# 2. 雛形具体化

**雛形具体化**を用いるグラフ簡約器

# 2.1 雛形具体化

Implementation of Functional Programming Languageの11章、12章

- 関数プログラムは**式を評価**によって実行される
- 式はグラフで表現される
- 評価は一連の簡約を行うことで実施される
- 簡約はグラフ中の簡約可能式(redex)を簡約しその結果で置き換える
- 評価は対象とする式が**正規形**(normal form)になれば終了
- 簡約系列な複数ありうるが停止したときには同じ正規形になる
- 正規形に到達する簡約系列があれば最外簡約戦略で必ず停止する

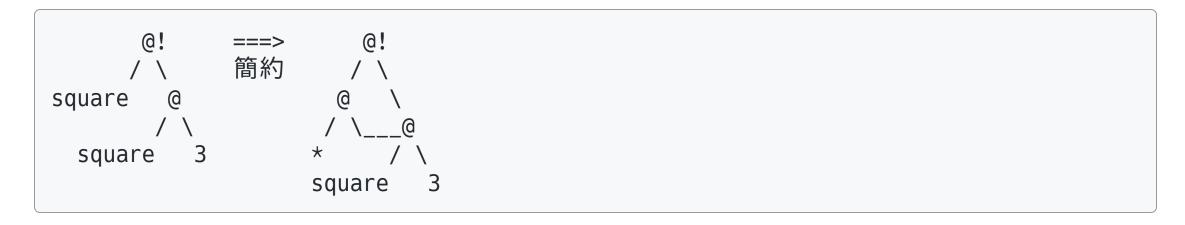
#### 簡約例

```
main = square (square 3) square x = x * x
```

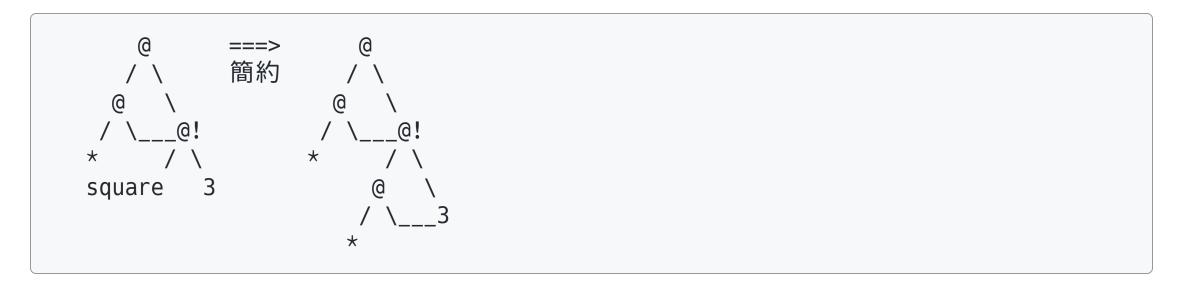
スーパーコンビネータ main には引数はなく、それ自身redexなので、ボディ部と置き換える。

```
main ===> @
簡約 /\
square @
/\
square 3
```

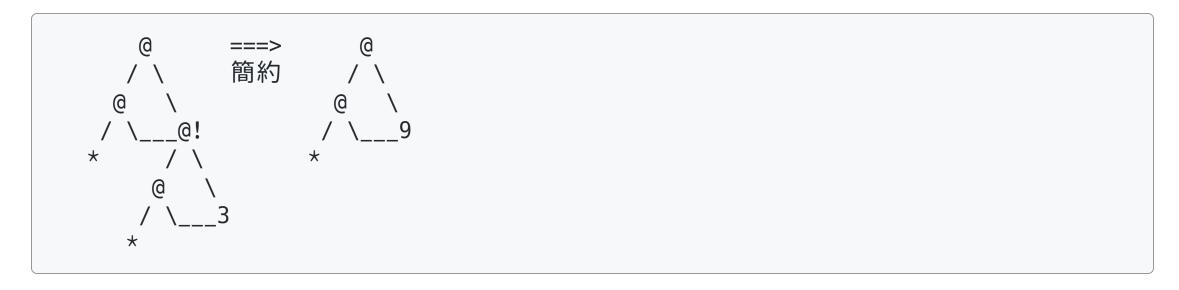
最外のredexは square の適用式。関数本体を具体化したものでredexを置き換える。仮引数の各出現を引数へのポインタで置き換える。



