Garbage Collection Based on a Linear Type System

Atsushi Igarashi and Naoki Kobayashi

University of Tokyo
igarashi@graco.c.u-tokyo.ac.jp
koba@is.s.u-tokyo.ac.jp

Abstract. We propose a type-directed garbage collection (GC) scheme for a programming language with static memory management based on a linear type system. Linear type systems, which can guarantee certain values (called *linear values*) to be used only once during program execution, are useful for memory management: memory space for linear values can be reclaimed immediately after they are used. However, conventional pointer-tracing GC does not work under such a memory management scheme: as linear values are used, dangling pointers to the memory space for them will be yielded.
This problem is solved by exploiting static type information during garbage collection in a way similar to tag-free GC. Type information in our linear type system represents not only the shapes of heap objects but also how many times the heap objects are accessed in the rest of computation. Using such type information at GC-time, our GC can avoid tracing dangling pointers; in addition, our GC can reclaim even reachable garbage. We formalize such a GC algorithm and prove its correctness.

1 Introduction

1.1 Memory Management and Linear Type Systems

Automatic memory management, one of the important features of modern high-level programming languages, releases programmers from burdens of correctly inserting explicit declarations of memory deallocation in programs. It is typically realized by garbage collection (GC), which is periodically invoked and finds unused memory space by traversing pointers. Although GC is indeed very useful run-time machinery, reclamation is delayed until the invocation of the garbage collector. Moreover, traditional tracing garbage collectors cannot reclaim memory space that is semantically garbage but reachable from the stack or registers.

To reuse memory space more eagerly, several techniques have been proposed based on region inference [3, 6, 25] or linear type systems[1] [7, 12, 14, 15, 17, 20,

[1] Strictly speaking, some of these type systems, including the one presented here, are classified as variants of *affine* linear type systems. Throughout this paper, we lump together type systems that can take into account how often values are used and refer to them as linear type systems.

27–29]. We follow the latter approach here. Linear (more precisely, affine linear) type systems can guarantee that certain values (called *linear values*) are used at most once, so we can reclaim memory space for such linear values immediately after they are used. The basic idea of linear type systems is to annotate type constructors with information about how often the memory space for values are accessed. For example, consider the expressions $x+1$ and $x+x$. In a conventional type system, both expressions are given type *Int* under the assumption that x is given type *Int*, written as follows:

$$x\colon \mathit{Int} \vdash x+1 : \mathit{Int}$$
$$x\colon \mathit{Int} \vdash x+x : \mathit{Int}$$

In a linear type system, each type *Int* is annotated with a *use*. It is either 0, 1, or ω and denotes how often a value is used for primitive operations such as $+$ during the execution. For example, an expression of the type Int^1 can be used at most once as an integer, that of the type Int^ω can be used an arbitrary number of times, and that of the type Int^0 cannot be used. Then, the type judgments below are both valid:

$$x\colon \mathit{Int}^1 \vdash x+1 : \mathit{Int}^\omega$$
$$x\colon \mathit{Int}^\omega \vdash x+x : \mathit{Int}^\omega$$

while the type judgment

$$x\colon \mathit{Int}^1 \vdash x+x : \mathit{Int}^\omega$$

is not because x is used twice in $x + x$. By using such a type system, linear values can be statically found and the compiler can statically insert deallocation code at certain primitive operations. Such deallocation, however, is performed independently of reachability of the memory space from the run-time stack or registers. Thus, it may create *dangling pointers* in the memory space, making conventional tracing GC fail.

1.2 Our Approach

In this paper we propose a GC scheme that can coexist with static memory management based on a linear type system. The basic idea is to exploit static type information during GC in a way similar to tag-free GC [4, 21–23, 26]. It may help to review the idea of tag-free GC (for languages with a monomorphic type system) first. The intuition behind tag-free GC was that "types represent the shapes of values." When we begin to trace a variable in the environment, its static type tells how to trace it: if x is given type $(\mathit{Int} \times \mathit{Int}) \times \mathit{Int}$, then we know the memory space for x stores (a pointer to) an integer pair and an integer. Thus, we can perform GC without run-time tags by recording the static type information on the free variables of function closures and of the program point where GC may happen.

In our GC, the above intuition is extended as follows: "types represent not only the shapes of values but also how often they are used in a certain context." For example, if x is given type Real^0, then we know not only that x points to a real number but also it is no longer used. Since variables corresponding to dangling pointers are always given the use 0 (if the type system gives correct types), the garbage collector can avoid tracing dangling pointers by ignoring variables with the use 0. In fact, it does not matter whether it is really a dangling pointer or not: even if there *is* a value at the address, the heap value need not be marked. As a result, our GC can collect semantic garbage reachable from the root set, as some GC schemes based on GC-time type inference can [9,11,13,22].

Unlike ordinary tracing GC, however, in our GC one object can be marked more than once; as a result, we need to extend the usual mechanism of mark bits that prevents verbose marking. For example, consider the following Standard ML program:

```
let val p = (1.5, 2.2)
    fun f x = x + (#1 p)
    fun g x = x + (#2 p)
in f 3.1 + g 4.3 end
```

Our linear type system can give the occurrence of `p` in the body of `f` type $\mathit{Real}^1 \times^1 \mathit{Real}^0$, which means the second element of `p` is not accessed in `f`, and that in the body of `g` type $\mathit{Real}^0 \times^1 \mathit{Real}^1$, which means the first element of `p` is not accessed in `g`. Now, suppose the garbage collector is invoked just before the execution of `f 3.1 + g 4.3`. Traverse from the closure of `f` marks only the first element of `p`; the second element is marked when traverse from the closure of `g` happens, marking the pair for the second time. (See Figure 1.) This kind of multiple traverses on one object may cause a lot of verbose markings, or even nontermination in the presence of cycles in the heap space. To prevent it, our garbage collector keeps track of the type, instead of the mark bit, of each marked object to remember *which part* of the object has already been marked. If the garbage collector reaches a marked object again, it compares the type of heap object derived from the current scan set with the one derived from the marked objects. The garbage collector does not go further if the access pattern expressed by the latter type subsumes that by the former: the condition means that all the heap objects that the current traverse tries to mark have been already marked (or scheduled to be marked) in the previous traverses. In the above example, GC compares `p`'s type in `g` ($\mathit{Real}^0 \times^1 \mathit{Real}^1$) with $\mathit{Real}^1 \times^1 \mathit{Real}^0$, used in the first marking; it is found that the previous traverse did not mark the second element but the current traverse requires it to be marked, making the garbage collector go further.

This extension is required when the underlying linear type system, like ours, can express non-uniform patterns of access to a single data structure in several contexts [17, 20]. On the other hand, the usual mark bits would work for earlier linear type systems [14, 27] that cannot express such non-uniformity. However, static memory management would be less effective since such type systems cannot ensure that, for example, the elements in `p` above are linear. (In our type system, `p` at its declaration can be given type $\mathit{Real}^1 \times^\omega \mathit{Real}^1$, which is obtained from the two types above by adding each use; thus, the memory space for both

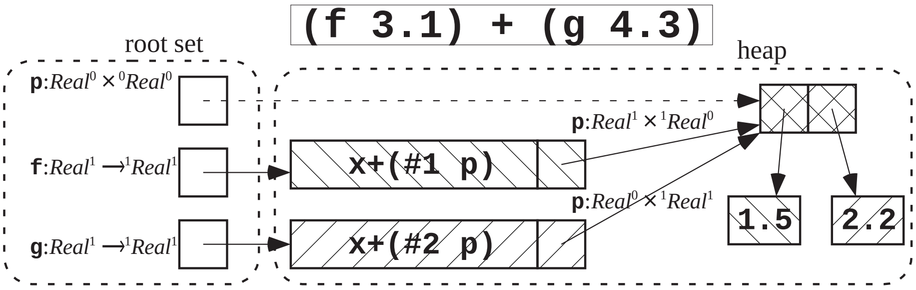

Fig. 1. The real numbers `1.5` and `2.2` are marked by the traverses from `f` and `g`, respectively. The pair is marked twice by both traverses.

elements can be reclaimed immediately after the execution of +.) See Section 5 for more discussion.

1.3 Our Contribution

The contributions of the present work are formalization of our GC algorithm for a language with a monomorphic linear type system and a proof of correctness of the algorithm. To prove the correctness, we also formalize operational semantics that takes account of immediate reclamation of memory space for linear values and prove type soundness of our linear type system with respect to the operational semantics; as in the previous work [21–23] on formalization of memory management, the operational semantics of our language makes run-time mechanisms such as stacks or heap space explicit. Here, only one particular instance of monomorphic linear type systems is dealt with, but our technique would be applicable to a language based on another variant of linear type system, even to an extension with the ML-style polymorphism (and use polymorphism [14, 29]). We could use techniques similar to the existing tag-free GC schemes [23, 26], which exploit run-time type passing.

1.4 Structure of the Paper

The rest of the paper is organized as follows: Section 2 introduces our target language and its operational semantics. Then, the type system is presented in Section 3; Section 4 presents the GC algorithm formally and proves its correctness. After discussing related work in Section 5, we conclude this paper in Section 6 with discussion on future work in Section 7.

2 Language $\lambda^{\kappa}_{\mathrm{gc}}$

In this section, we define a language called $\lambda^{\kappa}_{\mathrm{gc}}$ and give its operational semantics. The language $\lambda^{\kappa}_{\mathrm{gc}}$ is based on a call-by-value lambda-calculus equipped with integers, pairs, and recursive functions. We use a variant of A-normal form [8] so that the evaluation order is made explicit and all temporal results are bound to variables. In addition, each heap-allocated value is associated with a *use*, which denotes how often a value is used. Expressions and functions are annotated with type information on their free variables; they are required by the garbage collector, as mentioned in the previous section. Our operational semantics is given in a way similar to preceding work on abstract models of GC [21–23]: the heap, stack, and register file are made explicit in the reduction relation. Besides, deallocation of memory space for linear values is also explicit.

2.1 Syntax of Types

Before giving the syntax of expressions, we begin with the syntax of uses, types, and type environments.

Definition 1 (uses, types). *The set of* uses*, ranged over by the metavariable* κ*, and the set of* types*, ranged over by the metavariable* τ*, are given by the following syntax.*

$$\begin{array}{lcll} \kappa & ::= & 0 \mid 1 \mid \omega & (\textit{uses}) \\ \tau & ::= & \textit{Int} \mid \tau_1 \rightarrow^{\kappa} \tau_2 \mid \tau_1 \times^{\kappa} \tau_2 & (\textit{types}) \end{array}$$

The type *Int* denotes integers, the type $\tau_1 \rightarrow^{\kappa} \tau_2$ functions from τ_1 to τ_2, and $\tau_1 \times^{\kappa} \tau_2$ pairs of values of the types τ_1 and τ_2. Uses in a type denote how a value of the type can be used: a use 0 means that a value is never used, 1 means that a value can be used at most once, and ω means that a value can be used an arbitrary number of times. For example, $\textit{Int} \times^1 \textit{Int}$ denotes a type of pairs, from which we can extract integers at most once. Uses are not attached to *Int* because integers are not “boxed” in the operational semantics defined below and we are not concerned with uses of unboxed values.

Type environments defined below are not only used in the type system as usual but also attached to expressions to represent type information on the program point at which GC is invoked: from such type environments our garbage collector computes the types of the memory locations in the root set. They are also stored in function closures, making it possible to compute the types of the free variables in a closure.

Definition 2 (type environments). *The metavariables* x*,* y*,* z*, and* w *range over a countably infinite set of* variables*. A* type environment Γ *is a mapping from a finite set of variables to the set of types.*

We write $\textit{dom}(\Gamma)$ for the domain of Γ and $x_1{:}\,\tau_1, \ldots, x_n{:}\,\tau_n$, abbreviated to $\tilde{x}{:}\,\tilde{\tau}$, for the type environment Γ such that $\textit{dom}(\Gamma) = \{x_1, \ldots, x_n\}$ and $\Gamma(x_i) = \tau_i$ for each $i \in \{1, \ldots, n\}$. The empty type environment is written $\emptyset$. When $x \notin \textit{dom}(\Gamma)$, we write $\Gamma, x{:}\,\tau$ for the type environment Γ' such that $\textit{dom}(\Gamma') = \textit{dom}(\Gamma) \cup \{x\}$, $\Gamma'(x) = \tau$ and $\Gamma'(y) = \Gamma(y)$ if $x \neq y$. The type environment $\Gamma \setminus \{x_1, \ldots, x_n\}$ denotes the restriction of Γ to the domain $\textit{dom}(\Gamma) \setminus \{x_1, \ldots, x_n\}$.

2.2 Syntax of Expressions

The metavariable n ranges over the set of integers; the metavariable v ranges over the set of non-heap values; the metavariable e ranges over the set of expressions. The syntax of expressions is given by the following syntax:

$$\begin{array}{rcl} v & ::= & x \mid n \\ d & ::= & v \mid (v_1, v_2)^\kappa \mid {}^{\langle\Gamma\rangle}(\mathbf{fun}\, x(y) = e)^\kappa \\ e & ::= & {}^{\langle\Gamma\rangle}\,\mathbf{let}\ x = d\ \mathbf{in}\ e \mid {}^{\langle\Gamma\rangle}\,\mathbf{let}\ x = y\, v\ \mathbf{in}\ e \mid {}^{\langle\Gamma\rangle}\,\mathbf{let}\ x = v_1 + v_2\ \mathbf{in}\ e \\ & \mid & {}^{\langle\Gamma\rangle}\,\mathbf{let}\ (x, y) = z\ \mathbf{in}\ e \mid {}^{\langle\Gamma\rangle}\,\mathbf{if0}\ v\ \mathbf{then}\ e_1\ \mathbf{else}\ e_2 \mid {}^{\langle\Gamma\rangle}v \end{array}$$

The form $\mathbf{fun}\, x(y) = e$ is a recursive function taking y as an argument; x refers to the function itself. An expression e is a declaration of a value (which is a non-heap value, a pair, or a function), function application, addition of two integers, extraction from a pair[2], conditional branch, or return from a function call. Type environments are attached to functions and each step of an expression, recording types of variables in them. We call them *type environment annotations* and often use $TE(e)$ to denote that of e. When they are not important, we often omit to write them explicitly.

