Notes on PoPL

Apiros3

First Version : Apr 24, 2025
Last Update : Apr 24, 2025

Contents

1 Definitional Interpreter

In this section we aim to explain the concept and fundamentals of a definitional interpreter under various contexts.

We will base the language to be implemented to be based off the "Fun" programming language as specificed by the Principles of Programming Languages course at Oxford University.

1.1 Defining Fun

The first step in describing a fixed program is to specify the set of legal phrases (the concrete syntax), then by describing in the language that interprets the program a set of trees that capture the structure of legal phrases (the abstract syntax). For the rest of the section, we will form a Haskell datatype that captures the structure of these legal phrases.

At the simplest level, we have a function

```
parse :: String → Phrase
```

where the `String` is any line of text in the concrete syntax, then produces a corresponding tree in the abstract syntax, which we give the type `Phrase`.

Often the abstract syntax is much more simple than the concrete syntax, as the concrete syntax allows for convinient abbreviations (syntactic sugars).

Basic fun does not have type-checking, so we regard the set of valid expressions to be produced by a context free grammar, rejecting those which do not "make sense".

1.1.1 Abstract Syntax

The abstract syntax of a language can be expressed as a collection of mutually dependent datatype definitions. In Fun, there is `Expr` for expressions, `Defn` for definitions, and `Phrase` for top level phrases.

```
data Expr =
      Number Integer
    | Variable Ident
    | Apply Expr [Expr]
    | If Expr Expr Expr
    | Lambda [Ident] Expr
    | Let Defn Expr
```

Note that we can have functions with no arguemnts.

The definitions that appear after `let` also appear in the abstract syntax as

```
data Defn =
      Val Ident Expr
    | Rec Ident Expr
```

which correspond to giving variables denoted by `Ident` expressions in `Expr`.

In this way, the concrete form val $x(x_1, \ldots, x_n)$ e is syntactic sugar for val x = lambda$(x_1, \ldots, x_n)$ e. We use empty "()" if the function has no inputs, and the constructor `Rec` is for a definition that starts with a lambda (but is not enforced at the datatype level).

The top-level phrase that is typed in the prompt (or included as code in fun) is either an expression which is to evaluated, or a definition to be added into the environment. So,

```
data Phrase =
      Calculate Expr
    | Define Defn
```

Remark 1.1.1. In the abstract syntax for fun, identifiers (Ident) are represented by strings. This limits efficiencies, and in more optimized languages have indexing into a global list of identifiers (and thus can avoid string comparisons).

1.1.2 Interpreter for Fun

The main component of an interpreter is a function `eval` which takes an abstract syntax tree with an environment and turns it into a value of that expression. Specifically,

```
eval :: Expr → Env → Value
```

where `Env` is the type of environments which mapps identifiers to values, with `Value` representing possible values computed by Fun programs.

At the simplest level, values are denoted by

```
data Value =
      Function ([Value] → Value)
    | IntVal Integer
    | BoolVal Bool
    | Nil
    | Cons Value Value
```

and environments are just

```
type Env = Environment Value
```

Environments is just an abstract data type which maps identifiers to values of some type δ, where in Fun we take $\delta =$ `Value`.

We take the standard mapping constructors

```
type Environment δ
empty_env :: Environment δ
find ::      Environment δ → Ident → δ
define ::    Environment δ → Ident → δ → Environment δ
```

We write $find\ env\ x$ for the value to which x is bound in the environment env, and the interpretor gives an error if x is not bound to any value in env.

We also write `define env x v` for the environment that agrees with `env` apart from mapping x to v, hiding any binding of x from before. As notation, we write `env` $\oplus (x, v)$ for `define env x v`.

For convinience, we also define

```
make_env :: [(Ident, δ)] → Environment δ
```

such that

```
make_env [(x1, v1), ..., (xn, vn)] = empty_env ⊕ (x1, v1) ⊕···⊕ (xn, vn)
```

The function `eval` is one of the four main functions that makes the interpreter. The others are

```
apply :: Value → [Value] → Value
```

which applies a function to its list of arguemnts and produces the value returned by that function

```
abstract :: [Ident] → Expr → Env → Value
```

which forms a function value from a lambda expression

```
elab :: Defn → Env → Env
```

which elaborates a definition, producing a new environment in which the new name has been defined.

The entire process of interpretation starts with a call

```
eval exp init_env
```

where `exp` is the abstract syntax tree for an expression that has been input, and `init_env` is the initial environment.