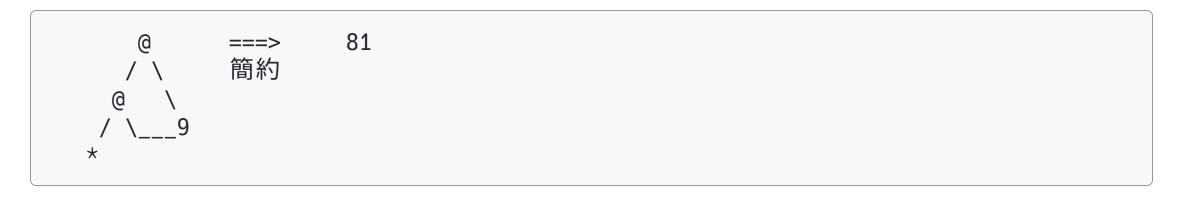
### ここでredexは内側の square の適用式だけなので、これを簡約する。



ここで内側のかけ算が、唯一のredexとなるのでこれを簡約する。



# 最後の簡約は簡単。



### 2.1.2 簡約の3ステップ

以下を正規形が得られるまで繰り返す

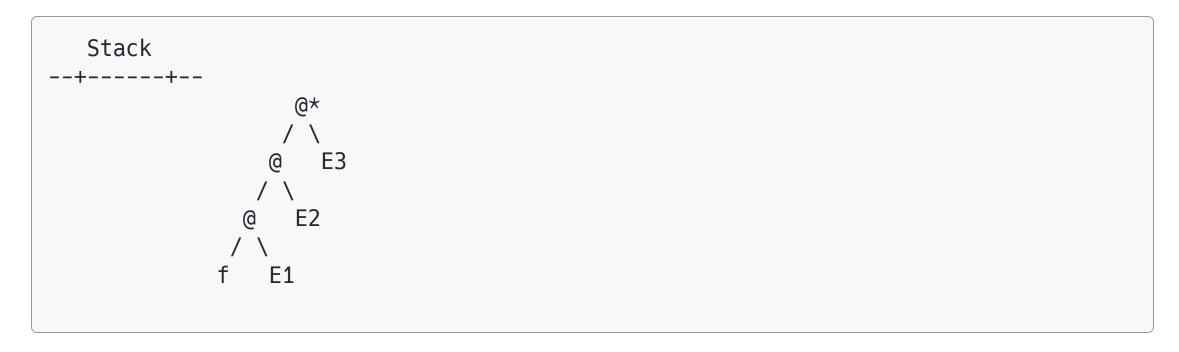
- 1. 次に簡約するredexを見つける
- 2. そのredexを簡約する
- 3. redexのルートを結果で更新する

#### 最外の関数適用が

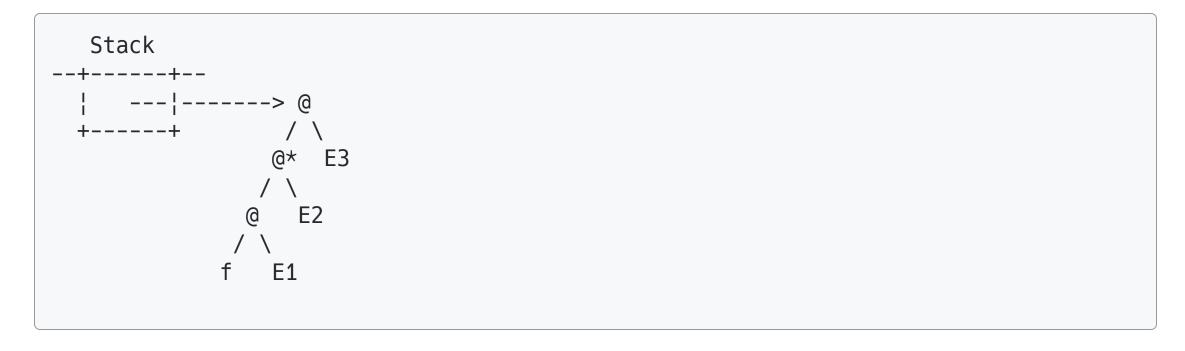
- スーパーコンビネータ適用の場合
  - この適用は必ずredexなので簡約(β簡約)
- 組込みのプリミティブ適用の場合
  - 2つの引数が共に正規形ならこの適用はredexなので簡約(δ簡約)
  - そうでないなら、引数を正規形にして(この適用がredexになって)から簡約 (δ簡約)