The bound variables of expressions are defined in a customary fashion: (1) the variable x is bound in e of $\mathbf{let}\ x = \cdots\ \mathbf{in}\ e$; and (2) the variables x and y are bound in e of $\mathbf{let}\ (x, y) = z\ \mathbf{in}\ e$ and $\mathbf{fun}\, x(y) = e$. A variable that is not bound will be called a free variable. We define substitutions of variables and α-conversion in a customary fashion and assume that the bound variables in an expression are pairwise distinct by α-conversion.

Remark 1. Note that programmers need not write explicit type and use annotations: $\lambda^\kappa_{\mathrm{gc}}$ is considered an intermediate form after type reconstruction. For type reconstruction, we can use techniques developed elsewhere [14, 15, 17, 20]; some of them can handle pair types that can express non-uniform access patterns.

2.3 Operational Semantics

In our semantics, we make run-time mechanisms of program execution, such as stack frames, explicit. To represent such run-time citizens, we introduce *environments* representing register files, *heaps* representing memory space, and *stacks* representing run-time stacks. A state of program execution is represented as a quadruple consisting of a heap, a stack, an environment and an expression; execution is represented as rewriting of states.

Definition 3 (environments). *An* environment *is a mapping from a finite set of variables to the set of non-heap values.*

[2] We use this form of simultaneous extraction rather than ordinary projection operations: if we could not extract two components simultaneously, a pair would be considered non-linear as its two elements are used.

We use the metavariable V for environments. We write $dom(V)$ for the domain of V and $\{x_1 = v_1, \ldots, x_n = v_n\}$ for the environment V such that $dom(V) = \{x_1, \ldots, x_n\}$ and $V(x_i) = v_i$ for each $i \in \{1, \ldots, n\}$. When $dom(V_1) \cap dom(V_2) = \emptyset$, we write $V_1 \uplus V_2$ for the environment V such that $dom(V) = dom(V_1) \cup dom(V_2)$ and $V(x) = V_i(x)$ for $x \in dom(V_i)$.

Definition 4 (heap values, heaps). *The set of* heap values*, ranged over by the metavariable h, is given by the following syntax:*

$$h \quad ::= \quad (v_1, v_2) \mid (V, {}^{\langle\Gamma\rangle}\mathbf{fun}\, x(y) = e)$$

A heap *is a mapping from a finite set of variables to the set of pairs, written h^κ, of a heap value h and a non-zero use κ (i.e., either 1 or ω).*

We use the metavariable H for a heap. The notations $dom(H)$, $\{x_1 = h_1^{\kappa_1}, \ldots, x_n = h_n^{\kappa_n}\}$, and $H_1 \uplus H_2$ are defined similarly to the corresponding notations for environments.

Definition 5 (stacks). *A* stack *is a sequence whose elements have the form $[V, \Gamma, \lambda x.e]$, called a* stack frame*.*

We use the metavariable S for a stack; we write $[\,]$ for the empty stack and $S[V, \Gamma, \lambda x.e]$ for a stack whose first element is $[V, \Gamma, \lambda x.e]$ and the remainder is S.

Definition 6 (programs, answers). *A* program *P is defined as a quadruple (H, S, V, e) of a heap, a stack, an environment and an expression. In particular, a program $(H, [\,], V, {}^{\langle\Gamma\rangle}v)$ consisting of the empty stack and a return expression is called an* answer *program.*

Then, we define rewriting rules for $\lambda^\kappa_{\mathrm{gc}}$ programs. We use several auxiliary notations defined below. The notation $\widehat{V}(v)$ looks up the environment V only if the argument is a variable.

$$\widehat{V}(x) = V(x)$$
$$\widehat{V}(n) = n$$

The notation $H \oplus \{x = h^\kappa\}$ extends H by allocating h^κ at x (if x is a fresh address); if the use κ is 0, the heap value will not be allocated. On the other hand, H^{-x} deallocates the heap value at x if its use is 1.

$$\begin{aligned}
H \oplus \{x = h^\kappa\} \quad &= \begin{cases} H & (\text{if } \kappa = 0 \text{ and } x \notin dom(H)) \\ H \uplus \{x = h^\kappa\} & (\text{if } \kappa \neq 0 \text{ and } x \notin dom(H)) \\ \text{undefined} & (\text{otherwise}) \end{cases} \\
(H \uplus \{x = h^1\})^{-x} &= H \\
(H \uplus \{x = h^\omega\})^{-x} &= H \uplus \{x = h^\omega\}
\end{aligned}$$

Definition 7. *The relation $P \longmapsto P'$ is the least relation closed under the rules in Figure 2.*

$$(H, S, V, {}^{\langle\Gamma\rangle}\,\textbf{let}\ x = v\ \textbf{in}\ e) \longmapsto (H, S, V \uplus \{x = \widehat{V}(v)\}, e) \qquad \text{(R-VAL)}$$

$$\frac{z \text{ fresh}}{\begin{array}{c}(H, S, V, {}^{\langle\Gamma\rangle}\,\textbf{let}\ x = (v_1, v_2)^\kappa\ \textbf{in}\ e') \\ \longmapsto (H \oplus \{z = (\widehat{V}(v_1), \widehat{V}(v_2))^\kappa\}, S, V \uplus \{x = z\}, e')\end{array}} \qquad \text{(R-PAIR)}$$

$$\frac{z \text{ fresh}}{\begin{array}{c}(H, S, V, {}^{\langle\Gamma\rangle}\,\textbf{let}\ x = ({}^{\langle\Gamma_0\rangle}\textbf{fun}\, y(w) = e_0)^\kappa\ \textbf{in}\ e') \\ \longmapsto (H \oplus \{z = (V, {}^{\langle\Gamma_0\rangle}\textbf{fun}\, y(w) = e_0)^\kappa\}, S, V \uplus \{x = z\}, e')\end{array}} \qquad \text{(R-FUN)}$$

$$\frac{H(V(y)) = (V_0, {}^{\langle\Gamma_0\rangle}\textbf{fun}\, z(w) = e_0)^\kappa}{\begin{array}{c}(H, S, V, {}^{\langle\Gamma\rangle}\,\textbf{let}\ x = y\, v\ \textbf{in}\ e) \\ \longmapsto (H^{-V(y)}, S[V, \mathit{TE}(e) \setminus \{x\}, \lambda x.e], V_0 \uplus \{z = V(y), w = \widehat{V}(v)\}, e_0)\end{array}} \qquad \text{(R-APP)}$$

$$\frac{n_1 = \widehat{V}(v_1) \qquad n_2 = \widehat{V}(v_2)}{(H, S, V, {}^{\langle\Gamma\rangle}\,\textbf{let}\ x = v_1 + v_2\ \textbf{in}\ e) \longmapsto (H, S, V \uplus \{x = \underline{n_1 + n_2}\}, e)} \qquad \text{(R-PLUS)}$$

$$\frac{H(V(z)) = (v_1, v_2)^\kappa}{(H, S, V, {}^{\langle\Gamma\rangle}\,\textbf{let}\ (x, y) = z\ \textbf{in}\ e) \longmapsto (H^{-V(z)}, S, V \uplus \{x = v_1, y = v_2\}, e)} \qquad \text{(R-EXT)}$$

$$\frac{\widehat{V}(v) = 0}{(H, S, V, {}^{\langle\Gamma\rangle}\,\textbf{if0}\ v\ \textbf{then}\ e_1\ \textbf{else}\ e_2) \longmapsto (H, S, V, e_1)} \qquad \text{(R-IFT)}$$

$$\frac{\widehat{V}(v) \neq 0}{(H, S, V, {}^{\langle\Gamma\rangle}\,\textbf{if0}\ v\ \textbf{then}\ e_1\ \textbf{else}\ e_2) \longmapsto (H, S, V, e_2)} \qquad \text{(R-IFF)}$$

$$(H, S[V_0, \Gamma_0, \lambda x.e_0], {}^{\langle\Gamma\rangle}v) \longmapsto (H, S, V_0 \uplus \{x = \widehat{V}(v)\}, e_0) \qquad \text{(R-RET)}$$

Fig. 2. $\lambda^\kappa_{\mathrm{gc}}$ reduction rules

The rules are fairly straightforward. In the rules R-PAIR and R-FUN, the heap value is allocated in the heap at a fresh address z; the execution continues after assigning z to x in the environment. When the program uses a heap value (R-APP and R-EXT), it is checked whether the value is linear and the memory space may be reclaimed (using H^{-x}). In the rule R-PLUS, the notation $\underline{n_1 + n_2}$ denotes the summation of the two integers n_1 and n_2 (not the syntactic expression "n_1+n_2"). The rules R-APP and R-RET are about function calls, involving manipulation of the stack. When a function is called (R-APP), the continuation $\lambda x.e$ after the call is pushed onto the stack together with its environment V and type environment $TE(e) \setminus \{x\}$; then, the execution continues with the function body e_0 under the closure's environment V_0 augmented with bindings of the actual argument $\widehat{V}(v)$ and the address $V(y)$ of the function itself. When execution reaches the end of a function (R-RET), the top of the stack is popped and the continuation is applied to the result $\widehat{V}(v)$. We write $\longmapsto^*$ for the reflexive and transitive closure of $\longmapsto$.

Remark 2. Notice that type environment annotations do not play a significant role during execution. They are just stored in closures and stack frames, and will not be used unless GC occurs. Thus, it is not necessary, in practice, to attach them at every step; they are required only at (1) every function definition, (2) the program point after every function call, and (3) program points at which GC may happen.

Remark 3. As we see in the typing rules, a non-linear value can be passed where a linear value is expected. Thus, to reclaim the memory space for linear values, we have to perform a dynamic check, requiring uses in heaps. The cost of the check can be, however, fairly cheap (without requiring extra memory space for use tags), as discussed elsewhere [17].

Example 1. Let

$$\begin{aligned}
e_1 &= \textbf{let } p_2 = (3,4)^1 \textbf{ in } e_2\\
e_2 &= \textbf{let } p = (2,p_2)^\omega \textbf{ in } e_3\\
e_3 &= \textbf{let } f = (\textbf{fun}\, f'(y) = e_f)^1 \textbf{ in let } g = (\textbf{fun}\, g'(z) = e_g)^1 \textbf{ in } e_4\\
e_4 &= \textbf{let } t_1 = f\ 1 \textbf{ in } e_5\\
e_5 &= \textbf{let } t_2 = g\ 2 \textbf{ in let } r = t_1 + t_2 \textbf{ in } r\\
e_f &= \textbf{let } (q_1,q_2) = p \textbf{ in let } (q_{21},q_{22}) = q_2 \textbf{ in let } r_f = y + q_{21} \textbf{ in } r_f\\
e_g &= \textbf{let } (r_1,r_2) = p \textbf{ in let } r_g = z + r_1 \textbf{ in } r_g
\end{aligned}$$

The program $(\{\}, [\,], \{\}, e_1)$ rewrites as follows:

$$\begin{aligned}
& (\{\}, [\,], \{\}, e_1)\\
\longmapsto\ & (\{x_2 = (3,4)^1\}, [\,], \{p_2 = x_2\}, e_2)\\
\longmapsto\ & (\{x_1 = (2,x_2)^\omega, x_2 = (3,4)^1\}, [\,], \{p = x_1, p_2 = x_2\}, e_3)
\end{aligned}$$

$$\longmapsto^* \left(\left\{ \begin{array}{l} x_1 = (2, x_2)^\omega, x_2 = (3,4)^1, \\ x_3 = (\{p = x_1, \ldots\}, \mathbf{fun}\, f'(y) = e_f)^1, \\ x_4 = (\{p = x_1, \ldots\}, \mathbf{fun}\, g'(z) = e_g)^1 \end{array} \right\}, [\,], \left\{ \begin{array}{l} p = x_1, p_2 = x_2, \\ f = x_3, g = x_4 \end{array} \right\}, e_4 \right)$$

$$\begin{array}{ll} \longmapsto & (\{x_1 = (2, x_2)^\omega, x_2 = (3,4)^1, x_4 = (\{p = x_1, \ldots\}, \mathbf{fun}\, g'(z) = e_g)^1\}, \\ & \quad [\{p = x_1, p_2 = x_2, \ldots\}, \Gamma, \lambda t_1.e_5], \{y = 1, f' = x_3, p = x_1, p_2 = x_2\}, e_f) \\ \longmapsto^* & (\{x_1 = (2, x_2)^\omega, x_4 = (\{p = x_1, \ldots\}, \mathbf{fun}\, g'(z) = e_g)^1\}, \\ & \quad [\{p = x_1, p_2 = x_2, \ldots\}, \Gamma, \lambda t_1.e_5], \{r_f = 4, q_{21} = 3, \ldots, \}, r_f) \\ \longmapsto^* & (\{x_1 = (2, x_2)^\omega, x_4 = (\{p = x_1, \ldots\}, \mathbf{fun}\, g'(z) = e_g)^1\}, \\ & \quad [\,], \{t_1 = 4, p = x_1, f = x_3, \ldots\}, e_5) \\ \longmapsto & \cdots \longmapsto (\{x_1 = (2, x_2)^\omega\}, [], \{r = 8, \ldots\}, r) \end{array}$$

3 Type System

In this section, we give a type system for $\lambda^\kappa_{\mathrm{gc}}$. Our type system can ensure the lack of not only illegal operations (such as application of a non-function value) but also illegal access to already deallocated heap space. We begin with several operations on uses, types, and type environments, used in the typing rules.