We can define env by pattern matching on `Expr`, with

```
eval (Number n) env = IntVal n
eval (Variable x) env = find env x
eval (If e1 e2 e3) env =
    case eval e1 env of
        BoolVal True → eval e2 env
        BoolVal False → eval e3 env
        _ → error "boolean required in conditional"
```

The error message helps with dealing with expressions `if` e_1 `then` e_2 `else` e_3 where e_1 does not evaluate to a Boolean. In the `Apply` case, we need to evaluate the arguemnts first, so we do this by

```
eval (Apply f es) env =
    apply (eval f env) (map ev es)
        where ev e1 = eval e1 env
```

The inner call `apply` is simply given by

```
apply (Function f) args = f args
apply _            args = error "applying a non-function"
```

The `Lambda` is a way of creating an abstraction, which should evaluate to a function. We can define this by

```
eval (Lambda xs e1) env = abstract xs e1 env
```

where `abstract` is a function that binds the parameters `xs` to the arguments `args` then evaluates the function body `e`. That is,

```
abstract xs e env =
    Function f
    where f args = eval e (defargs env xs args)
```

Where `defargs :: Env -> [Ident] -> [Value] -> Env` is defined such that

```
defargs env0 [x1, ..., xn] [v1, ..., vn]
    = env0 ⊕ (x1, v1) ⊕···⊕ (xn, vn)
```

where `env` $\oplus$ `(x, v)` = `define env x v`. Finally, we want to work with expressions of the form `let d in` e_1 (or `Let d e`), which can be given by extending the environment according to the definition `d`, then evaluate the expression `e`.

That is,

```
eval (Let d e1) env = eval e1 (elab d env)
```

Elaborating simply extends the definition, which we can do by

```
elab (Val x e) env = define env x (eval e env)
elab (Rec x (Lambda xs e1)) env =
    env' where env' = define env x (abstract xs e1 env')
elab (Rec x _) env =
    error "RHS of letrec must be a lambda"
```

Note the definition of `env'` uses recursion on itself, where recursion is modelled by recursion in Haskell.

Finally, we can define primitives which are definitions given in `init_env`, for instance the identifier "+" is bound to the function value `Function plus` where

```
plus [IntVal a, IntVal b] = IntVal (a + b)
```

and the Fun expression $e_1 + e_2$ has the abstract syntax

```
Apply (Variable "+") [e1, e2]
```

hence its value is obtained by looking up "+" in the environment and applying the resulting function to the values of the two arguments.

Remark 1.1.2. Consequently, by redefining "+", we can change our standard interpretation of the primitive operator, as there is no distinction between primitive in the environment and those which are overwritten.

We can apply d17n to the higher order interpreter for Fun to eliminate the use of functions in the Value type. There are two places in which functions are created, the value of a lambda expression is a function value `abstract xs e env` and each primitive is bound to a different function value specific to a primitive. We introduce two new constructors to Value, each depending on a type that lists all possible primtiives:

```
data Value =
    ...
    | Closure [Ident] Expr Env
    | Primitive Prim

data Prim =
    Plus | Minus ...
    deriving Show

type Env = Environment Value
```

Then `abstract` becomes a simple a simple call

```
abstract :: [Ident] → Expr → Env → Value
abstract xs e env = Closure xs e env
```

The original definition for `abstract` is placed as part of `apply`, used when the value being applied is one of the `Closure` objects. If it is called on a `Primitive`, we depend on a function `primapply` to be defined.

```
apply :: Value → [Value] → Value
apply (Closure xs e env) args =
    eval e (defargs env xs args)
apply (Primitive p) args =
    primapply p args
apply _ _ =
    error "applying to a non-function"
```

Before d17n, primitive names like '+' was bound in the environment to a value `Function plus` where `plus` was a specific function of type `[Value] -> Value`. Instead, we bind '+' to a value `Primitive Plus` where `Plus` is a primitive that is to be dealt with specifically through the `primapply` function.

We add in the initial environment:

```
init_env :: Env
init_env =
    make_env[
        primitive "+" Plus, ...
    ]
    where
        constant x v = (x, v)
        primitive x p = (x, Primitive p)
```

and primapply:

```
primapply :: Prim → [Value] → Value
primapply Plus [IntVal a, IntVal b] = IntVal (a + b)
...
primapply x args =
    error ("bad arguments to primitive " ++ show x ++ ": " ++ showlist
args)
```

Remark 1.1.3.

- We have saved the environment as part of the Closure, as we care about the environment when it is being abstracted.
- d17n can be applied to any whole program to eliminate higher order functions
- The cyclic structure for recursion becomes clear, as we are effectively making a loop in a pointer-linked data structure:

```
elab (Rec x (Lambda xs e)) env =
    env' where env' = define env x (Closure xs e env')
```

Adjusting instances for Show and Eq in Haskell, we solve the first and third problems mentioned at the start of the subsection. The third is solved as we elaborate based on Closure passing, where before we relied on wrapping it in a Function who was defined over Haskell.

Effectively, instead of passing a higher order function, we pass a token that represents the higher order function, and evaluate based on that information, which allows one to avoid making the main function itself to be a higher function.

