TECHNICAL REPORT TR98-02

Quasi-Linear Types

Naoki Kobayashi

September 1998

Department of Information Science
Faculty of Science, University of Tokyo

7–3–1 Hongo, Bunkyo-Ku Tokyo, 113 Japan

TITLE
Quasi-Linear Types

AUTHORS
Naoki Kobayashi

KEY WORDS AND PHRASES
linear types, memory management

ABSTRACT
Linear types (types of values that can be used just once) have been drawing great attentions because they are useful for memory management, in-place update of data structures, etc.: an obvious advantage is that a value of a linear type can be immediately deallocated after being used. However, the linear types have not been applied so widely in practice, probably because linear values (values of linear types) in the traditional sense do not so often appear in actual programs. In order to increase the applicability of linear types, we relax the condition of linearity by extending the types with information on an evaluation order and simple dataflow information. The extended type system, called a quasi-linear type system, is formalized and its correctness is proved. We have implemented a prototype type inference system for the core-ML that can automatically find out which value is linear in the relaxed sense. Promising results were obtained from preliminary experiments with the prototype system.

REPORT DATE	**WRITTEN LANGUAGE**
September 1998	English
TOTAL NO. OF PAGES	**NO. OF REFERENCES**
40	21

ANY OTHER IDENTIFYING INFORMATION OF THIS REPORT
A summary will be presented at ACM POPL'99. Up-to-date version of this report is available from http://www.yl.is.s.u-tokyo.ac.jp/ koba/pub/qlinear.ps.gz.

DISTRIBUTION STATEMENT
First issue 35 copies.

SUPPLEMENTARY NOTES

DEPARTMENT OF INFORMATION SCIENCE
Faculty of Science, University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan

Quasi-Linear Types

Naoki Kobayashi
Department of Information Science, University of Tokyo
7-3-1 Hongo,Bunkyo-ku, Tokyo 113-0033, Japan
Email:koba@is.s.u-tokyo.ac.jp

1 Introduction

1.1 Linear types

A number of type systems based on Girard's linear logic [6] have been proposed for functional languages [1, 3, 13, 21] and concurrent languages [12, 9]. They guarantee that certain data structures (called *linear* values) are accessed just once (or at most once). As suggested by many researchers [1, 13, 21], the distinction between linear values and other values brings us the following benefits:

- Improvement of memory management — If some data structure is statically known to be linear, it can be deallocated immediately after it is used.
- In-place update — If another data structure needs to be created immediately after a linear value is accessed, we can just replace the old data with the new data, instead of deallocating the old data and allocating a heap space for the new data. For instance, when the function $\lambda(x, y).(x, y + 1)$ is called, if the argument is expected to be a linear pair, the second element y of the pair can be destructively updated to $y + 1$.
- Safe inlining (in call-by-need languages) [21] — In general, replacing $(\lambda x.e_1)e_2$ with $[e_2/x]e_1$ in a call-by-need language may duplicate evaluation of e_2 and thus slow down the execution; however, if x is accessed just once in e_1, it is safe to perform such program transformation.

We are here interested in the first two of the above benefits. If there are many linear data, program execution (especially execution of functional programs, because many intermediate data structures like cons cells and closures tend to be created in naive functional programs) can get much benefit in both time and space.

1.2 Limitation of linear types

In spite of the above-mentioned benefit from linear type systems, they do not seem to have been used so widely, even after type inference algorithms [21, 9, 8] were proposed that can automatically find which values are linear.

We think one of the major reasons for this is that linear type systems are too naive for the above application: what is actually important for automatic deallocation and in-place update is to know which is the last access to a heap value; the condition of linearity is too strong for this purpose. For example, consider an expression $M \equiv \mathbf{let}\ x = 2.0\ \mathbf{in}\ \mathbf{let}\ y = x + 1.0\ \mathbf{in}\ y + x$:

it is easy to see that $y + x$ is the last access to x, but linear type systems cannot insert a code to deallocate x — just because x is used twice! This limitation also forces a programmer to follow a particular programming style. For instance, in $\lambda z.(\mathbf{fst}(z), \mathbf{snd}(z) + 1)$, the pair z is judged to be accessed twice; so, if one wants the pair to be deallocated, he or she must write $\lambda(x, y).(x, y + 1)$ instead.

1.3 Our proposal — almost linear types (quasi-linear types)

In order to remove the above-mentioned limitation of linear types, we relax the condition of linearity.

A key idea is to express information on the evaluation order and dataflow by using types. Recall the above expression M: if a type system can take it into account that $x+1.0$ is evaluated before $y+x$, it can ignore the use of x in $x+1.0$ and treat the above expression in the same way as **let** $x = 2.0$ **in let** $y = 1.0$ **in** $y + x$. In order to obtain such information on the evaluation order, the type system should also be able to deal with some dataflow information. Consider, for example, an expression **let** $y = f(x)$ **in** $g(x) + h(y)$. $f(x)$ is evaluated before $g(x)$ but it does not imply that the access to x in $g(x)$ is the last one: $f(x)$ may return x and it may be accessed in $h(y)$. In order to deal with this, we distinguish between the type of a value that must be used up in an expression and the type of a value that may be returned as a part of the evaluation result of an expression.

It would be possible to obtain the above kinds of information by using more sophisticated analyses, but we instead obtain it by a very simple method — by introducing a new *use* [21, 9]. Advantages of that approach include: (1) the whole analysis can be formalized uniformly in terms of a type system (this is especially advantageous when we introduce polymorphism) and (2) a program containing free variables can be analyzed as long as their type information is provided, which is useful for modular (or incremental) analysis and separate compilation of programs. We first review the ideas of previous linear type systems [21, 9] and then overview the ideas of our system below.

1.3.1 Review of linear type systems

A *use* was introduced to control how often values can be used [21, 9]: it was either 0 (not used at all), 1 (used at most once), or ω (used an arbitrary number of times) in [9] and was either 1 or ω in [21]. Types are annotated with such uses. For example, $real^1$ represents the type of real numbers that can be used at most once, and $int \rightarrow^{\omega} int$ is the type of functions on integers that can be called an arbitrary number of times. A type judgment can be accordingly refined and gives more useful information than the ordinary one: for example, $x : real^1 \vdash N : real^{\omega}$ means not only that x is used as a real number in N but also that it is used at most once. So, it is invalid if N is **let** $y = x + 1.0$ **in** $x + y$.

The following type derivation highlights a key point of the linear type system:

$$\frac{x : real^1 \vdash x + 1.0 : real^1 \qquad x : real^1, y : real^1 \vdash x + y : real^1}{x : real^1 + x : real^1 (= x : real^{\omega}) \vdash \mathbf{let}\ y = x + 1.0\ \mathbf{in}\ x + y : real^1}$$

The premises of the derivation imply that x is used once each in $x + 1.0$ and $x + y$. In order to know how x is totally used in the expression, we combine the type environments of subexpressions by *adding* corresponding uses as above. This is in contrast to the following rule of ordinary type systems, where type environments are *shared* between subexpressions:

$$\frac{\Gamma \vdash M_1 : \tau_1 \qquad \Gamma, y : \tau_1 \vdash M_2 : \tau_2}{\Gamma \vdash \mathbf{let}\ y = M_1\ \mathbf{in}\ M_2 : \tau_2}$$

By using the linear type system, we can infer how often each heap value is accessed, and annotate each allocation with a use. For instance, we can annotate the expression M as **let** $x^{\omega} = 2.0$ **in let** $y^1 = x + 1.0$ **in** $x + y$, so that y is deallocated after it is used in $x + y$ while x is not.

1.3.2 Overview of our type system

In order to express simple dataflow information, we introduce another use δ. Intuitively, a value with use δ (which we call a *δ-value* below) may be used many times but only *locally* — it cannot be returned as (a part of) a result.[1] So, x can have type real^{δ} in $x + 1.0$ and $(\lambda y.(y + 1.0))x$, but not in $\langle x, 1.0\rangle$ or $\lambda y.x + y$. A value with use 1 (hereafter, we call it a *quasi-linear* value or simply a *linear* value, and call a use-once value in the traditional sense a *strictly linear* value) can now be used in a more relaxed manner: It can be used many times as a δ-value, and then it can be used as a strictly linear value (i.e., it can be used once and deallocated).

In order to guarantee such usage of a value, it is crucial that the type system can take the evaluation order into account. The type derivation shown above is now changed as follows:

$$\frac{x : \mathit{real}^{\delta} \vdash x + 1.0 : \mathit{real}^1 \qquad x : \mathit{real}^1, y : \mathit{real}^1 \vdash x + y : \mathit{real}^1}{x : \mathit{real}^{\delta}; x : \mathit{real}^1 (= x : \mathit{real}^1) \vdash \mathbf{let}\ y = x + 1.0\ \mathbf{in}\ x + y : \mathit{real}^1}$$

Here, the new combination $(x : \mathit{real}^{\delta}); (x : \mathit{real}^1)$ of type environments captures the fact that x is first used as a δ-value in $x + 1.0$, and then it is used as a linear value in $x + y$. Thus, x can be deallocated after $x + y$ is evaluated. Similarly, even if one writes $\lambda z.(\mathbf{fst}(z), \mathbf{snd}(z) + 1)$ instead of $\lambda(x, y).(x, y + 1)$, z can be deallocated after $\mathbf{snd}(z)$ is evaluated.

Because a quasi-linear value may be used more than once, in addition to allocation of a heap value, each access to a heap value is also annotated with a use: it is for distinguishing between the last access of the value and other accesses. For example, the expression M is annotated as **let** $x^1 = 2.0$ **in let** $y^1 = x^{\delta} + 1.0$ **in** $y^1 + x^1$ (actually, we will annotate constructors and destructors of heap values instead of variables).

1.4 Main results

Main contributions of the present work are formalization of the new type system sketched above and a proof of its correctness. Since we must take the evaluation order into account, the proof is non-trivial and more involved than that of previous type systems.

Another contribution is implementation of a type inference system for the core-ML (i.e., Standard ML without modules [14]) based on the new type system, which inputs an unannotated core-ML expression and outputs a use-annotated expression. We have so far tested several simple programs such as sorting programs, Eratostene's sieve and the life, and obtained promising results: for the sorting programs and Eratostene's sieve, the type inference system could judge that all the dynamically allocated heap values except for top-level functions are linear (which indicates that they can be run almost without garbage collection). Our system could judge most heap values to be linear also for the life. Moreover, the annotated programs output by the system indicate that those linear cons cells can be updated in-place. More information on the prototype inference system is available through `http://www.yl.is.s.u-tokyo.ac.jp/~koba/research/qlinear/`.

[1] Intuition behind the name δ is that it is something between 0 and 1, and that it can be almost ignored compared with 1.

1.5 Structure of the paper

The rest of this paper is structured as follows. Section 2 introduces the syntax of the target language. Section 3 gives our type system for judging whether an expression is correctly annotated with uses. Section 4 shows the correctness of the type system and Section 5 briefly explains how to obtain a type inference algorithm that can automatically infer use annotation. Section 6 discusses several extensions of the target language. Section 7 explains our prototype inference system and the results of preliminary experiments. Section 8 discusses related work and Section 9 concludes this paper.

2 Syntax and Operational Semantics of the Target Language

In this section, we introduce the syntax of (use-annotated) expressions. Use annotation can be automatically recovered by a type inference algorithm discussed in Section 5.

2.1 Uses

As explained in Section 1, a *use* is introduced to control how often and in which way each heap value can be accessed.

Definition 2.1 [uses]: A use is 0, δ, 1, or ω.

We use metavariables $\kappa_1, \kappa_2, \ldots$ for uses.

As mentioned in Section 1, a new point here is the use δ: basically, if some heap value is assigned the use δ in an expression,[2] then it must be used *locally* inside the expression and cannot be returned as a part of the evaluation result. 0 means that the data cannot be accessed at all, 1 means that after the data is used in a restricted manner as a value of use δ, it can be either (1) used at most once and deallocated inside the current expression being evaluated or (2) put into the evaluation result as a quasi-linear value. ω means that the data can be used many times in any way.

A value of use ω can be regarded as a value of any other use, and a value of 1 can be regarded as a value of use δ or 0. In order to express this kind of relationship between uses, we define a total order $\leq$ on uses by $0 \leq \delta \leq 1 \leq \omega$. We write $\kappa_1 < \kappa_2$ if $\kappa_1 \neq \kappa_2$ and $\kappa_1 \leq \kappa_2$.

2.2 Expressions

We consider here a simply-typed λ-calculus with recursion and pairs as a target language. Its extension with polymorphism and other data structures will be briefly discussed in Section 6. Because our type system is sensitive to an evaluation order, we use the K-normal form [5] so that the evaluation order be explicit and each intermediate value be named.

Each allocation or read operation of a heap value is annotated with a use.

Definition 2.2 [expressions]:

$$
\begin{array}{rcl}
M & ::= & V \mid \textbf{let } x = \langle V_1, V_2 \rangle^{\kappa} \textbf{ in } M \mid \textbf{let } x = \textbf{fst}^{\kappa}(y) \textbf{ in } M \mid \textbf{let } x = \textbf{snd}^{\kappa}(y) \textbf{ in } M \\
 & & \mid \textbf{let } x = \lambda^{\kappa} y.M_1 \textbf{ in } M_2 \mid \textbf{let } x = \textbf{rec}^{\kappa}(f, y, M_1) \textbf{ in } M_2 \mid \textbf{let } x = \mathit{apply}^{\kappa}(y, V) \textbf{ in } M \\
 & & \mid \textbf{let } x = M_1 \textbf{ in } M_2 \mid \textbf{if0 } V \textbf{ then } M_1 \textbf{ else } M_2 \\
V & ::= & x \mid n
\end{array}
$$

[2] Since the heap value may be assigned a different use in another expression, it may be better to understand the use δ as a property of the reference (pointer) to the value, rather than the property of the value itself.

We write $\mathcal{E}$ for the set of expressions.

The metavariable V represents a value that can be represented in one word and need not be allocated in the heap. Here, it is either an integer constant (denoted by n) or a variable (which can be considered a heap address).

$\langle V_1, V_2\rangle^\kappa$ and $\lambda^\kappa y.M_1$ allocate a pair and a closure respectively in a heap, and records that their use is κ (which shall be restricted to $0, 1$, or ω by the type system given later). The recorded use represents how the value can be accessed in the rest of the computation: 0 means that the value cannot be accessed at all (so, $\langle V_1, V_2\rangle^0$ need not actually allocate the pair in the heap), 1 means that the value can be accessed as a linear value and may be deallocated later, and ω means that the value can be accessed in an arbitrary manner and can be deallocated only by other mechanisms such as garbage collection. $\mathbf{rec}^\kappa(f, y, M_1)$ creates a recursive function f such that $f(y) = M_1$.

There are three operations to access a heap: $\mathbf{fst}^\kappa(x)$ ($\mathbf{snd}^\kappa(x)$, resp.) extracts the first (second, resp.) element of the pair stored at x, and $\mathit{apply}^\kappa(y, V)$ reads the closure stored at y and applies it to V. κ attached to an access operation represents what access capability is used by the operation: δ represents only the capability to read the heap value and 1 represents the capability to deallocate the value also if it is linear. ω represents an arbitrary capability in principle but it is the same as δ in effect since the heap value must be an ω-value and cannot be deallocated (so, ω is unnecessary but included just for technical convenience). If κ is greater than the provided capability (i.e., the use recorded at the accessed address), it is an invalid access and causes an error (this check is actually unnecessary since the type system guarantees that no such invalid access occurs). Otherwise, if both κ and the recorded use are 1,[3] the accessed value is deallocated (formally, the deallocation is modeled as a change of the recorded use into 0). In the other cases, the heap is left unchanged.

Example 2.3: $\mathbf{let}\ x = \langle 1, 2\rangle^1\ \mathbf{in\ let}\ y = \mathbf{fst}^\delta(x)\ \mathbf{in\ let}\ z = \mathbf{snd}^1(x)\ \mathbf{in}\ z$ allocates a linear pair $\langle 1, 2\rangle$ in the heap, and then read its first element; at this time, x is not deallocated since **fst** is annotated with δ. After that, the second element is read and the pair is deallocated.

Remark 2.4: One may think that the above check of uses incurs a heavy overhead. However, because the type system guarantees that no invalid access occurs, all necessary operations on uses are actually to check whether the accessed heap value is annotated with 1 or ω when a heap access operation is annotated with 1. Moreover, the use of a heap value can be recorded not in a heap space itself but in the pointer to it by using a one-bit tag. So, we think the actual run-time overhead is quite small.

2.3 Operational semantics

In order to clarify how use annotation is used for allocating/deallocating heap values at run-time, we define an operational semantics as a rewriting relation of pairs of a heap and an expression [16, 21].

Definition 2.5 [evaluation contexts]: The set of *evaluation contexts* is given by the following syntax:

$$C[\,] ::= [\,] \mid \mathbf{let}\ x = C[\,]\ \mathbf{in}\ M.$$

We write $C[M]$ for the expression obtained from $C[\,]$ by replacing the hole $[\,]$ with M.

[3] The check of the latter use is necessary, because we allow an ω-value to be used as a linear value. If we forbid such coercion of an ω-value into a linear value as in [21], then the check is unnecessary. However, we prefer not to make such restriction because the check can be implemented efficiently. See Remark 2.4.

Definition 2.6 [heap]: A *heap value*, denoted by h, is a term of the form $\langle V_1, V_2\rangle$ or $\lambda x.M$. A *heap* is a mapping from a finite set of variables to pairs consisting of a use and a heap value. We write $\{x_1 \mapsto^{\kappa_1} h_1, \ldots, x_n \mapsto^{\kappa_n} h_n\}$ for a heap H such that $H(x_i) = (\kappa_i, h_i)$.

The use associated with a heap value is used for judging whether the value can be deallocated after it is accessed as a linear value: the value can be deallocated only if the use is 1. The use 0 means that the value has been already deallocated. We often write h^κ for $\langle V_1, V_2\rangle^\kappa$ or $\lambda^\kappa x.M$. We write $\mathcal{H}$ fot the set of heaps.

The reduction relation $\rightsquigarrow \in (\mathcal{H} \times \mathcal{E}) \times ((\mathcal{H} \times \mathcal{E}) \cup \{\mathbf{Error}\})$ is defined by the rules below. **Error** represents a state after invalid access to the heap has occurred (there are other kinds of errors such as trying to apply a non-function to a value, etc.; however, because the lack of those errors can be guaranteed by a type system in the usual way, this paper focuses on errors caused by invalid heap access).