### 2.1.3 背骨を巻き戻して次のredexをみつける

ルートから適用ノードの左側を辿る



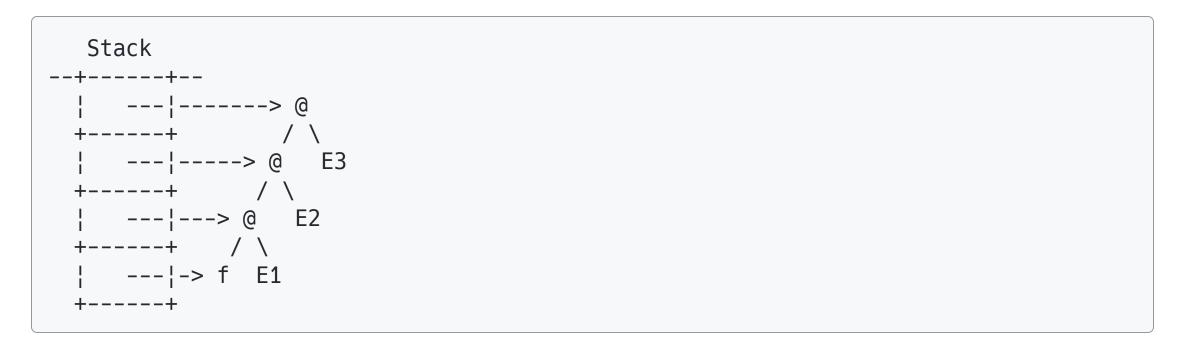
### ルートの適用ノードをスタックに積んで、左へ降りる



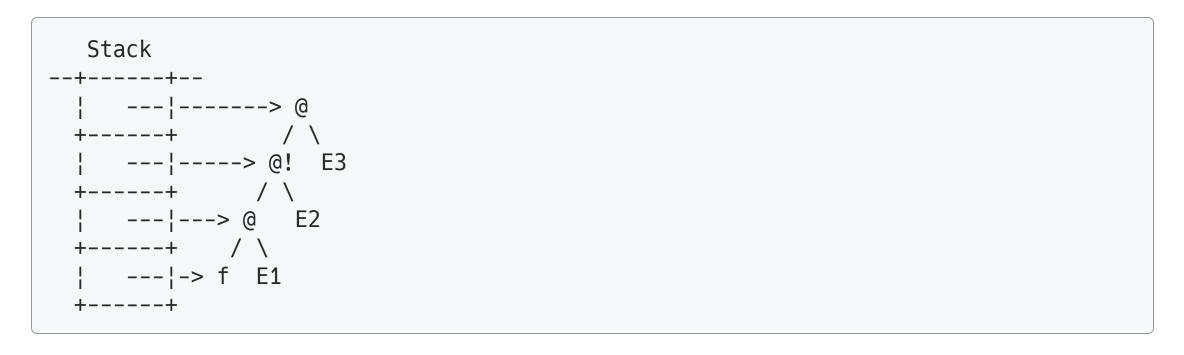
### 適用ノードをスタックに積んで、左に降りる

#### 適用ノードをスタックに積んで、左に降りる

f がスーパーコンビネータの場合: f をスタックに積んで、f のアリティ(ここでは2とする)を確認



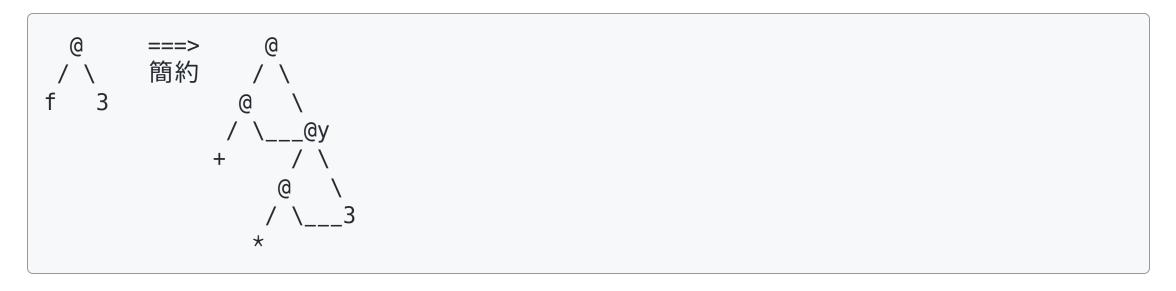
f がスーパーコンビネータの場合: アリティ(ここでは2とする)の分だけノードを上へ もどったところが、最外の簡約可能項のルートノード



f がプリミティブ(算術2項演算子)の場合:被演算子 E1 および E2 が共に正規形なら!のついたノードが最外の簡約可能項のルートノード。さもなければ、先に被演算子項を正規形にまで簡約する。

### 2.1.4 スーパーコンビネータの簡約可能項

- スーパーコンビネータの簡約可能項は、仮パラメータの出現位置を対応する実引数項へのポインタに置き換えたスーパーコンビネータ本体で置き換える
- 実引数項をコピーせず、ポインタを使い、共有していることに注意



# 2.1.5 更新

```
id x = x
f p = (id p) * p
main = f (sqrt 4)
f の簡約後
  id
    sqrt 4
                    sqrt
```

# は間接参照ノード

#### 2.1.6 CAF

fac20 = factorial 20

スーパーコンビネータ fac20 は CAF でかつ redex のルートなので、 fac20 の簡約結果 で上書きする

# 2.2 状態遷移システム

#### かけ算機械

#### 練習問題 2.1

かけ算マシンを手で走らせよ。初期状態 (2,3,0,0) からはじめ、各ステップで発火する規則を特定し、最終状態が (2,0,0,6) であることを確かめよ。