1.2 Memory

We add assignable variables, language sequencing, and while loops.

We have `new()`, which creates assignment variables, returning its address.

```
>>> val a = new();;
--- a = <address 1>
```

Then, we can set the contents of it by assignment

```
>>> a := 3;;
--> 3
```

The value of the assignment takes the expression `x := E` where `E` is the value of the expression `E`. We can then retrieve it's contents via the ! operator :

```
>>> !a;;
--> 3
```

Language sequencing will be written as $\texttt{e}_1\texttt{; e}_2$, where a while loop will be written $\texttt{while e}_1\texttt{ do e}_2$.

Then, we can write functions like factorial imperatively :

```
val fac(n) =
    let val k = new() in
    let val r = new() in
    k := n; r := 1;
    while !k > 0 do
        (r := !r * !k; k := !k - 1);
    !r;;
```

In the above code, `n` is a 'constant' whose value never changes, but `k` and `r` are mutable by the assignment construct.

We implement a memory to map vairables to contents. They support the following

```
type Memory α
type Location

contents :: Memory α → Location → α
update :: Memory α → Location → α → Memory α
fresh :: Memory α → (Location, Memory α)
```

Roughly, `contents` gives the content that is being stored at a location in memory. The `update` gives a new memory that is the same as the previous, updated to the location and α arguments. The function `fresh` creates and returns a fresh location such that if `(a, m') = fresh m`, then `a` is a location that is unused in `m`, and `m'` is a copy of `m` modified so that the location is now regarded as being in use. Thus if we have

```
let (a, m') = fresh m in
let (b, m'') = fresh m' in ...
```

then `a` and `b` are different locations. Note that as the implementation of memory is backed by a functional metalanguage, this imperative equivalent is indeed much slower than the functional equivalent.

In Fun, memories store items of type `Value`, written

```
type Mem = Memory Value
```

We also redefine the evaluation function :

```
eval :: Expr → Env → Mem → (Value, Mem)
eval (Number n) env mem = (IntVal n, mem)
eval (Variable x) env mem = (find env x, mem)
eval (If e1 e2 e3) env mem =
    let (b, mem') = eval e1 env mem in
    case b of
        BoolVal True → eval e2 env mem'
        BoolVal False → eval e3 env mem'
        _ → error "Boolean required in conditional"
eval (Apply f es) env mem =
    let (fv, mem') = eval f env mem in
    let (args, mem'') = evalargs es env mem' in
    apply fv args mem''
```

with helper functions

```
evalargs :: [Expr] → Env → Mem → ([Value], Mem)
evalargs [] env mem = ([], mem)
evalargs (e : es) env mem =
    let (v, mem1) = eval e env mem in
    let (vs, mem2) = evalargs es env mem1 in
    (v : vs, mem2)
```

Any declaration creates a new memory state. In particular, `elab` should take a memory state as an arguement and return another as part of its result, thus defining `let` should chain these results. Specifically,

```
eval (Let d e1) env mem =
    let (env', mem1) = elab d env mem in
        eval e1 env' mem1
```

Then, elaboration looks like

```
elab :: Defn → Env → Mem → (Env, Mem)
elab (Val x e) env mem =
    let (v, mem1)  eval e env mem in
        (define env x v, mem1)
elab (Rec x (Lambda xs e1)) env mem =
    (env', mem) where env' = define env x (abstract xs e1 env')
elab (Rec x _) env mem =
    error "RHS of letrec must be lambda"
```

Abstract does not need to touch memory, as it builds a function object, which if it requires the use of memory, will do so when it is called.

We adjust the Value type to accomodate primitives `new` and '!' as

```
data Value =
    ...
    | Addr Location
```

```
        | Function ([Value] → Mem → (Value, Mem))

eval (Assign e1 e2) env mem =
    let (v1, mem') = eval e1 env mem in
    case v1 of
        Addr a →
            let (v2, mem'') = eval e2 env mem' in
            (v2, update mem'' a v2)
        _ → error "assigning to a non-variable"
```

The assignment returns the value of the right hand side, with the effect that the location denoted by the left is updated to contain the same value.

Sequencing can be written like

```
eval (Sequence e1 e2) env mem =
    let (v1, mem') = eval e1 env mem in eval e2 env mem'
```

The while loop is written on the knowledge that

```
while e1 do e2
```

looks like

```
if e1 then (e2; while e1 do e2) else nil
```

with

```
eval (While e1 e2) env mem = f mem
where
    f mem =
        let (b, mem') = eval e1 env mem in
        case b of
            BoolVal True → let (v, mem'') = eval e2 env mem' in f mem''
            BoolVal False → (Nil, mem')
            _ → error "boolean required in while loop"
```

Where we have chosen the program to return `Nil` when the while loop terminates.