The following two rules are for allocating heap values. The use κ is recorded at a new address together with the allocated value.

$$(H, C[\mathbf{let}\ x = h^\kappa\ \mathbf{in}\ M]) \rightsquigarrow (H\{y \mapsto^\kappa h\}, C[[y/x]M])\ (y \text{ fresh}) \qquad \text{(R-Heap)}$$

$$(H, C[\mathbf{let}\ x = \mathbf{rec}^\kappa(f, y, M_1)\ \mathbf{in}\ M]) \rightsquigarrow (H\{z \mapsto^\kappa \lambda y.[z/f]M_1\}, C[[z/x]M])\ (z \text{ fresh}) \qquad \text{(R-Fix)}$$

If V is a one-word value, then reducing $\mathbf{let}\ x = V\ \mathbf{in}\ M$ does not change the heap:

$$(H, C[\mathbf{let}\ x = V\ \mathbf{in}\ M]) \rightsquigarrow (H, C[[V/x]M]) \qquad \text{(R-SVal)}$$

The following two rules are for function application. Since $apply^{\kappa'}(x, V)$ dereferences x as a value with use κ', the currently available use κ of x should be greater than or equal to κ': otherwise, the application causes an error. If $\kappa = \kappa' = 1$, the closure is deallocated after the application and it can no longer be called; so, the use of x should be changed to 0. It is expressed by the subtraction $\kappa - \kappa'$, whose definition is given in Figure 1.

$$\frac{\kappa \geq \kappa'}{(H\{x \mapsto^\kappa \lambda y.M\}, C[apply^{\kappa'}(x, V)]) \rightsquigarrow (H\{x \mapsto^{\kappa-\kappa'} \lambda y.M\}, C[[V/y]M])} \qquad \text{(R-App)}$$

$$\frac{\kappa < \kappa'}{(H\{x \mapsto^\kappa \lambda y.M\}, C[apply^{\kappa'}(x, V)]) \rightsquigarrow \mathbf{Error}} \qquad \text{(R-App-ERROR)}$$

Similarly, access to a pair $\mathbf{fst}^{\kappa'}(x)$ or $\mathbf{snd}^{\kappa'}(x)$ succeeds only if $\kappa \geq \kappa'$, and the use is reduced by κ' after the access.

$$\frac{\kappa \geq \kappa'}{(H\{x \mapsto^\kappa \langle V_1, V_2\rangle\}, C[\mathbf{fst}^{\kappa'}(x)]) \rightsquigarrow (H\{x \mapsto^{\kappa-\kappa'} \langle V_1, V_2\rangle\}, C[V_1])} \qquad \text{(R-Fst)}$$

$$\frac{\kappa < \kappa'}{(H\{x \mapsto^\kappa \langle V_1, V_2\rangle\}, C[\mathbf{fst}^{\kappa'}(x)]) \rightsquigarrow \mathbf{Error}} \qquad \text{(R-Fst-Error)}$$

$$\frac{\kappa \geq \kappa'}{(H\{x \mapsto^\kappa \langle V_1, V_2\rangle\}, C[\mathbf{snd}^{\kappa'}(x)]) \rightsquigarrow (H\{x \mapsto^{\kappa-\kappa'} \langle V_1, V_2\rangle\}, C[V_2])} \qquad \text{(R-Snd)}$$

$\kappa_1 \backslash \kappa_2$	0	δ	1	ω
0	0	undef	undef	undef
δ	δ	δ	undef	undef
1	1	1	0	undef
ω	ω	ω	ω	ω

Figure 1: Definition of $\kappa_1 - \kappa_2$

$$\frac{\kappa < \kappa'}{(H\{x \mapsto^{\kappa} \langle V_1, V_2 \rangle\}, C[\mathbf{snd}^{\kappa'}(x)]) \rightsquigarrow \mathbf{Error}} \qquad \text{(R-Snd-Error)}$$

The following are the rules for conditionals.

$$(H, C[\mathbf{if}\ 0\ \mathbf{then}\ M_1\ \mathbf{else}\ M_2\ ]) \rightsquigarrow (H, C[M_1]) \qquad \text{(R-IfT)}$$

$$\frac{n \neq 0}{(H, C[\mathbf{if}\ n\ \mathbf{then}\ M_1\ \mathbf{else}\ M_2\ ]) \rightsquigarrow (H, C[M_2])} \qquad \text{(R-IfF)}$$

Example 2.7: The expression in Example 2.3 is reduced as follows:

$$\begin{array}{ll} & (\{\}, \mathbf{let}\ x = \langle 1,2 \rangle^1\ \mathbf{in\ let}\ y = \mathbf{fst}^{\delta}(x)\ \mathbf{in\ let}\ z = \mathbf{snd}^1(x)\ \mathbf{in}\ z) \\ \rightsquigarrow & (\{u \mapsto^1 \langle 1,2 \rangle\}, \mathbf{let}\ y = \mathbf{fst}^{\delta}(u)\ \mathbf{in\ let}\ z = \mathbf{snd}^1(u)\ \mathbf{in}\ z) \\ \rightsquigarrow & (\{u \mapsto^1 \langle 1,2 \rangle\}, \mathbf{let}\ y = 1\ \mathbf{in\ let}\ z = \mathbf{snd}^1(u)\ \mathbf{in}\ z) \\ \rightsquigarrow & (\{u \mapsto^1 \langle 1,2 \rangle\}, \mathbf{let}\ z = \mathbf{snd}^1(u)\ \mathbf{in}\ z) \\ \rightsquigarrow & (\{u \mapsto^0 \langle 1,2 \rangle\}, \mathbf{let}\ z = 2\ \mathbf{in}\ z) \\ \rightsquigarrow & (\{u \mapsto^0 \langle 1,2 \rangle\}, 2) \end{array}$$

The heap value u is considered to be deallocated when the use is changed from 1 to 0.

Example 2.8: If the expression in Example 2.3 is wrongly annotated as:

$$\mathbf{let}\ x = \langle 1,2 \rangle^1\ \mathbf{in\ let}\ y = \mathbf{fst}^1(x)\ \mathbf{in\ let}\ z = \mathbf{snd}^{\delta}(x)\ \mathbf{in}\ z,$$

then it is reduced to an error as follows:

$$\begin{array}{ll} & (\{\}, \mathbf{let}\ x = \langle 1,2 \rangle^1\ \mathbf{in\ let}\ y = \mathbf{fst}^1(x)\ \mathbf{in\ let}\ z = \mathbf{snd}^{\delta}(x)\ \mathbf{in}\ z) \\ \rightsquigarrow & (\{u \mapsto^1 \langle 1,2 \rangle\}, \mathbf{let}\ y = \mathbf{fst}^1(u)\ \mathbf{in\ let}\ z = \mathbf{snd}^{\delta}(u)\ \mathbf{in}\ z) \\ \rightsquigarrow & (\{u \mapsto^0 \langle 1,2 \rangle\}, \mathbf{let}\ y = 1\ \mathbf{in\ let}\ z = \mathbf{snd}^{\delta}(u)\ \mathbf{in}\ z) \\ \rightsquigarrow & (\{u \mapsto^0 \langle 1,2 \rangle\}, \mathbf{let}\ z = \mathbf{snd}^{\delta}(u)\ \mathbf{in}\ z) \\ \rightsquigarrow & \mathbf{Error} \end{array}$$

3 Type System

If an expression is wrongly annotated with uses, a run-time error may occur. For example, consider an expression $\mathbf{let}\ x = \langle 1,2 \rangle^0\ \mathbf{in\ let}\ y = \mathbf{fst}^1(x)\ \mathbf{in}\ \cdots$: since $\langle 1,2 \rangle^0$ does not actually allocate the pair in the heap, an error occurs at $\mathbf{fst}^1(x)$. In order to reject such expressions, we introduce a type system. Our type system is a conservative extension of the usual type system for the simply typed λ-calculus in the sense that any expressions well-typed in the usual type system are well typed also in our type system.

3.1 Types

Types are used for expressing how often and in which way values can be used. Each instance of type constructor for product and arrow types is annotated with uses.

Definition 3.1 [types]: The set of types, ranged over by τ, is given by the following syntax.

$$\tau \quad ::= \quad int \mid \tau_1 \times^{\kappa} \tau_2 \mid \tau_1 \rightarrow^{\kappa_1,\kappa_2} \tau_2$$

κ of $\tau_1 \times^{\kappa} \tau_2$ and κ_1 of $\tau_1 \rightarrow^{\kappa_1,\kappa_2} \tau_2$ expresses how often and in which way a pair and a closure can be accessed. We require κ_2 of $\tau_1 \rightarrow^{\kappa_1,\kappa_2} \tau_2$ not to be δ. It represents how often (0, 1, or ω) the closure can be invoked in the sense of previous linear type systems [21, 9]. The closure itself is allocated and deallocated based on κ_1. The reason why we keep κ_2 will become clear when we present a typing rule for closure creations.

Example 3.2: $int \times^1 int$ is the type of a linear pair of integers (i.e., a pair that can be used many times in a restricted manner as a δ-value, and later used at most once in an arbitrary manner). $(int \times^{\delta} int) \rightarrow^{1,\omega} int$ is the type of a linear closure that takes a pair of integers as an argument, uses it locally, and returns an integer. The function may be invoked many times (thus its use is ω in the strict sense), but it can be used only linearly. For example, $\lambda^1 x.(\mathbf{fst}^{\delta}(x) + \mathbf{snd}^{\delta}(x))$ can have this type.

Example 3.3: $int \times^{\omega} (int \times^1 int)$ represents the type of a pair consisting of an integer and a pair of integers. The outer pair may be accessed many times in any way, but the inner pair of integers can be accessed only linearly. Here, unlike in previous linear type systems [21, 8], $int \times^{\omega} (int \times^1 int)$ does not mean that the inner pair can be accessed linearly each time the pair is extracted from the outer pair. So, it is different from the linear logic formula $!(int \times (int \times int))$. This departure from the linear logic is important for accurate use analysis. For example, consider an expression
let $f = \lambda^{\omega} z.(\mathbf{fst}^{\delta}(x) + z)$ **in let** $g = \mathbf{fst}^1(\mathbf{snd}^{\delta}(x))$ **in** M. If x does not appear in M, it can be assigned the type $int \times^{\omega} (int \times^1 int)$; so, the pair of integers can be deallocated at $\mathbf{fst}^1$.

3.2 Type judgment

A type judgement is of the form $\Gamma \vdash M : \tau$, where Γ (called a type environment) is a mapping from a finite set of variables to types. It expresses how each heap value is accessed during evaluation of M. For example, $x : int \times^{\delta} int \vdash M : int \times^1 int$ implies (1)M may access the heap address x, (2)the result is a linear pair of integers, and (3)x is used up in M and does not escape through the result.

Notation 3.4: We write $V_1 : \tau_1, \ldots, V_n : \tau_n$ for the type environment Γ such that $dom(\Gamma) = \{V_1, \ldots, V_n\} \cap \mathcal{V}$ and $\Gamma(V_i) = \tau_i$ for each $V_i \in dom(\Gamma)$, where $\mathcal{V}$ denotes the set of variables. For example, $(x : int, 2 : int)$ denotes the type environment that maps x to int. If $V \notin dom(\Gamma)$, we write $\Gamma, V : \tau$ for the type environment Γ' such that (1) $\Gamma' = dom(\Gamma) \cup (\{V\} \cap \mathcal{V})$ and (2) $\Gamma'(x) = \tau$ if $x \in (\{V\} \cap \mathcal{V})$ and $\Gamma'(x) = \Gamma(x)$ if $x \in dom(\Gamma)$.

3.3 Operations on uses, types, and type environments

In presenting typing rules, we use several operations on uses, types, and type environments given in Figure 2 and 3. As explained in Section 1, the operation $\kappa_1 + \kappa_2$ is used for computing the total use of a value when it is used as κ_1 in one place and as κ_2 in another place. When the order of the uses is statically known, $\kappa_1; \kappa_2$ is used instead. $\lceil \kappa \rceil$ is the least non-δ use that

is greater than or equal to κ. $\kappa_1 \cdot \kappa_2$ is used for computing the total use of a value when the value is used as a κ_2-value κ_1 times. These operations are extended to operations on types and type environments. We often omit $\cdot$ and write $\kappa_1\kappa_2$ for $\kappa_1 \cdot \kappa_2$ and $\kappa\Gamma$ for $\kappa \cdot \Gamma$. We give higher precedence to $\cdot$, $+$, and ; in this order. Motivations for these operations will become clearer when typing rules are given below.

Example 3.5: Let τ_1 be $int \times^\delta (int \times^\delta int)$ and τ_2 be $int \times^1 (int \times^0 int)$. Then,

$$\begin{array}{lcl} \tau_1 + \tau_2 & = & int \times^\omega (int \times^\delta int) \\ \tau_1; \tau_2 & = & int \times^1 (int \times^\delta int) \\ \lceil \tau_1 \rceil & = & int \times^1 (int \times^1 int) \\ \omega \cdot \tau_1 & = & int \times^\omega (int \times^\omega int). \end{array}$$

$$\begin{array}{lcl} (x : \tau_1, y : \tau_1) + (x : \tau_2) & = & x : (\tau_1 + \tau_2), y : \tau_1 \\ & = & x : int \times^\omega (int \times^\delta int), y : int \times^\delta (int \times^\delta int). \end{array}$$

3.4 Typing rules

3.4.1 Variables, constants, and weakening

The rule for variables and constants are given as:

$$V : \tau \vdash V : \tau \qquad \text{(T-Val)}$$

If an expression is well typed under some type environment, then it should also be well typed under a type environment that represents more access capabilities. For example, if $x : int \times^1 int \vdash M : int$, then $x : int \times^\omega int \vdash M : int$, since the former judgment requires x to be used as a linear pair, while the latter does not impose such a requirement. The following rule allows such weakening of an assumption:

$$\frac{\Gamma' \vdash M : \tau \qquad \Gamma' \leq \Gamma}{\Gamma \vdash M : \tau} \qquad \text{(T-Weak)}$$

The relation $\Gamma' \leq \Gamma$, which means "Γ represents more access capabilities than Γ'," is defined in Figure 3.

3.4.2 Allocation of heap values

Let us consider an expression **let** $x = \langle y, z \rangle^\kappa$ **in** M. In M, y and z can be accessed in two ways: either directly by using addresses y and z, or through the pair x (by extracting y and z). If $\Gamma, x : \tau_1 \times^\kappa \tau_2 \vdash M : \tau$, then the access in the former way is represented by the types $\Gamma(y)$ and $\Gamma(z)$, while the access in the latter way is represented by the types τ_1 and τ_2. Therefore, information on the total access is obtained by taking both access into account. For example, if $\Gamma(y) = int \times^1 int$ and $\tau_1 = int \times^1 int$, then the total access of y should be represented by $int \times^\omega int$ (since y may be totally accessed twice). In general, such information on the total access can be computed by using the summation '+' of types and type environments defined in Figure 3. For instance, the sum of $int \times^1 (int \times^0 int)$ and $int \times^1 (int \times^1 int)$ is calculated as $int \times^\omega (int \times^1 int)$. Thus, the rule for the allocation of a pair is given as:

$$\frac{\Gamma, x : \tau_1 \times^\kappa \tau_2 \vdash M : \tau_3 \qquad \kappa \neq \delta}{\Gamma + V_1 : \tau_1 + V_2 : \tau_2 \vdash \textbf{let } x = \langle V_1, V_2 \rangle^\kappa \textbf{ in } M : \tau_3} \qquad \text{(T-Pair)}$$

We require $\kappa \neq \delta$ because there is no use creating a δ-value: since it can never be deallocated, creating a δ-value is the same as creating an ω-value.

Next, the rule for the allocation of a closure is given as:

$$\frac{\Gamma_1, x : \tau_1 \vdash M_1 : \lceil \tau_2 \rceil \qquad \Gamma_2, f : \tau_1 \to^{\kappa_1,\kappa_2} \lceil \tau_2 \rceil \vdash M_2 : \tau_3 \qquad \kappa_1 \neq \delta}{\kappa_2 \cdot \lceil \Gamma_1 \rceil + \Gamma_2 \vdash \mathbf{let}\ f = \lambda^{\kappa_1} x.M_1\ \mathbf{in}\ M_2 : \tau_3} \quad \text{(T-Abs)}$$

Here, $\lceil \tau_2 \rceil$ denotes the least type (w.r.t. $\leq$) that is greater than or equal to τ_2 and does not contain δ at top-level. For example, $\lceil (int \times^{\delta} int \to^{\delta,1} int) \times^{\delta} int \rceil = (int \times^{\delta} int \to^{1,1} int) \times^{1} int$. $\lceil \tau_2 \rceil$ in the premise $\Gamma_1, x : \tau_1 \vdash M_1 : \lceil \tau_2 \rceil$ guarantees that when the closure $\lambda x.M$ is called, it does not return a δ-value as a result. $\Gamma_1, x : \tau_1 \vdash M_1 : \lceil \tau_2 \rceil$ also means that during each call of the closure, heap values are accessed as described by Γ_1. Because the other premise means that the closure will be called κ_2 times, the total access to heap values by M_1 can be represented by κ_2 copies of Γ_1, which is written as $\kappa_2 \cdot \Gamma_1$. Since the access happens only later (not when $\lambda x.M_1$ is evaluated and the closure is created), $\lceil _ \rceil$ is applied to it, so that it does not contain δ-values.

The following rule for recursive functions is similar to (T-Abs) except for the estimation of the use of the created function:

$$\frac{\Gamma_1, x : \tau_1, f : \tau_1 \to^{\kappa_1,\kappa_2} \lceil \tau_2 \rceil \vdash M_1 : \lceil \tau_2 \rceil \qquad \Gamma_2, f : \tau_1 \to^{\kappa_3,\kappa_4} \lceil \tau_2 \rceil \vdash M_2 : \tau_3}{\kappa_4(1+\kappa_2)\lceil \Gamma_1 \rceil + \Gamma_2 \vdash \mathbf{let}\ f = \mathbf{rec}^{\lceil \kappa_3(1+\kappa_1) \rceil}(f, x, M_1)\ \mathbf{in}\ M_2 : \tau_3} \quad \text{(T-Rec)}$$

The total number of calls of the function f is calculated by using the number κ_2 of calls in M_1 and the number κ_4 in M_2. Since each invocation of f results in other κ_2 invocations of f, the total calls are counted as $\kappa_4(1 + \kappa_2 + \kappa_2^2 + \cdots) = \kappa_4(1 + \kappa_2)$. The use of function f itself is similarly calculated by using its uses κ_1 and κ_3 in M_1 and M_2. The ceiling function is just to make sure that the use is not δ.

3.4.3 Let-expressions

The following rule for let-expressions best illustrates how the evaluation order is taken into account in our type system:

$$\frac{\Gamma_1 \vdash M_1 : \lceil \tau_1 \rceil \qquad \Gamma_2, x : \lceil \tau_1 \rceil \vdash M_2 : \tau}{\Gamma_1; \Gamma_2 \vdash \mathbf{let}\ x = M_1\ \mathbf{in}\ M_2 : \tau} \quad \text{(T-Let)}$$

Here, the premise means that heap values are accessed by M_1 as described by Γ_1 and by M_2 as described by $\Gamma_2, x : \lceil \tau_1 \rceil$. In ordinary linear type systems, the total use of heap values in **let** $x = M_1$ **in** M_2 was computed by adding Γ_1 and Γ_2. However, since M_2 is evaluated only after M_1 has been fully evaluated, we estimate the total access of heap values by *sequential composition* $\Gamma_1; \Gamma_2$. So, if $\Gamma_1 = x : int \times^{\delta} int$ and $\Gamma_2 = x : int \times^{1} int$, then the total use is calculated as $\Gamma_1; \Gamma_2 = x : int \times^{\delta;1} int = x : int \times^{1} int$, although x may be used many times. We require the type of M_1 not to contain δ at top-level, in order to make sure that every δ-value represented by Γ_1 is really "used up" during the evaluation of M_1 and does not escape to the rest of computation.

3.4.4 Access to heap values

The following is the rule for closure invocation:

Definition of $\kappa_1 + \kappa_2$

$\kappa_1 \backslash \kappa_2$	0	δ	1	ω
0	0	δ	1	ω
δ	δ	ω	ω	ω
1	1	ω	ω	ω
ω	ω	ω	ω	ω

Definition of $\kappa_1 ; \kappa_2$

$\kappa_1 \backslash \kappa_2$	0	δ	1	ω
0	0	δ	1	ω
δ	δ	δ	1	ω
1	1	ω	ω	ω
ω	ω	ω	ω	ω

Definition of $\kappa_1 \cdot \kappa_2$

$\kappa_1 \backslash \kappa_2$	0	δ	1	ω
0	0	0	0	0
δ	0	δ	δ	ω
1	0	δ	1	ω
ω	0	ω	ω	ω

Definition of $\lceil \kappa \rceil$

κ	0	δ	1	ω
$\lceil \kappa \rceil$	0	1	1	ω

Figure 2: Operations on uses

$$\frac{\Gamma_1, x : \tau_1 \vdash M : \tau_2 \qquad \kappa > 0}{(f : \tau_3 \to^{\kappa,1} \tau_1 + V : \tau_3); \Gamma_1 \vdash \mathbf{let}\ x = \mathit{apply}^{\kappa}(f, V)\ \mathbf{in}\ M : \tau_2} \quad \text{(T-App)}$$

The access to heap values by $\mathit{apply}^{\kappa}(f, V)$ is estimated as $f : \tau_3 \to^{\kappa,1} \tau_1 + V : \tau_3$. Note that the first use κ should not be zero, and that the second use of the type of f is 1, since the function is used once here. The access of heap values except for x by M is estimated by Γ_1. Since $\mathit{apply}^{\kappa}(f, V)$ is evaluated *before* M is evaluated, the total access can be estimated as their sequential composition $(f : \tau_3 \to^{\kappa,1} \tau_1 + V : \tau_3); \Gamma_1$.

The rule for reading the first element of a pair is given as:

$$\frac{\Gamma, y : \tau_1', x : \tau_1 \times^{\kappa'} \tau_2 \vdash M : \tau_3 \qquad \kappa > 0}{\Gamma, x : (\tau_1 + \tau_1') \times^{\kappa;\kappa'} \tau_2 \vdash \mathbf{let}\ y = \mathbf{fst}^{\kappa}(x)\ \mathbf{in}\ M : \tau_3} \quad \text{(T-Fst)}$$

The premise implies that the pair is used as a κ'-value in M after being used as a κ-value in $\mathbf{fst}^{\kappa}(x)$. So, the total use of x is given by $\kappa; \kappa'$. The first element may be accessed through either y or x, hence the total use of the first element is estimated as $\tau_1 + \tau_1'$.

