of decreasing histogram representation for $X_{(n)}$, respectively. Let l and d be a length and depth of X, respectively.

▶ **Lemma 13.** The following equations hold:

$$L_{n+1} = L_n + l - Z_n \tag{8}$$

$$D_{n+1} = \max\{D_n - 1, L_n + d\}. \tag{9}$$

Proof. From Lemma 8, we have $X_{(n)} \circ (B\ X) = (B\ X_{(n+1)}) \circ B^k$ for some k and the proof of the lemma implies $k = Z_n$. By Lemma 4, we have $L_n + l = L_{n+1} + Z_n$, hence equation (8) holds. By Lemma 12 and equation Lemma 4, we have

$$\begin{aligned}\max\{D_n, L_n + d + 1\} &= \max\{D_{n+1} + 1, L_{n+1} + Z_n - 1\} \\ &= \max\{D_{n+1} + 1, L_n + l - 1\}.\end{aligned}$$

Note that $d \geq l - 1$ holds from the definition of depth. Then we have $\max\{D_n, L_n + d + 1\} = D_{n+1} + 1$, hence equation (9) holds. ◀

From this lemma, the following theorem can be immediately proved.

▶ **Theorem 14.** Let h be a histogram representation of length at least 2. Then h does not have the ρ-property. In other words, B term does not have the ρ-property if it is not equivalent to $B^n\ B$ with any $n \geq 0$.

Proof. Let us define a measure M_n by $M_n = D_n + L_n$. If $l > 1$, we have

$$\begin{aligned} D_{n+1} + L_{n+1} &= \max\{D_n - 1, L_n + d\} + L_n + l - Z_n \\ &= \max\{D_n + L_n + l - 1, 2L_n + d + l - Z_n\} \\ &> D_n + L_n \end{aligned}$$

by Lemma 13. Therefore $h_{(n)}$ cannot agree by varying n and the statement holds as long as the length of h is at least 2. ◀

5 Concluding remark

We have investigated the ρ-property of B-terms in particular forms so far. While the B-terms equivalent to $B^n\ B$ with $n \leq 6$ have the ρ-property, B-terms not equivalent to $B^n\ B$ for any n do not. In this section, remaining problems related to these results are introduced and possible approaches to illustrate them are discussed.

5.1 Remaining problems

The ρ-property is defined any combinatory terms (and closed λ-terms). We investigates it only for B-terms as a simple but interesting instance in the present paper. Note that it is undecidable whether a given combinatory term X has the ρ-property. This is because our definition of the ρ-property implicitly assumes the normalizability of all terms in the sequence $\{X_{(n)}\}_{n=1}^{\infty}$.

Conversely, for any combinatory term X, we can discuss decidability of the ρ-property of X-terms as long as $\mathbf{CL}(X)$ is known normalizable. Since $\mathbf{CL}(B)$ is normalizable, we have investigated the ρ-property of B-terms in the present paper. The decidability of the ρ-property for $\mathbf{CL}(B)$ is still open, though. The author gives the following conjecture and still believes that decreasing histogram representation introduced in the paper will be helpful to solve the problem:

▶ **conjecture 15.** B-term e has the ρ-property if and only if e is equivalent to $B^n\ B$ with some $n \geq 0$. That is, ρ-property of B-terms is decidable.

The if-part is shown for $n \leq 6$ by computation and the only-if-part is shown by Theorem 14. Since BCK-terms are also normalizable because of linearity, one could consider the decidability of the ρ-property for $\mathbf{CL}(BCK)$ and $\mathbf{CL}(BCI)$ which is of course still open. The author surmises that the ρ-property of BCK- and BCI-terms is decidable. The decidability for S-terms and L-terms can be considered One could also consider S-terms because normalizability of S-terms is decidable [11]. None of S-term may have the ρ-property, though.