We add primitives

```
primitive "new" (λ [] mem →
    let (a, mem') = fresh mem in (Addr a, mem')
)
primitive "!" (λ [Addr a] mem → (contents mem a, mem))
```

The original primitives are adjusted such that memory passes straight through them. We adjust the environment:

```
init_env :: Env
init_env =
    make_env ...
    where
        constant x v = (x, v)
        primitive x f = (x, Function (primwrap x f))
        pureprim x f =
```

```
                (x, Function (primwrap x (λ args mem → (f args, mem))))
```

The top level for Fun with memory is

```
obey :: Phrase → GloState → (String, GloState)
```

where `GloState` is the type of global states that are passed from one evaluation to next. With memory, this is

```
type GloState = (Env, Mem)
```

Such that the memory with environment is passed between phrases. Then,

```
obey (Calculate exp) (env, mem) =
    let (v, mem') = eval exp env mem in
    (print_value v, (env, mem'))
obey (Define def) (env, mem) =
    let x = def_lhs def in
    let (env', mem') = elab def env mem in
    (print_defn env' x, (env', mem'))
```

With

```
main = dialog funParser obey (init_env, init_mem)
```

1.3 Output

We can add primitives for output such that evaluating `print(v)` would both yield the value and print it on the terminal. For instance:

```
>>> print(4) + 5;;
4
--> 9
```

Then, we would have

```
eval :: Expr → Env → (String, Value)
```

In the previous case we'd pass the memory to the next evaluation, but in this case a sequence of evaluations would result in a concatenation of the string outputs.

We have a binding for print as a primitive of the function:

```
print :: [Value] → (String, Value)
print [v] = (show v ++ "λn", v)
```

With

```
pureprin x f = (x, Function (primwrap x (λ args → ("", f args))))
```

2 Monads

2.1 Monad Laws

In the scope of Fun, a monad takes some $\alpha \to M\ \alpha$, equipped with an operator $\triangleright$ and a function `result` such that the following laws hold :

```
- (xm ▷ f) ▷ g = xm ▷ (λ x → f x ▷ g)
- (result x) ▷ f = f x
- xm ▷ result = xm
```

These outline associativity and identity rules.

Here, ▷ is an operator which takes the evaluated value of the left, and passes the return value to its right. Thus,

$$\triangleright : M\ \alpha \to (\alpha \to M\ \beta) \to M\ \beta$$

We can rewrite this in more succinct Haskell notation, by considering the function $^* : (\alpha \to M\beta) \to M\ \alpha \to M\ \beta$ by `* f xm = xm ▷ f`. Writing `f*` for shorthand, we can simplify our laws to

```
- g* · f* = (g* · f)*
- f* · result = f
- result* = id
```

2.2 Monads in Fun

In Fun, Monads allow structure in being able to pass some background information without explicitly referring to them. In the case of Memory and Output, we can pass these informations without explicitly writing their outputs by using ▷ to implicitly compute them being updated.

2.2.1 Memory and Output

We first note the types for `eval` and `elab`,

```
eval₁ :: Expr → Env → Mem → (Value, Mem)
elab₁ :: Defn → Env → Mem → (Env, Mem)
eval₂ :: Expr → Env → (String, Value)
elab₂ :: Defn → Env → (String, Env)
```

Using currying and higher-order types, we can reduce these to

```
eval :: Expr → Env → M Value
elab :: Defn → Env → M Env
```

where

```
type M₁ α = Mem → (α, Mem)
type M₂ α = (String, α)
```

Now consider some examples of the evaluation functions for the two languages.

In the ccase for numeric constants, we have

```
eval₁ (Number n) env = (λ mem → (IntVal n, mem))
eval₂ (Number n) env = ("", Intval n)
```

we can wrap this in a simple function such that

```
eval₁ (Number n) env = result (IntVal n)
```

where

```
result :: α → M α
result₁ x = (λ mem → (x, mem)) :: M₁ Value
result₂ x = ("", x)            :: M₂ Value
```

The `result` function essentially does what is considered doing nothing, when mapping to inside the monad. In the case of FunMem, this is mapping to the same memory, while in the case of FunOut, this is producing the empty string.

Similarly,

```
eval (Variable x) env = result (find env x)
```

Consider now a more complex example given by the conditional, where we have (omitting error messages)

```
eval₁ (If e₁ e₂ e₃) env mem =
    let (b, mem') = eval₁ e₁ env mem in
    case b of
        BoolVal True → eval₁ e₂ env mem'
        BoolVal False → eval₁ e₃ env mem'

eval₂ (If e₁ e₂ e₃) env =
    let (out₁, b) = eval₂ e₁ env in
    let (out₂, r) =
        case b of
            BoolVal True → eval₂ e₂ env
            BoolVal False → eval₂ e₃ env in
    (out₁ ++ out₂, r)
```

We can collapse this into a single operator,

```
eval (If e₁ e₂ e₃) env =
    eval₁ e₁ env ▷ (λ b →
        case b of
            BoolVal True → eval e₂ env
            BoolVal False → eval e₃ env
    )
```

where $\triangleright$ combines in sequence of calculations, passing the result of the first operand to the second.