#### 練習問題 2.2

状態列の不変条件とは、すべての状態で真となる述語である。 n および m の初期値 N および M と現在の n、m、d、t の値のとの関係を見つけよ。これにより、このかけ算機械が、かけ算を実行するものであることを証明せよ。すなわち、以下を示せ。

- 1. 初期状態で不変条件が成り立つ
- 2. ある状態で、不変条件が成り立てば、次の状態でも不変状態が成り立つ
- 3. 不変条件と停止条件が成り立てば、t=N imes M である
- 4. このかけ算機械は停止する

#### 状態遷移システムは、以下の点で便利

- 低レベルの詳細にわずらわされない程度に抽象的
- 隠れた詳細に依存していないことが確認できる程度に具体的
- 状態遷移システムは直截に実行可能なHaskellのコードに変換できる

#### 練習問題 2.3

状態が最終状態であるかを判断する述語 multFinal :: multState -> Bool を定義し、初期状態 (2,3,0,0) からかけ算機械を走らせると、最終状態が (2,0,0,6) になることを示せ。

# 2.3 Mark 1: 最小雛形具体化グラフ簡約器

マシン状態: (stack, dump, heap, globals) の4つ組

- *stack*: ヒープ上のノードを特定するアドレスのスタック
  - $\circ$   $a_1:s$  という記法は、 $a_1$  がスタックトップ、s がのこりのスタックであることを示す
- dump: 正格なプリミティブ演算の引数評価に先立ち、スパインのスタックを記録
- heap: タグ付きノードを集めたもの
  - 。 h[a:node] という記法は、ヒープ h において、a はノードnodeのアドレスであることを示す
- globals: スーパーコンビネータおよびプリミティブを表すノードへのアドレス

#### ノードの表現

```
data Node
= NAp Addr Addr
| NSupercomb Name [Name] CoreExpr
| NNum Int
```

- NAP  $a_1 \ a_2$  はアドレス $a_1$ にあるノードのアドレス $a_2$ にあるノードへの適用を表す
- NSupercomb  $args\ body$  は引数argsと本体bodyをもつスーパーコンビネータを表す
- ullet NNum n は整数nを表す