3.1 Notational Preliminaries

The relation $\tau_1 \geq \tau_2$ below means that an expression of type τ_1 can be more frequently used than that of τ_2; we allow an expression of type τ_1 to be coerced to that of type τ_2. Similarly, $\Gamma_1 \geq \Gamma_2$ means that the type bound in Γ_2 is less than that in Γ_1 for each variable in $dom(\Gamma_2)$.

Definition 8. *The binary relation $\geq$ between uses is the total order defined by $\omega \geq 1 \geq 0$. The binary relation $\tau_1 \geq \tau_2$ is defined by:*

$$\begin{array}{rl} \mathit{Int} \geq \mathit{Int} & \\ \tau_1 \rightarrow^{\kappa_1} \tau_2 \geq \tau_1 \rightarrow^{\kappa_2} \tau_2 & \text{if } \kappa_1 \geq \kappa_2 \\ \tau_{11} \times^{\kappa_1} \tau_{12} \geq \tau_{21} \times^{\kappa_2} \tau_{22} & \text{if } \kappa_1 \geq \kappa_2 \text{ and } \tau_{1i} \geq \tau_{2i} \text{ for } i = 1,2 \end{array}$$

We also write $\Gamma_1 \geq \Gamma_2$ if $dom(\Gamma_1) \supseteq dom(\Gamma_2)$ and $\Gamma_1(x) \geq \Gamma_2(x)$ for all $x \in dom(\Gamma_2)$.

The relation $\tau_1 \geq \tau_2$ would correspond to a subtyping relation $\tau_1 \preceq \tau_2$, meaning τ_1 is subtype of τ_2. Note that $\geq$ is the inverse of the usual subtyping relation.

Remark 4. We could adopt usual structural subtyping for function types, without which the use analysis gets rather coarser. It is not adopted simply to make the presentation simpler (while pair types require a structural rule, which is crucial to express non-uniform access patterns, mentioned in Section 1). In fact, introduction of the structural subtyping for function types would not affect our GC algorithm itself (even though it would complicate a proof of correctness of the GC): our GC algorithm would work as long as attached uses are correct with respect to the operational semantics.

Example 2. $(x\colon (\mathit{Int} \times^{\omega} \mathit{Int}) \times^{1} \mathit{Int},\ y\colon \mathit{Int} \rightarrow^{\omega} \mathit{Int},\ z\colon \mathit{Int}) \geq (x\colon (\mathit{Int} \times^{0} \mathit{Int}) \times^{1} \mathit{Int},\ y\colon \mathit{Int} \rightarrow^{1} \mathit{Int})$.

The summation of two types, defined below, is used to compute the total use of a variable. Suppose a variable is given type τ_1 in e_1 and τ_2 in e_2; if both expressions may be executed, then the total usage of the variable is represented by $\tau_1 + \tau_2$.

Definition 9. *The* summation *of two uses, written* $\kappa_1 + \kappa_2$*, is the commutative and associative operation that satisfies* $0+0=0$, $1+0=1$, *and* $1+1=\omega+0=\omega+1=\omega+\omega=\omega$. *The summation of two types, written* $\tau_1+\tau_2$*, is defined as follows:*

$$\begin{array}{ll} \mathit{Int} + \mathit{Int} & = \mathit{Int} \\ (\tau_1 \rightarrow^{\kappa_1} \tau_2) + (\tau_1 \rightarrow^{\kappa_2} \tau_2) & = \tau_1 \rightarrow^{\kappa_1+\kappa_2} \tau_2 \\ (\tau_{11} \times^{\kappa_1} \tau_{12}) + (\tau_{21} \times^{\kappa_2} \tau_{22}) & = (\tau_{11}+\tau_{21}) \times^{(\kappa_1+\kappa_2)} (\tau_{12}+\tau_{22}) \end{array}$$

The operation '+' on types are pointwise extended to type environments: the summation $\Gamma_1 + \Gamma_2$ *of two type environments is defined by:*

$$\begin{array}{l} \mathit{dom}(\Gamma_1+\Gamma_2) = \mathit{dom}(\Gamma_1) \cup \mathit{dom}(\Gamma_2) \\ (\Gamma_1+\Gamma_2)(x) = \begin{cases} \Gamma_1(x)+\Gamma_2(x) & \mathit{if}\ x \in \mathit{dom}(\Gamma_1) \cap \mathit{dom}(\Gamma_2) \\ \Gamma_1(x) & \mathit{if}\ x \in \mathit{dom}(\Gamma_1) \setminus \mathit{dom}(\Gamma_2) \\ \Gamma_2(x) & \mathit{if}\ x \in \mathit{dom}(\Gamma_2) \setminus \mathit{dom}(\Gamma_1) \end{cases} \end{array}$$

As mentioned in Section 1, the summation of two pair types are obtained by adding uses pointwise because a use inside a pair type constructor denotes how may times an element in the pair is used *in total*. Thus, for example, the summation of two types $(\mathit{Int} \times^{0} \mathit{Int}) \times^{1} \mathit{Int}$ and $(\mathit{Int} \times^{1} \mathit{Int}) \times^{1} \mathit{Int}$ is defined to be $(\mathit{Int} \times^{1} \mathit{Int}) \times^{\omega} \mathit{Int}$. The obtained type means that the inner pair can be used at most once in total even if the outer pair can be accessed arbitrarily many times.

Example 3. $(x\colon \mathit{Int},\ y\colon (\mathit{Int} \times^{1} \mathit{Int}) \times^{1} \mathit{Int}) + (y\colon (\mathit{Int} \times^{0} \mathit{Int}) \times^{1} \mathit{Int}) = x\colon \mathit{Int},\ y\colon (\mathit{Int} \times^{1} \mathit{Int}) \times^{\omega} \mathit{Int}$.

We also use the notation $\Gamma + v\colon \tau$, defined as follows:

$$\Gamma + v\colon \tau = \begin{cases} \Gamma + x\colon \tau & (\text{if } v = x) \\ \Gamma & (\text{if } v = n) \end{cases}$$

The product of a use κ and a type is defined below as the summation of κ copies of the type.

Definition 10. *The* product *of two uses, written* $\kappa_1 \cdot \kappa_2$*, is the commutative and associative operation that satisfies* $0 \cdot 0 = 0 \cdot 1 = 0 \cdot \omega = 0$, $1 \cdot 1 = 1$, *and* $1 \cdot \omega = \omega \cdot \omega = \omega$. *The product is extended to an operation on uses and types by:*

$$\begin{array}{ll} \kappa \cdot \mathit{Int} & = \mathit{Int} \\ \kappa \cdot (\tau_1 \rightarrow^{\kappa'} \tau_2) & = \tau_1 \rightarrow^{\kappa \cdot \kappa'} \tau_2 \\ \kappa \cdot (\tau_1 \times^{\kappa'} \tau_2) & = (\kappa \cdot \tau_1) \times^{\kappa \cdot \kappa'} (\kappa \cdot \tau_2) \end{array}$$

It is further extended to an operation on uses and type environments by:

$$\kappa \cdot (x_1 \colon \tau_1, \ldots, x_n \colon \tau_n) = x_1 \colon (\kappa \cdot \tau_1), \ldots, x_n \colon (\kappa \cdot \tau_n)$$

Example 4. $\omega \cdot (\mathit{Int} \times^1 (\mathit{Int} \times^0 \mathit{Int})) = (\mathit{Int} \times^\omega (\mathit{Int} \times^0 \mathit{Int}))$.

Example 5. $0 \cdot (x \colon \mathit{Int} \times^\omega \mathit{Int}, y \colon \mathit{Int}) = x \colon \mathit{Int} \times^0 \mathit{Int}, y \colon \mathit{Int}$.

If a heap value refers to another heap value through a variable whose use is 0, then the referred heap value actually need not exist in the heap. To ignore such potential dangling pointers, we discard bindings of types with the use 0 from a type environment, by using the *truncation*, defined below.

Definition 11. *The* truncation $\lfloor \Gamma \rfloor$ *of a type environment* Γ *is defined by:*

$$\lfloor \Gamma \rfloor = \Gamma \setminus \{x \mid \Gamma(x) = \tau_1 \times^0 \tau_2 \textit{ or } \tau_1 \rightarrow^0 \tau_2\}$$

Example 6. $\lfloor x \colon \mathit{Int}, y \colon \mathit{Int} \times^\omega \mathit{Int}, z \colon \mathit{Int} \rightarrow^0 \mathit{Int} \rfloor = x \colon \mathit{Int}, y \colon \mathit{Int} \times^\omega \mathit{Int}$.

3.2 Typing Rules

Typing rules for expressions. A type judgment for an expression is of the form $\Gamma \vdash e : \tau$. It means not only that e is well-typed in the ordinary sense, but also that each function and pair declared in e is used according to its use and each free variable is used according to the use of its type in Γ. For example, $\Gamma, x \colon \mathit{Int} \rightarrow^1 \mathit{Int} \vdash e : \tau$ means that e uses x *as a function on integers at most once*.

The formal rules are given in Figure 3. Since type environments contain information on how often variables are accessed, we need to take special care in merging type environments. For example, if $\Gamma_1, y \colon \mathit{Int} \rightarrow^1 \mathit{Int} \vdash d : \mathit{Int} \rightarrow^1 \mathit{Int}$ and $\Gamma_2, x \colon \mathit{Int} \rightarrow^1 \mathit{Int}, y \colon \mathit{Int} \rightarrow^1 \mathit{Int} \vdash e : \tau_2$, then y is used totally twice in **let** $x = d$ **in** e. Therefore, the total use of a variable in **let** $x = d$ **in** e should be obtained by adding the uses in two type environments as in the rule T-Dec. Similarly for the rules T-App, T-Plus, and T-Ext. On the other hand, in a conditional expression **if0** x **then** e_1 **else** e_2, either e_1 or e_2 is executed. Thus, the two branches should be typed under the same environment (T-If). In the rule T-App, the use of the type of the variable y should be 1 since the function stored in y is accessed there; similarly for the rule T-Ext. In the rule T-Fun, the two uses κ_1 and κ_2 represent the number of times of recursive calls and that of calls from outside the body, respectively. Thus, the type environment of the function body Γ is multiplied by $\kappa_2 \cdot (\kappa_1 + 1)$, an upper bound of the total number of times that the function is applied [15, 17]. (The type environment Γ for a single call is attached as the type environment annotation rather than $\kappa_2 \cdot (\kappa_1 + 1) \cdot \Gamma$; this is mainly for ease of our type soundness proof. Besides, for the purpose of GC, it does not matter whether we attach Γ or $\kappa_2 \cdot (\kappa_1 + 1) \cdot \Gamma$, as we will see later: the distinction between 1 and ω is not important in GC.) Annotated type environments must agree with ones from the type derivation to provide the garbage collector with correct type information.

Non-heap values, pairs, functions:

$$x{:}\,\tau \vdash x : \tau \qquad \text{(T-VAR)} \qquad\qquad \Gamma \vdash n : \mathit{Int} \qquad \text{(T-INT)}$$

$$v_1{:}\,\tau_1 + v_2{:}\,\tau_2 \vdash (v_1, v_2)^\kappa : \tau_1 \times^\kappa \tau_2 \qquad \text{(T-PAIR)}$$

$$\frac{\Gamma_1, y{:}\,\tau_1 \to^{\kappa_1} \tau_2, z{:}\,\tau_1 \vdash e_0 : \tau_2 \qquad \kappa = \kappa_2 \cdot (\kappa_1 + 1)}{\kappa \cdot \Gamma_1 \vdash (^{\langle \Gamma_1 \rangle}\mathbf{fun}\, y(z) = e_0)^\kappa : \tau_1 \to^{\kappa_2} \tau_2} \qquad \text{(T-FUN)}$$

Expressions:

$$\frac{\Gamma_1 \vdash d : \tau' \qquad \Gamma_2, x{:}\,\tau' \vdash e : \tau}{\Gamma_1 + \Gamma_2 \vdash {}^{\langle \Gamma_1 + \Gamma_2 \rangle}\,\mathbf{let}\ x = d\ \mathbf{in}\ e : \tau} \qquad \text{(T-DEC)}$$

$$\frac{\Gamma', x{:}\,\tau_2 \vdash e : \tau \qquad \Gamma = \Gamma' + (y{:}\,\tau_1 \to^1 \tau_2) + (v{:}\,\tau_2)}{\Gamma \vdash {}^{\langle \Gamma \rangle}\,\mathbf{let}\ x = y\, v\ \mathbf{in}\ e : \tau} \qquad \text{(T-APP)}$$

$$\frac{\Gamma', x{:}\,\mathit{Int} \vdash e : \tau \qquad \Gamma = \Gamma' + v_1{:}\,\mathit{Int} + v_2{:}\,\mathit{Int}}{\Gamma \vdash {}^{\langle \Gamma \rangle}\,\mathbf{let}\ x = v_1 + v_2\ \mathbf{in}\ e : \tau} \qquad \text{(T-PLUS)}$$

$$\frac{\Gamma', x{:}\,\tau_1, y{:}\,\tau_2 \vdash e : \tau \qquad \Gamma = \Gamma' + z{:}\,\tau_1 \times^1 \tau_2}{\Gamma \vdash {}^{\langle \Gamma \rangle}\,\mathbf{let}\ (x, y) = z\ \mathbf{in}\ e : \tau} \qquad \text{(T-EXT)}$$

$$\frac{\Gamma' \vdash e_1 : \tau \qquad \Gamma' \vdash e_2 : \tau \qquad \Gamma = \Gamma' + v{:}\,\mathit{Int}}{\Gamma \vdash {}^{\langle \Gamma \rangle}\,\mathbf{if0}\ v\ \mathbf{then}\ e_1\ \mathbf{else}\ e_2 : \tau} \qquad \text{(T-IF)}$$

$$\Gamma + v{:}\,\tau \vdash {}^{\langle \Gamma + v:\tau \rangle} v : \tau \qquad \text{(T-RET)}$$

Fig. 3. $\lambda^\kappa_{\mathrm{gc}}$ typing rules for expressions

Typing Rules for Programs. We also give a type system for environments, stacks, heaps, and programs. The type system for programs is important to show correctness of both static and dynamic memory management. As proved below, we can show that a well-typed program cannot go wrong under an execution without GC (Theorem 3), by the standard technique based on subject reduction and progress [31]. This means that, at least, static memory management based on uses is correct. In the next section, correctness of GC will also be proved in terms of well-typedness of a program: it is shown that, given a well-typed program, our garbage collection always succeeds and preserves well-typedness of the program after GC (with the garbage-collected heap). Then, the results of executions with and without GC agree. This means that GC, the dynamic memory management, is also correct.