For the first case, we note that the memory after the first evaluation is being passed, such that

```
xm ▷₁ f =
    let (x, mem') = xm mem in f x mem'
```

For the second case, we simply add the string produced through the output, which we can do by

```
xm ▷₂ f =
    let (out₁, x) = xm in
    let (out₂, y) = f x in
    (out₁ ++ out₂, y)
```

As another instance, consider

```
eval₁ (Let d e₁) env mem =
    let (env', mem') = elab₁ d env mem in
    eval₁ e₁ env' mem'
```

```
eval₂ (Let d e₁) env =
    let (env', out₁) = elab₂ d env in
    let (x, out₂) = eval₂ e₁ env' in
    (x, out₁ ++ out₂)
```

The evaluation can be viewed as a function of the form `Env` $\to M$ `Value`, which passes the new environment into the evaluation. Thus, we can write the evaluation for let as,

```
eval (Let d e) env =
    elab d env ▷ (λ env' → eval e₁ env')
```

2.3 Monad of Memory in Fun

The monad of memeory can be written simply as

```
type M α = Mem → (α, Mem)

result :: α → M α
result x mem = (x, mem)

▷ :: M α → (α → M β) → M β
(xm ▷ f) mem =
    let (x, mem₁) = xm mem in f x mem₁
```

2.3.1 Principal Functions

Additionally, a choice of monad usually comes with a few operations to implement specific language features. For a memory, we have `get` that retrieves the value stored in a location, `put` that modifies the contents of a location, and `new` that allocates a fresh location.

Explicitly,

```
get :: Location → M Value
get a mem = (contents mem a, mem)

put :: Location → Value → M ()
put a v mem = ((), update mem a v)

new :: M Location
new mem = let (a, mem') = fresh mem in (a, mem')
```

Alternatively, `new` = `fresh`. These operations are placed so that the details of the computational model (eg exception + memory) can be implmented by only changing the inner implementation of the above functions.

The term "Semantic Domains" refers to the types that are used in the interpreter.

The semantic domains for assignment variables can be written as

```
data Value =
    // Original Alternatives
    | Addr Location
    | Function ([Value] → M Value)
```

where the existence of the output type of the function being `M Value` represents the fact the input body can interact with the memory. We also fix the decision that `Values` are what names are bound to in the environment, and the same domain of vlaues that can be stored in memory. That is,

```
type Env = Environment Value
type Mem = Memory Value
```

We also redefine the standard functions `eval`, `abstract`, `apply`, and `elab`.
That is,

```
eval :: Expr → Env → M Value
eval (Number n) env = result (IntVal n)
eval (Variable x) env = result (find env x)
eval (Apply f es) env =
    eval f env ▷ (λ fv →
        evalargs es env ▷ (λ args →
            apply fv args
        )
    )
eval (lambda xs e1) env =
    result (abstract xs e1 env)
eval (If e1 e2 e3) env =
    eval e1 env ▷ (λ b →
        case b of
            BoolVal True → eval e2 env
            BoolVal False → eval e3 env
            _ → error "Boolean required in conditional"
    )
eval (Let d e1) env =
    elab d env ▷ (λ env' → eval e1 env')
```

The above is enough to define a language that is purely functional, but we can add imperative features like sequencing and while loops by

```
eval (Sequence e1 e2) env =
    eval e1 env ▷ (λ v → eval e2 env)
eval (While e1 e2) env = u
    where
        u = eval e1 env ▷ (λ v1 →
            case v1 of
                BoolVal True → eval e2 env ▷ (λ v2 u)
                BoolVal False → result Nil
                _ → error "Boolean required in while loop"
        )
```

where we also have standard helper functions:

```
evalargs :: [Expr] → Env → M [Value]
evalargs [] env = result []
evalargs (e : es) env =
    eval e env ▷ (λ v → evalargs es env ▷ (λ vs → result (v :: vs)))
```

Alternatively,

```
mapm :: (α → M β) → [α] → M [β]
mapm f [] = result []
mapm f (e : es) = f e ▷ (λ v → mapm f es ▷ (λ vs → result (v :: vs)))

evalargs xs env = mapm (λ e → eval e env) xs
```

The clause for `While` defines the meaning of loops as recursion.

Languages also have constructs that are unique to them, shared only with closely related ones. In Fun with memory, the only such construct is assignment, which uses the `put` operation.

```
eval (Assign e1 e2) env =
    eval e1 env ▷ (λ v1 →
        case v1 of
            Addr a →
                eval e2 env ▷ (λ v2 → put a v2 ▷ (λ () → result v2))
            _ → error "assigning to a non-variable"
    )
```

Note that in the assignment $e_1 := e_2$, e_1 must be an address, updated with the value of e_2, and the same value is yielded as the value of the assignment itself.

