Prettyprinting Intermediate Computations from a Bytecode

January 18, 2017

1 Programs

Programs are defined like this. Variable accesses have been converted to deBruijn indices when the program was converted from an OCaml one.

```
type op = Add | Sub | Mul | Div
type prog =
 Int of int
| Bool of bool
| Var of int
| Eq of prog * prog
| Op of prog * op * prog
| Apply of prog * prog
| Lambda of prog
| Let of prog * prog
| If of prog * prog * prog
For example, the OCaml program
let x = 5 in if x = 4 then 1 else (fun x \rightarrow x + 1) 2
may be represented as:
Let(Int 5,
    If(Eq(Var 1, Int 4),
       Int 1,
       Apply(Lambda(Op(Var 1, Add, Int 1), Int 2))))
```

2 Compilation Scheme

The abstract machine instructions are as followed (Leroy plus BOOL, IF, EQ)

```
EMPTY
INT(integer)
BOOL(boolean)
OP(op)
EQ
ACCESS(integer)
CLOSURE(instructions)
LET
ENDLET
```

```
APPLY
RETURN
IF
```

Here is the compilation scheme, again extended from Leroy:

```
C(Int(i)) = INT(i)
                                     \mathcal{C}(\mathsf{Bool}(b)) = \mathsf{BOOL}(b)
                                  \mathcal{C}(\mathsf{Op}(a, \oplus, b)) = \mathcal{C}(a); \mathcal{C}(b); \mathsf{OP}(\oplus)
                                      C(Eq(a,b)) = C(a); C(b); EQ
                                       C(Var(n)) = ACCESS(n)
                                 C(Lambda(a)) = CLOSURE(C(a); RETURN)
                                    C(\text{Let}(a,b)) = C(a); \text{LET}; C(b); \text{ENDLET}
                                C(Apply(a,b)) = C(a); C(b); APPLY
                                   \mathcal{C}(\mathrm{If}(a,b,c)) = \mathcal{C}(\mathrm{Lambda}(b)); \mathcal{C}(\mathrm{Lambda}(c)); \mathcal{C}(a); \mathrm{IF}
let x = 5 in if x = 4 then 1 else (fun x \rightarrow x + 1) 2
compiles to (including an EMPTY at the end):
```

INT 5

LET

CLOSURE

So our example

INT 1

RETURN

CLOSURE

CLOSURE

ACCESS 1

INT 1

OP +

RETURN

INT 2

APPLY

RETURN

ACCESS 1

INT 4

EQ

BRANCH

ENDLET

EMPTY

Evaluation Scheme

Here is the evaluation scheme \mathcal{E} , again extended from Leroy.

Machine state before			Machine state after		
Code	Env	Stack	Code	Env	Stack
INT(i); c	e	s	c	e	i.s
BOOL(b); c	e	s	c	e	b.s
$\mathrm{OP}(\oplus);c$	e	i.i'.s	c	e	$\oplus (i,i').s$
EQ; c	e	i.i'.s	c	e	(i=i').s
ACCESS(n); c	e	s	c	e	e(n).s
CLOSURE(c'); c	e	s	c	e	c'[e].s
$\operatorname{LET}; c$	e	v.s	c	v.e	s
ENDLET; c	v.e	s	c	e	s
APPLY;c	e	v.c'[e'].s	c'	v.e'	c.e.s
RETURN;c	e	v.c'.e'.s	c'	e'	v.s
IF;c	e	T.c'[e'].c''[e''].s	c'	e'	c[e].s
IF;c	e	F.c'[e'].c''[e''].s	c''	e''	c[e].s

The final result is at the top of the stack when the code is EMPTY.

4 Decompilation Scheme

We need to be able to decompile:

- Any program which has been compiled by the compilation scheme above.
- Certain incomplete evaluations under the evaluation scheme above. That is to say, given (c, s) we can decompile a program which represents the evaluation at that stage. We need not be able to decompile arbitrary (c, e, s) triples.