Similarly, the rule for extraction of the second element of the pair is given as:

$$\frac{\Gamma, y : \tau_2', x : \tau_1 \times^{\kappa'} \tau_2 \vdash M : \tau_3 \qquad \kappa > 0}{\Gamma, x : \tau_1 \times^{\kappa;\kappa'} (\tau_2 + \tau_2') \vdash \mathbf{let}\ y = \mathbf{snd}^{\kappa}(x)\ \mathbf{in}\ M : \tau_3} \quad \text{(T-Snd)}$$

3.4.5 Rules for conditionals

In the following rule for conditional expressions, we require that the then-part and the else-part has the same typing.

$$\frac{\Gamma \vdash M_1 : \tau_1 \qquad \Gamma \vdash M_2 : \tau_1}{V : \mathit{int} + \Gamma \vdash \mathbf{if0}\ V\ \mathbf{then}\ M_1\ \mathbf{else}\ M_2 : \tau_1} \quad \text{(T-If)}$$

- Relation $\leq$

$$
\begin{array}{l}
int \leq int \\
(\tau_1 \times^{\kappa} \tau_2) \leq (\tau_1' \times^{\kappa'} \tau_2') \text{ if } (\tau_1 \leq \tau_1' \wedge \tau_2 \leq \tau_2' \wedge \kappa \leq \kappa') \\
(\tau_1 \rightarrow^{\kappa_1,\kappa_2} \tau_2) \leq (\tau_1 \rightarrow^{\kappa_1',\kappa_2'} \tau_2) \text{ if } (\kappa_1 \leq \kappa_1' \wedge \kappa_2 \leq \kappa_2') \\
\Gamma \leq \Gamma' \text{ if } dom(\Gamma) \subseteq dom(\Gamma') \text{ and } \Gamma(x) \leq \Gamma'(x) \text{ for each } x \in dom(\Gamma).
\end{array}
$$

- Binary operations ($\mathbf{op} = +, ;$)

$$
\begin{array}{lcl}
int \mathbf{op} int & = & int \\
(\tau_1 \times^{\kappa} \tau_2)\mathbf{op}(\tau_1' \times^{\kappa'} \tau_2') & = & \left\{ \begin{array}{l} (\tau_1 \mathbf{op} \tau_1') \times^{\kappa \mathbf{op} \kappa'} (\tau_2 \mathbf{op} \tau_2') \\ \qquad \text{if } \tau_1 \mathbf{op} \tau_1' \text{ and } \tau_2 \mathbf{op} \tau_2' \text{ are well defined.} \\ \text{undefined} \qquad \text{otherwise} \end{array} \right. \\
(\tau_1 \rightarrow^{\kappa_1,\kappa_2} \tau_2)\mathbf{op}(\tau_1 \rightarrow^{\kappa_1',\kappa_2'} \tau_2) & = & \tau_1 \rightarrow^{\kappa_1 \mathbf{op} \kappa_1', \kappa_2 \mathbf{op} \kappa_2'} \tau_2 \\
(\Gamma_1 \mathbf{op} \Gamma_2)(x) & = & \left\{ \begin{array}{ll} \Gamma_1(x)\mathbf{op}\Gamma_2(x) & \text{if } x \in dom(\Gamma_1) \cap dom(\Gamma_2) \\ \Gamma_1(x) & \text{if } x \in dom(\Gamma_1) \backslash dom(\Gamma_2) \\ \Gamma_2(x) & \text{if } x \in dom(\Gamma_2) \backslash dom(\Gamma_1) \end{array} \right.
\end{array}
$$

- Unary operations ($\mathbf{op} = \lceil _ \rceil, \kappa \cdot _, \lfloor _ \rfloor$)

$$
\begin{array}{l}
\mathbf{op}(int) = int \\
\mathbf{op}(\tau_1 \times^{\kappa} \tau_2) = \mathbf{op}(\tau_1) \times^{\mathbf{op}(\kappa)} \mathbf{op}(\tau_2) \\
\mathbf{op}(\tau_1 \rightarrow^{\kappa_1,\kappa_2} \tau_2) = \tau_1 \rightarrow^{\mathbf{op}(\kappa_1),\mathbf{op}(\kappa_2)} \tau_2 \\
(\mathbf{op}(\Gamma))(x) = \mathbf{op}(\Gamma(x))
\end{array}
$$

Figure 3: Definitions of a relation and operations on types and type environments

Example 3.6: A type judgment

$$\emptyset \vdash \mathbf{let}\ x = \langle 1, 2\rangle^1\ \mathbf{in\ let}\ y = \mathbf{fst}^\delta(x)\ \mathbf{in\ let}\ z = \mathbf{snd}^1(x)\ \mathbf{in}\ z : \mathit{int}$$

is obtained from the following derivation:

$$\dfrac{\dfrac{\dfrac{\dfrac{}{z : \mathit{int} \vdash z : \mathit{int}}\ (\text{T-Weak})}{x : \mathit{int} \times^0 \mathit{int}, y : \mathit{int}, z : \mathit{int} \vdash z : \mathit{int}}\ (\text{T-Snd})}{x : \mathit{int} \times^1 \mathit{int}, y : \mathit{int} \vdash \mathbf{let}\ z = \mathbf{snd}^1(x)\ \mathbf{in}\ z : \mathit{int}}\ (\text{T-Fst})}{x : \mathit{int} \times^1 \mathit{int} \vdash \mathbf{let}\ y = \mathbf{fst}^\delta(x)\ \mathbf{in\ let}\ z = \mathbf{snd}^1(x)\ \mathbf{in}\ z : \mathit{int}}\ (\text{T-Fst})}{\emptyset \vdash \mathbf{let}\ x = \langle 1, 2\rangle^1\ \mathbf{in\ let}\ y = \mathbf{fst}^\delta(x)\ \mathbf{in\ let}\ z = \mathbf{snd}^1(x)\ \mathbf{in}\ z : \mathit{int}}\ (\text{T-Pair})$$

4 Type soundness

In this section, we show that no well-typed closed expression is reduced to **Error**, i.e., that no invalid heap access occurs during evaluation of well-typed closed expressions.

We want to show it, as usual [21, 8], by defining a typing system for a run-time state (H, M) and proving that the reduction relation preserves typing and that no well-typed state (H, M) causes an error immediately. Unfortunately, however, the reduction relation $\rightsquigarrow$ defined in Section 2.3 is not suitable for this purpose: a key feature of our type system is to capture information on the order of heap access (recall (T-Let), for example) , whereas the flat representation of a heap loses that information. Therefore, we define an alternative operational semantics that represents a heap space by using nested letrec-expressions, and show the soundness of the type system with respect to that semantics. Although it includes rather strange reduction rules, it is easy to see that the alternative semantics is essentially equivalent to the ordinary semantics given in Section 2.3.

In Section 4.1, we define the alternative semantics and show a theorem that intuitively says: (1) there is a correspondence between the reduction in the ordinary semantics and that in the alternative semantics, and (2) if an expression is reduced to **Error** in the ordinary semantics, then it is also reduced to **Error** in the alternative semantics. These properties imply that if an expression is not reduced to **Error** in the alternative semantics, then it is not either in the ordinary semantics and the result of evaluation is the same. Section 4.2 is devoted to showing that no well-typed expression indeed is reduced to **Error** in the alternative semantics.

4.1 Alternative Operational Semantics

Before presenting an alternative semantics, we explain why the reduction relation $\rightsquigarrow$ fails to preserve typing. Let us consider the following configuration:

$$(\{w \mapsto^1 \langle 1, 2\rangle, x \mapsto^1 \langle w, 3\rangle\}, \mathbf{let}\ v = (\mathbf{let}\ y = \mathbf{fst}^\delta(x)\ \mathbf{in\ let}\ z = \mathbf{fst}^\delta(y)\ \mathbf{in}\ z)\ \mathbf{in}\ M).$$

Then, $x : (\mathit{int} \times^\delta \mathit{int}) \times^\delta \mathit{int} \vdash \mathbf{let}\ y = \mathbf{fst}^\delta(x)\ \mathbf{in\ let}\ z = \mathbf{fst}^\delta(y)\ \mathbf{in}\ z : \mathit{int}$. So, if $x : (\mathit{int} \times^1 \mathit{int}) \times^1 \mathit{int} \vdash M : \mathit{int}$, then by (T-Let),

$$x : (\mathit{int} \times^1 \mathit{int}) \times^1 \mathit{int} \vdash \mathbf{let}\ v = (\mathbf{let}\ y = \mathbf{fst}^\delta(x)\ \mathbf{in\ let}\ z = \mathbf{fst}^\delta(y)\ \mathbf{in}\ z)\ \mathbf{in}\ M : \mathit{int},$$

since $((\mathit{int} \times^\delta \mathit{int}) \times^\delta \mathit{int}); ((\mathit{int} \times^1 \mathit{int}) \times^1 \mathit{int}) = (\mathit{int} \times^1 \mathit{int}) \times^1 \mathit{int}$. So it is fine that w is a linear pair. However, if we reduce the configuration by (R-Fst), the resulting configuration:

$$(\{w \mapsto^1 \langle 1, 2\rangle, x \mapsto^1 \langle w, 3\rangle\}, \mathbf{let}\ v = \mathbf{let}\ z = \mathbf{fst}^\delta(w)\ \mathbf{in}\ z\ \mathbf{in}\ M)$$

requires w to be an ω-pair: w is accessed in M (through x) as a linear pair and in $\mathbf{fst}^{\delta}(w)$ as a δ-pair, and the type system cannot capture the order between the two uses (because the access is made through different variables x and w).

One way to avoid the above problem would be to extend the type system so that an order of access to different variables can be expressed, as in [10]. Since the resulting typing rule seems to become rather complex, we instead redefine the operational semantics. The idea is, before losing information on the order of heap access, to split the accessed heap space in advance by taking the order into account. For example, the above configuration is first converted to something like:

$$\begin{array}{l}\mathbf{let}\ v = (\{w \mapsto^{\delta} \langle 1,2\rangle, x \mapsto^{\delta} \langle w,3\rangle\}, \mathbf{let}\ y = \mathbf{fst}^{\delta}(x)\ \mathbf{in}\ \mathbf{let}\ z = \mathbf{fst}^{\delta}(y)\ \mathbf{in}\ z) \\ \mathbf{in}\ (\{w \mapsto^{1} \langle 1,2\rangle, x \mapsto^{1} \langle w,3\rangle\}, M).\end{array}$$

Since now the heap space has been split for the definition part and for the body part of the let-expression, we can safely throw away information that the definition part is evaluated before the body part. Thus it can be reduced to:

$$\mathbf{let}\ v = (\{w \mapsto^{\delta} \langle 1,2\rangle, x \mapsto^{\delta} \langle w,3\rangle\}, \mathbf{let}\ z = \mathbf{fst}^{\delta}(w)\ \mathbf{in}\ z)\ \mathbf{in}\ (\{w \mapsto^{1} \langle 1,2\rangle, x \mapsto^{1} \langle w,3\rangle\}, M).$$

In order to express the above kind of *nested* heaps and expressions, we introduce a new class of expressions called *dynamic expressions* in Section 4.1.1. We then define a reduction relation for dynamic expressions in Section 4.1.2 and discuss correspondence between the original semantics and the new semantics in Section 4.1.3.

4.1.1 Dynamic Expressions

Dynamic expressions given below are the same as static expressions except for letrec-expressions of the form $\mathbf{letrec}\ x = h^{\kappa}\ \mathbf{in}\ D$. $\mathbf{letrec}\ x = h^{\kappa}\ \mathbf{in}\ D$ means that the heap value h with use κ is available at the address x in the heap; D represents other heap values and the current expression to be evaluated.

Definition 4.1 [dynamic expressions]: The set $\mathcal{D}$ of dynamic expressions, ranged over by D, is given by the following syntax:

$$\begin{array}{rcl} D & ::= & V \mid \mathbf{let}\ x = \langle V_1, V_2\rangle^{\kappa}\ \mathbf{in}\ M \mid \mathbf{let}\ x = \lambda^{\kappa} y.M_1\ \mathbf{in}\ M_2 \\ & & \mid \mathbf{let}\ x = \mathbf{fst}^{\kappa}(y)\ \mathbf{in}\ M \mid \mathbf{let}\ x = \mathbf{snd}^{\kappa}(y)\ \mathbf{in}\ M \\ & & \mid \mathbf{let}\ x = \mathbf{rec}^{\kappa}(f, y, M_1)\ \mathbf{in}\ M_2 \mid \mathbf{let}\ x = apply^{\kappa}(y, V)\ \mathbf{in}\ M \\ & & \mid \mathbf{if0}\ V\ \mathbf{then}\ M_1\ \mathbf{else}\ M_2 \mid \mathbf{let}\ x = D_1\ \mathbf{in}\ D_2 \\ & & \mid \mathbf{letrec}\ x = \langle V_1, V_2\rangle^{\kappa}\ \mathbf{in}\ D \mid \mathbf{letrec}\ x = \lambda^{\kappa} y.M\ \mathbf{in}\ D \\ V & ::= & n \mid x \end{array}$$

Here, unlike x in a let-expression $\mathbf{let}\ x = M_1\ \mathbf{in}\ M_2$, we regard x in a letrec-expression $\mathbf{letrec}\ x = h^{\kappa}\ \mathbf{in}\ D$ as a *free* variable.

Example 4.2:

$$\mathbf{letrec}\ x = \langle 1,2\rangle^{\omega}\ \mathbf{in}\ \mathbf{letrec}\ f = \lambda^{1} y.(\mathbf{let}\ z = \mathbf{fst}^{1}(x)\ \mathbf{in}\ z)\ \mathbf{in}\ \mathbf{let}\ w = apply^{1}(f, 0)\ \mathbf{in}\ w$$

represents the configuration:

$$(\{x \mapsto^{\omega} \langle 1,2\rangle, f \mapsto^{1} \lambda y.(\mathbf{let}\ z = \mathbf{fst}^{1}(x)\ \mathbf{in}\ z)\}, \mathbf{let}\ w = apply^{1}(f, 0)\ \mathbf{in}\ w).$$

Example 4.3: Heaps and expressions are nested in the following expression:

$$\mathbf{let}\ x = (\mathbf{letrec}\ y = \langle 1,2\rangle^{\delta}\ \mathbf{in}\ \mathbf{let}\ z = \mathbf{fst}^{\delta}(y)\ \mathbf{in}\ z)\ \mathbf{in}\ \mathbf{letrec}\ y = \langle 1,2\rangle^{1}\ \mathbf{in}\ \mathbf{let}\ w = \mathbf{snd}^{\delta}(y)\ \mathbf{in}\ w$$

It means that the expression to be evaluated is
$\mathbf{let}\ x = (\mathbf{let}\ z = \mathbf{fst}^{\delta}(y)\ \mathbf{in}\ z)\ \mathbf{in}\ \mathbf{let}\ w = \mathbf{snd}^{\delta}(y)\ \mathbf{in}\ w$, and when $\mathbf{let}\ z = \mathbf{fst}^{\delta}(y)\ \mathbf{in}\ z$ is evaluated, y is viewed as a δ-value while it is viewed as a linear value when $\mathbf{let}\ w = \mathbf{snd}^{\delta}(y)\ \mathbf{in}\ w$ is evaluated.

4.1.2 Reduction Semantics

We define an operational semantics by using a reduction relation $D \xrightarrow{l} E$. D is a dynamic expression and E is either a dynamic expression or a special constant **Error** representing invalid heap access. The label l shows which heap value is accessed in the reduction step. It is either ϵ, which indicates that an internal heap value is accessed, or $x = h^{\kappa}$, which indicates that the heap address x is accessed as a value with use κ, or ϵ', which indicates that a heap is splitted as explained above.

The following three are the rules for heap allocation, which correspond to the rules (R-Pair), (R-Abs), and (R-Fix).

$$\mathbf{let}\ x = \langle V_1, V_2\rangle^{\kappa}\ \mathbf{in}\ M \xrightarrow{\epsilon} \mathbf{letrec}\ z = \langle V_1, V_2\rangle^{\kappa}\ \mathbf{in}\ [z/x]M\ (z \text{ fresh}) \quad \text{(DR-Pair)}$$

$$\mathbf{let}\ x = \lambda^{\kappa} y.M_1\ \mathbf{in}\ M \xrightarrow{\epsilon} \mathbf{letrec}\ z = \lambda^{\kappa} y.M_1\ \mathbf{in}\ [z/x]M\ (z \text{ fresh}) \quad \text{(DR-Abs)}$$

$$\mathbf{let}\ y = \mathbf{rec}^{\kappa}(f, x, M_1)\ \mathbf{in}\ M_2 \xrightarrow{\epsilon} \mathbf{letrec}\ z = \lambda^{\kappa} x.[z/f]M_1\ \mathbf{in}\ [z/y]M_2\ (z \text{ fresh}) \quad \text{(DR-Fix)}$$

The following three are the rules for heap access. A label is used to record which heap address was accessed as a value of which use.

$$\frac{\kappa > 0}{\mathbf{let}\ u = \mathbf{fst}^{\kappa}(x)\ \mathbf{in}\ M \xrightarrow{x=\langle V_1,V_2\rangle^{\kappa}} [V_1/u]M} \quad \text{(DR-Fst)}$$

$$\frac{\kappa > 0}{\mathbf{let}\ u = \mathbf{snd}^{\kappa}(x)\ \mathbf{in}\ M \xrightarrow{x=\langle V_1,V_2\rangle^{\kappa}} [V_2/u]M} \quad \text{(DR-Snd)}$$

$$\frac{\kappa > 0}{\mathbf{let}\ u = apply^{\kappa}(x, V)\ \mathbf{in}\ M' \xrightarrow{x=\lambda^{\kappa} z.M} \mathbf{let}\ u = [V/z]M\ \mathbf{in}\ M'} \quad \text{(DR-App)}$$

The following two rules check whether each heap access is valid or not. If the use κ of the heap value is less than the access mode κ', then the access is invalid and therefore the expression is reduced to **Error**. Otherwise, the evaluation succeeds and the consumed use κ' is removed from κ.

$$\frac{D \xrightarrow{x=h^{\kappa'}} D' \qquad \kappa \geq \kappa'}{\mathbf{letrec}\ x = h^{\kappa}\ \mathbf{in}\ D \xrightarrow{\epsilon} \mathbf{letrec}\ x = h^{\kappa-\kappa'}\ \mathbf{in}\ D'} \quad \text{(DR-Read)}$$

$$\frac{D \xrightarrow{x=h^{\kappa'}} D' \qquad \kappa' > \kappa}{\mathbf{letrec}\ x = h^{\kappa}\ \mathbf{in}\ D \xrightarrow{\epsilon} \mathbf{Error}} \quad \text{(DR-Error)}$$

The following are the rules for conditional expressions.

$$\textbf{if0}\ 0\ \textbf{then}\ M_1\ \textbf{else}\ M_2 \xrightarrow{\epsilon} M_1 \qquad \text{(DR-IfT)}$$

$$\frac{n \neq 0}{\textbf{if0}\ n\ \textbf{then}\ M_1\ \textbf{else}\ M_2 \xrightarrow{\epsilon} M_2} \qquad \text{(DR-IfF)}$$

The following two rules specify which part of an expression can be evaluated. (DR-Rec) says that if an expression can be reduced, it can also be reduced in the same way even when an irrelevant heap has been added. (DR-Let) says that, in a let-expression, the definition part should be reduced first.

$$\frac{D \xrightarrow{l} D' \qquad x \text{ does not appear in } l}{\textbf{letrec}\ x = h^{\kappa}\ \textbf{in}\ D \xrightarrow{l} \textbf{letrec}\ x = h^{\kappa}\ \textbf{in}\ D'} \qquad \text{(DR-Rec)}$$

$$\frac{D_1 \xrightarrow{l} D_1' \qquad l \text{ is } \epsilon \text{ or } \epsilon'}{\textbf{let}\ x = D_1\ \textbf{in}\ D_2 \xrightarrow{l} \textbf{let}\ x = D_1'\ \textbf{in}\ D_2} \qquad \text{(DR-Let)}$$

The following are the rules for error propagation.

$$\frac{D_1 \xrightarrow{\epsilon} \textbf{Error}}{\textbf{let}\ x = D_1\ \textbf{in}\ D_2 \xrightarrow{\epsilon} \textbf{Error}} \qquad \text{(DR-Error2)}$$

$$\frac{D \xrightarrow{\epsilon} \textbf{Error}}{\textbf{letrec}\ x = h^{\kappa}\ \textbf{in}\ D \xrightarrow{\epsilon} \textbf{Error}} \qquad \text{(DR-Error3)}$$

The following rule (DR-Split) is a key rule. As explained above, before a let-expression is evaluated, a heap value is "split" into two: one for evaluation of the definition-part and the other for evaluation of the body-part.

$$\textbf{letrec}\ x = h^{\kappa_1;\kappa_2}\ \textbf{in let}\ y = D_1\ \textbf{in}\ D_2 \xrightarrow{\epsilon'} \textbf{let}\ y = \textbf{letrec}\ x = h^{\kappa_1}\ \textbf{in}\ D_1\ \textbf{in letrec}\ x = h^{\kappa_2}\ \textbf{in}\ D_2 \qquad \text{(DR-Split)}$$

If the definition-part of a let-expression has been fully evaluated, the split heap values can be merged, with δ-values being thrown away.

$$\begin{aligned}\textbf{let}\ x = (\textbf{letrec}\ \vec{z} = \vec{h}^{\vec{\kappa_1}}\ \textbf{in letrec}\ \vec{w} = \vec{h'}^{\vec{\kappa_3}}\ \textbf{in}\ V)\ \textbf{in}\ (\textbf{letrec}\ \vec{z} = \vec{h}^{\vec{\kappa_2}}\ \textbf{in}\ M)\\ \xrightarrow{\epsilon} \textbf{letrec}\ \vec{z} = \vec{h}^{\lfloor\vec{\kappa_1}\rfloor;\vec{\kappa_2}}\ \textbf{in letrec}\ \vec{w} = \vec{h'}^{\lfloor\vec{\kappa_3}\rfloor}\ \textbf{in}\ [V/x]M\end{aligned} \qquad \text{(DR-Var)}$$

Here, $\textbf{letrec}\ \vec{z} = \vec{h}^{\vec{\kappa}}\ \textbf{in}\ D$ is a shorthand for $\textbf{letrec}\ z_1 = h_1^{\kappa_1}\ \textbf{in} \cdots \textbf{letrec}\ z_n = h_n^{\kappa_n}\ \textbf{in}\ D$

Example 4.4: The expression in Example 2.3 is reduced as follows:

$$\begin{aligned}
&\textbf{let}\ x = \langle 1,2\rangle^1\ \textbf{in let}\ y = \textbf{fst}^{\delta}(x)\ \textbf{in let}\ z = \textbf{snd}^1(x)\ \textbf{in}\ z\\
\xrightarrow{\epsilon}\ &\textbf{letrec}\ x = \langle 1,2\rangle^1\ \textbf{in let}\ y = \textbf{fst}^{\delta}(x)\ \textbf{in let}\ z = \textbf{snd}^1(x)\ \textbf{in}\ z\\
\xrightarrow{\epsilon'}\ &\textbf{let}\ y = (\textbf{letrec}\ x = \langle 1,2\rangle^{\delta}\ \textbf{in}\ \textbf{fst}^{\delta}(x))\ \textbf{in}\ (\textbf{letrec}\ x = \langle 1,2\rangle^1\ \textbf{in let}\ z = \textbf{snd}^1(x)\ \textbf{in}\ z)\\
\xrightarrow{\epsilon}\ &\textbf{let}\ y = (\textbf{letrec}\ x = \langle 1,2\rangle^{\delta}\ \textbf{in}\ 1)\ \textbf{in}\ (\textbf{letrec}\ x = \langle 1,2\rangle^1\ \textbf{in let}\ z = \textbf{snd}^1(x)\ \textbf{in}\ z)\\
\xrightarrow{\epsilon}\ &\textbf{letrec}\ x = \langle 1,2\rangle^{\delta;1}\ \textbf{in let}\ z = \textbf{snd}^1(x)\ \textbf{in}\ z\\
\xrightarrow{\epsilon'}\ &\textbf{let}\ z = (\textbf{letrec}\ x = \langle 1,2\rangle^1\ \textbf{in}\ \textbf{snd}^1(x))\ \textbf{in}\ (\textbf{letrec}\ x = \langle 1,2\rangle^0\ \textbf{in}\ z)\\
\xrightarrow{\epsilon}\ &\textbf{let}\ z = (\textbf{letrec}\ x = \langle 1,2\rangle^0\ \textbf{in}\ 2)\ \textbf{in}\ (\textbf{letrec}\ x = \langle 1,2\rangle^0\ \textbf{in}\ z)\\
\xrightarrow{\epsilon}\ &\textbf{letrec}\ x = \langle 1,2\rangle^0\ \textbf{in}\ 2
\end{aligned}$$

4.1.3 Correspondence between the heap-based operational semantics and the alternative operational semantics

Although the alternative operational semantics given above may look strange, it is essentially equivalent to the heap-based operational semantics. Basically, each rule (DR-XX) corresponds to (R-XX), and the rules (DR-Read), (DR-Rec) and (DR-Let) are introduced instead of using context $C[\ ]$. So, the only special rules are (DR-Split) and (DR-Var). The role of (DR-Split) is just to force the access capability of each heap value to be split into two; so, it may causes more errors (due to inappropriate splitting), but it does not change the result of a successful evaluation. We formally state it in Theorem 4.7 below, after introducing several definitions.