Moving to `apply` and `abstract`,

```
abstract :: [Ident] → Expr → Env → Value
abstract xs e env =
    Function (λ args → eval e (defargs env xs args))

apply :: Value → [Value] → M Value
apply (Function f) args = f args
apply _ args = error "applying a non-function"
```

The type of `abstract` does not change, as forming a function value does not require interactions with the memory. The type of `apply` does change to reflect the different type of the function that is wrapped in `Function`.

The `elab` processes declaration, which should have similar definition in most interpreters. Elaborating may involve an interaction with the memory, as in `val x = e`, evaluation of the right side may need to use the memory. Naturally,

```
elab :: Defn → Env → M Env
elab (Val x e) env =
    eval e env ▷ (λ v → result (define env x v))
elab (Rec x (Lambda xs e1)) env =
    result env' where env' = define env x (abstract xs e1 env')
elab (Rec x _) env =
    error "RHS of letrec must be a lambda"
```

2.3.2 Primitives

Standard primitives are shared with the purely functional language, but we also need to insert a call to result, reflecting the fact primitives do not need to interact with the memory.

The primitives specific to the language with assignment variables are "!x" which fetches the contents of the the memory cell named by x, and new(), which allocates a fresh, uninitialised cell.

We thus have

```
primitive "!" (λ [Addr a] → get a)
primitive "new" (λ [] → new ▷ (λ a → result (Addr a)))
```

The syntax allows the expression !x to be written without parenthesis. Note this is a Haskell feature that allows lambda on a specific constructor without casing.

The initial environment is given by

```
init_env :: Env
init_env =
    make_env [
        // Usual constants + primitives
        // primitives from above
    ]
    where
        constant x v = (x, v)
        primitive x f = (x, Function (primwrap x f))
        pureprim x f = primitive x (result · f)
```

2.3.3 Main Program

We give the global state and obey as follows

```
type GloState = (Env, Mem)
obey :: Phrase → GloState → (String, GloState)
obey (Calculate exp) (env, mem) =
    let (v, mem') = eval exp env mem in
        (print_value v, (env, mem'))
obey (Define def) (env, mem) =
    let x  = def_lhs def in
    let (env', mem') = elab def env mem in
    (print_defn env' x, (env', mem'))
```

The main code is then given by

```
module FunMonad(main) where
    // common imports
import Memory
infixl 1 ▷
```

We import Memory in order to implement assignable variables. We add in the memory monad, principle functions (including catch-all for eval), primitives, initial environment, instance declarations (of Eq and Show).

The main program based on obey completes the program :

```
main = dialog funParser obey (init_env, init_mem)
```

2.4 Monadic Equivalence

With pure fun, for any expression `e`, we expect the program

```
let val x = e in x
```

to be equivalent to `e`.

To prove this, first note the meaning of the constructs :

```
eval (Let d e1) env = elab d env ▷ (λ env' → eval e1 env')
eval (Variable x) env = result (find env x)
elab (Val x e) env = eval e env ▷ (λ v → result (define env x v))
```

Writing `LHS` to represent the code above, we have

```
eval LHS env = (vm ▷ f) ▷ g
    where
        vm = eval e env
        f v = result (define env x v)
        g env' = result (find env' x)
```

Noting this is just an expansion of `eval (Let (Val x e) (Variable x)) env`.

Applying the associative law, we have

```
(vm ▷ f) ▷ g = vm ▷ (λ v → f v ▷ g)
```

Now, given `v` and `env`,

```
f v ▷ g = result env' ▷ g = g env' = result (find env' x) = result v
```

on the assumption that `find (define env x v) x = v`. Now, applying the right identity,

```
eval LHS env = eval e env ▷ (λ v → result v) = eval e env
```

2.5 Exceptions

We define a notion of "`orelse`", which runs the first argument, returns it if successive, and calls the second if it fails. For instance, this is useful when one wants to treat -1 as the error term.

Consider

```
val index(x, xs) =
    let rec search(ys) =
        if ys = nil then fail()
        else if head(ys) = x then 0
        else search(tail(ys)) + 1 in
    search(xs) orelse -1;;
```

Then, `index(x,xs)` returns -1 if the value does not appear in the list.