We add names to VarAccess, Lambda and Let:

```
type prog =
   Int of int
| Bool of bool
| VarAccess of name * int
| Eq of prog * prog
| Op of prog * op * prog
| Apply of prog * prog
| Lambda of name * prog
| Let of name * prog * prog
| If of prog * prog * prog
```

Similarly, we add names to the ACCESS, CLOSURE and LET instructions (not required for evaluation, but only for decompilation).

```
EMPTY
INT(integer)
BOOL(boolean)
OP(op)
EQ
ACCESS(name, integer)
CLOSURE(name, instructions)
LET(name)
```

ENDLET APPLY RETURN IF

Decompilation is performed by going through the instructions in order, holding a stack a little like the evaluation stack, but which may also contain decompiled program fragments – the empty stack is written {}. When we have gone through all the instructions, the final program is at the top of the stack. We do not need the environment, since we are not running the code, just decompiling it.

```
\mathcal{D}(\mathsf{EMPTY}, v.s) = v
\mathcal{D}(\mathsf{INT}(i); c, s) = \mathcal{D}(c, \mathsf{Int}(i).s)
\mathcal{D}(\mathsf{BOOL}(i); c, s) = \mathcal{D}(c, \mathsf{Bool}(b).s)
\mathcal{D}(\mathsf{OP}(\oplus); c, i.i'.s) = \mathcal{D}(c, \mathsf{Op}(i, \oplus, i').s)
\mathcal{D}(\mathsf{EQ}; c, i.i'.s) = \mathcal{D}(c, \mathsf{Eq}(i, i').s)
\mathcal{D}(\mathsf{ACCESS}(n, l); c, s) = \mathcal{D}(c, \mathsf{VarAccess}(n, l).s)
\mathcal{D}(\mathsf{CLOSURE}(n, c'); c, s) = \mathcal{D}(c, c'[n, \{\}].s)
\mathcal{D}(\mathsf{LET}(n); c, v.s) = \mathsf{Let}(n, v, \mathcal{D}(c, s))
\mathcal{D}(\mathsf{ENDLET}; c, s) = \mathcal{D}(c, s)
\mathcal{D}(\mathsf{APPLY}; c, v.c'[n, e'].s) = \mathsf{Apply}(\mathsf{Lambda}(n, \mathcal{D}(c', \{\})), v)
\mathcal{D}(\mathsf{RETURN}; c, v.c'.e'.s) = \mathcal{D}(c', v.s)
\mathcal{D}(\mathsf{RETURN}; c, c'.e'.s) = \mathcal{D}(c, s)
\mathcal{D}(\mathsf{IF}; c, e.c'[e'].c''[e''].s) = \mathcal{D}(c, \mathsf{If}(e, \mathcal{D}(c', s), \mathcal{D}(c'', s)).s)
```

This decompiler works for:

- Any program-stack pair $(P, \{\})$ where P was compiled by C above.
- program, stack pair (P, S) which is an intermediate state of the evaluation procedure \mathcal{E} (minus the environment) where P begins with OP or APPLY.

Our example program decompiles properly from bytecode.

5 Worked examples

The following pages contain a worked example of the compilation C, the evaluation E, and full-program and partial-evaluation invocations of the decompiler D.

Compilation under C:

C(Let(Int 5, If(Eq(Var 1, Int 4), Int 1, Apply(Lambda(Op(Var 1, Add, Int 1), Int 2))))
Rule C-Let
C(Int 5); LET; C(If(Eq(Var 1, Int 4), Int 1, Apply(Lambda(Op(Var 1, Add, Int 1), Int 2)))); ENDLET
Rule C-Int
INT 5; LET; C(If(Eq(Var 1, Int 4), Int 1, Apply(Lambda(Op(Var 1, Add, Int 1), Int 2)))); ENDLET
Rule C-If
INT 5; LET; C(Lambda (Int 1)); C(Lambda(Apply(Lambda(Op(Var 1, Add, Int 1), Int 2)))); C(Eq(Var 1, Int 4)); IF; ENDLET
Rule C-Eq then Rule C-Eq then Rule C-Eq
INT 5; LET; C(Lambda (Int 1)); C(Lambda(Apply(Lambda(Op(Var 1, Add, Int 1), Int 2)))); ACCESS 1; INT 4; EQ; IF; ENDLET
Rule C-Lambda then Rule C-Int
INT 5; LET; CLOSURE [INT 1; RETURN]; C(Lambda(Apply(Lambda(Op(Var 1, Add, Int 1), Int 2)))); ACCESS 1; INT 4; EQ; IF; ENDLET
Rule C-Lambda
INT 5; LET; CLOSURE [INT 1; RETURN]; CLOSURE [C(Apply(Lambda(Op(Var 1, Add, Int 1), Int 2)))); RETURN]; ACCESS 1; INT 4; EQ; IF; ENDLET