Definition 4.5 [translation of dynamic expressions]: A translation function $(_)^* \in \mathcal{D} \to (\mathcal{H} \times \mathcal{E})$ is defined by:

$$\begin{array}{rcl}
V^* & = & (\{\}, V) \\
(\textbf{let } x = h^\kappa \textbf{ in } M)^* & = & (\{\}, \textbf{let } x = h^\kappa \textbf{ in } M) \\
(\textbf{let } x = \textbf{rec}^\kappa(f, y, M_1) \textbf{ in } M_2)^* & = & (\{\}, \textbf{let } x = \textbf{rec}^\kappa(f, y, M_1) \textbf{ in } M_2) \\
(\textbf{let } x = \textbf{fst}^\kappa(y) \textbf{ in } M)^* & = & (\{\}, \textbf{let } x = \textbf{fst}^\kappa(y) \textbf{ in } M) \\
(\textbf{let } x = \textbf{snd}^\kappa(y) \textbf{ in } M)^* & = & (\{\}, \textbf{let } x = \textbf{snd}^\kappa(y) \textbf{ in } M) \\
(\textbf{let } x = apply^\kappa(y, V) \textbf{ in } M)^* & = & (\{\}, \textbf{let } x = apply^\kappa(y, V) \textbf{ in } M) \\
(\textbf{if0 } V \textbf{ then } M_1 \textbf{ else } M_2)^* & = & (\{\}, \textbf{if0 } V \textbf{ then } M_1 \textbf{ else } M_2) \\
(\textbf{let } x = D_1 \textbf{ in } D_2)^* & = & (H_1; H_2, \textbf{let } x = M_1 \textbf{ in } M_2) \\
 & & \text{where } (H_1, M_1) = D_1^* \wedge (H_2, M_2) = D_2^* \\
(\textbf{letrec } x = h^\kappa \textbf{ in } D)^* & = & (\{x \mapsto^\kappa h\} + H, M) \text{ where } (H, M) = D^*
\end{array}$$

where $H_1; H_2$ and $H_1 + H_2$ are defined by:

$$(H_1; H_2)(x) = \begin{cases} (\kappa_1; \kappa_2, h) & \text{if } H_1(x) = (\kappa_1, h) \text{ and } H_2(x) = (\kappa_2, h) \\ H_1(x) & \text{if } x \in dom(H_1) \text{ and } x \notin dom(H_2) \\ H_2(x) & \text{if } x \in dom(H_2) \text{ and } x \notin dom(H_1) \end{cases}$$

$$(H_1 + H_2)(x) = \begin{cases} H_2(x) & \text{if } x \in dom(H_2) \\ H_1(x) & \text{if } x \in dom(H_1) \text{ and } x \notin dom(H_2) \end{cases}$$

Definition 4.6: The binary relation $\geq$ on heaps is defined by:

$$H_1 \geq H_2 \iff (dom(H_1) = dom(H_2)) \wedge ((H_1(x) = (\kappa_1, h_1) \wedge H_2(x) = (\kappa_2, h_2)) \Rightarrow (h_1 = h_2 \wedge \kappa_1 \geq \kappa_2)).$$

We write $(H_1, M_1) \geq (H_2, M_2)$ if $M_1 \equiv M_2$ and $H_1 \geq H_2$.

The correspondence between the new semantics and the heap-based semantics in Section 2 is formally stated in the theorem below. Intuitively, it means (1) the new semantics and the heap-based semantics are equivalent as long as an error does not occur, and (2) if an error occurs in the heap-based semantics, so in the new semantics (the converse may not hold). We write $D \xrightarrow{\hat{\epsilon}} E$ if $D \xrightarrow{\epsilon'}{}^* \xrightarrow{\epsilon} E$ and write $D \uparrow$ if D is always reduced to an error, i.e., if $D \xrightarrow{\hat{\epsilon}} \textbf{Error}$ and there is no $D' \in \mathcal{D}$ such that $D \xrightarrow{\hat{\epsilon}} D'$. Here, we restrict the syntax of dynamic expressions so that, if D contains an expression of the form $\textbf{let } x = D \textbf{ in letrec } \vec{x} = \vec{h}^{\vec{\kappa}} \textbf{ in } M$, then D is of the form $\textbf{letrec } \vec{x} = \vec{h}^{\vec{\kappa}} \textbf{ in letrec } \vec{y} = \vec{h'}^{\vec{\kappa'}} \textbf{ in } M'$ and $\vec{y}$ do not appear in M. We also require that there is no forward reference, i.e., if $D = \textbf{letrec } x = h^\kappa \textbf{ in } D_1$ and h contains free occurences of y, then D_1 does not contain a sub-expression of the form $\textbf{letrec } y = h'^{\kappa'} \textbf{ in } D_2$.

Theorem 4.7:

1. If $D \xrightarrow{\hat{\epsilon}} D'$, then $D^* \rightsquigarrow (H, M)$ and $(H, M) \geq D'^*$ for some H, M.
2. If $(H, M) \rightsquigarrow \mathbf{Error}$ and $(H, M) \geq D^*$, then $D \uparrow$.
3. If $(H, M) \rightsquigarrow (H', M')$ and $(H, M) \geq D^*$, then either $D \uparrow$ or there exists D' such that $D \xrightarrow{\hat{\epsilon}} D'$ and $(H', M') \geq D'^*$.

Proof: See Appendix A.4. □

4.2 Type Soundness

We show soundness of the type system in this section. By the soundness, we mean that no invalid heap access occurs in evaluation of a well-typed expression. In other words, any well-typed expression is correctly annotated with uses.

By Theorem 4.7 above, it suffices to show the soundness of the type system with respect to the alternative semantics. We show it by proving that reductions in the alternative semantics preserves typing and that no well-typed expression accesses an invalid heap address immediately. We extend the type system for dynamic expressions by introducing the following rules:

$$\frac{x \notin \{V_1, V_2\}}{V_1 : \tau_1 + V_2 : \tau_2 \vdash x = \langle V_1, V_2 \rangle^\kappa : \tau_1 \times^\kappa \tau_2} \quad \text{(DT-Pair)}$$

$$\frac{\Gamma_1, x : \tau_1, f : \tau_1 \to^{\kappa_1, \kappa_2} \lceil \tau_2 \rceil \vdash M_1 : \tau_2 \qquad \kappa \geq \kappa_3(1 + \kappa_1)}{\kappa_4(1 + \kappa_2)\lceil \Gamma_1 \rceil \vdash f = \lambda^\kappa x.M_1 : \tau_1 \to^{\kappa_3, \kappa_4} \lceil \tau_2 \rceil} \quad \text{(DT-Abs)}$$

$$\frac{\Gamma_1 \vdash x = h^\kappa : \tau_1 \qquad \Gamma_2, x : \tau_1 \vdash D : \tau_2}{\Gamma_1 + \Gamma_2 \vdash \mathbf{letrec}\ x = h^\kappa\ \mathbf{in}\ D : \tau_2} \quad \text{(DT-Heap)}$$

The type soundness is guaranteed by the two theorems below. Proofs are given in Appendix A.

Theorem 4.8: If $\Gamma \vdash D : \tau$, then $D \not\xrightarrow{\epsilon} \mathbf{Error}$.

Theorem 4.9 [Subject Reduction]:

- If $\Gamma \vdash D : \tau$ and $D \xrightarrow{\epsilon} D'$, then $\Gamma \vdash D' : \tau$; and
- If $\Gamma \vdash D : \tau$ and $D \xrightarrow{\epsilon'} D'$, then there exists D'' such that $D \xrightarrow{\epsilon'} D''$ and $\Gamma \vdash D'' : \tau$.

The complication of the above statement comes from the fact that even a well-typed dynamic expression may be reduced to an ill-typed expression if a heap value is split in an inapproriate way. For example, $\mathbf{letrec}\ x = \langle 1, 2 \rangle^1\ \mathbf{in}\ \mathbf{let}\ z = \mathbf{snd}^1(x)\ \mathbf{in}\ z$ can be reduced by (DR-Split) to $\mathbf{let}\ z = (\mathbf{letrec}\ x = \langle 1, 2 \rangle^\delta\ \mathbf{in}\ \mathbf{snd}^1(x))\ \mathbf{in}\ (\mathbf{letrec}\ x = \langle 1, 2 \rangle^1\ \mathbf{in}\ z)$, which is not well typed and will be reduced to **Error**. The second statement says that, if heap splitting (the rule (DR-Heap)) can be applied to a well-typed expression, there is always a *good* splitting that preserves the well-typeness of the expression.

Corollary 4.10: If $\emptyset \vdash M : \tau$, then $(\{\}, M) \not\rightsquigarrow^* \mathbf{Error}$.

Proof: Suppose $\emptyset \vdash M : \tau$ and $(\{\}, M) \rightsquigarrow^* (H, M') \rightsquigarrow \mathbf{Error}$. By Theorem 4.7 and Theorem 4.9, there is D such that $M \longrightarrow^* D$, $\emptyset \vdash D : \tau$, $(H, M') \geq D^*$, and $D \uparrow$. $\emptyset \vdash D : \tau$ and $D \uparrow$ contradict with Theorem 4.8 and Theorem 4.9. □

5 Type Reconstruction

We can automatically infer use annotations from unannotated expressions. Since it is done basically in the same way as that for previous linear type systems [9, 21], we explain it only briefly.

The basic idea is to introduce variables ranging over uses and types, and constraints on them, so that we can express the most general typing (principal typing) of an expression. A new type judgment is of the form $\Gamma, C \vdash M : \tau$, where C specifies which set of uses/types each use/type variable (in Γ, M, or τ) can range over. For instance, **let** $x = \mathbf{fst}(y)$ **in** x is typed as: $y : \alpha \times^i \beta, \{i \geq j \geq \delta, \alpha \geq \gamma\} \vdash \mathbf{let}\ x = \mathbf{fst}^j(y)\ \mathbf{in}\ x : \gamma$. It is easy to modify typing rules in Section 3 for the new type judgment form.

Given an unannotated expression M', we can infer use annotations in the following two steps.

1. Compute the most general quadruple Γ, C, M, τ such that $\Gamma, C \vdash M : \tau$ and $\mathit{Erase}(M) = M'$ (here, $\mathit{Erase}(M)$ is a term obtained from M by removing all use annotations).
2. Solve the constraint C and substitute a solution for Γ, M, τ. (We can always compute the least solution by a simple iterative method [9].)

One difference from type reconstruction for previous type systems is that the least solution (i.e., assignment of the least possible uses to use variables) is not necessarily the best solution from the viewpoint of memory management. For example, consider an expression **let** $y = \langle 1, 2\rangle$ **in** (**let** $x = \mathbf{fst}(y)$ **in** x). The least annotation is **let** $y = \langle 1, 2\rangle^1$ **in** (**let** $x = \mathbf{fst}^\delta(y)$ **in** x). In this expression, y cannot be deallocated since the only use of the pair y ($\mathbf{fst}(y)$) is annotated with δ. Therefore, it is better to annotate the expression as **let** $y = \langle 1, 2\rangle^1$ **in** (**let** $x = \mathbf{fst}^1(y)$ **in** x). In general, such a solution can be obtained by computing the least solution and then replacing occurrences of δ with 1 as long as we do not need to change occurrences of 1 in other places into ω. Although there is no guarantee that the resulting solution is the best one, we think that it is enough good a solution in practice.

The computational cost for type reconstruction is exponential in the size of an input in the worst case. The reason is that the number of use variables can be linear in the size of the type of a variable appearing in an input expression and that the size of a type can be exponential in the size of an input expression. For example, if a variable x is required to have type $(\mathit{int} \times \mathit{int}) \times (\mathit{int} \times \mathit{int})$ by the context, its typing is:

$$x : (\mathit{int} \times^{i_1} \mathit{int}) \times^{i_3} (\mathit{int} \times^{i_2} \mathit{int}), \{i_1 \geq i'_1, i_2 \geq i'_2, i_3 \geq i'_3\} \vdash x : (\mathit{int} \times^{i'_1} \mathit{int}) \times^{i'_3} (\mathit{int} \times^{i'_2} \mathit{int}).$$

So, the number of use variables is linear to the size 3 of the type $(\mathit{int} \times \mathit{int}) \times (\mathit{int} \times \mathit{int})$. Next, consider an expression

$$\mathbf{let}\ x_0 = 1\ \mathbf{in}\ \mathbf{let}\ x_1 = \langle x_0, x_0\rangle\ \mathbf{in}\ \mathbf{let}\ x_2 = \langle x_1, x_1\rangle\ \mathbf{in}\ \cdots\ \mathbf{let}\ x_n = \langle x_{n-1}, x_{n-1}\rangle\ \mathbf{in}\ x_n.$$

The size of the type of the expression is exponential in n.

In spite of the above fact, we expect that the reconstruction can be performed efficiently for realistic programs for the following reasons. First, the above problem occurs only when tuple or record types are extraordinarily nested; a huge flat tuple type expression $\mathit{int} \times \mathit{int} \times \cdots \times \mathit{int}$ or an arrow type expression $\tau_1 \to \tau_2$ do not cause any problem. Second, if the computational cost is really problematic for some program, we can save the cost by sacrificing the accuracy of the analysis. For example, we can use a more inaccurate version of the $+$ operator defined by $\tau_1 \times^\kappa \tau_2 + \tau_1 \times^{\kappa'} \tau_2 = \tau_1 \times^{\kappa+\kappa'} \tau_2$, as in [21, 9]; this forces many type expressions to be shared and saves the time and space of the analysis (in fact, the previous analysis in [9] is performed in polynomial time for the monomorphic type system).

6 Extensions

We have so far considered a simply-typed λ-calculus with recursion and pairs. In this section, we discuss two kinds of extensions: one is to introduce polymorphism in order to improve the analysis, and the other is to introduce other kinds of data such as recursive data structures and reference cells.

6.1 Polymorphism

Because the type system presented in Section 3 is monomorphic in both uses and types, the analysis of (quasi-)linear values is often rough. Consider the following expression:

$$\textbf{let } f = \lambda x.\langle x, x\rangle \textbf{ in let } z = \mathit{apply}(f, 2) \textbf{ in let } w = \mathit{apply}(f, 3) \textbf{ in } M.$$

If either z or w is used as an ω-value in M, the other is also forced to be an ω-value, because the code for the pair creation (i.e., the body of the function f) is shared.

In order to avoid the above problem, we can introduce polymorphism on uses. Then, f can be given a type $\forall i \neq \delta.(\mathit{int} \to^{\omega,\omega} (\mathit{int} \times^{i} \mathit{int}))$ and the above expression can be annotated, for example, as

$$\textbf{let } f = \Lambda i.\lambda x.\langle x, x\rangle^{i} \textbf{ in let } z = \mathit{apply}(f[1], 2) \textbf{ in let } w = \mathit{apply}(f[\omega], 3) \textbf{ in } M.$$

Here, uses are explicitly passed as parameters to polymorphic functions.

In general, we need to introduce the following *constrained type scheme*:

Definition 6.1:

$$\begin{array}{l} C \text{ (constraints)} ::= \mathit{true} \mid C_1 \wedge C_2 \mid i \geq u \mid i \neq \delta \mid \alpha \geq \rho \mid \mathit{nonzero}(\alpha) \mid \mathit{nondelta}(\alpha) \\ \rho ::= \alpha \mid \rho_1 + \rho_2 \mid \rho_1; \rho_2 \mid \lceil \rho \rceil \\ \sigma \text{ (type schemes)} ::= \tau \mid \forall \vec{\alpha}, \vec{i} :: C.\sigma \end{array}$$

Here, $\mathit{nonzero}(\alpha)$ means that top-level uses of the type α cannot be 0, and $\mathit{nondelta}(\alpha)$ means that top-level uses of the type α cannot be δ. A new rule for polymorphic let-expressions (used indeed in the prototype analyzer) roughly looks like:

$$\frac{\begin{array}{c}\Gamma_1, C_1 \wedge C_2 \vdash M_1 : \lceil \tau_1 \rceil \qquad \Gamma_2, x : \forall \vec{\alpha}, \vec{i} :: C_2.\lceil \tau_1 \rceil, C_3 \vdash M_2 : \tau_2 \\ \vec{i}, \vec{\alpha} \notin \mathit{CVar}(\Gamma_1, C_1) \qquad \{\vec{i}, \vec{\alpha}\} \subseteq \mathit{Fuv}(C_2)\end{array}}{\Gamma_1; \Gamma_2, C_1 \wedge C_3 \vdash \textbf{let } x = M_1 \textbf{ in } M_2 : \tau_2} \quad (\text{T-LetPoly})$$

Here, $\mathit{Fuv}(C)$ denotes the set of type/use variables appearing in C, and $\mathit{CVar}(\Gamma, C)$ denotes the set of type/use variables that may influence the value of some use/type variable in Γ or C.

6.2 Extension with recursive data structures, reference cells, etc.

6.2.1 Large constants

So far, we have considered only integers as constants. In order to deal with large constants (i.e., constants that cannot be represented in one word and must be allocated in a heap), we need to annotate each allocation of a large constant. For example, allocation of a linear real number 2.0 is annotated as $\textbf{let } x = 2.0^{1} \textbf{ in } \cdots$. The type of a large constant is also annotated with a use. For instance, the type of linear real numbers is represented by real^{1}. The typing rule for real numbers is given as follows:

$$\frac{\Gamma, x : \mathit{real}^{\kappa} \vdash M : \tau}{\Gamma \vdash \textbf{let } x = r^{\kappa} \textbf{ in } M : \tau}$$

6.2.2 Primitive functions

Primitive functions can be treated as constants of polymorphic types, and each occurrence of a primitive function can be annotated with uses. For example, the primitive + on real numbers can be assigned a type

$$\forall i_1, i_2, i_3, i_4 :: \{i_1 \geq \delta, i_2 \geq \delta, i_3 \geq \delta, i_4 \neq \delta\}.((\mathit{real}^{i_1} \times^{i_3} \mathit{real}^{i_2}) \rightarrow^{\omega,\omega} \mathit{real}^{i_4})$$

and can be annotated like: $\mathbf{let}\ x = +[1,1,1,1]\langle y, z\rangle^1\ \mathbf{in}\ \cdots$.