We can define a monad that captures failure as follows

```
data M α = Ok α | Fail
```

We make this into a monad by defining the standard functions associated with it

```
result :: α → M α
result x = Ok x
```

```
(▷) :: M α → (α → M β) → M β
(Ok x) ▷ f = f x
Fail ▷ f = Fail
```

with associated operations to implement the `fail()` primitive and the `orelse` construct as follows

```
failure :: M α
failure = Fail

orelse :: M α → M α → M α
orelse (Ok x) ym = Ok m
orelse Fail ym = ym
```

The semantic domain `Value` contains the same kinds of values as in pure Fun, as there are no values introduced, just the possibility of a failing evaluation. The type used to represent functions changes, to incorporate the monad to reflect the fact the function body may fail. So,

```
data Value =
    // Standard values
    | Function ([Value] → M Value)
```

The `eval` function then comes with a clause for the `orelse` construct, with the abstract syntax

```
data Expr = ...
    | OrElse Expr Expr
```

Then, the evaluation for this is simply

```
eval (OrElse e1 e2) env =
    orelse (eval e1 env) (eval e2 env)
```

The primitive fail is simple :

```
primitive "fail" (λ [] → failure)
```

Finally, at the top-level, we have

```
obey :: Phrase → Env → (String, Env)
obey (Calculate exp) env =
    case (eval exp env) of
        Ok v → (print_value v, env)
        Fail → ("*failed*", env)
obey (Define def) env =
    let x = def_lhs def in
    case elab def env of
        Ok env' → (print_defn env' x, env')
        Fail → ("*failed*", env)
```

Putting the parser together in the standard way.

The `orelse` definition is dependent on the lazy evaluation of Haskell, as we expect the second arguemnt to not be evaluated unless the first ends in failure. For instance, the epxression

```
3 orelse (let rec loop(n) = loop(n+1) in loop(0))
```

should evaluate to 3 and not infinite recursion. The solution to this is to use **continuations** such that the type `M α` becomes

```
type M α = (α → Answer) → (() → Answer) → Answer
```

for some type Answer, with the idea that `xm ks kf` calls the success continuation `ks` to signal success and failure continuation `kf` if it fails.

Alternatively, we can avoid dependence on Haskell's evaluation by making `M α` into a function type :

```
type M α = () → Maybe α
```

with the standard `Maybe α = Just α | Nothing`. Then,

```
result :: α → M α
result x = (λ () → Just x)

(▷) :: M α → (α → M β) → M β
xm ▷ f =
    (λ () →
        case xm() of
            Just x → f x ()
            Nothing → Nothing
    )

failure :: M α
failure = (λ () → Nothing)

orelse :: M α → M α → M α
orelse xm ym =
    (λ () →
        case xm() of
            Ok x → Ok x
            Nothing → ym ()
    )
```

This forces arguments to be functions that are called only if their results are needed, allowing expressions to yield an answer without evaluation.

Whether the metalanguage is lazy or not, we must have

```
failure ▷ f = failure
```

thus, `orelse` must be added as an expression, not a primitive, as in the evaluation for `Apply` expressions, we evaluate the arguments, which means if any fail, the entire call fails. Thus, we have no such primitive.

3 Metalanguage Independent Machines

3.1 Defunctionalization (d17n)

There are a few problems to cover with the interpreter, as many parts still rely on native features of Haskell. These include

- The Fun interpreter uses higher order functions and type of values include functions from values to values, which make no sense as a mathematical set. Specifically, we have

```
data Value = ... | Function ([Value] → Value) | ...
```

  but there is no injection from $[X] \to X$ into X.

- The interpreter uses recursion to deal with the recursive syntax
- The implementation of recursive function definitions involves a recursive value definition in Haskell (as opposed to a recursive function)
- The evaluation strategy (including laziness) is not clear.

Defunctionalization is a whole program transformation that turns higher-roder functional programs into first order ones. This can be split up into steps, although it is a single transformation.

Start with an original program

```
prog x ys = (map aug ys, map sqr ys)
    where
        aug y = x + y
        sqr y = y * y
```

The first step is to **decode the dataflow**, isolating expressions that create or use or pass the value of higher order arguemnts or results, together with a trivial function apply that unwraps them when necessary.

For instance we can do

```
data Func = Func (Integer → Integer)

prog x ys = (map (Func aug) ys, map (Func sqr) ys)
    where
        aug y = x + y
        sqr y = y * y

map :: Func → [Integer] → [Integer]
map f [] = []
map f (x : xs) = apply f x : map f x

apply :: Func → Integer → Integer
apply (Func f) x = f x
```

We now remove lambdas, identifying free variables, creating constructors for each lambda expression (which will be treated by the apply function) so that we have

```
data Func = Sqr | Aug Integer
```

Then, we can rewrite prog as

```
prog x ys =
    (map (Aug x) ys, map Sqr ys)
```

We then adjust apply as an interpreter for the tiny language the Func datatype has become, associating each constructor to the expression it represents. In our case, we have

```
apply :: Func → Integer → Integer
apply (Aug x) y = x + y
apply Sqr y = y * y
```

3.2 Continuation Passing Style

This takes a recursive program and gives it an iteratively controlled behavior. It introduces higher-order functions, but this can be removed via d17n. In particular, we show that semantics are independent of recursion / evaluation strategy of the metalanguage.