We use the following type judgments for programs:

$\Gamma \vdash V : \Gamma'$	(V is a well-typed environment providing Γ' under Γ)
$\Gamma \vdash S : \tau_1 \rightarrow^1 \tau_2$	(S is a well-typed stack of type $\tau_1 \rightarrow^1 \tau_2$)
$\Gamma \vdash H : \Gamma_1; \Gamma_2$	(H is a well-typed heap described by Γ_1 and Γ_2)
$\vdash P : \tau$	(P is a well-typed program of type τ)

The typing rules are given in Figure 4. In the rule T-Env, the type environment $x_1 \colon \tau_1, \ldots, x_n \colon \tau_n$ denotes how $V = \{x_1 = v_1, \ldots, x_n = v_n, \ldots\}$ can be accessed from an expression; the type environment $v_1 \colon \tau_1 + \cdots + v_n \colon \tau_n$ denotes the heap addresses (and their types) accessed through the environment V. Since variables other than $x_1, \ldots, x_n$ are not actually used, V may contain arbitrary garbage bindings ("…" after $x_n = v_n$). To deal with possible aliasing (different variables x_i and x_j may maps to the same address z), the summation is required. The rules T-HPair and T-HFun are similar to the rules T-Pair and T-Fun, except that, in T-HFun, free variables of the function refer outside through the environment V. The empty stack is given type $\tau \rightarrow^1 \tau$ for any τ since it can be regarded as an identity function. If a top stack frame is a function of type $\tau_1 \rightarrow^1 \tau_2$ and the rest of the stack is given type $\tau_2 \rightarrow^1 \tau_3$, then the whole stack, regarded as composition of the functions, is given type $\tau_1 \rightarrow^1 \tau_3$. Since each stack frame will be applied at most once, a stack is given the use 1. The type judgment form $\Gamma \vdash H : \Gamma_1; \Gamma_2$ for heaps and the rule T-Heap are explained as follows. The type environment Γ_2 describes how each heap value can be used; thus, provided that $\Gamma_{x_i} \vdash H(x_i) : \tau_i$ for each $x_i \in \mathit{dom}(H)$, the type environment Γ_2 is $x_1 \colon \tau_1, \ldots, x_n \colon \tau_n$. On the other hand, Γ_1 describes how the heap can be referred to by the stack and environment of a program. Since each Γ_{x_i} describes how the heap value $H(x_i)$ refers to other heap values, their total use $\Gamma_1 + \Gamma_{x_1} + \cdots + \Gamma_{x_n}$ should be less than Γ_2. The truncation in the rule T-Heap represents the fact that the stack, environment, and heap values may include dangling pointers if they are given type with the use 0. The type environment Γ (on the left of $\vdash$) describes references to outside of the heap. In a well-typed program, the heap should be closed (except for dangling pointers) and Γ is empty. On the other hand, in a proof of correctness of our GC, it plays an important role to describe references from the collected heap values to the rest of heap values to be collected. Since the type environment Γ_2 is required mainly for conciseness of correctness of our GC and not important

Environments, heap values:

$$v_1\!:\tau_1 + \cdots + v_n\!:\tau_n \vdash \{x_1 = v_1, \ldots, x_n = v_n, \ldots\} : x_1\!:\tau_1, \ldots, x_n\!:\tau_n \qquad \text{(T-Env)}$$

$$v_1\!:\tau_1 + v_2\!:\tau_2 \vdash (v_1, v_2)^\kappa : \tau_1 \times^\kappa \tau_2 \qquad \text{(T-HPair)}$$

$$\frac{\Gamma, y\!:\tau_1, x\!:\tau_1 \to^{\kappa_1} \tau_2 \vdash e : \tau_2 \qquad \Gamma' \vdash V : \Gamma \qquad \kappa = \kappa_2 \cdot (\kappa_1 + 1)}{\kappa \cdot \Gamma' \vdash (V, {}^{\langle\Gamma\rangle}\mathbf{fun}\, x(y) = e)^\kappa : \tau_1 \to^{\kappa_2} \tau_2} \qquad \text{(T-HFun)}$$

Stacks, heaps, programs:

$$\Gamma \vdash [\,] : \tau \to^1 \tau \qquad \text{(T-Empty)}$$

$$\frac{\Gamma_1 \vdash S : \tau_2 \to^1 \tau_3 \qquad \Gamma_2 \vdash V : \Gamma_3 \qquad \Gamma_3, x\!:\tau_1 \vdash e : \tau_2}{\Gamma_1 + \Gamma_2 \vdash S[V, \Gamma_3, \lambda x.e] : \tau_1 \to^1 \tau_3} \qquad \text{(T-Push)}$$

$$\frac{\Gamma_1 \vdash h_1^{\kappa_1} : \tau_1 \quad \cdots \quad \Gamma_n \vdash h_n^{\kappa_n} : \tau_n \qquad \Gamma' + (\tilde{x}\!:\tilde{\tau}) \geq \lfloor \Gamma + \Gamma_1 + \cdots + \Gamma_n \rfloor}{\Gamma' \vdash \{x_1 = h_1^{\kappa_1}, \ldots, x_n = h_n^{\kappa_n}\} : \Gamma; \tilde{x}\!:\tilde{\tau}} \qquad \text{(T-Heap)}$$

$$\frac{\emptyset \vdash H : \Gamma_1 + \Gamma_2; \Gamma \qquad \Gamma_1 \vdash S : \tau' \to^1 \tau \qquad \Gamma_2 \vdash V : \Gamma_3 \qquad \Gamma_3 \vdash e : \tau'}{\vdash (H, S, V, e) : \tau} \qquad \text{(T-Prog)}$$

Fig. 4. $\lambda^\kappa_{\mathrm{gc}}$ typing rules for environments, stacks, heaps, and programs

here, we often abbreviate $\Gamma \vdash H : \Gamma_1; \Gamma_2$ to $\Gamma \vdash H : \Gamma_1$. The rule T-Prog is straightforward: if a heap H is closed and well typed, a stack S is given type $\tau' \to^1 \tau$, an environment is well typed, and e is given type τ', then the program (H, S, V, e) is given type τ.

Example 7. Let the heap

$$H = \left\{ \begin{array}{l} x_1 = (2, x_2)^\omega, x_2 = (3, 4)^1, \\ x_3 = (\{p = x_1, \ldots\}, {}^{\langle p:Int\times^1(Int\times^1 Int)\rangle}\mathbf{fun}\, f'(y) = e_f)^1, \\ x_4 = (\{p = x_1, \ldots\}, {}^{\langle p:Int\times^1(Int\times^0 Int)\rangle}\mathbf{fun}\, g'(y) = e_g)^1 \end{array} \right\}$$

(where e_f and e_g are from Example 1) and the environment

$$V = \{p = x_1, p_2 = x_2, f = x_3, g = x_4\}$$

and the expression

$$\begin{aligned} e_4 = {}^{\langle f:Int\to^1 Int, g:Int\to^1 Int\rangle}\, &\mathbf{let}\ t_1 = f\ 1\ \mathbf{in} \\ {}^{\langle g:Int\to^1 Int, t_1:Int\rangle}\, &\mathbf{let}\ t_2 = g\ 2\ \mathbf{in} \\ {}^{\langle t_1:Int, t_2:Int\rangle}\, &\mathbf{let}\ r = t_1 + t_2\ \mathbf{in}\ r. \end{aligned}$$

Then, by the typing rules, the following judgments are derived.

$$\emptyset \vdash H : (x_1 \colon (\mathit{Int} \times^0 \mathit{Int}) \times^\omega \mathit{Int}, x_2 \colon \mathit{Int} \times^0 \mathit{Int}, x_3 \colon \mathit{Int} \to^1 \mathit{Int}, x_4 \colon \mathit{Int} \to^1 \mathit{Int})$$
$$x_3 \colon \mathit{Int} \to^1 \mathit{Int}, x_4 \colon \mathit{Int} \to^1 \mathit{Int} \vdash V : (f \colon \mathit{Int} \to^1 \mathit{Int}, g \colon \mathit{Int} \to^1 \mathit{Int})$$
$$f \colon \mathit{Int} \to^1 \mathit{Int}, g \colon \mathit{Int} \to^1 \mathit{Int} \vdash e : \mathit{Int}.$$

Thus, by the rule T-Prog, $\vdash (H, [\,], V, e) : \mathit{Int}$.

3.3 Soundness of Type System

The type system introduced in this section guarantees that a well-typed program can cause neither run-time type errors nor illegal memory access by dereferencing dangling pointers.

Theorem 1 (Subject Reduction). *If $\emptyset \vdash P : \tau$ and $P \longmapsto P'$, then $\emptyset \vdash P' : \tau$.*

Theorem 2 (Progress). *If $\vdash P : \tau$ and P is not an answer, then there exists P' such that $P \longmapsto P'$.*

Theorem 3 (Type Soundness). *If $\vdash P : \tau$ and $P \longmapsto^* P'$ with P' being a normal form, then P' is an answer and $\vdash P' : \tau$.*

We prove these theorems below; we begin with developing several lemmas required.

Lemma 1 (Heap Allocation). *If $\Gamma' \vdash H : \Gamma_1 + \Gamma_2$ and $\Gamma_2 \vdash h^\kappa : \tau$ and $x \notin \mathit{dom}(H)$, then $\Gamma' \vdash H \oplus \{x = h^\kappa\} : \Gamma_1 + x \colon \tau$.*

Proof. By a case analysis on whether $\kappa = 0$ or not. The case where $\kappa = 0$ is immediate since the outermost use of τ should also be 0. The other case is also easy. □

Lemma 2 (Deallocation).

1. *If $\emptyset \vdash H : (\Gamma + x \colon \tau_1 \to^1 \tau_2)$ and $H(x) = (V, {}^{\langle \Gamma' \rangle}\mathbf{fun}\, y(z) = e)^\kappa$ with $\kappa \geq 1$, then $\emptyset \vdash H^{-x} : \Gamma + (\Gamma'', x \colon \tau_1 \to^{\kappa'} \tau_2)$ and $\Gamma'' \vdash V : \Gamma'$ and $\Gamma', y \colon \tau_1 \to^{\kappa'} \tau_2, z \colon \tau_1 \vdash e : \tau_2$ for some κ' and Γ''.*
2. *If $\Gamma' \vdash H : (\Gamma + x \colon \tau_1 \times^1 \tau_2)$ and $H(x) = (v_1, v_2)^\kappa$ with $\kappa \geq 1$, then $\Gamma' \vdash H^{-x} : \Gamma + v_1 \colon \tau_1 + v_2 \colon \tau_2$.*

Proof.

1. By the rules T-Heap and T-HFun,

$$H = \{x = (V, {}^{\langle \Gamma' \rangle}\mathbf{fun}\, y(z) = e)^{\kappa_0}, x_1 = h_1^{\kappa_1}, \ldots, x_n = h_n^{\kappa_n}\}$$
$$\Gamma_1 \vdash h_1^{\kappa_1} : \tau_{x_1} \qquad \ldots \qquad \Gamma_n \vdash h_n^{\kappa_n} : \tau_{x_n}$$
$$x \colon \tau_1 \to^{\kappa_0} \tau_2, x_1 \colon \tau_{x_1}, \ldots, x_n \colon \tau_{x_n} \geq \lfloor (\Gamma + x \colon \tau_1 \to^1 \tau_2) + \Gamma_1 + \cdots + \Gamma_n + \kappa \cdot \Gamma'_0 \rfloor$$
$$\kappa \cdot \Gamma'_0 \vdash (V, {}^{\langle \Gamma' \rangle}\mathbf{fun}\, y(z) = e)^\kappa : \tau_1 \to^{\kappa_0} \tau_2 \qquad \kappa = \kappa_0 \cdot (\kappa_1 + 1)$$
$$\Gamma'_0 \vdash V : \Gamma' \qquad \Gamma', y \colon \tau_1 \to^{\kappa_1} \tau_2, z \colon \tau_1 \vdash e : \tau_2$$

We have two cases according to the value of κ. We show the case where $\kappa = 1$; the other case $\kappa = \omega$ is similar.
It must be the case that $\kappa_0 = 1$ and $\kappa_1 = 0$. Then, for each type environment Γ, Γ_{x_1}, ..., Γ_{x_n}, it must be the case that either x is not in the domain or the type of x is of the form $\tau_1 \to^0 \tau_2$. Thus,

$$x_1\colon \tau_{x_1}, \dots, x_n\colon \tau_{x_n} \geq \lfloor \Gamma + \Gamma_1 + \dots + \Gamma_n + \Gamma'_0 \rfloor.$$

Then, by the rule T-Heap, $\emptyset \vdash H^{-x} : \Gamma + (\Gamma'_0, x\colon \tau_1 \to^0 \tau_2)$. Letting $\Gamma'' = \Gamma'_0$ and $\kappa' = 0$ finishes the case.