Rule $\mathcal{C} ext{-}$ Apply

INT 5; LET; CLOSURE [INT 1; RETURN]; CLOSURE [$\mathcal{C}(Lambda (Op (Var 1, Add, Int 1)); \mathcal{C}(Int 2); APPLY; RETURN]; ACCESS 1; INT 4; EQ; IF; ENDLET$

Rule $\mathcal C$ -Int then Rule $\mathcal C$ -Lambda then Rule $\mathcal C$ -Op then Rule $\mathcal C$ -Var then Rule $\mathcal C$ -Int

INT 5; LET; CLOSURE [INT 1; RETURN]; CLOSURE [CLOSURE [ACCESS 1; INT 1; OP +; RETURN]; INT 2; APPLY; RETURN]; ACCESS 1; INT 4; EQ; IF; ENDLET

Evaluation under \mathcal{E} . Stacks and environments are written {items}, and a closure on the stack is written [instructions]{environment}. Environments may be put on the stack.

	Machine state after					
Instruction	Code	Env	Stack			
-	INT 5; LET; CLOSURE [INT 1; RETURN]; CLOSURE [CLO-	{}	 {}			
	SURE [ACCESS 1; INT 1; OP +; RETURN]; INT 2; APPLY;					
	RETURN]; ACCESS 1; INT 4; EQ; IF; ENDLET					
INT	LET; CLOSURE [INT 1; RETURN]; CLOSURE [CLOSURE	{}	{5}			
	[ACCESS 1; INT 1; OP +; RETURN]; INT 2; APPLY; RE-					
	TURN]; ACCESS 1; INT 4; EQ; IF; ENDLET					
LET	CLOSURE [INT 1; RETURN]; CLOSURE [CLOSURE [AC-	{5}	{}			
	CESS 1; INT 1; OP +; RETURN]; INT 2; APPLY; RETURN];					
	ACCESS 1; INT 4; EQ; IF; ENDLET					
CLOSURE	CLOSURE [CLOSURE [ACCESS 1; INT 1; OP +; RETURN];	{5}	{[INT 1; RETURN]{5}}			
	INT 2; APPLY; RETURN]; ACCESS 1; INT 4; EQ; IF; ENDLET					
CLOSURE	ACCESS 1; INT 4; EQ; IF; ENDLET	{5}	{[CLOSURE [ACCESS 1; INT 1; OP +; RETURN]{5}; INT 2;			
			APPLY; RETURN]; [INT 1; RETURN]{5}}			
ACCESS	INT 4; EQ; IF; ENDLET	{5}	{5; [CLOSURE [ACCESS 1; INT 1; OP +; RETURN]{5}; INT			
			2; APPLY; RETURN]; [INT 1; RETURN]{5}}			
INT	EQ; IF; ENDLET	{5}	{4; 5; [CLOSURE [ACCESS 1; INT 1; OP +; RETURN]{5}; INT			
			2; APPLY; RETURN]; [INT 1; RETURN]{5}}			
EQ	IF; ENDLET	{5}	{false; [CLOSURE [ACCESS 1; INT 1; OP +; RETURN]{5};			
			INT 2; APPLY; RETURN]; [INT 1; RETURN]{5}}			
IF	CLOSURE [ACCESS 1; INT 1; OP +; RETURN]; INT 2; APPLY;	{5}	{[ENDLET]; {5}}			
	RETURN					
CLOSURE	INT 2; APPLY; RETURN	{5}	{[ACCESS 1; INT 1; OP +; RETURN]{5}; [ENDLET]; {5}}			
INT	APPLY; RETURN	{5}	{2; [ACCESS 1; INT 1; OP +; RETURN]{5}; [ENDLET]; {5}}			
APPLY	ACCESS 1; INT 1; OP +; RETURN	{2; 5}	{[RETURN]; {5}; [ENDLET]; {5}}			
ACCESS	INT 1; OP +; RETURN	{2; 5}	{2; [RETURN]; {5}; [ENDLET]; {5}}			
INT	OP +; RETURN	{2; 5}	{1; 2; [RETURN]; {5}; [ENDLET]; {5}}			
OP	RETURN	{2; 5}	{3; [RETURN]; {5}; [ENDLET]; {5}}			
RETURN	EMPTY	{2; 5}	{3; [RETURN]; {5}; [ENDLET]; {5}}			
RETURN	EMPTY	{5}	{3; [ENDLET]; {5}}			
ENDLET	EMPTY	{}	{3}			
EMPTY						