6.2.3 Recursive data structures

Recursive data structures such as lists and trees can be treated in the same way as pairs. The type of a list of values of type τ is expressed as $\tau\ \mathit{list}^\kappa$, where κ expresses the use of each cons cell in the list. For example, $\mathit{real}^1\ \mathit{list}^\omega$ is the type of a list whose cons cell may be accessed many times in an arbitrary manner, but whose element is a real number that can be accessed only linearly. Primitives for lists can be given the following types:

$$\begin{array}{lcl}
:: & : & \forall \alpha, i.((\alpha \times^i \alpha\ \mathit{list}^i) \rightarrow^{\omega,\omega} \alpha\ \mathit{list}^i) \\
\mathit{nil} & : & \forall \alpha, i.\alpha\ \mathit{list}^i \\
\mathit{null} & : & \forall \alpha, i.(\alpha\ \mathit{list}^i \rightarrow^{\omega,\omega} \mathit{bool}) \\
\mathit{hd} & : & \forall \alpha, i :: \{i \geq \delta\}.(\alpha\ \mathit{list}^i \rightarrow^{\omega,\omega} \alpha) \\
\mathit{tl} & : & \forall \alpha, i :: \{i \geq \delta\}.(\alpha\ \mathit{list}^i \rightarrow^{\omega,\omega} \alpha\ \mathit{list}^i)
\end{array}$$

With these types, the function *append* is given the following type:

$$\forall \alpha, \beta, \gamma, i, j, k :: \{\alpha \geq \gamma, \beta \geq \gamma, i \geq \delta, j \geq \delta, k \neq \delta, l \geq \delta.((\alpha\ \mathit{list}^i \times^l \beta\ \mathit{list}^j) \rightarrow^{\omega,\omega} \gamma\ \mathit{list}^k).$$

In general, we can generate the types of constructors and destructors from datatype declarations of Standard ML [14].

6.2.4 Reference Cells

In order to deal with reference cells, we need to introduce a new type constructor. The type of a reference cell is of the form $\tau\ \mathit{ref}^{\kappa_1,\kappa_2}$, where τ is the type of a stored value, κ_1 is the use of the reference cell, and κ_2 denotes how many times each stored value may be read.

The primitives for reference cells are annotated as follows:

$$M ::= \cdots \mid \mathbf{let}\ x = \mathbf{ref}^\kappa(y)\ \mathbf{in}\ M \mid \mathbf{let}\ x = !^\kappa y\ \mathbf{in}\ M \mid \mathbf{let}\ x :=^\kappa V\ \mathbf{in}\ M$$

Here, $\mathbf{ref}^\kappa(y)$ creates a reference cell x and store y into it; x can be accessed as a κ-value afterwards. $!^\kappa y$ reads the value stored in the reference cell y; at the same time, if κ is 1 and the reference y is linear, it deallocates the reference cell. $x :=^\kappa V$ stores the value V into the reference cell x; at the same time, if κ is 1 and the reference y is linear, it deallocates the reference cell.

Typing rules can be given, for example, as follows:

$$\frac{\Gamma, x : \lceil\tau\rceil\ \mathit{ref}^{\kappa_1,\kappa_2} \vdash M : \tau' \qquad \kappa_1 \neq \delta}{\Gamma + y : \kappa_2\lceil\tau\rceil \vdash \mathbf{let}\ x = \mathbf{ref}^{\kappa_1}(y)\ \mathbf{in}\ M : \tau'} \quad \text{(T-Ref)}$$

$$\frac{\Gamma, y : \tau\ \mathit{ref}^{\kappa_1,\kappa_2}, x : \tau \vdash M : \tau' \qquad \kappa > 0}{\Gamma, y : \tau\ \mathit{ref}^{\kappa;\kappa_1,1+\kappa_2} \vdash \mathbf{let}\ x = !^\kappa y\ \mathbf{in}\ M : \tau'} \quad \text{(T-Deref)}$$

$$\frac{\Gamma, x : \tau \ \mathit{ref}^{\kappa_1,\kappa_2} \vdash M : \tau' \qquad \kappa > 0}{\Gamma + V : \kappa_2\tau, x : \tau \ \mathit{ref}^{\kappa;\kappa_1,0} \vdash \mathbf{let}\ x :=^{\kappa} V\ \mathbf{in}\ M : \tau'} \qquad \text{(T-UPDATE)}$$

In the rule (T-REF), $x : \lceil\tau\rceil \ \mathit{ref}^{\kappa_1,\kappa_2}$ means that the created reference cell may be read in M κ_2 times before the cell is updated and that each time the value is read, it is used as a value of type τ. So, the stored value y may be totally used as a value of type $\kappa_2\tau$. A tricky point about the rule (T-UPDATE) is that the read count of x is 0 in the conclusion. This is because the currently stored value is replaced by V and therefore it is not read in $\mathbf{let}\ x :=^{\kappa} V\ \mathbf{in}\ M$ at all.

7 Preliminary Experiments

We have implemented a prototype type reconstruction system for the core-ML (i.e., SML [14] without modules) based on our quasi-linear type system and a simple profiler that executes the output program and counts the number of allocated heap values. We first describe the current status of the system in Section 7.1 and then report the results of simple experiments in Section 7.2. The prototype system will be available from `http://www.yl.is.s.u-tokyo.ac.jp/~koba/research/qlinear/`.

7.1 A prototype type reconstruction system

Following the extensions discussed in Section 6, we have implemented a prototype analyzer. It takes an expression of the core-ML as an input, performs type (and use) reconstruction and produces a use-annotated expression. It has been implemented by using the ML kit version 1 [4] as a front end, and supports full features of the core-ML: records, datatype declarations, references, and exceptions. Polymorphism on types and uses is based on the let-polymorphism discussed in Section 6.

Polymorphic recursion on uses would be sound (just as polymorphic recursion on regions is sound in the region inference [19]), but it is not supported currently. That is because the algorithm would become complex and also because the polymorphic recursion does not seem so critical for our analysis.

The current system has the following limitations. It produces the least use annotation, not the best one (see Section 5); The analysis of the use of a reference cell is less accurate than the one presented in Section 6; Unnecessary uses are passed as parameters to (use-)polymorphic functions (they can be removed in the same manner as unnecessary region parameters are removed in a post-path of the region inference [5]); The current system is very slow for several known reasons (it is just because we preferred rapid prototyping and the system will be optimized in the future); A compiler that utilizes the analyzed information has not been implemented yet.

7.2 Results of preliminary experiments

Table 1 shows the result of experiments. The column 'zero,' 'linear,' and 'omega' respectively show how many heap values of use 0, 1, and ω are allocated dynamically in each program. The rightmost column shows how many percentages of heap values are zero or linear. For a use-polymorphic function like $f : \forall i.(\mathit{real}^i \to^{\omega,\omega} \mathit{real}^i)$, the allocation was counted only once; in other words, creations of instances like $f[1]$ and $f[\omega]$ are not counted as the heap allocation (this is because the current analyzer infers not the use of each instance of a use-polymorphic function but the total use of all the instances).

	zero	linear	omega	(zero+linear)/total
sumlist10000	0	40,001	2	100.0%
qsort20	0	448	4	99.1%
msort20	0	496	4	99.2%
sieve2000	0	256,455	8	100.0%
life	0	2,353,116	26,600	98.9%
Knuth-Bendix	0	10,339,410	2,949,292	77.8%
boyer	0	1,163,786	342,642	77.3%
mandelbrot	0	5,672,170	920,944	86.0%
dangle	20,000	20,032	34	99.9%
reynolds3	0	21,515	2,059	91.3%
reynolds3u	0	21,514	13	99.9%

Table 1: The number of allocated heap values

"sumlist10000" is a program that makes a list of integers from 0 to 10,000 and computes the sum of them. "qsort20" and "msort20" sort a list of 20 real numbers by using the quick sort and merge sort algorithms. "sieve2000" finds prime numbers less than 2,000 by using Eratosthene's sieve algorithm. "life" computes the first 10 generations of lives generated from the initial 44 lives. "dangle" and "reynolds3" are programs taken from [5] (with parameters being changed). "reynolds3u" is a slightly modified version of "reynolds3," in which one function is uncurried so that our analysis works better. The quick sort was written in the following naive way:

```
fun quick ([]) = []
  | quick (x::l) =
        let val (l1,l2)=divide(x,l)
        in append(quick(l1), x::quick(l2)) end
```

Other programs are found at
`http://www.yl.is.s.u-tokyo.ac.jp/~koba/research/qlinear/`.

In the cases of sumlist, merge/quick sorts and Eratosthene's sieve, almost all of the heap values are linear: non-linear values were only top-level functions. Moreover, by looking at the produced use-annotated expression, we can find that the cons cells cells generated in merge/quick sorts and Eratosthene's sieve can all be updated in-place (the application of linear types to the in-place quick sort has been suggested by Baker [2] but a remarkable point here is that it can be performed for naive programs with no programmer's annotation).

In the case of the life (slightly modified from the original version: see the above WWW page and discussions in Section 9), dangle, and reynolds3, not all heap values are linear, but the number of linear values is still huge enough to greatly reduce the need for garbage collection. In the Knuth-Bendix and boyer programs, the percentage of linear values is relatively smaller. This is probably because unification of first-order terms performed in the program causes sharing of many heap values.

One should, however, note that the percentages shown in the rightmost column do not directly indicate how much memory space can be saved by our analysis: compared with memory management using conventional garbage collection, extra memory space is required for representing use-polymorphic functions and its instances, and it is not yet clear how soon linear values can indeed be deallocated. Also, the current analysis is performed on unoptimized programs; other optimizations such as inlining and hoisting may reduce the ratio of linear values.

Therefore, we know for sure the real impact of our analysis only after more serious experiments are carried out (for that purpose, some of the future work described in Section 9 must be completed).

8 Related work

8.1 Previous linear type systems

To our knowledge, among previous linear type systems [21, 9, 3] that can infer the usage of values, only Barendsen and Smetsers's uniqueness typing [3] takes an evaluation order into account. The uniqueness typing [3] is different from ours in many ways: the evaluation order is call-by-need,[4] their type system does not have "δ-type" of values that must be used up locally, and instead it relies on a separate analysis for analyzing the order of memory access (so, the analysis is not integrated as a use-type system, and it is rather complex yet inaccurate in many cases).

Another, rather technical novelty of our analysis is the definition of the operation "+." In the previous type system [8], $\tau_1 + \tau_2$ is defined only if τ_1 and τ_2 are identical except for their outermost uses. This restriction made the analysis inaccurate: for example, if a pair of real numbers is used as $real^1 \times^1 real^0$ in one place and as $real^0 \times^1 real^1$ in another place, both uses are coerced to $real^1 \times^1 real^1$ and the total use was counted as $real^\omega \times^\omega real^\omega$.

For comparison of the accuracy of the analyses, Table 2 shows the least type each analysis can assign to the argument of a function (we require that the result of the function must be at least a linear value). The second column "without δ" shows what happens if we improve [8] by using the definition of + in this paper. So, the difference between the first and second column shows the effect of the improved definition of "+," and the difference between the second and third column shows the pure effect of introducing δ-values. The functions listed in the table are:

```
fun add(x:real) = let val y=x+1.0 in x+y end;
fun swap(z:real*real) = (#2(z), #1(z));
fun twice(f) = f(f(1))
fun append(l1,l2) = if null(l1) then l2
                    else let val x::l1' = l1 in
                           x::append(l1',l2)
fun filter(p, l) = if null(l) then []
                   else let val x::l' = l in
                          if p(x) then x::filter(p,l')
                          else filter(p, l')
                        end
```

As for the uniqueness typing, as far as we can see from the description in the recent language report (p.62 of [18]) of Concurrent Clean, it seems to assign $real^\omega$, $real^\omega \times^1 real^\omega$, and $int \rightarrow^{\omega,\omega} int$ respectively to the arguments of (call-by-value versions of) `add`, `swap`, and `twice`.

Odersky [17] also proposed an extension of a linear type system that allows a linear value to be used many times for non-destructive operations before being used for destructive operations. Unlike ours, destructive operations are assumed to be explicitly specified by a programmer, not inferred automatically.

[4] Recently, they dealt also with strict let expressions, but the analysis for them seems to be more naive than our analysis (p.62 of [18])

	[21, 8]	without δ	this paper
`add`	$real^{\omega}$	$real^{\omega}$	$real^{\delta}$
`swap`	$real^{\omega} \times^{\omega} real^{\omega}$	$real^{1} \times^{\omega} real^{1}$	$real^{1} \times^{\delta} real^{1}$
`twice`	$int \rightarrow^{\omega,\omega} int$	$int \rightarrow^{\omega,\omega} int$	$int \rightarrow^{\delta,\omega} int$
`append`	$real^{\omega}\ list^{\omega} \times^{1} real^{\omega}\ list^{1}$	$real^{1}\ list^{1} \times^{1} real^{1}\ list^{1}$	$real^{1}\ list^{\delta} \times^{1} real^{1}\ list^{1}$
`filter`	$(int \rightarrow^{\omega,\omega} bool) \times^{1} int\ list^{\omega}$	$(int \rightarrow^{\omega,\omega} bool) \times^{1} int\ list^{1}$	$(int \rightarrow^{\delta,\omega} bool) \times^{\delta} int\ list^{\delta}$

Table 2: Comparison with previous linear type systems

8.2 Region inference

There is an alternative approach to static memory management: region inference [19, 5]. The region inference abstracts a bunch of memory addresses as a region, and estimates its life-time by performing a kind of type inference. The region inference infers, in addition to the usual type information, which region each expression may access, and automatically inserts allocation and deallocation of regions.

Our type system and the region inference have both advantages and disadvantages. Because the life-time of data is estimated region-wise, it is difficult to use the region inference for in-place update (although it is possible in some cases by using multiplicity inference and storage mode analysis [5]). Another shortcoming of the region inference is that too many data are often merged into the same region (especially when one uses recursive data structures, higher-order functions, and reference cells), and as a result, the estimation of the life-time of data becomes very rough. On the other hand, the advantages of the region inference are that it is completely independent of how often values are accessed, and that the deallocation of data in the same region can be performed in a constant time. So, it may be interesting to use our type system and the region inference as complementary methods for static memory management.

The goal of our analysis is also similar to that of the storage mode analysis [5], which was proposed as a complementary to the region inference. It analyzes whether each access to a value is the last access to the region of the value, and if so, it inserts a code to deallocate all values in the region. Thus, unlike our analysis, the storage mode analysis can deallocate a value only after all the data in the region become garbage. The storage mode analysis is not based on a type system, and to our knowledge, its correctness has not been proved yet [5].

We give below examples for which both our analysis and the region inference work well, those for which our analysis works well but the region inference does not, and those for which the region inference works well but our analysis does not.

8.2.1 A program for which both our analysis and the region inference work well

Consider the following program `dangle` taken from [5]:

```
fun mklist 0 = []
  | mklist n = n::mklist(n-1);

fun cycle(m,f) =
    if m=0 then (m,f)
    else cycle(m-1, let val x=[(m,mklist 2000)]
                in fn ()=> #1(hd x)+f() end);
```

```
val r = cycle(100, fn()=>0);
```

In both the region inference and our analysis, the list generated by "`mklist 2000`" can be collected before `cycle` is called, although `x` is live at this moment and `x` has a pointer to the list. (In the case of our analysis, the list is not allocated in the first place, since the list creation is annotated with use 0.)

8.2.2 Programs for which our analysis works but the region inference does not

Consider the following program `reynolds3` taken from [5]:

```
datatype 'a tree = Lf | Br of 'a * 'a tree * 'a tree;

fun mk_tree 0 = Lf
  | mk_tree n = let val t=mk_tree(n-1) in Br(n,t,t) end;

fun member(x,[]) = false
  | member(x,x'::rest) = x=x' orelse member(x,rest);

fun search p Lf = false
  | search p (Br(x,t1,t2)) =
     if member(x,p) then true
     else search (x::p) t1 orelse search (x::p) t2;

val it = search [] (mk_tree 20)
```

In the region inference, the cons cell created by "`(x::p)`" can be deallocated only after "`search (x::p) t1 orelse search (x::p) t2`" is fully evaluated, while in our analysis, it can be deallocated after "`search (x::p) t1`" is evaluated.

Consider also the following function `merge`:

```
fun merge ([]:int list, l) = l
  | merge (l, []) = l
  | merge (x::l1, y::l2) = if x<y then x::merge(l1,y::l2)
                           else y::merge(x::l1,l2)
```

In the region inference, no cons cells of the lists $l1$ and $l2$ can be deallocated after `merge`$(l1, l2)$ is computed, because all cons cells are merged into the same region. On the other hand, unnecessary cons cells (cons cells matched to the patterns `x::l1` and `y::l2` in the above program) can be immediately deallocated in our analysis.

Consider another example:

```
let
  val r = ref (fn x=>x+1.0)
in
  ((!r)(2.0), !r)
end
```

In the region inference, 2.0 cannot be deallocated after the above expression is evaluated because it is stored in the region accessed by the function `fn x=>x+1.0`, which is still alive. On the other hand, in our analysis, 2.0 can be deallocated immediately after `(!r)(2.0)` is evaluated, because the type of `!r` is inferred as $real^1 \rightarrow^{\omega,\omega} real^\omega$.

Reasons why the region inference does not work In order to know why the region inference does not work for the above examples, one needs to know ideas of the region inference. In the region inference, a type is extended with information on where a value of the type is stored. For example, $(real, \rho)$ is the type of a real number that is stored in the region ρ. $((real, \rho_1) \rightarrow^{\{\rho_1,\rho_2\}} (real, \rho_2), \rho_3)$ is the type of a function that takes an argument stored in the region ρ_1, accesses the regions ρ_1 and ρ_2, and returns a real number stored in the region ρ_2. The function itself is stored in the region ρ_3. By performing type inference based on such a type system, operations for allocating/deallocating regions can be inserted. For example, if a closed expression e is inferred to have a type $(real, \rho_1)$ and a value is allocated in the region ρ_2 inside e, we know that the value is not alive and can be deallocated after e has been evaluated.

The main reason why the region inference does not work for the first and second example is that all the cons cells of a list must be stored in the same region. Because we must represent each type by using a finite expression, it is impossible to distinguish the region of each cell that constitutes a recursive data structure. So, for example, the type of a list of real numbers must be something like $((real, \rho_1)\ list, \rho_2)$, where ρ_1 is the region in which the elements of the list are stored, and ρ_2 is the region in which the cons cells are stored. Thus, cons cells and elements that are once put into the same list can be deallocated only after all of them become garbage.

The reason for the third example is that without region polymorphism [19], region-annotated types can express only poor dataflow information. A region-polymorphic type:

$$\forall \rho_1, \rho_2.((real, \rho_1) \rightarrow^{\{\rho_1,\rho_2\}} (real, \rho_2), \rho_3)$$

means that a function of that type takes its argument stored in some region ρ_1 and returns a result stored in another region ρ_2; so, it subsumes information represented by a use-annotated type $real^{\delta} \rightarrow^{\omega,\omega} real^{\omega}$ in our type system. However, in the third example, `!r` can be assigned only a monomorphic type: $((real, \rho_1) \rightarrow^{\{\rho_1,\rho_2\}} (real, \rho_2), \rho_3)$. So, if `!r` is used in another place (and indeed this may be the case because `!r` escapes), the argument must be stored in the same region ρ_1, which means that the region ρ_1 can be deallocated only after it becomes certain that there is no more call to the function `!r`.

8.2.3 A program for which the region inference works well but our analysis does not

Consider the following program

```
fun f (x:real) = fn y=> x+y
val it = let val g = f(1.0) in g(2.0)+g(3.0) end
```

In our analysis, the use of `1.0` is judged to be ω and cannot be deallocated (because upon each invocation, `g` uses `x` as a value of use of at least 1, and `g` is called twice). On the other hand, the region inference can deallocate `1.0` after the whole expression is evaluated.

Consider also the following program:

```
fun choose(b, x, y) = if b then x else y
val it = let val z=1.0 in choose(true, z, z)+2.0 end
```

1.0 is inferred to be an ω-value and cannot be deallocated in our analysis, while it can be deallocated in the region inference.

The reason why our analysis does not work for the first example is that the dataflow information represented by the use δ is too naive. The second example reveals a limitation of our analysis in counting value access: without knowing the value of `b`, our analysis can only find that both `x` and `y` may be used at most once (if the result is used once), and therefore, it infers that `z` may be used twice.

8.3 Single-threaded lambda calculus

Guzmán and Hudak's single threaded lambda calculus [7] also seems to be related with our type system. They proposed a type system to check that destructive updates do not destroy the referential transparency of functional languages. For the type system, they introduced a *use*, which is similar to but somewhat more general than our use. Major differences are: their goal is to *check* whether destructive updates specified by a programmer are safe, while (one of) our goal is to *find* places where destructive updates can be safely performed; and (probably because of this difference in the goals) a use is associated to an identifier in their type system in order to express how the entire data (referred to by the identifier) is accessed, while in our type system, a use is associated to each instance of type constructor in order to express how each memory in a heap is accessed.

9 Concluding Remarks

9.1 Summary

We proposed an extended linear type system that can elegantly take an evaluation order into account, and proved its correctness. The new type system can judge more values to be (quasi-)linear than ordinary linear type systems, and thus, it is expected to improve memory space utilization and enable more in-place update of data structures. We have already implemented a prototype type inference system for the core-ML, and obtained promising results.