2. Similar to the first part. □

Lemma 3 (Canonical Form). *Suppose* $\emptyset \vdash H : \Gamma$.

1. *If* $\Gamma(x) = \tau_1 \to^\kappa \tau_2$ *with* $\kappa \neq 0$, *then* $H(x) = (V, {}^{\langle \Gamma' \rangle}\mathbf{fun}\, y(z) = e)^{\kappa'}$ *for some* V, Γ', y, z, e, *and* κ' *with* $\kappa' \geq \kappa$.
2. *If* $\Gamma(x) = \tau_1 \times^\kappa \tau_2$ *with* $\kappa \neq 0$, *then* $H(x) = (v_1, v_2)^{\kappa'}$ *for some* v_1, v_2, *and* κ' *with* $\kappa' \geq \kappa$.

Proof. Immediate from the rules T-Heap, T-HFun, T-HPair. □

Proof of Theorem 1. By a case analysis on the rule used to derive $P \longmapsto P'$. We show only the main cases.

Case R-Pair: $P = (H, S, {}^{\langle \Gamma \rangle}\mathbf{let}\ x = (v_1, v_2)^\kappa\ \mathbf{in}\ e)$
$P' = (H \oplus \{y = (\widehat{V}(v_1), \widehat{V}(v_2))^\kappa\}, S, V \uplus \{x = y\}, e)$ $\quad$ y is fresh
By the rules T-Prog, T-Dec and T-Pair,

$$\begin{array}{lll} \emptyset \vdash H : \Gamma_1 + \Gamma_2 & \Gamma_1 \vdash S : \tau' \to^1 \tau & \Gamma_2 \vdash V : \Gamma_3 + v_1\colon \tau_1 + v_2\colon \tau_2 \\ \Gamma_3, x\colon \tau_1 \times^\kappa \tau_2 \vdash e : \tau' & \Gamma = \Gamma_3 + v_1\colon \tau_1 + v_2\colon \tau_2 & \end{array}$$

By the rule T-Env, there exists Γ'_2 such that $\Gamma_2 = \Gamma'_2 + \widehat{V}(v_1)\colon \tau_1 + \widehat{V}(v_2)\colon \tau_2$ and $\Gamma'_2 \vdash V : \Gamma_3$. By the rule rnT-HPair,

$$\widehat{V}(v_1)\colon \tau_1 + \widehat{V}(v_2)\colon \tau_2 \vdash (\widehat{V}(v_1), \widehat{V}(v_2))^\kappa : \tau_1 \times^\kappa \tau_2.$$

Since y is fresh, $y \notin dom(H)$. Then, by Lemma 1,

$$\emptyset \vdash H \oplus \{y = (\widehat{V}(v_1), \widehat{V}(v_2))^\kappa\} : \Gamma_1 + \Gamma'_2 + y\colon \tau_1 \times^\kappa \tau_2.$$

Also, by the rule T-Env, $\Gamma'_2 + y\colon \tau_1 \times^\kappa \tau_2 \vdash V \uplus \{x = y\} : \Gamma_3, x\colon \tau_1 \times^\kappa \tau_2$. Finally, the rule T-Prog finishes the case.

Case R-Ext: $P = (H, S, V, {}^{\langle \Gamma \rangle}\mathbf{let}\ (x, y) = z\ \mathbf{in}\ e)$ $\quad$ $H(V(z)) = (v_1, v_2)^\kappa$
$P' = (H^{-V(z)}, S, V \uplus \{x = v_1, y = v_2\}, e)$
By the rules T-Prog and T-Ext,

$$\begin{array}{lll} \emptyset \vdash H : \Gamma_1 + \Gamma_2 & \Gamma_1 \vdash S : \tau' \to^1 \tau & \Gamma_2 \vdash V : \Gamma_3 + z : \tau_1 \times^1 \tau_2 \\ \Gamma_3, x\colon \tau_1, y\colon \tau_2 \vdash e : \tau' & \Gamma = \Gamma_3 + z : \tau_1 \times^1 \tau_2 & \end{array}$$

By the rule T-Env, there exists Γ_2' such that $\Gamma_2 = \Gamma_2' + V(z){:}\,\tau_1 \times^1 \tau_2$ and $\Gamma_2' \vdash V : \Gamma_3$. By Lemma 2, $\emptyset \vdash H^{-z} : \Gamma_1 + \Gamma_2' + v_1{:}\,\tau_1 + v_2{:}\,\tau_2$. By the rule T-Env,

$$\Gamma_2' + v_1{:}\,\tau_1 + v_2{:}\,\tau_2 \vdash V \uplus \{x = v_1, y = v_2\} : \Gamma_4, x{:}\,\tau_1, y{:}\,\tau_2.$$

The rule T-Prog finishes the case.

Case R-App: $P = (H, S, V, {}^{\langle\Gamma\rangle}\,\mathbf{let}\ x = y\,v\ \mathbf{in}\ e)$
$H(V(y)) = (V', {}^{\langle\Gamma'\rangle}\mathbf{fun}\,z(w) = e')^{\kappa}$
$P' = (H^{-V(y)}, S[V, \mathit{TE}(e) \setminus \{x\}, \lambda x.e], V' \uplus \{z = V(y), w = \widehat{V}(v)\}, e')$
By the rules T-Prog and T-App

$$\begin{array}{lll} \emptyset \vdash H : \Gamma_1 + \Gamma_2 & \Gamma_1 \vdash S : \tau' \to^1 \tau & \Gamma_2 \vdash V : \Gamma_3 + y{:}\,\tau_1 \to^1 \tau_2 + v{:}\,\tau_1 \\ \Gamma = \Gamma_3 + y{:}\,\tau_1 \to^1 \tau_2 + v{:}\,\tau_1 & \Gamma_3, x{:}\,\tau_2 \vdash e : \tau' & \mathit{TE}(e) = \Gamma_3, x{:}\,\tau_2 \end{array}$$

By Lemma 2,

$$\begin{array}{l} \emptyset \vdash H^{-V(y)} : \Gamma_1 + \Gamma_2' + \widehat{V}(v){:}\,\tau_1 + (\Gamma'', V(y){:}\,\tau_1 \to^{\kappa'} \tau_2) \\ \Gamma'' \vdash V' : \Gamma' \qquad \Gamma', z{:}\,\tau_1 \to^{\kappa'} \tau_2, w{:}\,\tau_1 \vdash e' : \tau_2 \end{array}$$

for some κ'. Then, by the rule T-Env,

$$\widehat{V}(v){:}\,\tau_1 + (\Gamma'', V(y){:}\,\tau_1 \to^{\kappa'} \tau_2) \vdash V' \uplus \{z = V(y), w = \widehat{V}(v)\} : \Gamma', z{:}\,\tau_1 \to^{\kappa'} \tau_2, w{:}\,\tau_1.$$

Also, by the rule T-Env, there exists Γ_2' such that $\Gamma_2 = \Gamma_2' + V(y){:}\,\tau_1 \to^1 \tau_2 + \widehat{V}(v){:}\,\tau_1$ and $\Gamma_2' \vdash V : \Gamma_3$. Then, by the rule T-Push,

$$\Gamma_1 + \Gamma_2' \vdash S[V, \mathit{TE}(e) \setminus \{x\}, \lambda x.e] : \tau_2 \to^1 \tau.$$

Finally, the rule T-Prog finishes the case.

Case R-Ret: $P = (H, S[V', \Gamma', \lambda x.e_0], V, {}^{\langle\Gamma\rangle}v)$
$P' = (H, S, V' \uplus \{x = \widehat{V}(v)\}, e_0)$
By the rules T-Prog, T-Push, T-Ret, we have

$$\begin{array}{lll} \emptyset \vdash H : \Gamma_1 + \Gamma_2 + \Gamma_3 & \Gamma_1 \vdash S : \tau_2 \to^1 \tau & \Gamma_2 \vdash V' : \Gamma' \\ \Gamma', x{:}\,\tau_1 \vdash e_0 : \tau_2 & \Gamma_3 \vdash V : \Gamma & \Gamma = \Gamma' + v{:}\,\tau_1 \end{array}$$

By the rule T-Env, and T-Var, $\widehat{V}(v) \vdash \widehat{V}(v) : \tau_1$. Then, there exists Γ_4 such that $\Gamma_2 + \widehat{V}(v){:}\,\tau_1 \vdash V' \uplus \{x = \widehat{V}(v)\} : \Gamma', v{:}\,\tau_1$. Since $\Gamma_2 + \Gamma_2 \geq \Gamma_2 + \widehat{V}(v){:}\,\tau_1$, the rule T-Prog finishes the case. □

Proof of Theorem 2. By a case analysis on the form of P. Most cases are easy. We show a few main cases below:

Case: $P = (H, S, V, {}^{\langle\Gamma\rangle}\,\mathbf{let}\ x = y\,v\ \mathbf{in}\ e)$
By the rules T-Prog and T-App,

$$\begin{array}{lll} \emptyset \vdash H : \Gamma_1 + \Gamma_2 & \Gamma_1 \vdash S : \tau' \to^1 \tau & \Gamma_2 \vdash V : \Gamma_3 + y{:}\,\tau_1 \to^1 \tau_2 + v{:}\,\tau_1 \\ \Gamma_3, x{:}\,\tau_2 \vdash e : \tau' & \Gamma = \Gamma_3 + y{:}\,\tau_1 \to^1 \tau_2 + v{:}\,\tau_1 & \end{array}$$

Since y is given a function type, $V(y)$ must be some variable such that $\Gamma_2(V(y)) \geq \tau_1 \to^1 \tau_2$. By Lemma 3, $H(V(y)) = (V', {}^{\langle\Gamma'\rangle}\mathbf{fun}\,z(w) = e')^{\kappa'}$ for some V', Γ', w, z, e', and κ' with $\kappa' \geq 1$. Then, $H^{-V(y)}$ is well defined, finishing the case.

Case: $P = (H, S, V, {}^{\langle\Gamma\rangle}\,\mathbf{let}\ (x, y) = z\ \mathbf{in}\ e)$
By the rules T-Prog and T-Ext

$$\begin{array}{lll} \emptyset \vdash H : \Gamma_1 + \Gamma_2 & \Gamma_1 \vdash S : \tau' \to^1 \tau & \Gamma_2 \vdash V : \Gamma_3 + z{:}\,\tau_1 \times^1 \tau_2 \\ \Gamma_3, x{:}\,\tau_1, y{:}\,\tau_2 \vdash e : \tau' & \Gamma = \Gamma_3 + z{:}\,\tau_1 \times^1 \tau_2 & \end{array}$$

Since z is given a pair type, $V(z)$ must be some variable such that $\Gamma_2(V(z)) \geq \tau_1 \times^1 \tau_2$. By Lemma 3, $H(V(z)) = (v_1, v_2)^\kappa$ for some v_1, v_2, and κ such that $\kappa \geq 1$. Then, H^{-z} is well defined, finishing the case. □

Proof of Theorem 3. Immediate from Theorems 1 and 2. □

4 Garbage Collection

In this section we describe our GC algorithm for $\lambda^{\kappa}_{\mathrm{gc}}$ formally and prove its correctness.

4.1 GC Algorithm

First, we define three auxiliary functions TL_{env}, TL_{stack} and TL_{hval} to collect the type information on the addresses referred to by an environment, a stack, and a heap value, respectively. The first two are used when the garbage collector is invoked: the initial scan set is computed by $TL_{\mathrm{stack}}[S] + TL_{\mathrm{env}}[V, TE(e)]$ where S, V, e are respectively the stack, environment, and expression from the program for which GC is invoked. Besides, the function TL_{env} is used to compute type information on (continuation) closures stored in a heap or a stack. The last one is used when a heap value is marked and the new scan set is computed.

Definition 12. *The functions TL_{env}, TL_{stack} and TL_{hval} are defined as follows:*

$$\begin{array}{lcl} TL_{\mathrm{env}}[V, \Gamma] & = & \omega \cdot \sum_{x \in dom(\Gamma)} V(x) : \Gamma(x) \\ TL_{\mathrm{stack}}[[\,]] & = & \emptyset \\ TL_{\mathrm{stack}}[S[V, \Gamma, \lambda x.e]] & = & TL_{\mathrm{stack}}[S] + TL_{\mathrm{env}}[V, \Gamma] \\ TL_{\mathrm{hval}}[(V, {}^{\langle\Gamma\rangle}\mathbf{fun}\,x(y) = e)^\kappa, \tau_1 \to^{\kappa'} \tau_2] & = & \omega \cdot TL_{\mathrm{env}}[V, \Gamma] \\ TL_{\mathrm{hval}}[(v_1, v_2)^\kappa, \tau_1 \times^{\kappa'} \tau_2] & = & \omega \cdot (v_1{:}\,\tau_1 + v_2{:}\,\tau_2) \end{array}$$

Note that, the garbage collector need not distinguish between the uses 1 and ω in the scan set (and type information on the marked heap values). Hence, types are multiplied by ω to "normalize" type information.

Our GC algorithm is represented as rewriting of a quadruple $(H_f, \Gamma_s, H_t, \Gamma_t)$, consisting of two heaps and two type environments. Intuitively, H_f corresponds to a "from-space," H_t to a "to-space," and Γ_s to a "scan-set," which maintains locations to be traced together with their types. The type environment Γ_t, which maintains types of H_t, is used for avoiding verbose tracing, mentioned in Section 1. This algorithm could be implemented by extending either copying or mark-and-sweep GC, but some implementation details such as forwarding pointers or the sweeping phase are abstracted out from the model.