Decompilation under \mathcal{D} of a program compiled under \mathcal{C} , with an empty stack to begin, since the program is unexecuted.

 $\mathcal{D}(\text{INT 5}; \text{LET x}; \text{CLOSURE [INT 1}; \text{RETURN]}; \text{CLOSURE [CLOSURE [ACCESS (x, 1); INT 1; OP +; RETURN]}; \text{INT 2}; \text{APPLY}; \text{RETURN]}; \text{ACCESS (x, 1)}; \text{INT 4}; \text{EQ; IF; ENDLET, } \})$

Rule \mathcal{D} -CONST

D(LET x; CLOSURE [INT 1; RETURN]; CLOSURE [CLOSURE [ACCESS (x, 1); INT 1; OP +; RETURN]; INT 2; APPLY; RETURN]; ACCESS (x, 1); INT 4; EQ; IF; ENDLET, {Int 5})

Rule \mathcal{D} -LET

Let $(x, Int 5, D(CLOSURE [INT 1; RETURN]; CLOSURE [CLOSURE [ACCESS (x, 1); INT 1; OP +; RETURN]; INT 2; APPLY; RETURN]; ACCESS (x, 1); INT 4; EQ; IF; ENDLET, <math>\{\}$)

Rule \mathcal{D} -CLOSURE

Let $(x, Int 5, \mathcal{D}(CLOSURE [ACCESS (x, 1); INT 1; OP +; RETURN]; INT 2; APPLY; RETURN]; ACCESS (x, 1); INT 4; EQ; IF; ENDLET, {CLOSURE [INT 1; RETURN]}))$

Rule \mathcal{D} -CLOSURE

Let $(x, Int 5, D(ACCESS(x, 1); INT 4; EQ; IF; ENDLET, \{CLOSURE [ACCESS(x, 1); INT 1; OP +; RETURN; INT 2; APPLY; RETURN]; CLOSURE [INT 1; RETURN]\}))$

Rule \mathcal{D} -ACCESS

Let $(x, Int 5, D(INT 4; EQ; IF; ENDLET, {VarAccess}(x, 1); CLOSURE [ACCESS}(x, 1); INT 1; OP +; RETURN; INT 2; APPLY; RETURN]; CLOSURE [INT 1; RETURN])))$

Rule \mathcal{D} -INT

Let $(x, Int 5, D(EQ; IF; ENDLET, \{Int 4; VarAccess(x, 1); CLOSURE [ACCESS(x, 1); INT 1; OP +; RETURN]; INT 2; APPLY; RETURN; CLOSURE [INT 1; RETURN]\}))$

Rule \mathcal{D} -EO

Let $(x, Int 5, D(IF; ENDLET, \{Eq(VarAccess (x, 1), Int 4); CLOSURE [ACCESS (x, 1); INT 1; OP +; RETURN]; INT 2; APPLY; RETURN; CLOSURE [INT 1; RETURN]\}))$

Rule \mathcal{D} -IF

Let $(x, Int 5, D(ENDLET, \{If (Eq (VarAccess (x, 1), Int 4), D(CLOSURE [ACCESS (x, 1); INT 1; OP +; RETURN]; INT 2; APPLY; RETURN, {}), D(CLOSURE [INT 1; RETURN], {}))$

Rule \mathcal{D} -ENDLET

Let $(x, Int 5, If (Eq (VarAccess (x, 1), Int 4), \mathcal{D}(CLOSURE [ACCESS (x, 1); INT 1; OP +; RETURN]; INT 2; APPLY; RETURN, {}))$