By modifying the definition of equivalence of terms, we can consider different problems on the ρ-property. As mentioned in Section 2, Smullyan's lark combinator $L = \lambda x.\lambda y.x\ (y\ y)$ does not have the ρ-property because the λ-term corresponding to $L_{(4)}$ is not normalizable. However, equivalence can be defined without λ-terms. Statman [10] showed that equivalence of L-terms is decidable where the equivalence is given by a congruence relation induced by $L\ e_1\ e_2 \to e_1\ (e_2\ e_2)$. It would be interesting to investigate the ρ-property of L-terms in this setting.

5.2 Possible approaches

The present paper introduces a histogram representation to make equivalence check of B-terms easier. The idea of the representation is based on that we can lift all $\circ$'s (2-argument B) to outside of B (1-argument B) by equation (4). One may consider it the other way around. Using the equation, we can lift all B's (1-argument B) to outside of $\circ$ (2-argument B). Then one of arguments of $\circ$ becomes B. By equation (5), we can move all B's to right. Thereby we find another canonical representation for B-terms given by

$$e ::= B \mid B\ e \mid e \circ B$$

whose uniqueness should be formally proved, though.

Waldmann [12] suggests that the ρ-property of $B^n\ B$ may be checked even without converting B-terms into canonical forms. He simply defines B-terms by

$$e ::= B^k \mid e\ e$$

and regards B^k as a constant which has a rewrite rule $B^k\ e_1\ e_2\ \ldots\ e_{k+2} \to e_1\ (e_2\ \ldots\ e_{k+2})$. He implemented the check program in Haskell to confirm the ρ-property. Even in the restriction on rewriting rules, he found that $(B^1\ B)_{(36)} = (B^1\ B)_{(56)}$, $(B^2\ B)_{(274)} = (B^2\ B)_{(310)}$ and $(B^3\ B)_{(4267)} = (B^3\ B)_{(10063)}$, in which it requires a bit more right applications to find the ρ-property than the case of a canonical representation. If the ρ-property of $B^n\ B$ for any $n \geq 0$ is shown under the restricted equivalence given by rewriting rules, then we can conclude the if-part of Conjecture 15.

It might be possible to directly use a syntax of λ-terms instead of combinatory forms. It is not necessary to consider the canonical representation. This is a natural approach to investigate the ρ-property of general λ-terms. However, it is difficult to grasp how terms change by repetitive right application in general due to the β-reduction.

Another possible approach is to observe the change of (principal) types by right repetitive application. Although there are many distinct λ-terms of the same type, we can consider a desirable subset of typed λ terms. As shown by Hirokawa [1], each BCK-term can be characterized by its type, that is, any two λ-terms in $\mathbf{CL}(BCK)$ of the same principal type are identical up to β-equivalence.

References

1. Sachio Hirokawa. Principal types of BCK-lambda-terms. *Theoretical Computer Science*, 107(2):253–276, Jan 1993.
2. Donald E. Knuth. *The Art of Computer Programming, Volume 2 (3rd Ed.): Seminumerical Algorithms*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1997.
3. E.S. Ljapin. *Semigroups*. Semigroups. American Mathematical Society, 1968.
4. Keisuke Nakano. Observation on function composition. RIMS/CS regular meeting, Kyoto University, June 1998.
5. Keisuke Nakano. ρ-property of combinators. 29th TRS Meeting in Tokyo, 2008. Presentation only.
6. Keisuke Nakano. Repetitive right applications of B terms. 3rd Workshop on Proof Theory and Rewriting, March 2013. Presentation only.
7. Chris Okasaki. Flattening combinators: surviving without parentheses. *Journal of Functional Programming*, 13(4):815–822, July 2003.
8. R.M. Smullyan. *To Mock a Mocking Bird*. Knopf Doubleday Publishing Group, 2012.
9. Richard Statman. To type a mockingbird. A draft paper is available `http://tlca.di.unito.it/PAPER/TypeMock.pdf`, 2011.

10 Rick Statman. The word problem for Smullyan's lark combinator is decidable. *Journal of Symbolic Computation*, 7(2):103–112, February 1989.

11 Johannes Waldmann. The combinator S. *Information and Computation*, 159(1-2):2–21, May 2000.

12 Johannes Waldmann. Personal communication, March 2013.