### 状態遷移規則

(2.1)

#### 状態遷移規則

(2.2)

$$a_0:a_1:\cdots:a_n:s$$
  $d$   $h[a_0: exttt{NSupercomb}\,[x_1,\ldots,x_n]\,\,body]$   $f$   $a_r:s$   $d$   $h'$   $f$ 

#### 関数 instatiate の引数は、

- 1. 具体化する式
- 2. ヒープ
- 3. 名前からヒープ上のアドレスへのグローバルマッピング f をスタックにある引数名からヒープアドレスへのマッピングで拡張したもの

#### 返り値は、

• 新しいヒープと、新しく構成されたインスタンスの(ルートノードの)アドレス

# 2.3.2 実装の構造

```
run :: String -> String
run = showResult . eval . compile . parse
```

1. parse はソースコード(:: String)を構文解析して CoreProgram を構成する関数(教科書とは意味を変更)

```
parse :: String -> CoreProgram
```

2. compile は CoreProgram を雛形具体化機械の初期状態に変換

```
compile :: CoreProgram -> TiState
```

3. eval はプログラム実行関数、初期状態から状態遷移を繰り返し、最終状態にまで 遷移させ、結果は通過したすべての状態のリスト

```
eval :: TiState -> [TiState]
```

4. showResult は最終結果を整形して表示する

```
showResult :: [TiState] -> String
```

# 2.3.3 パーザ

Language モジュールをインポートする

## 2.3.4 コンパイラ

TiState

type TiState = (TiStack, TiDump, TiHeap, TiGlobals, TiStats)

1. TiStack スパインスタック、ヒープアドレス Addr のスタック

type TiStack = [Addr]

2. TiDump ダンプ、2.6節までは不要。ダミー

```
data TiDump = DummyTiDump
initalTiDump :: TiDump
initalTiDump = DummyTiDump
```

3. TiHeap は Node を格納するヒープ

```
type TiHeap = Heap Node
type Heap a = (Int, [Addr], [(Addr, a)])
```

Heap は使用アドレス数、未使用アドレス集合、アドレスと内容の2つ組のリスト、の3つを組にしたもの

4. TiGlobal はスーパーコンビネータ名とその定義が納められているヒープ上のアドレスの連想リスト

```
type TiGlobals = Assoc Name Addr
type Assoc a b = [(a, b)]
```

5. TiStats 実行時性能統計のためのデータ収集用、ひとまずステップカウント

```
type TiStats = Int
tiStatInitial :: TiStats
tiStatInitial = 0
tiStatIncSteps :: TiStats -> TiStats
tiStatIncSteps s = s + 1
tiStatGetSteps :: TiStats -> Int
tiStatGetSteps s = s
applyToStats :: (TiStats -> TiStats) -> TiState -> TiState
applyToStats f (stack, dump, heap, scDefs, stats)
  = (stack, dump, heap, scDefs, f stats)
```

#### compile

buildInitialHeap プログラムから NSupercomb ノードを含むヒープと、スーパーコンビネータ名とヒープ上のアドレスの対応を示す連想リストを構成する。

## 2.3.5 評価器

#### 最終状態の判定

• 計算はスタックに単一の数またはデータオブジェクトが含まれる状態になったときにの み停止する

## step ある状態から1つ次の状態への遷移

```
step :: TiState -> TiState
step state = case state of
  (stack, dump, heap, globals, stats) -> dispatch (hLookup heap (head stack))
where
  dispatch (NNum n) = numStep state n
  dispatch (NAp a1 a2) = apStep state a1 a2
  dispatch (NSupercomb sc args body) = scStep state sc args body
```

## NNum ノードと NAp ノードはやさしい

```
numStep :: TiState -> Int -> TiState
numStep state n = error "Number applied as a function"

apStep :: TiState -> Addr -> Addr -> TiState
apStep state a1 a2 = case state of
  (stack, dump, heap, globals, stats) -> (a1:stack, dump, heap, globals, stats)
```

#### スーパーコンビネータの適用:

- 1. 本体を具体化、引数名をスタックにあるアドレスと結びつける(規則2.2)
- 2. 簡約可能項のルートを含む引数をスタックから除去、簡約結果をスタックにプッシュ