Definition 13. *The relation* $(H_f,\, \Gamma_s,\, H_t,\, \Gamma_t) \Longrightarrow (H'_f,\, \Gamma'_s,\, H'_t,\, \Gamma'_t)$ *is the least relation closed under the following rules:*

$$\begin{array}{c}(H_f \uplus \{x = h^\kappa\},\, (\Gamma_s,\, x{:}\,\tau),\, H_t,\, \Gamma_t) \Longrightarrow \\ (H_f,\, \Gamma_s + \lfloor TL_{\mathrm{hval}}[h^\kappa,\, \tau]\rfloor,\, H_t \uplus \{x = h^\kappa\},\, (\Gamma_t,\, x{:}\,\tau))\end{array} \qquad \text{(GC-Mark)}$$

$$\frac{\Gamma_t(x) \not\geq \tau}{\begin{array}{c}(H_f,\, (\Gamma_s,\, x{:}\,\tau),\, H_t,\, \Gamma_t) \Longrightarrow \\ (H_f,\, \Gamma_s + \lfloor TL_{\mathrm{hval}}[H_t(x),\, \tau]\rfloor,\, H_t, \Gamma_t + x{:}\,\tau)\end{array}} \qquad \text{(GC-Remark)}$$

$$\frac{\Gamma_t(x) \geq \tau}{(H_f,\, (\Gamma_s,\, x{:}\,\tau),\, H_t,\, \Gamma_t) \Longrightarrow (H_f,\, \Gamma_s,\, H_t,\, \Gamma_t)} \qquad \text{(GC-Skip)}$$

The rule GC-Mark moves the heap value at x to the to-space, computes the new scan set $\Gamma_s + \lfloor TL_{\mathrm{hval}}[h^\kappa,\, \tau]\rfloor$, and adds the type of the variable to Γ_t. Truncation is applied since types with the use 0 need not be marked (or should not as the memory space at the address has been deallocated). Note that, by truncation, the garbage collector may skip marking reachable garbage. As mentioned in Section 1, the garbage collector may mark one heap value more than once; the rule GC-Remark is applied when the heap value itself has been already marked but something reachable from it is neither marked nor scheduled to be marked yet. The premise $\Gamma_t(x) \not\geq \tau$, which judges existence of such unmarked values, holds if and only if, for the use 0 at some position in $\Gamma_t(x)$, the use ω resides at the corresponding position in the type τ. For example, suppose τ in the scan set is $(\mathit{Int}\times^\omega \mathit{Int})\times^\omega \mathit{Int}$ and $\Gamma_t(x)$ is $(\mathit{Int}\times^0 \mathit{Int})\times^\omega \mathit{Int}$. Then, the pair pointed to by x has to be marked again since the previous traverses did not mark the inner pair of type $\mathit{Int} \times^0 \mathit{Int}$. After this traverse the type $\Gamma_t(x)$ becomes $(\mathit{Int} \times^\omega \mathit{Int}) \times^\omega \mathit{Int}$ and the rule GC-Skip explained below will always be applied whenever x is selected from the scan set. The rule GC-Skip is applied when the values that the current traverse tries to copy are already marked or scheduled to be marked ($\Gamma_t(x) \geq \tau$).

Finally, the whole GC process, which computes a garbage-collected heap called *collection* from a program, is defined below. Given a program, it computes an initial scan set by using TL_{env} and TL_{stack}, begins rewriting defined above with the given heap and the initial scan set until it reaches the empty scan set; then, the to-space is the collection:

Definition 14 (collection). *A heap H' is a* collection *of a program* $(H,\ S,\ V,\ e)$ *if and only if*

$$(H,\ \lfloor TL_{\text{stack}}[S] + TL_{\text{env}}[V,\ TE(e)]\rfloor,\ \emptyset,\ \emptyset) \Longrightarrow^* (H'',\ \emptyset,\ H',\ \Gamma_t)$$

where $\Longrightarrow^*$ *is the reflexive and transitive closure of* $\Longrightarrow$.

Example 8. A collection of a program P

$$P = (H, [\,], \{p = x_1, p_2 = x_2, f = x_3, g = x_4\}, {}^{\langle f:Int\to^1 Int, g:Int\to^1 Int\rangle}e_5)$$

where

$$H = \left\{ \begin{array}{l} x_1 = (2, x_2)^\omega, x_2 = (3,4)^1, \\ x_3 = (\{p = x_1\}, {}^{\langle p:Int\times^1(Int\times^1 Int)\rangle}\mathbf{fun}\, f'(y) = e_f)^1, \\ x_4 = (\{p = x_1\}, {}^{\langle p:Int\times^1(Int\times^0 Int)\rangle}\mathbf{fun}\, g'(z) = e_g)^1 \end{array} \right\}$$

(where e_4, e_f, and e_g are taken from Example 7) is H itself. A possible sequence of GC rewriting steps is given as follows:

$$\begin{array}{rl}
& (H, (x_3{:}\, Int \to^\omega Int, x_4{:}\, Int \to^\omega Int), \emptyset, \emptyset) \\
\text{(GC-Mark)} & \Longrightarrow (H \setminus \{x_4\}, (x_1{:}\, Int \times^\omega (Int \times^0 Int), x_3{:}\, Int \to^\omega Int), \\
& \quad \{x_4 = H(x_4)\}, x_4{:}\, Int \to^\omega Int) \\
\text{(GC-Mark)} & \Longrightarrow (H \setminus \{x_1, x_4\}, x_3{:}\, Int \to^\omega Int + \lfloor x_2{:}\, Int \times^0 Int \rfloor, \\
& \quad \{x_1 = (2, x_2)^\omega, x_4 = H(x_4)\}, (x_1{:}\, Int \times^\omega (Int \times^0 Int), x_4{:}\, Int \to^\omega Int)) \\
& = (H \setminus \{x_1, x_4\}, x_3{:}\, Int \to^\omega Int, \\
& \quad \{x_1 = (2, x_2)^\omega, x_4 = H(x_4)\}, (x_1{:}\, Int \times^\omega (Int \times^0 Int), x_4{:}\, Int \to^\omega Int)) \\
\text{(GC-Mark)} & \Longrightarrow (\{x_2 = (3,4)^1\}, x_1{:}\, Int \times^\omega (Int \times^\omega Int), H \setminus \{x_2\}, \\
& \quad (x_1{:}\, Int \times^\omega (Int \times^0 Int), x_3{:}\, Int \to^\omega Int, x_4{:}\, Int \to^\omega Int)) \\
\text{(GC-Remark)} & \Longrightarrow (\{x_2 = (3,4)^1\}, x_2{:}\, Int \times^\omega Int, H \setminus \{x_2\}, \\
& \quad (x_1{:}\, Int \times^\omega (Int \times^\omega Int), x_3{:}\, Int \to^\omega Int, x_4{:}\, Int \to^\omega Int)) \\
\text{(GC-Mark)} & \Longrightarrow (\emptyset, \emptyset, H, \\
& \quad (x_1{:}\, Int \times^\omega (Int \times^\omega Int), x_2{:}\, Int \times^\omega Int, x_3{:}\, Int \to^\omega Int, x_4{:}\, Int \to^\omega Int))
\end{array}$$

Notice how the rule GC-Remark is applied and the type of x_1 is changed afterwards, and that the use of the rule GC-Remark could be dispensed with if x_3 were traversed first.

4.2 Correctness of GC

Correctness of the GC algorithm is proved by showing that a collection H' of a well-typed program $(H,\ S,\ V,\ e)$ is always obtained and both heaps before and after GC are well typed with respect to the same type environment, i.e.,

$\emptyset \vdash H : \Gamma$ and $\emptyset \vdash H' : \Gamma$. If the above condition holds, we can safely replace the heap in a well-typed program with its collection and obtain the same result (if exists) because the collection is a subset of the original heap and the program after GC does not cause run-time type errors. A more formal argument would be similar to the discussions found in [22, 23].

The main lemmas (6, 7, and 8) ensure that a certain invariant (the well-formedness condition in Definition 16 below) holds during GC and that GC always terminates with the empty scan set without failure. Then, the main theorem (4) is an easy consequence from the three lemmas.

Before the definition of well-formedness, we introduce an operation to recover the actual types of the scan set Γ_s and the marked set Γ_t.

Definition 15. *The* filtered *use of* κ_1 *by* κ_2*, written* $\kappa_1 \sharp \kappa_2$*, is defined by:* $\kappa \sharp \omega = \kappa$ *and* $\kappa \sharp 0 = 0$*. It is pointwise extended to types and type environments:*

$$\begin{array}{ll} \mathit{Int} \sharp \mathit{Int} & = \mathit{Int} \\ (\tau_1 \to^{\kappa_1} \tau_2) \sharp (\tau_1 \to^{\kappa_2} \tau_2) & = \tau_1 \to^{\kappa_1 \sharp \kappa_2} \tau_2 \\ (\tau_{11} \times^{\kappa_1} \tau_{12}) \sharp (\tau_{21} \times^{\kappa_2} \tau_{22}) & = (\tau_{11} \sharp \tau_{21}) \times^{\kappa_1 \sharp \kappa_2} (\tau_{21} \sharp \tau_{22}) \end{array}$$

$$\begin{array}{l} \mathit{dom}(\Gamma_1 \sharp \Gamma_2) = \mathit{dom}(\Gamma_1) \cap \mathit{dom}(\Gamma_2) \\ (\Gamma_1 \sharp \Gamma_2)(x) = \Gamma_1(x) \sharp \Gamma_2(x) \quad \text{for each } x \in \mathit{dom}(\Gamma_1) \cap \mathit{dom}(\Gamma_2) \end{array}$$

Definition 16 (well-formedness). *Suppose* $\emptyset \vdash H : \Gamma; \Gamma_H$*. The quadruple* $(H_f, \Gamma_s, H_t, \Gamma_t)$ *is* well formed *with respect to* $\emptyset \vdash H : \Gamma; \Gamma_H$ *iff:*

1. $H = H_f \uplus H_t$
2. $\omega \cdot \lfloor \Gamma \rfloor + \sum_{x \in \mathit{dom}(\Gamma_t)} \lfloor \mathit{TL}_{\mathrm{hval}}[H(x), \Gamma_H(x)] \rfloor \geq \Gamma_s + \Gamma_t$
3. $(\Gamma_H \sharp \Gamma_s) \vdash H_t : \Gamma; (\Gamma_H \sharp \Gamma_t)$

Intuitive meanings of the conditions are as follows: (1) every heap value is in either the from-space or the to-space; (2) the type of a variable in the scan set is compatible with the corresponding heap value's actual type (from this condition, we can show that $\omega \cdot \Gamma_H \geq \Gamma_s + \Gamma_t$); and (3) the scan-set holds all free variables used by the heap values in the to-set. The filtered type environments $\Gamma_H \sharp \Gamma_s$ and $\Gamma_H \sharp \Gamma_t$ are used to recover the actual type information on the scan set and the types of marked heap values from the normalized ones Γ_s and Γ_t.

Lemma 4 (Properties of $\mathit{TL}_{\mathrm{env}}$, $\mathit{TL}_{\mathrm{stack}}$, and $\mathit{TL}_{\mathrm{hval}}$).

1. *If* $\Gamma_1 \vdash V : \Gamma_2$*, then* $\Gamma_1 \sharp \mathit{TL}_{\mathrm{env}}[V, \Gamma_2] \vdash V : \Gamma_2$ *with* $\Gamma_1 = \Gamma_1 \sharp \mathit{TL}_{\mathrm{env}}[V, \Gamma_2]$ *and* $\omega \cdot (\Gamma_1 \sharp \mathit{TL}_{\mathrm{env}}[V, \Gamma_2]) = \mathit{TL}_{\mathrm{env}}[V, \Gamma_2]$.
2. *If* $\Gamma \vdash S : \tau_1 \to^1 \tau_2$*, then* $\Gamma \sharp \mathit{TL}_{\mathrm{stack}}[S] \vdash S : \tau_1 \to^1 \tau_2$ *with* $\Gamma = \Gamma \sharp \mathit{TL}_{\mathrm{stack}}[S]$ *and* $\omega \cdot (\Gamma \sharp \mathit{TL}_{\mathrm{stack}}[S]) = \mathit{TL}_{\mathrm{stack}}[S]$.
3. *If* $\Gamma \vdash h^\kappa : \tau$ *and* $\omega \cdot \tau \geq \tau'$*, then* $\Gamma \sharp \mathit{TL}_{\mathrm{hval}}[h^\kappa, \tau'] \vdash h^\kappa : \tau \sharp \tau'$ *with* $\Gamma \geq \Gamma \sharp \mathit{TL}_{\mathrm{hval}}[h^\kappa, \tau']$ *and* $\omega \cdot (\Gamma \sharp \mathit{TL}_{\mathrm{hval}}[h^\kappa, \tau']) = \mathit{TL}_{\mathrm{hval}}[h^\kappa, \tau']$.