Static memory management like the one proposed here and the region inference is currently less popular than conventional garbage collection. However, we think it will become very important in the near future. First, static memory management is especially attractive in parallel/distributed environments, since it requires much less communications/synchronizations than conventional garbage collection. Second, with the increasing importance of cache memory, saving the memory space for a program is also likely to save its execution time.

9.2 Future work

Much work is left to be done on the type system. In order to judge a real impact of the type system, we need to carry out more serious benchmarks. In order to implement a compiler, we need to work on the following issues (though they seem fairly straightforward except for the last issue):

- Refinement of the type inference system. As discussed in Section 5, the least use annotation may not be *optimal*; so, we need to modify (the constraint solver part of) the current prototype system so that it can produce a better use annotation.
- Reduction of the cost for use passing. Since passing uses to use-polymorphic functions take some extra cost, we need to reduce it by dropping unnecessary uses (just as unnecessary region parameters in region inference [5]) and by a similar technique to type lifting [20].
- Transformation for in-place updates. We need to implement a compilation path that finds places where deallocation of a heap value is immediately followed by allocation of another value of the same type and replaces it with in-place update.
- Combination of conventional garbage collection with linear-type-based memory management. Since the linear-type-based memory management can automatically deallocate only quasi-linear values, we need a conventional garbage collector as well. There is, however, a

subtle problem: since the linear-type-based memory management is not based on pointer traversal information, it may create a dangling pointer (just as the region-based memory managements does [19]). We are currently working (with Atsushi Igarashi) to deal with this problem along similar lines to tag-free garbage collection [20, 15]. The idea is to utilize use-annotated types for tracing heap values at run-time, instead of conventional types.

Other future work includes applications of our type system to parallel/distributed programming languages, for which linear-type-based static memory management seems much more attractive. As for the theoretical part, we have not yet proved the correctness of the polymorphic version of the type system (though we do not think it is difficult, it would be cumbersome).

Improvement of the accuracy of the analysis is also future work. The major weak points of the current analysis are:

- Weakness in dealing with variable aliasing: Because the type system cannot deal with the order of access to different variables, the analysis can be rough when variables are aliased. For example, consider an expression **let** $y = x$ **in let** $\mathbf{fst}^{\delta}(x)$ **in** $\mathbf{fst}^{1}(y)$. Since x and y refer to the same heap value, we could regard x as a linear value; however, the current type system judges x to be an ω-value.

- Weakness in dealing with curried functions: Consider an expression $(\lambda z.(\mathbf{fst}(z) + 1.0))(2.0, 3.0)$ and its curried version $((\lambda x.\lambda y.(x + 1.0))2.0)3.0$. For the former expression, δ can be assigned to 2.0, while for the latter one, 1 must be assigned to it. This is the reason why we had to modify the original life program for the preliminary experiments as mentioned in Section 7. We rewrote the curried member function (that takes a list as the first parameter) to its uncurried version.

The second problem may not be so serious because the ordinary compiler optimization uncurries functions as far as possible, but it may be interesting to improve the analysis by using more accurate dataflow information like the one obtained by the region inference. We can probably overcome the first problem by representing a type environment as a poset of type bindings in order to express the order of access to different variables, as in an earlier version [10] of Kobayashi's type system for deadlock-freedom [11].

Acknowledgement

We would like to thank Atsushi Igarashi and Mads Tofte for useful discussions and comments.

References

[1] Henry G. Baker. Lively linear lisp – look ma, no garbage! *ACM Sigplan Notices*, 27(8):89–98, 1992.

[2] Henry G. Baker. A linear logic quicksort. *ACM Sigplan Notices*, 29(2):13–18, 1994.

[3] Erik Barendsen and Sjaak Smetsers. Conventional and uniqueness typing in graph rewrite systems. Technical Report CSI-R9328, Computer Science Institute, University of Nijmegen, 1993. An extended abstract appeared in Proc. of FST&TCS 12, Springer LNCS 761, pp.41-51.

[4] Lars Birkedal, Nick Rothwell, Mads Tofte, and David N. Turner. The ML Kit (Version 1). Technical Report 93/14, Department of Computer Science, University of Copenhagen, 1993.

[5] Lars Birkedal, Mads Tofte, and Magnus Vejlstrup. From region inference to von neumann machines via region representation inference. In *Proceedings of ACM SIGPLAN/SIGACT Symposium on Principles of Programming Languages*, pages 171–183, 1996.

[6] Jean-Yves Girard. Linear logic. *Theoretical Computer Science*, 50:1–102, 1987.

[7] Juan G. Guzmán and Paul Hudak. Single-threaded polymorphic lambda calculus. In *Proceedings of IEEE Symposium on Logic in Computer Science*, pages 333–343, 1990.

[8] Atsushi Igarashi. Type-based analysis of usage of values for concurrent programming languages. Master's thesis, Department of Information Science, University of Tokyo, 1997.

[9] Atsushi Igarashi and Naoki Kobayashi. Type-based analysis of usage of communication channels for concurrent programming languages. In *Proceedings of International Static Analysis Symposium (SAS'97)*, volume 1302 of *Lecture Notes in Computer Science*, pages 187–201. Springer-Verlag, 1997.

[10] Naoki Kobayashi. A Partially Deadlock-free Typed Process Calculus (I) – A Simple System –. Technical Report 96-02, Department of Information Science, University of Tokyo, September 1996.

[11] Naoki Kobayashi. A partially deadlock-free typed process calculus. *ACM Transactions on Programming Languages and Systems*, 20(2):436–482, 1998. A preliminary summary appeared in Proceedings of LICS'97, pages 128–139.

[12] Naoki Kobayashi, Benjamin C. Pierce, and David N. Turner. Linearity and the pi-calculus. In *Proceedings of ACM SIGPLAN/SIGACT Symposium on Principles of Programming Languages*, pages 358–371, January 1996.

[13] Ian Mackie. Lilac : A functional programming language based on linear logic. *Journal of Functional Programming*, 4(4):1–39, October 1994.

[14] Robin Milner, Mads Tofte, Robert Harper, and David MacQueen. *The Definition of Standard ML (Revised)*. The MIT Press, 1997.

[15] Greg Morrisett. *Compiling with Types*. PhD thesis, School of Computer Science, Carnegie Mellon University, 1995.

[16] Greg Morrisett, Matthias Felleisen, and Robert Harper. Abstract models of memory management. In *Proceedings of Functional Programming Languages and Computer Architecture*, pages 66–76, 1995.

[17] Martin Odersky. Observers for linear types. In *Proceedings of 4th European Symposium on Programming (ESOP'92)*, volume 582 of *Lecture Notes in Computer Science*, pages 390–407. Springer-Verlag, 1992.

[18] Rinus Plasmeijer and Marko van Eekelen. Concurrent Clean ver.1.3 language report. available from `http://www.cs.kun.nl/~clean/`, 1997.

[19] Mads Tofte and Jean-Pierre Talpin. Implementing the call-by-value lambda-calculus using a stack of regions. In *Proceedings of ACM SIGPLAN/SIGACT Symposium on Principles of Programming Languages*, pages 188–201, 1994.

[20] Andrew Tolmach. Tag-free garbage collection using explicit type parameters. In *Proceedings of ACM Conference on Lisp and Functional Programming*, pages 1–11, 1994.

[21] David N. Turner, Philip Wadler, and Christian Mossin. Once upon a type. In *Functional Programming Languages and Computer Architecture*, San Diego, California, 1995.

A Proofs

A.1 Laws on operations on uses, types, and type environments

We first list several laws on operations defined in Section 3. They are used frequently in the proofs below.

Lemma A.1:

1. The operation + (on uses, types, and type environments) is associative and commutative.
2. The operation ; (on uses, types, and type environments) is associative.
3. The operation $\cdot$ on uses is associative and commutative.
4. The unary operations $\lceil_\rceil$ and $\lfloor_\rfloor$ are monotonic.
5. $\kappa_1 + \kappa_2 \geq \kappa_1;\kappa_2$, $\tau_1 + \tau_2 \geq \tau_1;\tau_2$, and $\Gamma_1 + \Gamma_2 \geq \Gamma_1;\Gamma_2$.
6. $(\kappa_1+\kappa_2)-\kappa_3 \geq (\kappa_1-\kappa_3)+\kappa_2$, $(\tau_1+\tau_2)-\tau_3 \geq (\tau_1-\tau_3)+\tau_2$, and $(\Gamma_1+\Gamma_2)-\Gamma_3 \geq (\Gamma_1-\Gamma_3)+\Gamma_2$.
7. $(\kappa_1;\kappa_2) - \kappa_1 \geq \kappa_2$, $(\tau_1;\tau_2) - \tau_1 \geq \tau_2$, and $(\Gamma_1;\Gamma_2) - \Gamma_1 \geq \Gamma_2$.
8. $(\kappa_1;\kappa_2)\cdot\kappa_3 = (\kappa_1\cdot\kappa_3);(\kappa_2\cdot\kappa_3)$, $(\kappa_1;\kappa_2)\cdot\tau = (\kappa_1\cdot\tau);(\kappa_2\cdot\tau)$, and $(\kappa_1;\kappa_2)\cdot\Gamma = (\kappa_1\cdot\Gamma);(\kappa_2\cdot\Gamma)$
9. $\lceil\kappa_1\rceil + \kappa_2 = \lceil\kappa_1\rceil;\kappa_2$, $\lceil\tau_1\rceil + \tau_2 = \lceil\tau_1\rceil;\tau_2$, and $\lceil\Gamma_1\rceil + \Gamma_2 = \lceil\Gamma_1\rceil;\Gamma_2$.
10. $(\kappa_1 + \kappa_2);(\kappa_3 + \kappa_4) \leq (\kappa_1;\kappa_3) + (\kappa_2;\kappa_4)$, $(\tau_1 + \tau_2);(\tau_3 + \tau_4) \leq (\tau_1;\tau_3) + (\tau_2;\tau_4)$, and $(\Gamma_1 + \Gamma_2);(\Gamma_3 + \Gamma_4) \leq (\Gamma_1;\Gamma_3) + (\Gamma_2;\Gamma_4)$.

Proof: Trivial from the definitions of the operations. □

A.2 Proof of Theorem 4.8

Definition A.2 [outermost use]: The *outermost use* of a type τ, written $Use(\tau)$, is defined by: $Use(\tau_1 \times^\kappa \tau_2) = \kappa$ and $Use(\tau_1 \to^{\kappa_1,\kappa_2} \tau_2) = \kappa_1$. $Use(int)$ is undefined.

Lemma A.3: If $\Gamma \vdash D : \tau$ and $D \stackrel{x=h^\kappa}{\longrightarrow} D'$, then $Use(\Gamma(x)) \geq \kappa$.

Proof: Trivial induction on the derivation of $D \stackrel{x=h^\kappa}{\longrightarrow} D'$. □

Proof of Theorem 4.8: Suppose $\Gamma \vdash D : \tau$ and $D \stackrel{\epsilon}{\longrightarrow}$ **Error**. Then, D has a subterm of the form **letrec** $x = h^\kappa$ **in** D' such that $\Gamma' \vdash$ **letrec** $x = h^\kappa$ **in** $D' : \tau'$, $D' \stackrel{x=h^{\kappa'}}{\longrightarrow} D''$, and $\kappa' > \kappa$. By typing rules for dynamic expressions, there must be Γ'' such that $\Gamma'', x:\tau'' \vdash D' : \tau'$ and $\kappa \geq Use(\tau'')$. By Lemma A.3, it must be that $\kappa \geq Use(\tau'') \geq \kappa'$, which contradicts with $\kappa' > \kappa$. □

A.3 Proof of Theorem 4.9

We will show the theorem by induction on the derivation of $D \longrightarrow D'$. Before proving the theorem, we need several lemmas. The first lemma says that we can obtain syntax-directed typing rules by eliminating (T-Weak).

Lemma A.4 [Elimination of weakening]: If $\Gamma \vdash D : \tau$ is derivable, then it is derivable also by using (T-Weak) only after the rule (T-Var).

Proof: This follows from the fact that (T-Weak) applied below rules except for (T-Var) can be permuted upwards. □

Lemma A.5 [Renaming]: If $\Gamma, x : \tau_1 \vdash D : \tau_2$ and $\Gamma + V : \tau_1$ is well defined, then $\Gamma + V : \tau_1 \vdash [V/x]D : \tau_2$.

Proof: Trivial induction on the derivation of $\Gamma, x : \tau_1 \vdash D : \tau_2$. □

The following two lemmas show how typing is changed by reductions labeled with $x = h^{\kappa}$.

Lemma A.6: If $\Gamma, x : \tau_1 \times^{\kappa} \tau_2 \vdash D : \tau_3$ and $D \stackrel{x=\langle V_1, V_2 \rangle^{\kappa'}}{\longrightarrow} D'$, then $\tau_1 = \tau_1' + \tau_1'', \tau_2 = \tau_2' + \tau_2''$, and $\Gamma + V_1 : \tau_1' + V_2 : \tau_2', x : \tau_1'' \times^{\kappa - \kappa'} \tau_2'' \vdash D' : \tau_3$ for some $\tau_1', \tau_1'', \tau_2', \tau_2''$.

Proof: We show the lemma by induction on the derivation of $D \stackrel{x=\langle V_1, V_2 \rangle^{\kappa'}}{\longrightarrow} D'$, with case analysis on the last rule.

- Case for (DR-Fst):
  In this case, $D = \mathbf{let}\ u = \mathbf{fst}^{\kappa'}(x)\ \mathbf{in}\ D_1$ and $D' = [V_1/u]D_1$. Suppose $(1)\Gamma, x : \tau_1 \times^{\kappa} \tau_2 \vdash D : \tau_3$. By the rule (T-Fst), we have

  $(2)\Gamma, u : \tau_1', x : \tau_1'' \times^{\kappa''} \tau_2'' \vdash D_1 : \tau_3$
  $(3)\tau_1 = \tau_1' + \tau_1''$
  $(4)\tau_2 = \tau_2'' = null(\tau_2'') + \tau_2''$
  $(5)\kappa = \kappa'; \kappa''$

  By applying Lemma A.5 to (2), we obtain (6) $\Gamma + V_1 : \tau_1', x : \tau_1'' \times^{\kappa''} \tau_2'' \vdash [V_1/u]D_1 : \tau_3$. By the fact $\kappa - \kappa' \geq \kappa''$ (which follows from (5)) and the rule (T-Weak), we have $\Gamma + V_1 : \tau_1' + V_2 : null(\tau_2''), x : \tau_1'' \times^{\kappa - \kappa'} \tau_2'' \vdash [V_1/u]D_1 : \tau_3$.

- Case for (DR-Snd): Similar to the case for (DR-Fst)

- Case for (DR-Rec):
  In this case, $(1)D = \mathbf{letrec}\ w = h^{\kappa_1}\ \mathbf{in}\ D_1$, $(2)D' = \mathbf{letrec}\ w = h^{\kappa_1}\ \mathbf{in}\ D_1'$, and $(3)D_1 \stackrel{x=\langle V_1, V_2 \rangle^{\kappa'}}{\longrightarrow} D_1'$. Suppose $(4)\Gamma, x : \tau_1 \times^{\kappa} \tau_2 \vdash D : \tau_3$.
  By (DT-Heap),

  $(5)\Gamma_1, x : \tau_3 \times^{\kappa_2} \tau_4 \vdash w = h^{\kappa_1} : \tau_7$
  $(6)\Gamma_2, x : \tau_5 \times^{\kappa_3} \tau_6, w : \tau_7 \vdash D_1 : \tau$
  $(7)\Gamma = \Gamma_1 + \Gamma_2$
  $(8)(\tau_3 \times^{\kappa_2} \tau_4) + (\tau_5 \times^{\kappa_3} \tau_6) = \tau_1 \times^{\kappa} \tau_2$

  for some $\Gamma_1, \Gamma_2, \tau_3, \tau_4, \tau_5, \tau_6, \tau_7, \kappa_2, \kappa_3$. By (3), (6), and induction hypothesis,

$(9)\Gamma_2 + V_1 : \tau_5' + V_2 : \tau_6', x : \tau_5'' \times^{\kappa_3 - \kappa'} \tau_6'', w : \tau_7 \vdash D_1' : \tau$
$(10)\tau_5' + \tau_5'' = \tau_5$
$(11)\tau_6' + \tau_6'' = \tau_6$

for some $\tau_5', \tau_5'', \tau_6', \tau_6''$. By (5), (9) and (DT-HEAP),

$$(12)\Gamma_1 + \Gamma_2 + V_1 : \tau_5' + V_2 : \tau_6', x : (\tau_3 + \tau_5'') \times^{\kappa_2 + (\kappa_3 - \kappa')} (\tau_4 + \tau_6'') \vdash D' : \tau.$$

Let $\tau_1' = \tau_5', \tau_2' = \tau_6', \tau_1'' = \tau_3 + \tau_5''$, and $\tau_2'' = \tau_4 + \tau_6''$. Then, by (7), (9), (10), and (11), we have $\tau_1' + \tau_1'' = \tau_1$, $\tau_2' + \tau_2'' = \tau_2$, and $\Gamma + V_1 : \tau_1' + V_2 : \tau_2', x : \tau_1'' \times^{\kappa_2 + (\kappa_3 - \kappa')} \tau_2'' \vdash D' : \tau$. Since $\kappa - \kappa' = (\kappa_2 + \kappa_3) - \kappa' \geq \kappa_2 + (\kappa_3 - \kappa')$, by using (T-WEAK), we obtain $\Gamma + V_1 : \tau_1' + V_2 : \tau_2', x : \tau_1'' \times^{\kappa - \kappa'} \tau_2'' \vdash D' : \tau$ as required.

□

Lemma A.7: If $\Gamma, x : \tau_1 \to^{\kappa_1, \kappa_2} \tau_2 \vdash D : \tau$ and $D \overset{x = \lambda^{\kappa'} y. M_1}{\longrightarrow} D'$, then $\Gamma + \Gamma_1, x : \tau_1 \to^{(\kappa_1 + \kappa_3) - \kappa', (\kappa_2 + \kappa_4) - 1} \tau_2 \vdash D' : \tau$ holds for any Γ_1 such that $\Gamma_1, x : \tau_1 \to^{\kappa_3, \kappa_4} \tau_2, y : \tau_1 \vdash M_1 : \tau_2$ and $\Gamma + \Gamma_1$ is well defined.

Proof: Induction on the derivation of $D \overset{x = \lambda^{\kappa'} y. M_1}{\longrightarrow} D'$, with case analysis on the last rule.

- Case for (DR-APP):
  In this case, $(1)D = \textbf{let } u = apply^{\kappa'}(x, V) \textbf{ in } D_2$, and $(2)D' = \textbf{let } u = [V/y]M_1 \textbf{ in } D_2$. Suppose $(3)\Gamma, x : \tau_1 \to^{\kappa_1, \kappa_2} \tau_2 \vdash D : \tau$ and $(4)\Gamma_1, y : \tau_1, x : \tau_1 \to^{\kappa_3, \kappa_4} \tau_2 \vdash M_1 : \tau_2$. By (3), we have

  $(5)\Gamma_2, x : \tau_1 \to^{\kappa_1', \kappa_2'} \tau_2, u : \tau_2 \vdash D_2 : \tau$
  $(6)\kappa_1 = \kappa'; \kappa_1'$
  $(7)\kappa_2 = 1; \kappa_2'$
  $(8)\Gamma = (V : \tau_1); \Gamma_2$

  for some $\Gamma_2, \kappa_1', \kappa_2'$. By applying Lemma A.5 to (4), we have

  $$(9)\Gamma_1 + V : \tau_1, x : \tau_1 \to^{\kappa_3, \kappa_4} \tau_2 \vdash [V/y]M_1 : \tau_2.$$

  By applying (T-LET) to (5) and (9), we obtain

  $$(10)(\Gamma_1 + V : \tau_1, x : \tau_1 \to^{\kappa_3, \kappa_4} \tau_2); (\Gamma_2, x : \tau_1 \to^{\kappa_1', \kappa_2'} \tau_2) \vdash D' : \tau.$$

  Here, we have $\kappa_3; \kappa_1' \leq \kappa_1 + \kappa_3 - \kappa'$ by the following inequalities:

  $$\begin{array}{rcl} (\kappa_1 + \kappa_3) - \kappa' & = & ((\kappa'; \kappa_1') + \kappa_3) - \kappa' \\ & \geq & (\kappa'; \kappa_1') - \kappa' + \kappa_3 \\ & \geq & \kappa_1' + \kappa_3 \\ & \geq & \kappa_3; \kappa_1' \end{array}$$

  (Here, we used the fact $(\kappa_1; \kappa_2) - \kappa_1 \geq \kappa_2$.) Similarly, $\kappa_4; \kappa_2' \leq \kappa_2 + \kappa_4 - 1$. So,

  $$\begin{array}{rcl} & (\Gamma_1 + V : \tau_1, x : \tau_1 \to^{\kappa_3, \kappa_4} \tau_2); (\Gamma_2, x : \tau_1 \to^{\kappa_1', \kappa_2'} \tau_2) \\ = & ((\Gamma_1 + V : \tau_1); \Gamma_2), x : \tau_1 \to^{\kappa_3; \kappa_1', \kappa_4; \kappa_2'} \tau_2 \\ \leq & \Gamma_1 + (V : \tau_1; \Gamma_2), x : \tau_1 \to^{\kappa_1 + \kappa_3 - \kappa', \kappa_2 + \kappa_4 - 1} \tau_2 \\ = & \Gamma_1 + \Gamma, x : \tau_1 \to^{\kappa_1 + \kappa_3 - \kappa', \kappa_2 + \kappa_4 - 1} \tau_2 \end{array}$$

By applying (T-Weak) to (10), we obtain

$$\Gamma_1 + \Gamma, x : \tau_1 \rightarrow^{\kappa_1+\kappa_3-\kappa', \kappa_2+\kappa_4-1} \tau_2 \vdash D' : \tau$$

as required.