(Mark 1 では更新は行わない)

```
scStep :: TiState -> Name -> [Name] -> CoreExpr -> TiState
scStep state scName argNames body = case state of
  (stack, dump, heap, globals, stats)
    -> (stack', dump, heap', globals, stats)
    where
        stack' = resultAddr : drop (length argNames + 1) stack
        (heap', resultAddr) = instantiate body heap env
        env = argBindings ++ globals
        argBindings = zip argNames (getargs heap stack)
```

# getargs スタックにある引数名に対応するヒープ上の実引数ノードのアドレスを取得する

```
getargs :: TiHeap -> TiStack -> [Addr]
getargs heap stack = case stack of
  sc:stack' -> map getarg stack'
  where
    getarg addr = arg
    where
        NAp fun arg = hLookup heap addr
[] -> error "Empty stack"
```

#### instantiate

```
-> TiHeap -- Heap before instatiation
         -> Assoc Name Addr -- Association of names to address
         -> (TiHeap, Addr) -- Heap after instatiation, and address of root of instance
instantiate expr heap env = case expr of
 ENum n
       -> hAlloc heap (NNum n)
 EAp e1 e2
                 -> hAlloc heap2 (NAp a1 a2)
  where
    (heap1, a1) = instantiate e1 heap env
    (heap2, a2) = instantiate e2 heap1 env
 EVar v
               -> (heap, aLookup env v (error ("Undefined name " ++ show v)))
 EConstr tag arity -> instantiateConstr tag arity heap env
 ELet isrec defs body -> instantiateLet isrec defs body heap env
 ELam vs e -> error "Can't instantiate lambda abstractions"
```

## 2.3.6 結果の整形

#### showResults

#### showState

#### showStack

```
showStack :: TiHeap -> TiStack -> IseqRep
showStack heap stack
 = iConcat
    [ iStr "Stack ["
    , iIndent (iInterleave iNewline (map showStackItem stack))
    , iStr " ]"
   where
      showStackItem addr
        = iConcat [ showFWAddr addr, iStr ": "
                  , showStkNode heap (hLookup heap addr)
```

#### showStkNode

#### showNode

#### showAddr

```
showAddr :: Addr -> IseqRep
showFWAddr :: Addr -> IseqRep
showFWAddr addr = iStr (space (4 - length str) ++ str)
where
   str = show addr
```

#### showStats

#### ここまでの実装をテストせよ

showState を改変してヒープの内容をすべて表示するようにせよ。

```
showState :: TiState -> IseqRep
showState (stack, dump, heap, globals, stats)
 = iConcat [ showStack heap stack, iNewline
            , showHeap heap, iNewline
showHeap :: TiHeap -> IseqRep
showHeap heap = case heap of
 (_,_,contents) -> iConcat
   [ iStr "Heap ["
    , iIndent (iInterleave iNewline (map showHeapItem contents))
    , iStr " 1"
 where
    showHeapItem (addr, node)
      = iConcat [ showFWAddr addr, iStr ": "
                , showNode node
```

scStep が引数が充満でない場合に適切なエラーを表示するようにせよ。

```
scStep :: TiState -> Name -> [Name] -> CoreExpr -> TiState
scStep state scName argNames body = case state of
  (stack, dump, heap, globals, stats)
    | length stack < length argNames + 1</pre>
      -> error "Too few argments given"
    ! otherwise
      -> (stack', dump, heap', globals, stats)
   where
      stack' = resultAddr : drop (length argNames + 1) stack
      (heap', resultAddr) = instantiate body heap env
      env = argBindings ++ globals
      argBindings = zip argNames (getargs heap stack)
```