Proof. Easy. □

Lemma 5. *If a quadruple* $(H_1, \Gamma_1, H_2, \Gamma_2)$ *is well formed with respect to* $\emptyset \vdash H : \Gamma; \Gamma_H$*, then* $\omega \cdot \Gamma_H \geq \Gamma_1 + \Gamma_2$*.*

Proof. By the rule T-Heap,

$$\begin{array}{l} H = \{x_1 = h_1^{\kappa_1}, \ldots, x_n = h_n^{\kappa_n}\} \\ \Gamma_{x_i} \vdash h_i^{\kappa_i} : \tau_i \quad (\text{for } i \in \{1, \ldots, n\}) \qquad \Gamma_H = \tilde{x}\colon \tilde{\tau} \geq \lfloor \Gamma + \Gamma_{x_1} + \cdots + \Gamma_{x_n} \rfloor \end{array}$$

and by the second well-formedness condition,

$$\omega \cdot \lfloor \Gamma \rfloor + \textstyle\sum_{x \in dom(\Gamma_2)} \lfloor TL_{\text{hval}}[H(x), \Gamma_H(x)] \rfloor \geq \Gamma_1 + \Gamma_2.$$

For each x_i, if $x_i \in dom(\Gamma_2)$, then, by Lemma 4,

$$\omega \cdot \Gamma_{x_i} \geq \omega \cdot (\Gamma_{x_i} \sharp\, TL_{\text{hval}}[H(x_i), \Gamma_H(x_i)]) = TL_{\text{hval}}[H(x_i), \Gamma_H(x_i)].$$

Then,

$$\begin{aligned} \omega \cdot \Gamma_H &\geq \omega \cdot \lfloor \Gamma + \Gamma_{x_1} + \cdots + \Gamma_{x_n} \rfloor = \omega \cdot \lfloor \Gamma \rfloor + \omega \cdot \lfloor \Gamma_{x_1} \rfloor + \cdots + \omega \cdot \lfloor \Gamma_{x_n} \rfloor \\ &\geq \omega \cdot \lfloor \Gamma \rfloor + \textstyle\sum_{x \in dom(\Gamma_2)} \lfloor TL_{\text{hval}}[H(x), \Gamma_H(x)] \rfloor \geq \Gamma_1 + \Gamma_2 \end{aligned}$$

finishing the proof. ◻

Lemma 6 (GC Well-Formedness Preservation). *Suppose* $\emptyset \vdash H : \Gamma; \Gamma_H$*. If* $(H_1, \Gamma_1, H_2, \Gamma_2)$ *is well formed with respect to* $\emptyset \vdash H : \Gamma; \Gamma_H$ *and* $(H_1, \Gamma_1, H_2, \Gamma_2) \Longrightarrow (H_1', \Gamma_1', H_2', \Gamma_2')$*, then* $(H_1', \Gamma_1', H_2', \Gamma_2')$ *is well formed with respect to* $\emptyset \vdash H : \Gamma; \Gamma_H$*.*

Proof. By a case analysis on the rule used to derive $(H_1, \Gamma_1, H_2, \Gamma_2) \Longrightarrow (H_1', \Gamma_1', H_2', \Gamma_2')$.

Case GC-Mark*:*

$$\begin{array}{ll} H_1 = H_f \uplus \{x = h^\kappa\} & \Gamma_1 = \Gamma_s, x\colon \tau \\ H_1' = H_f & \Gamma_1' = \Gamma_s + \lfloor TL_{\text{hval}}[h^\kappa, \tau] \rfloor \\ H_2' = H_2 \uplus \{x = h^\kappa\} & \Gamma_2' = \Gamma_2, x\colon \tau \end{array}$$

From the well-formedness conditions,

$$\begin{array}{l} H = (H_f \uplus \{x = h^\kappa\}) \uplus H_2 \\ \omega \cdot \lfloor \Gamma \rfloor + \sum_{y \in dom(\Gamma_2)} \lfloor TL_{\text{hval}}[H(y), \Gamma_H(y)] \rfloor \geq (\Gamma_s, x\colon \tau) + \Gamma_2 \\ (\Gamma_H \sharp\, \Gamma_s, x\colon \tau) \vdash H_2 : \Gamma; (\Gamma_H \sharp\, \Gamma_2). \end{array}$$

Then, it suffices to show

$$\begin{array}{l} H = H_f \uplus (H_2 \uplus \{x = h^\kappa\}) \\ \omega \cdot \lfloor \Gamma \rfloor + \sum_{y \in dom(\Gamma_2)} \lfloor TL_{\text{hval}}[H(y), \Gamma_H(y)] \rfloor + \lfloor TL_{\text{hval}}[H(x), \Gamma_H(x)] \rfloor \\ \quad \geq (\Gamma_s + \lfloor TL_{\text{hval}}[h^\kappa, \tau] \rfloor) + (\Gamma_2, x\colon \tau) \\ \Gamma_H \sharp\, (\Gamma_s + \lfloor TL_{\text{hval}}[h^\kappa, \tau] \rfloor) \vdash (H_2 \uplus \{x = h^\kappa\}) : \Gamma; (\Gamma_H \sharp\, (\Gamma_2, x\colon \tau)). \end{array}$$

The first condition is trivial. The second condition is also easily shown as follows. By Lemma 5, $\omega \cdot \Gamma_H \geq \Gamma_1 + \Gamma_2$ and so $\omega \cdot \Gamma_H(x) \geq \tau$; then, since $H(x) = h^\kappa$, we have $TL_{\text{hval}}[H(x), \Gamma_H(x)] \geq TL_{\text{hval}}[h^\kappa, \tau]$. The third condition is shown by using the rule T-Heap and the fact that $\Gamma_H \sharp\, TL_{\text{hval}}[h^\kappa, \tau] \vdash h^\kappa : \Gamma_H(x) \sharp\, \tau$, shown by Lemma 4.

Case GC-REMARK*:* Similar to the case above.

Case GC-SKIP*:* $\quad \Gamma_1 = \Gamma_s, x{:}\tau \qquad \Gamma_2(x) \geq \tau$
Easy because every use in $\Gamma_2(x)$ is either 0 or ω, making $\Gamma_2(x) + \tau = \Gamma_2(x)$. $\square$

Lemma 7 (GC Progress). *If* $\emptyset \vdash H : \Gamma; \Gamma_H$ *and* $(H_1, \Gamma_1, H_2, \Gamma_2)$ *is well-formed with respect to* $\emptyset \vdash H : \Gamma; \Gamma_H$*, then either* Γ_1 *is empty or* $(H_1, \Gamma_1, H_2, \Gamma_2) \Longrightarrow (H_1', \Gamma_1', H_2', \Gamma_2')$ *for some* $(H_1', \Gamma_1', H_2', \Gamma_2')$*.*

Proof. Take some $x \in \mathit{dom}(\Gamma_1)$. We first show either $x \in \mathit{dom}(H_1)$ or $x \in \mathit{dom}(H_2)$ below. By Lemma 5, $\omega \cdot \Gamma_H \geq \Gamma_1 + \Gamma_2$. Thus, $\mathit{dom}(\Gamma_1) \subseteq \mathit{dom}(\Gamma_H)$. Now, the fact that $H = H_1 \uplus H_2$ and $\mathit{dom}(H) = \mathit{dom}(\Gamma_H)$ concludes that either $x \in \mathit{dom}(H_1)$ or $x \in \mathit{dom}(H_2)$.

Case: $\quad H_1 = H_f \uplus \{x = h^\kappa\}$
It suffices to show the rule GC-MARK can be applied. First, well-definedness of $H_1' = H_f$ and $H_2' = H_2 \uplus \{x = h^\kappa\}$ and $\Gamma_2' = \Gamma_2, x{:}\Gamma_1(x)$ is immediate. Well-definedness of $\Gamma_1' = (\Gamma_1 \setminus \{x\}) + \lfloor \mathit{TL}_{\mathrm{hval}}[h^\kappa, \Gamma_1(x)] \rfloor$ is shown as follows. By an argument similar to the one found in the proof of Lemma 5, $\omega \cdot \Gamma_H \geq \lfloor \mathit{TL}_{\mathrm{hval}}[H(x), \Gamma_H(x)] \rfloor$. Since $\omega \cdot \Gamma_H \geq \Gamma_1$, we have $\mathit{TL}_{\mathrm{hval}}[H(y), \Gamma_H(y)] \geq \mathit{TL}_{\mathrm{hval}}[H(y), \Gamma_1(y)]$ for each $y \in \mathit{dom}(\Gamma_1)$. Then, $\omega \cdot \Gamma_H \geq \lfloor \mathit{TL}_{\mathrm{hval}}[H(x), \Gamma_1(x)] \rfloor$. Since well-definedness of $\Gamma \geq \Gamma'$ and $\Gamma \geq \Gamma''$ implies well-definedness of $\Gamma' + \Gamma''$, the type environment $\Gamma_1 \setminus \{x\} + \lfloor \mathit{TL}_{\mathrm{hval}}[H(x), \Gamma_1(x)] \rfloor$ is well defined.

Case: $\quad H_2 = H_t \uplus \{x = h^\kappa\}$
If $\Gamma_2(x) \geq \Gamma_1(x)$, then the rule GC-SKIP will be applied. Otherwise, the rule GC-REMARK is applied. $\Gamma_2 + x{:}\Gamma_1$ is well defined because $\Gamma_1 + \Gamma_2$ is well defined (the second well-formed condition). The argument about well-definedness of $(\Gamma_1 \setminus \{x\}) + \mathit{TL}_{\mathrm{hval}}[H_2(x), \Gamma_1(x)]$ is similar to the above case. $\square$

Lemma 8 (GC Termination). *If* $(H_1, \Gamma_1, H_2, \Gamma_2)$ *is well-formed with respect to* $\emptyset \vdash H : \Gamma; \Gamma_H$*, then there is no infinite sequence beginning with* $(H_1, \Gamma_1, H_2, \Gamma_2)$*.*

Proof. We define a binary relation $<$ by $(H_1, \Gamma_1, H_2, \Gamma_2) < (H_1', \Gamma_1', H_2', \Gamma_2')$ iff (1) $\Gamma_1 \subset \Gamma_1'$ with $\Gamma_2 = \Gamma_2'$ or (2) $\omega \cdot \Gamma_H \geq \Gamma_2 \geq \Gamma_2'$ with $\Gamma_2 \neq \Gamma_2'$. The ordering $<$ is well-founded because it is defined as an lexicographic ordering derived from two well-founded orderings. Then, we show that $\Longrightarrow$ generates a monotonically decreasing sequence with respect to $<$. Suppose $(H_f, \Gamma_s, H_t, \Gamma_t) \Longrightarrow (H_f', \Gamma_s', H_t', \Gamma_f')$. We have a case analysis on the rewriting rule used.

Case GC-MARK*:* We show that the condition (2) is satisfied. First, obviously, $\Gamma_t' \geq \Gamma_t$. The quadruple after rewriting is well formed by Lemma 6; so, by Lemma 5, $\omega \cdot \Gamma_H \geq \Gamma_s' + \Gamma_t' \geq \Gamma_t'$.

Case GC-REMARK*:* Since $\Gamma_t(x) \not\geq \Gamma_s(x)$, it must be the case that $\Gamma_t' \geq \Gamma_t$ with $\Gamma_t' \neq \Gamma_t$. Finally, $\omega \cdot \Gamma_H \geq \Gamma_t'$ is shown similarly to the above case.

Case GC-SKIP*:* Immediate from the fact that $\Gamma'_s \subset \Gamma_s$ and $\Gamma_t = \Gamma'_t$. $\square$

Theorem 4 (Correctness of GC Algorithm). *If* $\vdash (H, S, V, e) : \tau$*, then there exists a collection* H' *of the program* (H, S, V, e) *and* $\vdash (H', S, V, e) : \tau$*.*

Proof. By the rule T-HEAP,

$$\emptyset \vdash H : \Gamma_1 + \Gamma_2; \Gamma_H \qquad \Gamma_1 \vdash S : \tau_1 \rightarrow^1 \tau \qquad \Gamma_2 \vdash V : \mathit{TE}(e) \qquad \mathit{TE}(e) \vdash e : \tau_1.$$

By Lemma 4,

$$\begin{array}{ll} \Gamma_1 \sharp \mathit{TL}_{\text{stack}}[S] \vdash S : \tau_1 \rightarrow^1 \tau & \Gamma_1 = \Gamma_1 \sharp \mathit{TL}_{\text{stack}}[S] \\ \Gamma_2 \sharp \mathit{TL}_{\text{env}}[V, \mathit{TE}(e)] \vdash V : \mathit{TE}(e) & \Gamma_2 = \Gamma_2 \sharp \mathit{TL}_{\text{env}}[V, \mathit{TE}(e)]. \end{array}$$

Thus, $\emptyset \vdash H : ((\Gamma_1 \sharp \mathit{TL}_{\text{stack}}[S]) + (\Gamma_2 \sharp \mathit{TL}_{\text{env}}[V, \mathit{TE}(e)])); \Gamma_H$.

We first show that $(H, \lfloor \mathit{TL}_{\text{stack}}[S] + \mathit{TL}_{\text{env}}[V, \mathit{TE}(e)] \rfloor, \emptyset, \emptyset)$ is well formed with respect to $\emptyset \vdash H : (\Gamma_1 \sharp \mathit{TL}_{\text{stack}}[S]) + (\Gamma_2 \sharp \mathit{TL}_{\text{env}}[V, \mathit{TE}(e)]); \Gamma_H$. The first and third conditions are trivial. The second condition is also easy because, by Lemma 4, $\omega \cdot \lfloor (\Gamma_1 \sharp \mathit{TL}_{\text{stack}}[S]) + (\Gamma_2 \sharp \mathit{TL}_{\text{env}}[V, \mathit{TE}(e)]) \rfloor = \lfloor \mathit{TL}_{\text{stack}}[S] + \mathit{TL}_{\text{env}}[V, \mathit{TE}(e)] \rfloor$.

Then, by Lemmas 6–8, it is shown that the rewriting will terminate in a well-formed state with the empty scan set. By the third well-formedness condition, we have $\Gamma_H \sharp \emptyset \vdash H_t : (\Gamma_1 \sharp \mathit{TL}_{\text{stack}}[S]) + (\Gamma_2 \sharp \mathit{TL}_{\text{env}}[V, \mathit{TE}(e)]); (\Gamma_H \sharp \Gamma_t)$, i.e., $\emptyset \vdash H_t : (\Gamma_1 \sharp \mathit{TL}_{\text{stack}}[S]) + (\Gamma_2 \sharp \mathit{TL}_{\text{env}}[V, \mathit{TE}(e)])$, finishing the proof. $\square$

5 Related Work

Linear type systems. In most of the existing linear type systems [7, 27–29], the uses 0 and ω here are not distinguished. Thus, there would be no dangling pointers created in the heap space and conventional tracing GC could be applied. Mogensen [20] and the authors [14, 15] independently introduced 0 to the use information[3]. Under such a linear type system with the use 0, the garbage collector has to avoid tracing dangling pointers. In [14, 15], the summation of two pair types is defined only when the element types are the same, i.e., elements of a data structure have to be uniformly accessed in every context where the data structure is accessed. Thus, unlike our system, the type information on marked objects would not be required and the usual mark-bit mechanism would be enough. Furthermore, Mogensen's type system and Kobayashi's quasi-linear type system [17] removed the restriction on summation of types as in this paper. The new summation operator together with the use 0 plays a significant role to refine the analysis to detect linear values and so improve effectiveness of the static memory management. On the other hand, as we have studied, GC has to know which part of the marked object has been marked.