- Case for (DR-Rec): In this case, $(1) D = \textbf{letrec}\ w = h^{\kappa_5}\ \textbf{in}\ D_2$, $(2) D' = \textbf{letrec}\ w = h^{\kappa_5}\ \textbf{in}\ D_2'$, and $(3) D_2 \stackrel{x=\lambda^{\kappa'} y.M_1}{\longrightarrow} D_2'$. Suppose $(4) \Gamma, x : \tau_1 \rightarrow^{\kappa_1,\kappa_2} \tau_2 \vdash D : \tau$ and $(5) \Gamma_1, x : \tau_1 \rightarrow^{\kappa_3,\kappa_4} \tau_2, y : \tau_1 \vdash M_1 : \tau_2$. We need to show $\Gamma + \Gamma_1, x : \tau_1 \rightarrow^{\kappa_1+\kappa_3-\kappa', \kappa_2+\kappa_4-1} \tau_2 \vdash D' : \tau$. By (DT-Heap) and (4), it must be that

$$\begin{array}{l}(6) \Gamma_2, x : \tau_1 \rightarrow^{\kappa_6,\kappa_7} \tau_2 \vdash w = h^{\kappa_5} : \tau_3 \\ (7) \Gamma_3, x : \tau_1 \rightarrow^{\kappa_8,\kappa_9} \tau_2, w : \tau_3 \vdash D_2 : \tau \\ (8) \Gamma = \Gamma_2 + \Gamma_3 \\ (9) \kappa_1 = \kappa_6 + \kappa_8 \\ (10) \kappa_2 = \kappa_7 + \kappa_9\end{array}$$

for some $\Gamma_2, \Gamma_3, \kappa_4, \kappa_5, \kappa_6, \kappa_7, \kappa_8, \kappa_9$. By (3), (6) and induction hypothesis, $\Gamma_3 + \Gamma_1, w : \tau_3, x : \tau_1 \rightarrow^{\kappa_8+\kappa_3-\kappa', \kappa_9+\kappa_4-1} \tau_2 \vdash D_2' : \tau$. By using (DT-Heap), we obtain

$$\Gamma_2 + \Gamma_3 + \Gamma_1, x : \tau_1 \rightarrow^{\kappa_6+(\kappa_8+\kappa_3-\kappa'), \kappa_7+(\kappa_9+\kappa_4-1)} \tau_2 \vdash D' : \tau.$$

Since $\kappa_6+(\kappa_8+\kappa_3-\kappa') \leq (\kappa_6+\kappa_8+\kappa_3)-\kappa' = \kappa_1+\kappa_3-\kappa'$ and $\kappa_7+(\kappa_9+\kappa_4-1) \leq \kappa_2+\kappa_4-1$, by using (T-Weak) we obtain

$$\Gamma + \Gamma_1, x : \tau_1 \rightarrow^{\kappa_1+\kappa_3-\kappa', \kappa_2+\kappa_4-1} \tau_2 \vdash D' : \tau$$

as required.

□

Next, we shall show that the reduction (DR-Var) preserves typing (Lemma A.11 below). We need the three technical lemmas (Lemma A.8, A.9, and A.10) below.

Lemma A.8: If $\Gamma_1 \vdash x = h^{\kappa_1} : \tau_1$ and $\Gamma_2 \vdash x = h^{\kappa_2} : \tau_2$, and if both $\Gamma_1 + \Gamma_2$ and $\tau_1 + \tau_2$ are well defined, then $\Gamma_1 + \Gamma_2 \vdash x = h^{\kappa_1+\kappa_2} : \tau_1 + \tau_2$.

Proof: Trivial from the rules (DT-Pair) and (DT-Abs). □

Lemma A.9: If $\Gamma \vdash y = h^{\kappa} : \tau$ then $\Gamma' \vdash y = h^{\lfloor\kappa\rfloor} : \lfloor\tau\rfloor$ and $\Gamma' \leq \lfloor\Gamma\rfloor$ for some Γ'.

Proof: We show this by case analysis on h.

- Case where h is of the form $\langle V_1, V_2 \rangle$.
  It must be that $\Gamma = V_1 : \tau_1 + V_2 : \tau_2$ and $\tau = \tau_1 \times^{\kappa} \tau_2$ for some τ_1 and τ_2. By (DT-Pair), we have $V_1 : \lfloor\tau_1\rfloor + V_2 : \lfloor\tau_2\rfloor \vdash y = h^{\lfloor\kappa\rfloor} : \lfloor\tau_1\rfloor \times^{\lfloor\kappa\rfloor} \lfloor\tau_2\rfloor$. Let $\Gamma' = V_1 : \lfloor\tau_1\rfloor + V_2 : \lfloor\tau_2\rfloor$. Then $\Gamma' \vdash y = h^{\lfloor\kappa\rfloor} : \lfloor\tau\rfloor$ and $\Gamma' \leq \lfloor\Gamma\rfloor$ as required.

- Case where h is of the form $\lambda x.M$.
  It must be that

$$\begin{array}{l}\Gamma_1, x : \tau_1, y : \tau_1 \rightarrow^{\kappa_1,\kappa_2} \tau_2 \vdash M : \tau_2 \\ \tau = \tau_1 \rightarrow^{\kappa_3,\kappa_4} \tau_2 \\ \kappa \geq \kappa_3(1 + \kappa_1) \\ \Gamma = \kappa_4(1 + \kappa_2)\lceil\Gamma_1\rceil\end{array}$$

for some $\Gamma_1, \tau_1, \tau_2, \kappa_1, \kappa_2, \kappa_3, \kappa_4$. By (DT-ABS), we have $\lfloor\kappa_4\rfloor(1+\kappa_2)\lceil\Gamma_1\rceil \vdash y = h^{\lfloor\kappa_3\rfloor(1+\kappa_1)} : \tau_1 \to^{\lfloor\kappa_3\rfloor,\lfloor\kappa_4\rfloor} \tau_2$. Since $\lfloor\kappa_4\rfloor(1+\kappa_2)\lceil\Gamma_1\rceil = \lfloor\kappa_4\rfloor\lfloor 1+\kappa_2\rfloor\lceil\Gamma_1\rceil = \lfloor\Gamma\rfloor$ and $\lfloor\kappa_3\rfloor(1+\kappa_1) = \lfloor\kappa_3(1+\kappa_1)\rfloor \leq \lfloor\kappa\rfloor$, we have $\lfloor\Gamma\rfloor \vdash y = h^{\lfloor\kappa\rfloor} : \lfloor\tau\rfloor$ as required.

□

Lemma A.10: If $\Gamma \vdash \textbf{letrec } \vec{y} = \vec{h}^{\vec{\kappa}} \textbf{ in } V : \lceil\tau\rceil$, then $\lfloor\Gamma\rfloor \vdash \textbf{letrec } \vec{y} = \vec{h}^{\lfloor\vec{\kappa}\rfloor} \textbf{ in } V : \lceil\tau\rceil$.

Proof: Induction on the length of $\vec{y}$.

- Base case:
  Suppose $\Gamma \vdash V : \lceil\tau\rceil$. Since it must be derived by (T-VAR) and (T-WEAK), it must be that $\Gamma \geq V : \lceil\tau\rceil$. Since $\lfloor\Gamma\rfloor \geq V : \lceil\tau\rceil$, we have $\lfloor\Gamma\rfloor \vdash V : \lceil\tau\rceil$.

- Induction step: Suppose $\Gamma \vdash \textbf{letrec } y' = h'^{\kappa'}, \vec{y} = \vec{h}^{\vec{k}} \textbf{ in } V : \lceil\tau\rceil$. By (DT-HEAP), it must be that
  $$\Gamma_1 \vdash y' = h'^{\kappa'} : \tau_1$$
  $$\Gamma_2, y' : \tau_1 \vdash \textbf{letrec } \vec{y} = \vec{h}^{\vec{k}} \textbf{ in } V : \lceil\tau\rceil$$
  $$\Gamma = \Gamma_1 + \Gamma_2.$$
  By Lemma A.9 and induction hypothesis, we have
  $$\Gamma_1' \vdash y' = h'^{\lfloor\kappa'\rfloor} : \lfloor\tau_1\rfloor$$
  $$\lfloor\Gamma_2\rfloor, y' : \lfloor\tau_1\rfloor \vdash \textbf{letrec } \vec{y} = \vec{h}^{\lfloor\vec{k}\rfloor} \textbf{ in } V : \lceil\tau\rceil$$
  $$\Gamma_1' \leq \lfloor\Gamma_1\rfloor$$
  By using (DT-HEAP) and the fact that $\lfloor\Gamma\rfloor = \lfloor\Gamma_1 + \Gamma_2\rfloor \geq \lfloor\Gamma_1\rfloor + \lfloor\Gamma_2\rfloor \geq \Gamma_1' + \lfloor\Gamma_2\rfloor$, we have $\lfloor\Gamma\rfloor \vdash \textbf{letrec } y' = h'^{\lfloor\kappa'\rfloor}, \vec{y} = \vec{h}^{\lfloor\vec{k}\rfloor} \textbf{ in } V : \lceil\tau\rceil$ as required.

□

Lemma A.11: If $\Gamma \vdash \textbf{let } x = (\textbf{letrec } \vec{z} = \vec{h}^{\vec{\kappa_1}}, \vec{w} = \vec{h'}^{\vec{\kappa_3}} \textbf{ in } V) \textbf{ in } (\textbf{letrec } \vec{z} = \vec{h}^{\vec{\kappa_2}} \textbf{ in } M) : \tau$, then $\Gamma \vdash \textbf{letrec } \vec{z} = \vec{h}^{\lfloor\vec{\kappa_1}\rfloor;\vec{\kappa_2}}, \vec{w} = \vec{h'}^{\lfloor\vec{\kappa_3}\rfloor} \textbf{ in } [V/x]M : \tau$.

Proof: It is sufficient to show the case where $\vec{w}$ is an empty sequence, because $\Gamma \vdash \textbf{let } x = (\textbf{letrec } \vec{z} = \vec{h}^{\vec{\kappa_1}}, \vec{w} = \vec{h'}^{\vec{\kappa_3}} \textbf{ in } V) \textbf{ in } (\textbf{letrec } \vec{z} = \vec{h}^{\vec{\kappa_2}} \textbf{ in } M) : \tau$ implies $\Gamma \vdash \textbf{let } x = (\textbf{letrec } \vec{z} = \vec{h}^{\vec{\kappa_1}}, \vec{w} = \vec{h'}^{\vec{\kappa_3}} \textbf{ in } V) \textbf{ in } (\textbf{letrec } \vec{z} = \vec{h}^{\vec{\kappa_2}}, \vec{w} = \vec{h'}^{\vec{0}} \textbf{ in } M) : \tau$ (by the rules (DT-PAIR) and (DT-ABS)). We show by induction on the length of the sequence $\vec{z}$.

- Base case: Suppose $\Gamma \vdash \textbf{let } x = V \textbf{ in } M : \tau$. Then, (1)$\Gamma_1 \vdash V : \lceil\tau_1\rceil$, (2)$\Gamma_2, x : \lceil\tau_1\rceil \vdash M : \tau$, (3)$\Gamma = \Gamma_1;\Gamma_2$ and (4)$\Gamma_1 \geq V : \lceil\tau_1\rceil$. By applying Lemma A.5 to (2), we obtain $\Gamma_2 + V : \lceil\tau_1\rceil \vdash [V/x]M : \tau$. Since $\Gamma = \Gamma_1;\Gamma_2 \geq (V : \lceil\tau_1\rceil);\Gamma_2 = \Gamma_2 + V : \lceil\tau_1\rceil$, we have $\Gamma \vdash [V/x]M : \tau$ by (T-WEAK).

- Induction step: Suppose
  $$\Gamma \vdash \textbf{let } x = (\textbf{letrec } z = h^{\kappa}, \vec{w} = \vec{h}^{\vec{\kappa_1}} \textbf{ in } V) \textbf{ in letrec } z = h^{\kappa'}, \vec{w} = \vec{h}^{\vec{\kappa_1'}} \textbf{ in } M : \tau.$$

By typing rules, we have

$$(1)\Gamma_1 \vdash \textbf{letrec } z = h^{\kappa}, \vec{w} = \vec{h}^{\vec{\kappa_1}} \textbf{ in } V : \lceil \tau_1 \rceil$$
$$(2)\Gamma_2, x : \lceil \tau_1 \rceil \vdash \textbf{letrec } z = h^{\kappa'}, \vec{w} = \vec{h}^{\vec{\kappa_1'}} \textbf{ in } M : \tau$$
$$(3)\Gamma = \Gamma_1; \Gamma_2$$

By (1), (2) and typing rules, it must be that

$$(4)\Gamma_1', z : \tau_z \vdash \textbf{letrec } \vec{w} = \vec{h}^{\vec{\kappa_1}} \textbf{ in } V : \lceil \tau_1 \rceil$$
$$(5)\Gamma_1'' \vdash z = h^{\kappa} : \tau_z$$
$$(6)\Gamma_2', x : \lceil \tau_1 \rceil, z : \tau_z' \vdash \textbf{letrec } \vec{w} = \vec{h}^{\vec{\kappa_1'}} \textbf{ in } M : \tau$$
$$(7)\Gamma_2'' \vdash z = h^{\kappa'} : \tau_z'$$
$$(8)\Gamma_1' + \Gamma_1'' = \Gamma_1$$
$$(9)\Gamma_2' + \Gamma_2'' = \Gamma_2$$

By (4) and Lemma A.10, we have

$$(10)\lfloor \Gamma_1' \rfloor, z : \lfloor \tau_z \rfloor \vdash \textbf{letrec } \vec{w} = \vec{h}^{\lfloor \vec{\kappa_1} \rfloor} \textbf{ in } V : \lceil \tau_1 \rceil.$$

By applying (T-Let) to (6) and (10), we obtain

$$\lfloor \Gamma_1' \rfloor; \Gamma_2', z : (\lfloor \tau_z \rfloor; \tau_z') \vdash \textbf{let } x = (\textbf{letrec } \vec{w} = \vec{h}^{\lfloor \vec{\kappa_1} \rfloor} \textbf{ in } V) \textbf{ in letrec } \vec{w} = \vec{h}^{\vec{\kappa_1'}} \textbf{ in } M : \tau.$$

By induction hypothesis and (T-Weak), it follows that

$$(11)\lfloor \Gamma_1' \rfloor; \Gamma_2', z : (\lfloor \tau_z \rfloor + \tau_z') \vdash \textbf{letrec } \vec{w} = \vec{h}^{\lfloor \vec{\kappa_1} \rfloor; \vec{\kappa_1'}} \textbf{ in } [V/x]M : \tau.$$

By (5), (7), (11), Lemma A.8 and Lemma A.9, we have

$$(\lfloor \Gamma_1' \rfloor; \Gamma_2') + \lfloor \Gamma_1'' \rfloor + \Gamma_2'' \vdash \textbf{letrec } z = h^{\lfloor \kappa \rfloor + \kappa'}, \vec{w} = \vec{h}^{\lfloor \vec{\kappa_1} \rfloor; \vec{\kappa_1'}} \textbf{ in } [V/x]M : \tau.$$

By $(\lfloor \Gamma_1' \rfloor; \Gamma_2') + \lfloor \Gamma_1'' \rfloor + \Gamma_2'' = \lfloor \Gamma_1' \rfloor + \Gamma_2' + \lfloor \Gamma_1'' \rfloor + \Gamma_2'' \leq \lfloor \Gamma_1' + \Gamma_1'' \rfloor + (\Gamma_2' + \Gamma_2'') = \lfloor \Gamma_1 \rfloor + \Gamma_2 = \lfloor \Gamma_1 \rfloor; \Gamma_2 \leq \Gamma$, we have

$$\Gamma \vdash \textbf{letrec } z = h^{\lfloor \kappa \rfloor; \kappa'}, \vec{w} = \vec{h}^{\lfloor \vec{\kappa_1} \rfloor; \vec{\kappa_1'}} \textbf{ in } [V/x]M : \tau$$

as required.

□

Proof of Theorem 4.9: We show this by induction on the derivation of $D \xrightarrow{l} D'$ with case analysis on the last rule.

- (DR-Pair), (DR-Fix): Trivial from the typing rules.
- (DR-Abs):
  In this case, $D = \textbf{let } x = \lambda^{\kappa_1} y.M_1 \textbf{ in } M_2$, $D' = \textbf{letrec } x = \lambda^{\kappa_1} y.M_1 \textbf{ in } M_2$, and $l = \epsilon$. Suppose $\Gamma \vdash D : \tau$. By the rule (T-Abs),

  $$(1)\Gamma_1, y : \tau_1 \vdash M_1 : \tau_2$$
  $$(2)\Gamma_2, x : \tau_1 \to^{\kappa_1, \kappa_2} \tau_2 \vdash M_2 : \tau$$
  $$(3)\Gamma = \kappa_2 \lceil \Gamma_1 \rceil + \Gamma_2$$

Without loss of generality, we can assume that $x \notin dom(\Gamma_1)$. By applying (T-Weak) to (1), we have (4)$\Gamma_1, y : \tau_1, x : \tau_1 \to^{0,0} \tau_2 \vdash M_1 : \tau_2$. By (2), (4), (DT-Abs) and (DT-Heap), we have $\kappa_2(1+0)\lceil\Gamma_1\rceil + \Gamma_2 \vdash D' : \tau$, that is, $\Gamma \vdash D' : \tau$.

- (DR-Read): In this case, (1)$D = \mathbf{letrec}\ x = h^{\kappa}\ \mathbf{in}\ D_1$, (2)$D' = \mathbf{letrec}\ x = h^{\kappa-\kappa'}\ \mathbf{in}\ D_1'$, and (3)$D_1 \stackrel{x=h^{\kappa'}}{\longrightarrow} D_1'$ for $h = \langle V_1, V_2\rangle$ or $h = \lambda y.M_2$. Suppose (4)$\Gamma \vdash D : \tau$.
  - If $h = \langle V_1, V_2\rangle$, then (5)$\Gamma = \Gamma_1 + V_1 : \tau_1 + V_2 : \tau_2$ and (6)$\Gamma_1, x : \tau_1 \times^{\kappa} \tau_2 \vdash D_1 : \tau$ for some Γ_1, τ_1, and τ_2. By Lemma A.6, there are some $\tau_1', \tau_1'', \tau_2', \tau_2''$ such that

    (7)$\tau_1 = \tau_1' + \tau_1''$
    (8)$\tau_2 = \tau_2' + \tau_2''$
    (9)$\Gamma_1 + V_1 : \tau_1' + V_2 : \tau_2', x : \tau_1'' \times^{\kappa-\kappa'} \tau_2'' \vdash D_1' : \tau$.

    By (9), (DT-Pair) and (DT-Heap), (10)$\Gamma_1 + V_1 : \tau_1' + V_2 : \tau_2' + V_1 : \tau_1'' + V_2 : \tau_2'' \vdash \mathbf{letrec}\ x = \langle V_1, V_2\rangle^{\kappa-\kappa'}\ \mathbf{in}\ D_1' : \tau$. Because $\Gamma_1 + V_1 : \tau_1' + V_2 : \tau_2' + V_1 : \tau_1'' + V_2 : \tau_2'' = \Gamma$ by (5), (7), and (8), $\Gamma \vdash \mathbf{letrec}\ x = \langle V_1, V_2\rangle^{\kappa-\kappa'}\ \mathbf{in}\ D_1'$.
  - If $h = \lambda^{\kappa} y.M_2$, then it must be that

    (11)$\Gamma_1, y : \tau_1, x : \tau_1 \to^{\kappa_1,\kappa_2} \tau_2 \vdash M_2 : \tau_2$
    (12)$\Gamma_2, x : \tau_1 \to^{\kappa_3,\kappa_4} \tau_2 \vdash D_1 : \tau$
    (13)$\Gamma = \kappa_4(1+\kappa_2)\lceil\Gamma_1\rceil + \Gamma_2$
    (14)$\kappa \geq \kappa_3(1+\kappa_1), \kappa_3 \geq \delta, \kappa_4 \geq 1$

    for some $\Gamma_1, \Gamma_2, \tau_1, \tau_2, \kappa_1, \kappa_2, \kappa_3$, and κ_4. By (3), (11), (12), and Lemma A.7, we have

    (15)$\Gamma_2 + \Gamma_1, x : \tau_1 \to^{\kappa_1+\kappa_3-\kappa', \kappa_2+\kappa_4-1} \tau_2 \vdash D_1' : \tau_3$

    By (11), (15), (DT-Abs) and (DT-Heap),

    $$(\kappa_2+\kappa_4-1)(1+\kappa_2)\lceil\Gamma_1\rceil + (\Gamma_2+\Gamma_1) \vdash \mathbf{letrec}\ x = \lambda^{(\kappa_1+\kappa_3-\kappa')(1+\kappa_1)} y.M_2\ \mathbf{in}\ D_1' : \tau$$

    The required condition $\Gamma \vdash \mathbf{letrec}\ x = \lambda^{\kappa-\kappa'} y.M_2\ \mathbf{in}\ D_1' : \tau$ follows if we show (a) $(\kappa_2+\kappa_4-1)(1+\kappa_2)+1 \leq \kappa_4(1+\kappa_2)$ and (b) $(\kappa_1+\kappa_3-\kappa')(1+\kappa_1) \leq \kappa - \kappa'$. They can be shown as follows:
    (a) The inequality is trivial if $\kappa_4(1+\kappa_2) = \omega$. Otherwise, κ_2 must be 0 by (14). Then, $(\kappa_2+\kappa_4-1)(1+\kappa_2)+1 = \kappa_4 - 1 + 1 = \kappa_4 = \kappa_4(1+\kappa_2)$.
    (b) The inequality is trivial if $\kappa = \omega$. Otherwise κ_1 must be 0 by (14). Then, $(\kappa_1+\kappa_3-\kappa')(1+\kappa_1) = \kappa_3 - \kappa' = \kappa_3(1+\kappa_1) - \kappa' \leq \kappa - \kappa'$.