より多くの実行時情報を収集するようにせよ。

- プリミティブの簡約とスーパーコンビネータの簡約を分けて計数
- ヒープ操作(とくにアロケーション)の回数
- スタックの最大深さ

instantiate に渡す環境 env は、

env = argBindings ++ globals

と定義されているが、これを

env = globals ++ argBindings

とするとどうなるか

eval の定義を

としたほうが、わかりやすそうにみえるが、この定義には欠点がある。それはどのようなものか。

# Mark 2: let(rec) 式

instantiate を拡張して、ELet 項に対応する

instantiate の定義に非再帰的 let 式に対応する等式を加えよ。

ELet nonRecursive defs body を具体化する

- 1. defs の各定義の右辺を具体化する
- 2. defs の各定義の左辺(名前)と新たに具体化されたものとを結びつけて環境を拡張する
- 3. 拡張された環境と式本体を渡して instantiate を呼ぶ

再帰的 let に instantiate が対応できるようにせよ

(ヒント)前問のステップ 1 で instantiate に既存の環境を渡していたが、代りにステップ 2 で拡張した環境を渡すようにする

## テストプログラム

# 結果は 4 である

以下のプログラムを実行するとどうなるか?

main = letrec f = f x in f

Haskellのような強い型付けを行う言語でも同じ問題が起きるか?

Mark 3: 更新の追加

# 遷移規則(2.2)の代わりに(2.3)を使う

(2.3)

$$egin{aligned} a_0:a_1:\cdots:a_n:s & d & h[a_0: exttt{NSupercomb}\,[x_1,\ldots,x_n]\,\,body] & f \ \Longrightarrow & a_r:s & d & h'[a_n: exttt{NInd}\,\,a_r] & f \end{aligned}$$

ここで、 $(h', a_r) = instantiate\ body\ h\ f[x_1\mapsto a_1, \ldots, x_n\mapsto a_n]$ である。instantiateが返す結果のルートへの間接参照 $a_r$ で、ノード $a_n$ を上書きする。

スパインをアンワインドするときに間接参照に出会す可能性があるので、その場合に対応する規則(2.4)を追加する。

(2.4)

$$egin{array}{lll} a:s & d & h[a: \mathtt{NInd}\ a_1] & f \ \Longrightarrow & a_1:s & d & h & f \end{array}$$

## 新しい規則を実装するには以下を行う

1. Node 型に新しいデータ構成子 NInd を追加。それに伴い showNode もこれに対応する必要がある

```
data Node = NAp Addr Addr
| NSupercomb Name [Name] CoreExpr -- Supercombinator
| NNum Int -- Number
| NInd Addr -- Indirection
```

- 2. 規則(2.3)対応: リデックスのルートを hUpdate をつかって、結果の間接参照に置き換えるように、 scStep を変更する
- 3. 規則(2.4)対応: dispatch の定義に間接参照を扱う等式を追加

間接参照による更新を実装せよ。以下のプログラムの実行を、Mark 1 と Mark 3 で走らせ効果を比較せよ。

```
id x = x;
main = twice id 3
```

まず、手で簡約を行ってどう計算されるか確認せよ

```
main = twice twice id 3
```

と定義した場合どうなるか。

## 2.5.1 間接参照の低減

あきらかに不要な間接参照を構成しないようにする。

```
instantiateAndUpdate
:: CoreExpr -- ^ スーパーコンビネータの本体
-> Addr -- ^ 更新するノードのアドレス
-> TiHeap -- ^ 具体化前のヒープ
-> Assoc Name Addr -- ^ 仮引数からアドレスへの連想リスト
-> TiHeap -- ^ 具体化語のヒープ
```

# 式が関数適用であったときの instantiateAndUpdate の定義

```
instantiateAndUpdate (EAp e1 e2) updAddr heap env
= hUpdate heap2 updAddr (NAp a a2)
where
    (heap1, a1) = instantiate e1 heap env
    (heap2, a2) = instantiate e2 heap1 env
```

更新の必要があるのはルートノードだけなので、 e1 および e2 に対しては元の instantiate にする

instantiateAndUpdate の定義を完成せよ。以下の点に留意すること

- 具体化する式が単なる変数なら、間接参照が必要である。(なぜか)
- let(rec) 式に対する再帰的な具体化の部分は注意深く考えること。

scStep を instantiateAndUpdate を呼ぶように変更せよ

• scStep でおこなっていた更新のコードは取り除け。

簡約段数とヒープアロケーション数を計測し、効果を確かめよ。

# 2.6 Mark 4: 算術演算の追加

### 2.6.1 算術に対する遷移規則

符号反転の遷移規則(2.5) 引数が評価済みの場合

符号反転の遷移規則(2.6) 引数未評価の場合

スタックをダンプに積んでから、引数の評価にはいる。評価が済んだら、ダンプに積んだ スタックを復帰する(2.7)