Rule \mathcal{D} -CLOSURE then \mathcal{D} -INT then \mathcal{D} -RETURN

Let $(x, Int 5, If (Eq (VarAccess (x, 1), Int 4), \mathcal{D}(CLOSURE [ACCESS (x, 1); INT 1; OP +; RETURN]; INT 2; APPLY; RETURN, {}), Int 1))$

Rule \mathcal{D} -CLOSURE

Let $(x, Int 5, If (Eq (VarAccess (x, 1), Int 4), \mathcal{D}(INT 2; APPLY; RETURN, {CLOSURE [ACCESS (x, 1); INT 1; OP +; RETURN]}), Int 1))$ Rule \mathcal{D} -INT

Let $(x, Int 5, If (Eq (VarAccess (x, 1), Int 4), \mathcal{D}(APPLY; RETURN, \{Int 2; CLOSURE [ACCESS (x, 1); INT 1; OP +; RETURN]\}), Int 1))$ Rule \mathcal{D} -APPLY

Let $(x, Int 5, If (Eq (VarAccess (x, 1), Int 4), Apply (\mathcal{D}(ACCESS (x, 1); INT 1; OP +; RETURN, {}), Int 2), Int 1))$ Rule \mathcal{D} -ACCESS then \mathcal{D} -INT then \mathcal{D} -OP then \mathcal{D} -RETURN

Let(x, Int 5, If(Eq(VarAccess (x, 1), Int 4), Apply(Lambda(Op(Var 1, Add, Int 1)), Int 2), Int 1))

D(IF; ENDLET, false; [CLOSURE [ACCESS(x, 1); INT 1; OP +; RETURN]; INT 2; APPLY; RETURN]; [INT 1; RETURN]) Rule \mathcal{D} -IF $\mathcal{D}(\text{ENDLET}, \text{If } (\text{false}, \mathcal{D}([\text{INT 1}; \text{RETURN}], \{\})) \mathcal{D}([\text{CLOSURE}[\text{ACCESS}(x, 1); \text{INT 1}; \text{OP} +: \text{RETURN}]; \text{INT 2}; \text{APPLY}; \text{RETURN}], \{\}), \{\})$ Rule \mathcal{D} -ENDLET If (Bool false, $\mathcal{D}([INT 1; RETURN], \{\})$, $\mathcal{D}([CLOSURE [ACCESS(x, 1); INT 1; OP +: RETURN]; INT 2; APPLY; RETURN] \{\})$,) Rule \mathcal{D} -INT then \mathcal{D} -RETURN If (Bool false, Int 1, D([CLOSURE [ACCESS(x, 1); INT 1; OP +: RETURN]; INT 2; APPLY; RETURN]{}),) Rule \mathcal{D} -CLOSURE If (Bool false, Int 1, $\mathcal{D}([INT 2; APPLY; RETURN], \{[ACCESS(x, 1); INT 1; OP +: RETURN]\}))$ Rule \mathcal{D} -INT If (Bool false, Int 1, $\mathcal{D}([APPLY; RETURN], \{Int 2; [ACCESS(x, 1); INT 1; OP +: RETURN]\}))$ Rule \mathcal{D} -APPLY If (Bool false, Int 1, Apply (Lambda(x, $\mathcal{D}(ACCESS(x, 1); INT 1; OP +: RETURN, \{\}), Int 2)))$ Rule \mathcal{D} -ACCESS If (Bool false, Int 1, Apply (Lambda(x, $\mathcal{D}(INT 1; OP +; RETURN, \{VarAccess(x, 1)\}), Int 2)))$ Rule \mathcal{D} -INT If (Bool false, Int 1, Apply (Lambda(x, $\mathcal{D}(OP +; RETURN, \{Int 1; VarAccess(x, 1)\}), Int 2))))$ Rule \mathcal{D} -OP If (Bool false, Int 1, Apply (Lambda(x, $\mathcal{D}(RETURN, \{Op(VarAccess(x, 1), Add, Int 1)\}), Int 2)))$ Rule \mathcal{D} -RETURN If (Bool false, Int 1, Apply (Lambda(x, Op(VarAccess(x, 1), Add, Int 1), Int 2)))

Decompilation under \mathcal{D} of a program compiled under \mathcal{C} and partially evaluated with \mathcal{D} .

This is the program if false then 1 else (fun $x \rightarrow x + 1$) 2, as required.