[3] The original idea of the use 0 is attributed to Bierman [5] and the implicit idea of the use 0 is also found in the preceding paper on a linear type system for the π-calculus [18].

Chirimar, Gunter, and Riecke [7] formalized memory management based on reference counting for a language with a linear type system. There are no dangling pointers in the memory space and the memory management algorithm itself was fairly straightforward.

Type-directed GC. There have been two approaches towards tag-free GC for ML-style polymorphic languages. In order to recover the actual type arguments of a polymorphic function, in one approach, explicit type arguments are passed at run-time [4, 23, 26] and, in the other, type reconstruction is performed at GC-time [1, 9–11, 13, 22]. Some type inference GC [9, 11, 13, 22] for polymorphic languages can collect reachable garbage as our GC also can. In some cases, our GC can collect more garbage than the type inference GC schemes proposed so far. For example, consider an expression $f(x) + g(x) + g(y)$ and suppose the function f uses its argument but g does not. If both f and g are used in a monomorphic context (as they are function arguments, for example), then y must be given the same type as f's domain type, which is concrete (i.e., not a type variable); thus, type inference GC cannot collect an object at y. In our type system, on the other hand, the use of the type of y can be 0 even if the use of f's domain type is more than 0. Thus, our garbage collector can collect the object at y. We leave further comparison of our GC and type inference GC for future work.

Agesen, Detlefs, and Moss [2] showed that, by a liveness analysis of local variables in a Java virtual machine, a garbage collector can avoid tracing useless references, thus reducing the required heap size. In fact, use of a liveness analysis seems to become fairly common in real compilers such as several Java JIT compilers. Since our garbage collector can avoid tracing useless variables (with the use 0) not only in the *stack* but also the *heap*, our technique can be more effective.

Region-based static memory management. Tofte et al. have proposed another technique for static memory management, based on region inference [6, 25]; it analyzes the lifetime of *regions*, which are fragments of memory space with nested lifetime, by using an effect-based type system and inserts explicit allocation/deallocation primitives into programs. The region-based memory management may also deallocate elements of a data structure before the deallocation of itself, creating dangling pointers. Kariya and Kobayashi [16] have developed an algorithm of tracing GC under the region-based static memory management. The idea is similar to ours: since the static type information tells the garbage collector which regions are not used by a certain value, it can be avoided to trace dangling pointers.

6 Conclusions

We have studied a GC scheme for a programming language with static memory management based on a linear type system. Since it allows linear values to be reclaimed before the reclamation of values pointing to them, dangling pointers may occur in the heap space; in order to deal with them, we exploited static type information during GC; our GC scheme can not only avoid tracing dangling pointers but also collect some of reachable garbage. We have formalized our GC algorithm and proved its correctness.

As discussed in the next section, efficient implementation of our GC will require much effort and the cost per invocation may be more expensive than ordinary GC. Nevertheless, it would be likely to benefit by *less frequent* invocation of GC, thanks to the static memory management. In particular, our GC is expected to be more effective in distributed systems, where (global) GC is rather expensive, or embedded systems, which typically have tight memory restriction.

7 Future Work

The work presented here is rather preliminary; much work is left to be done to adopt our technique to a real programming language like ML and evaluate a real impact.

First of all, integration with a polymorphic type system will be crucial. But, in fact, our GC scheme itself can be extended in a fairly straightforward manner: the technique of run-time type passing [23, 26] can be used to obtain actual type argument information on type variables, which the garbage collector will require. Moreover, this technique would also be extended to polymorphism on uses.

However, naive implementation of our memory management scheme described here will not be very effective. As in real implementation of region-based memory management [6], auxiliary techniques will be needed to reduce the overheads. We briefly discuss main issues and possible solutions below:

Effectiveness of use analysis. The present type system, apart from the lack of polymorphism, would not be very useful for static memory management: as discussed elsewhere [17], the condition of use-onceness is too restrictive. Kobayashi's quasi-linear type system [17] (extended with polymorphism), a refinement of a linear type system for memory management, would be a good base of our system. We should also explore the design space about polymorphism, involving many engineering tradeoffs between the power of the type system and the overhead of type reconstruction [29, 30].

Reduction of run-time cost. One of the main potential run-time overheads of our memory management would lie in deallocation of linear values involving dynamic check of uses, even though the check can be implemented without extra memory space for tags. To omit those checks, we might be able to modify the type system so that it also classifies primitive operations into two: those accepting only linear values with performing deallocation, and those accepting only non-linear values without deallocation. However, type reconstruction for such a type system would be impractical; we expect the complexity to be exponential. Instead, we will be able to benefit from flow analysis to determine at which deallocation points dynamic checks can be omitted. In the presence of polymorphism, the technique of type lifting [19,24,26] can be used to reduce the cost of run-time type passing.

Reduction of GC-time cost. Our garbage collector is presented so that it keeps track of type information on the heap values already marked. However, in real implementation, the full type information is *not* needed: for example, for a function closure, only one bit information (0 or ω), corresponding to the outermost use of the function type, is sufficient. Concerning pair types, the required information can be represented by a bit vector of length $1 + n_1 + n_2$ where n_i is the length of such a bit vector for the i-th component type. Thus, without deeply nested pair types (or large tuple types), the cost could be comparable to conventional tracing GC, which uses a mark bit for each heap block. Moreover, such bit vectors may even be omitted by analyzing aliasing in the heap space. Besides the cost of maintenance of type information, the GC cost will heavily depend on how often remarking happens; some heuristics of the tracing strategy may be effective.

As other theoretical issues, it is interesting to generalize our system so that the GC algorithm is parameterized by the underlying linear type system. Also, formal connection to type inference GC would be also worth investigating.

Acknowledgment

Comments from anonymous referees of TIC2000 helped us improve the presentation.

References

[1] Shail Aditya, Christine H. Flood, and James E. Hicks. Garbage collection for strongly-typed languages using run-time type reconstruction. In *Proceedings of the ACM Conference on Lisp and Functional Programming (LFP)*, pages 12–23, Orlando, FL, June 1994.

[2] Ole Agesen, David Detlefs, and J. Eliot B. Moss. Garbage collection and local variable type-precision and liveness in Java virtual machines. In *Proceedings of the ACM Conference on Programming Language Design and Implementation (PLDI)*, pages 269–279, Montreal, Canada, June 1998.

[3] Alexander Aiken, Manuel Fähndrich, and Raph Levien. Better static memory management: Improving region-based analysis of higher-order languages. In *Proceedings of the ACM Conference on Programming Language Design and Implementation (PLDI)*, pages 174–185, San Diego, CA, June 1995.

[4] Andrew W. Appel. Runtime tags aren't necessary. *Lisp and Symbolic Computation*, 2(7):153–162, 1989.

[5] Gavin M. Bierman. Type systems, linearity and functional languages. In *Proceedings of the CLICS Workshop*, pages 71–92, March 1992. Available as Aarhus University Technical Report DAIMI PB 397-I.

[6] Lars Birkedal, Mads Tofte, and Magnus Vejlstrup. From region inference to von Neumann machines via region representation inference. In *Proceedings of the 23rd ACM SIGACT-SIGPLAN Symposium on Principles of Programming Languages (POPL)*, pages 171–183, St. Petersburg Beach, FL, January 1996.

[7] Jawahar Chirimar, Carl A. Gunter, and Jon G. Riecke. Proving memory management invariants for a language based on linear logic. In *Proceedings of the ACM Conference on Lisp and Functional Programming (LFP)*, pages 139–150, San Francisco, CA, July 1992.

[8] Cormac Flanagan, Amr Sabry, Bruce F. Duba, and Matthias Felleisen. The essence of compiling with continuations. In *Proceedings of the ACM Conference on Programming Language Design and Implementation (PLDI)*, pages 237–247, Albuquerque, NM, June 1993.

[9] Pascal Fradet. Collecting more garbage. In *Proceedings of the ACM Conference on Lisp and Functional Programming (LFP)*, pages 24–33, Orlando, FL, June 1994.

[10] Benjamin Goldberg. Tag-free garbage collection for strongly typed programming languages. In *Proceedings of the ACM Conference on Programming Language Design and Implementation (PLDI)*, pages 165–176, Toronto, Canada, June 1991.

[11] Benjamin Goldberg and Michael Gloger. Polymorphic type reconstruction for garbage collection without tags. In *Proceedings of the ACM Conference on Lisp and Functional Programming (LFP)*, pages 53–65, San Francisco, CA, June 1992.

[12] Juan C. Guzmán and Paul Hudak. Single-threaded polymorphic lambda calculus. In *Proceedings of the 5th IEEE Symposium on Logic in Computer Science (LICS)*, pages 333–343, Philadelphia, PA, June 1990.

[13] Haruo Hosoya and Akinori Yonezawa. Garbage collection via dynamic type inference — a formal treatment —. In *Proceedings of the 2nd International Workshop on Types in Compilation (TIC)*, volume 1473 of *Lecture Notes on Computer Science*, pages 215–239, Kyoto, Japan, March 1998. Springer-Verlag.

[14] Atsushi Igarashi. Type-based analysis of usage of values for concurrent programming languages. Master's thesis, University of Tokyo, Tokyo, Japan, February 1997.

[15] Atsushi Igarashi and Naoki Kobayashi. Type reconstruction for linear π-calculus with I/O subtyping. *Information and Computation*, 161:1–44, August 2000. A preliminary summary was presented in *Proceedings of SAS'97* under the title "Type-based Analysis of Communication for Concurrent Programming Languages," Springer LNCS 1302, pages 187–201.

[16] Hideki Kariya and Naoki Kobayashi. Combining region-based memory management with conventional garbage collection. *JSSST Computer Software*, 16(3):66–70, May 1999. (In Japanese).

[17] Naoki Kobayashi. Quasi-linear types. In *Proceedings of the 26th ACM SIGACT-SIGPLAN Symposium on Principles of Programming Languages (POPL)*, pages 29–42, San Antonio, TX, January 1999.

[18] Naoki Kobayashi, Benjamin C. Pierce, and David N. Turner. Linearity and the pi-calculus. *ACM Transactions on Programming Languages and Systems*, 21(5):914–947, September 1999. A preliminary summary appeared in *Proceedings of ACM POPL'96*, pp. 358–371.

[19] Yasuhiko Minamide. Compilation based on a calculus for explicit type passing. In *Proceedings of the Second Fuji International Workshop on Functional and Logic Programming*, 1996.

[20] Torben Æ. Mogensen. Types for 0, 1 or many uses. In *Proceedings of the 9th International Workshop on Implementation of Functional Languages*, number 1467 in Lecture Notes on Computer Science, pages 157–165, St. Andrews, Scotland, September 1997. Springer-Verlag.

[21] Greg Morrisett. *Compiling with Types.* PhD thesis, Carnegie Mellon University, December 1995.

[22] Greg Morrisett, Matthias Felleisen, and Robert Harper. Abstract models of memory management. In *Proceedings of the 7th International Conference on Functional Programming Languages and Computer Architecture (FPCA)*, pages 66–77, San Diego, CA, June 1995.

[23] Greg Morrisett and Robert Harper. Semantics of memory management for polymorphic languages. Technical Report CMU-CS-96-176, Carnegie Mellon University, Pittsburgh, PA, September 1996.

[24] Bratin Saha and Zhong Shao. Optimal type lifting. In *Proceedings of the 2nd International Workshop on Types in Compilation (TIC)*, volume 1473 of *Lecture Notes on Computer Science*, pages 156–177, Kyoto, Japan, March 1998. Springer-Verlag.

[25] Mads Tofte and Jean-Pierre Talpin. Implementation of the typed call-by-value λ-calculus using a stack of regions. In *Proceedings of the 21st ACM SIGACT-SIGPLAN Symposium on Principles of Programming Languages (POPL)*, pages 188–201, Portland, OR, January 1994.

[26] Andrew Tolmach. Tag-free garbage collection using explicit type parameters. In *ACM Conference on Lisp and Functional Programming (LFP)*, pages 1–11, Orlando, FL, June 1994.

[27] David N. Turner, Philip Wadler, and Christian Mossin. Once upon a type. In *Proceedings of the 7th International Conference on Functional Programming Languages and Computer Architecture (FPCA)*, pages 1–11, San Diego, CA, June 1995.

[28] Philip Wadler. Linear types can change the world! In M. Broy and C. Jones, editors, *Programming Concepts and Methods*, Sea of Galilee, Israel, April 1990. North Holland. IFIP TC 2 Working Conference.

[29] Keith Wansbrough and Simon Peyton Jones. Once upon a polymorphic type. In *Proceedings of the 26th ACM SIGACT-SIGPLAN Symposium on Principles of Programming Languages (POPL)*, pages 15–28, San Antonio, TX, January 1999.

[30] Keith Wansbrough and Simon Peyton Jones. Simple usage polymorphism. In *Proceedings of the 3rd International Workshop on Types in Compilation (TIC)*, Montreal, Canada, September 2000.

[31] Andrew K. Wright and Matthias Felleisen. A syntactic approach to type soundness. *Information and Computation*, 115(1):38–94, November 1994.