引数のルートノードが間接参照になっている場合がある。 (2.1)の特殊な場合として、 (2.8)を導入する

(2.8)

$$egin{array}{lll} a:s & d & h \left[egin{array}{ccccc} a: & ext{NAp} \ a_1 & a_2 \ a_2 & : & ext{NInd} \ a_3 \end{array}
ight] & f \ \Longrightarrow & a:s & d & h \left[a: ext{NAp} \ a_1 \ a_3 \end{array}
ight] & f \end{array}$$

(2.8)が機能するよう、(2.6)を変更して(2.9)にする必要がある

(2.9)

### 練習問題

加算に対する遷移規則を書け。(ほかの二項算術演算子も実質同じである。)

## 2.6.2 算術演算式の実装

- 1. TiDump をスタックのスタックとして再定義
- 2. 名前 n 値 p のプリミティブを表すヒープノード NPrim Name Primitive を追加
- 3. showNode を拡張
- 4. Primitive を定義
- 5. NPrim ノードを初期ヒープにわりあてるよう buildInitialHeap を拡張
- 6. プリミティブの名前と値の連想リスト primitives
- 7. allocatePrim を allocateSc を参考に実装
- 8. step の dispatch を NPrim に対応して、 primStep を呼ぶものとして拡張

- 9. primStep を実装
  - primNeg を実装
- 10. 遷移規則(2.7)を実装するために numStep を変更
- 11. 同様に遷移規則(2.8)を実装するために apStep を変更する
- 12. tiFinal を変更する

符号反転演算を実装し以下のプログラムでテストせよ。

### 算術演算の実装には 2項算術演算汎用の

```
primArith :: TiState -> (Int -> Int -> Int) -> TiState
```

を用意して、これを使う

primArith を実装し、テストせよ

## 2.7 Mark 5: 構造をもつデータ

case 式が実装できるといいのだが、雛形具体化機械の枠組みでは難しい代りに if 、casePair 、caseList などの組み込み関数を使い特定の構造をもつデータ型を扱えるようにする

なぜ雛形具体化機械で case 式を導入するのは難しいのか (ヒント) case 式に対して instantiate がすべきことを考えよ

## 構造をもつデータの構築

状態遷移規則(2.10)

```
a:a_1:\cdots:a_n:[] \quad d \qquad h \left[egin{array}{ccccc} a &:& 	exttt{NPrim} & (	exttt{PrimConstr} \ t \ n) \ a_1 &:& 	exttt{NAp} \ a \ b_1 \ \cdots \ a_n &:& 	exttt{NAp} \ a_{n-1} \ b_n \ \end{array}
ight] \quad f \ \Longrightarrow \qquad a_n:[] \quad d \qquad h \left[a_n: 	exttt{NData} \ t \left[b_1, \ldots, b_n
ight]
ight] \qquad f \ \end{array}
```

# 2.7.2 条件式

条件式に関する状態遷移規則を書け

コア言語で、 or 、 xor 、 not の定義を与えよ。 これらの定義を extraPreludeDefs に 追加せよ。

# 2.6.3 構造をもつデータの実装

これまでの変更をすべて行え。これで意味のある再帰定義関数がかける。

```
fac n = if (n == 0) 1 (n * fac (n - 1));
main = fac 3
```

# 2.7.4 対 (ペア)

casePair に対応する状態遷移規則を書け。

#### 必要な規則は

- 第1引数が評価済みの場合、簡約を実行するための規則
- 第1引数が未評価の場合、評価を開始するための規則

実装した上で以下のプログラムでテストせよ。

main = fst (snd (fst (MkPair (MkPair 1 (MkPair 2 3)) 4)))

# 2.7.5 リスト

コア言語で Cons 、Nil 、head 、tail の定義を与えよ。空リストに対して head やtail を適用したときに返るものとして、新たなプリミティブ abort を導入し、Haskellの error を呼び出して、プログラムを停止するようにせよ。