- (DR-IfT):
  In this case, $M = \mathbf{if0}\ 0\ \mathbf{then}\ M_1\ \mathbf{else}\ M_2$. Suppose $\Gamma \vdash M : \tau$. Then

  $\Gamma_1 \vdash M_1 : \tau$
  $\Gamma_1 \vdash M_2 : \tau$
  $\Gamma_2 \vdash 0 : int$
  $\Gamma = \Gamma_1 + \Gamma_2$

  By (T-Weak), we have $\Gamma_1 + \Gamma_2 \vdash M_1 : \tau$, i.e., $\Gamma \vdash M' : \tau$.

- (DR-IfF): Similar to the case for (T-IfT).

- (DR-Rec): In this case, $D = \mathbf{letrec}\ x = h^{\kappa}\ \mathbf{in}\ D_1$, $D' = \mathbf{letrec}\ x = h^{\kappa}\ \mathbf{in}\ D_1'$, and $D_1 \stackrel{l}{\longrightarrow} D_1'$. Suppose $\Gamma \vdash D : \tau$. Then, $\Gamma_1, x : \tau_1 \vdash D_1 : \tau$, $\Gamma_2 \vdash x = h^{\kappa} : \tau_1$, and $\Gamma = \Gamma_1 + \Gamma_2$ for some Γ_1 and Γ_2. Case analysis on l.

- Case $l = \epsilon$: By $D_1 \xrightarrow{\epsilon} D_1'$ and induction hypothesis, $\Gamma_1, x : \tau_1 \vdash D_1' : \tau$. So, by using (DT-Heap), we obtain $\Gamma \vdash D' : \tau$.
- Case $l = \epsilon'$: By $D_1 \xrightarrow{\epsilon'} D_1'$ and induction hypothesis, there exists D_1'' such that $\Gamma_1, x : \tau_1 \vdash D_1'' : \tau$ and $D_1 \xrightarrow{\epsilon'} D_1''$. Let $D'' = \mathbf{letrec}\ x = h^{\kappa}\ \mathbf{in}\ D_1''$ Then, $D \xrightarrow{\epsilon'} D''$ and $\Gamma \vdash D'' : \tau$.

- (DR-Let): In this case, $D = \mathbf{let}\ x = D_1\ \mathbf{in}\ D_2$, $D' = \mathbf{let}\ x = D_1'\ \mathbf{in}\ D_2$, and $D_1 \xrightarrow{l} D_1'$. Suppose $\Gamma \vdash D : \tau$. Then, by (T-Let), $\Gamma_1 \vdash D_1 : \tau_1$, $\Gamma_2, x : \tau_1 \vdash D_2 : \tau$, and $\Gamma = \Gamma_1; \Gamma_2$ for some Γ_1, Γ_2, and τ_1. Case analysis on l.
  - Case $l = \epsilon$: By $D_1 \xrightarrow{l} D_1'$ and induction hypothesis, $\Gamma_1 \vdash D_1' : \tau_1$. By using (T-Let), we obtain $\Gamma \vdash D' : \tau$.
  - Case $l = \epsilon'$: By $D_1 \xrightarrow{\epsilon'} D_1'$ and induction hypothesis, $\Gamma_1 \vdash D_1'' : \tau_1$ and $D_1 \xrightarrow{\epsilon'} D_1''$ for some D_1''. Let $D'' = \mathbf{let}\ x = D_1''\ \mathbf{in}\ D_2$. Then $\Gamma \vdash D'' : \tau$ and $D \xrightarrow{\epsilon'} D''$.

- (DR-Split): In this case, $D = \mathbf{letrec}\ x = h^{\kappa}\ \mathbf{in}\ \mathbf{let}\ y = D_1\ \mathbf{in}\ D_2$,
  $D' = \mathbf{let}\ y = (\mathbf{letrec}\ x = h^{\kappa_1}\ \mathbf{in}\ D_1)\ \mathbf{in}\ \mathbf{letrec}\ x = h^{\kappa_2}\ \mathbf{in}\ D_2$, and $\kappa = \kappa_1; \kappa_2$. Suppose $\Gamma \vdash D : \tau$. We show $\Gamma \vdash D'' : \tau$ and $D \xrightarrow{\epsilon'} D''$ for some D'', by case analysis on h.
  - Case where h is of the form $\langle V_1, V_2 \rangle$.
    By (DT-Pair), (DT-Heap) and (T-Let), we have

    $$\Gamma_1, x : \tau_1 \times^{\kappa_3} \tau_2 \vdash D_1 : \tau_3$$
    $$\Gamma_2, x : \tau_4 \times^{\kappa_4} \tau_5, y : \tau_3 \vdash D_2 : \tau$$
    $$\Gamma = (\Gamma_1; \Gamma_2) + V_1 : (\tau_1; \tau_4) + V_2 : (\tau_2; \tau_5)$$
    $$\kappa = \kappa_3; \kappa_4$$

    Let $D'' = \mathbf{let}\ y = (\mathbf{letrec}\ x = h^{\kappa_3}\ \mathbf{in}\ D_1)\ \mathbf{in}\ \mathbf{letrec}\ x = h^{\kappa_4}\ \mathbf{in}\ D_2$. Then, $D \xrightarrow{\epsilon'} D''$ and

    $$(\Gamma_1 + V_1 : \tau_1 + V_2 : \tau_2); (\Gamma_2 + V_1 : \tau_4 + V_2 : \tau_5) \vdash D'' : \tau.$$

    Because $(\Gamma_1 + V_1 : \tau_1 + V_2 : \tau_2); (\Gamma_2 + V_1 : \tau_4 + V_2 : \tau_5) \leq (\Gamma_1; \Gamma_1) + V_1 : (\tau_1; \tau_4) + V_2 : (\tau_2; \tau_5) = \Gamma$, we have $\Gamma \vdash D'' : \tau$ by (T-Weak). (Here we used the law $(\Gamma_1 + \Gamma_2); (\Gamma_3 + \Gamma_4) \leq (\Gamma_1; \Gamma_3) + (\Gamma_2; \Gamma_4)$.
  - Case where h is of the form $\lambda z.M_3$.
    By (DT-Abs), (DT-Heap) and (T-Let), we have

    $$\Gamma_1, x : \tau_1 \to^{\kappa_3, \kappa_4} \tau_2 \vdash D_1 : \tau_3$$
    $$\Gamma_2, x : \tau_1 \to^{\kappa_5, \kappa_6} \tau_2, y : \tau_3 \vdash D_2 : \tau$$
    $$\Gamma_3, x : \tau_1 \to^{\kappa_7, \kappa_8} \tau_2, z : \tau_1 \vdash M_3 : \tau$$
    $$\Gamma = (\kappa_4; \kappa_6)(1 + \kappa_8)\lceil \Gamma_3 \rceil + (\Gamma_1; \Gamma_2)$$
    $$\kappa \geq (\kappa_3; \kappa_5)(1 + \kappa_7)$$

    Since $(\kappa_3; \kappa_5)(1 + \kappa_7) = (\kappa_3(1 + \kappa_7)); (\kappa_5(1 + \kappa_7))$, we have κ_9, κ_{10} such that $\kappa = \kappa_9; \kappa_{10}$, $\kappa_9 \geq \kappa_3(1 + \kappa_7)$, and $\kappa_{10} \geq \kappa_5(1 + \kappa_7)$. Let $D'' = \mathbf{let}\ y = (\mathbf{letrec}\ x = h^{\kappa_9}\ \mathbf{in}\ D_1)\ \mathbf{in}\ \mathbf{letrec}\ x = h^{\kappa_{10}}\ \mathbf{in}\ D_2$. Then, $D \xrightarrow{\epsilon'} D''$. Moreover, by (T-Let), (DT-Abs) and (DT-Heap),

    $$(\kappa_4(1 + \kappa_8)\lceil \Gamma_3 \rceil + \Gamma_1); (\kappa_6(1 + \kappa_8)\lceil \Gamma_3 \rceil + \Gamma_2) \vdash D'' : \tau$$

    Because $(\kappa_4(1 + \kappa_8)\lceil \Gamma_3 \rceil + \Gamma_1); (\kappa_6(1 + \kappa_8)\lceil \Gamma_3 \rceil + \Gamma_2) \leq (\kappa_4(1 + \kappa_8)\lceil \Gamma_3 \rceil; \kappa_6(1 + \kappa_8)\lceil \Gamma_3 \rceil) + (\Gamma_1; \Gamma_2) = \Gamma$, we have $\Gamma \vdash D'' : \tau$ by (T-Weak).

- (DR-VAR): This case follows from Lemma A.11.

□

A.4 Proof of Theorem 4.7

Lemma A.12: If $D \xrightarrow{x=h^{\kappa'}} D'$, $D^* = (H, M)$ and $D'^* = (H', M')$, then $(H + \{x \mapsto^{\kappa} h\}, M) \rightsquigarrow (H' + \{x \mapsto^{\kappa-\kappa'} h\}, M')$ for any κ such that $\kappa \geq \kappa'$.

Proof: Straightforward induction on the derivation of $D \xrightarrow{x=h^{\kappa'}} D'$. □

Lemma A.13: If $(H_1, M) \rightsquigarrow (H_1', M')$, then $(H_1; H_2, M) \rightsquigarrow (H, M')$ and $H \geq (H_1'; H_2)$ for some H.

Proof: Trivial from the fact that $H_1 = H_1'$ or $H_1 = H_3\{x \mapsto^{\kappa} h\} \wedge H_1' = H_3\{x \mapsto^{\kappa-\kappa'} h\}$. □

Lemma A.14: If **letrec** $\vec{x} = \vec{h}^{\vec{\kappa}}$ **in** $D_1 \uparrow$, then **letrec** $\vec{x} = \vec{h}^{\vec{\kappa}}$ **in let** $y = D_1$ **in** $D_2 \uparrow$.

Proof: Suppose **letrec** $\vec{x} = \vec{h}^{\vec{\kappa}}$ **in** $D_1 \uparrow$. Then, **letrec** $\vec{x} = \vec{h}^{\vec{\kappa}}$ **in let** $y = D_1$ **in** $D_2 \xrightarrow{\epsilon'}{}^{*}$ **let** $y =$ (**letrec** $\vec{x} = \vec{h}^{\vec{\kappa}}$ **in** D_1) **in** (**letrec** $\vec{x} = \vec{h}^{\vec{0}}$ **in** D_2) $\xrightarrow{\hat{\epsilon}}$ **Error**. Moreover, suppose **letrec** $\vec{x} = \vec{h}^{\vec{\kappa}}$ **in let** $y = D_1$ **in** $D_2 \xrightarrow{\epsilon'}{}^{*} D' \xrightarrow{\epsilon} D''$. Then, it must be that

(1)$D' \equiv$ **letrec** $\vec{x_1} = \vec{h_1}^{\vec{\kappa_1}}$ **in let** $y =$ **letrec** $\vec{x_2} = \vec{h_2}^{\vec{\kappa_{21}}}$ **in** D_1 **in letrec** $\vec{x_2} = \vec{h_2}^{\vec{\kappa_{22}}}$ **in** D_2
(2)$D'' \equiv$ **letrec** $\vec{x_1} = \vec{h_1}^{\vec{\kappa_1}}$ **in let** $y =$ **letrec** $\vec{x_2} = \vec{h_2}^{\vec{\kappa'_{21}}}$ **in** D_1' **in letrec** $\vec{x_2} = \vec{h_2}^{\vec{\kappa_{22}}}$ **in** D_2
(3)**letrec** $\vec{x_2} = \vec{h_2}^{\vec{\kappa_{21}}}$ **in** $D_1 \xrightarrow{\epsilon}$ **letrec** $\vec{x_2} = \vec{h_2}^{\vec{\kappa'_{21}}}$ **in** D_1'
(4)$\vec{x} = \vec{h}^{\vec{\kappa}} \equiv \vec{x_1} = \vec{h_1}^{\vec{\kappa_1}}, \vec{x_2} = \vec{h_2}^{\vec{\kappa_{21}};\vec{\kappa_{22}}}$.

The conditions (3) and (4) contradicts with **letrec** $\vec{x} = \vec{h}^{\vec{\kappa}}$ **in** $D_1 \uparrow$. (We can easily construct a derivation for **letrec** $\vec{x} = \vec{h}^{\vec{\kappa}}$ **in** $D_1 \xrightarrow{\epsilon}$ **letrec** $\vec{x} = \vec{h}^{\vec{\kappa'}}$ **in** D_1' from (3).) □

Proof of Theorem 4.7:

1. Because $D \xrightarrow{\epsilon'} D'$ implies $D^* = D'^*$, it suffices to show that $D \xrightarrow{\epsilon} D'$ implies that $D^* \rightsquigarrow (H, M)$ and $(H, M) \geq D'^*$ hold for some H, M. We show it by induction on the derivation of $D \xrightarrow{\epsilon} D'$, with case analysis on the last rule.

   Cases for (DR-PAIR), (DR-ABS), (DR-FIX), (DR-IFT), (DR-IFF), (DR-REC), and (DR-VAR) are trivial. Case for (DR-READ) follows from Lemma A.12.

   If the last rule is (DR-LET), we have (1)$D \equiv$ **let** $x = D_1$ **in** D_2, (2)$D' \equiv$ **let** $x = D_1'$ **in** D_2, (3)$D_1 \xrightarrow{\epsilon} D_1'$, (4)$D_1^* = (H_1, M_1)$, (5)$D_2^* = (H_2, M_2)$, and (6)$D^* = (H_1; H_2,$ **let** $x = M_1$ **in** $M_2)$. By (3), (4), and induction hypothesis, there are H_1', M_1' such that (7)$(H_1, M_1) \rightsquigarrow (H_1', M_1')$ and (8)$D_1'^* \leq (H_1', M_1')$. By (6), (7), and Lemma A.13, $D^* \rightsquigarrow (H',$**let** $x = M_1'$ **in** $M_2)$ and (9)$H' \geq (H_1'; H_2)$ for some H'. By (2), (5), (8), and (9), we have $D'^* \leq (H',$ **let** $x = M_1'$ **in** $M_2)$ as required.

2. Suppose (1)$(H, M) \rightsquigarrow$ **Error** and (2)$(H, M) \geq D^*$. By reduction rules, (3)$M \equiv C[I]$ for some context $C[\,]$ and expression I of the form $apply^{\kappa}(x, V)$, $\mathbf{fst}^{\kappa}(x)$, or $\mathbf{snd}^{\kappa}(x)$. We can easily show $D \uparrow$ by induction on the structure of $C[\,]$.

- Case $C[\,] \equiv [\,]$: Straightfoward case analysis on I.
- Case $C[\,] \equiv \mathbf{let}\ x = C'[\,]\ \mathbf{in}\ M_2$:

$$(5) D \equiv \mathbf{letrec}\ \vec{y} = \vec{h_y}^{\vec{\kappa_y}}\ \mathbf{in}\ (\mathbf{let}\ x = \mathbf{letrec}\ \vec{z} = \vec{h_z}^{\vec{\kappa_{z1}}}, \vec{w} = \vec{h_w}^{\vec{\kappa_w}}\ \mathbf{in}\ C'[I]\ \mathbf{in}\ \mathbf{letrec}\ \vec{z} = \vec{h_z}^{\vec{\kappa_{z2}}}\ \mathbf{in}\ M_1)$$
$$(6) H \geq \{\vec{y} \mapsto^{\vec{\kappa_y}} \vec{h_y}, \vec{z} \mapsto^{\vec{\kappa_{z1}};\vec{\kappa_{z2}}} \vec{h_z}, \vec{w} \mapsto^{\vec{\kappa_w}} \vec{h_w}\}.$$

  Let $D_1 = \mathbf{letrec}\ \vec{y} = \vec{h_y}^{\vec{\kappa_y}}, \vec{z} = \vec{h_z}^{\vec{\kappa_{z1}}}, \vec{w} = \vec{h_w}^{\vec{\kappa_w}}\ \mathbf{in}\ C'[I]$. Then by induction hypothesis, $D_1 \uparrow$. By Lemma A.14, $D \uparrow$.

3. Suppose $(1)(H, M) \rightsquigarrow (H', M')$ and $(2)(H, M) \geq D^*$. By reduction rules, $(3)M \equiv C[I]$ and $(4)M' \equiv C[M_1]$ for some context $C[\,]$ and expression I of the form $\mathbf{let}\ x = h^{\kappa}\ \mathbf{in}\ M_2$, $\mathbf{let}\ x = \mathbf{rec}^{\kappa}(f, y, M_2)\ \mathbf{in}\ M_3$, $\mathbf{let}\ x = V\ \mathbf{in}\ M$, $apply^{\kappa}(x, V)$, $\mathbf{fst}^{\kappa}(x)$, $\mathbf{snd}^{\kappa}(x)$, or $\mathbf{if0}\ n\ \mathbf{then}\ M_2\ \mathbf{else}\ M_3$. We show that either $D \uparrow$ holds or there exists D' such that $D \xrightarrow{\hat{\epsilon}} D'$ and $(H', M') \geq D'^*$ by induction on the structure of $C[\,]$.

- Case $C[\,] = [\,]$: Straightfoward case analysis on I.
- Case $C[\,] = \mathbf{let}\ x = C'[\,]\ \mathbf{in}\ M_2$: By (2), it must be that

$$(5) D \equiv \mathbf{letrec}\ \vec{y} = \vec{h_y}^{\vec{\kappa_y}}\ \mathbf{in}\ (\mathbf{let}\ x = \mathbf{letrec}\ \vec{z} = \vec{h_z}^{\vec{\kappa_{z1}}}, \vec{w} = \vec{h_w}^{\vec{\kappa_w}}\ \mathbf{in}\ C'[I]\ \mathbf{in}\ \mathbf{letrec}\ \vec{z} = \vec{h_z}^{\vec{\kappa_{z2}}}\ \mathbf{in}\ M_1)$$
$$(6) H \geq \{\vec{y} \mapsto^{\vec{\kappa_y}} \vec{h_y}, \vec{z} \mapsto^{\vec{\kappa_{z1}};\vec{\kappa_{z2}}} \vec{h_z}, \vec{w} \mapsto^{\vec{\kappa_w}} \vec{h_w}\}.$$

  Let $H_1 = \{\vec{y} \mapsto^{\vec{\kappa_y}} \vec{h_y}, \vec{z} \mapsto^{\vec{\kappa_{z1}}} \vec{h_z}, \vec{w} \mapsto^{\vec{\kappa_w}} \vec{h_w}\}$ and $D_1 = \mathbf{letrec}\ \vec{y} = \vec{h_y}^{\vec{\kappa_y}}, \vec{z} = \vec{h_z}^{\vec{\kappa_{z1}}}, \vec{w} = \vec{h_w}^{\vec{\kappa_w}}\ \mathbf{in}\ C'[I]$ Then, by (3) and (4), either $(H_1, C'[I]) \rightsquigarrow (H_1', C'[M_1])$ and $H' \geq H_1'; \{\vec{z} \mapsto^{\vec{\kappa_{z2}}} \vec{h_z}\}$, or $(H_1, C'[I]) \rightsquigarrow \mathbf{Error}$. In the latter case, $D_1 \uparrow$ by the second property of the theorem. By Lemma A.14, $D \uparrow$. In the former case, by the induction hypothesis, $D_1 \xrightarrow{\hat{\epsilon}} D_1'$ and $(H_1', C'[M_1]) \geq D_1'^*$. Let $D' = \mathbf{let}\ x = D_1'\ \mathbf{in}\ \mathbf{letrec}\ \vec{y} = \vec{h_y}^{\vec{0}}, \vec{z} = \vec{h_z}^{\vec{\kappa_{z2}}}\ \mathbf{in}\ M_1$. Then, $D \xrightarrow{\hat{\epsilon}} D'$ and $(H', M') \geq D'^*$.

□