Example 3.2.1. Consider the factorial function defined by

```
fac :: Int → Int
fac n = if n = 0 then 1 else fac (n - 1) * n
```

The calculation grows and shrinks, as it needs to deal with unfolding until termination.

Alternatively, we can define

```
fac :: Int → Int
fac = faciter n 1

faciter :: Int → Int → Int
factier n f =
    if n = 0 then f else factier (n-1) (n * f)
```

The control information space does not increase with n in this case. Such functions can be written as a loop

```
f := 1
while n <> 0 do
    f := n * f; n := n-1
end
```

And we say that the faciter function has **iterative control behavior**.

Definition 3.2.2. *A Continuation Passing Style (CPS) takes any functional program and makes it have iterative control behavior.*

The general idea is that control context is needed for evaluating arguments and not for calling procedures (we care about evaluation order, but after that is simply plain application). We therefore wrap such context as an extra argument called the **continuation**. It is a function that takes the result of the function being called, and from it calculates the final answer of the whole program.

Example 3.2.3. We take the above for the factorial function. Suppose the main program is

```
main n = show (fac n)
```

In particular, we have

```
fack :: Int → (Int → Answer) → Answer
fack n k =
    if n = 0 then k 1 else fack (n - 1) (λ r → k (r * n))
```

The input on the continuation is the calculated value for $(n-1)!$, and we multiply that by n, passing it to the original continuation k. Taking the identity as the second argument, this should

represent our understanding of the normal factorial function. If this is instead replaced by a general function `show`, the recursion deals with recursive functions that must be represented somehow in space. Running, we get

```
fack 5 show
    = fack 4 k4 where k4 r = show (r * 5)
    = fack 3 ...
    ...
    = fack 0 k0 where k0 r = k1 (r * 1)
    = k0 1
    = k1 1
    = ...
    = k4 24
    = show 120
```

To d17n this, we can introduce datatypes as

```
data Cont =
    Show
  | Mult Int Cont
```

Then the apply function is

```
appcont :: Cont → Int → Answer
appcont Show r = r
appcont (Mult n k) r = appcont k (r * n)
```

Then,

```
fack n k =
    if n = 0 then appcont k 1
    else fack (n - 1) (Mult n k)
```

Calculating with this has no growing control context for subsequent recursive calls, but there is a growing data structure that acts like a stack.

Example 3.2.4. Continuations allow one to control the evaluation strategy. Consider the construction

```
let x = e1 in e2
```

where if e_1 never terminates but is not used in e_2, then the evaluation depends on the metalanguage. By taking the expression

$$(\lambda x \to e_2)e_1$$

into the CPS form such that

```
    let y = f x in g y
```

is replaced by

$$f\ x\ (\lambda y \to g\ y\ k)$$

then the let expression is in a form where it is independent of the evaluation strategy.

Example 3.2.5. Consider the Fibonacci function. We have

```
fibk :: Int → (Int → a) → a
fibk n k =
    if n >= 2 then
        fibk (n-1) (λ n' → fibk (n-2) (λ n'' → k (n' + n'')))
    else k 1
```

Intuitively, `fibk n` takes a continuation k' and passes the evaluation inside, based on the construction of the base cases. In a sense, we have `fibk n k` $=$ `k (fib n)`.

To d17n this, we do

```
data Cont = Fib Cont | Fib2 Int Cont | Add Int Cont | Result

apply :: Cont → Int → Int
apply (Fib k) n =
    if n >= 2 then
        apply (Fib (Fib2 n k)) (n-1)
    else
        apply k 1
apply (Fib2 n k) (n') = apply (Fib (Apply n' k)) (n-2)
apply (Add n k) (n') = apply k (n + n')
apply Result n = n
```

(Intuitively treat as 'takes result of application and passes it into the continuation')

Remark 3.2.6. d17n gives concrete data spaces to the control information space (the 'rest of computation').

The evaluation space is the 'places in which evaluations can occur'

3.3 CEK Machine

Consider the factorial function as before that we applied CPS and then applied d17n. Note first that the recursive calls are tail-recursive (there are no further applications to the recursive calls). The type `Cont` is isomorphic to the type of lists of integers (alternatively stacks).

Now consider a machine that has configurations that represent whether we are reducing an `appcont` or `fack` as

- $[k, r]$ means `appcont k r`
- $\langle x, k\rangle$ means `fack x k`

Where we treat k as a stack.

Then the rules corresponding to the reductions are

$$\begin{aligned} [x : k, r] &\to [k, x * r] \\ \langle 0, k\rangle &\to [k, 1] \\ \langle x, k\rangle &\to \langle x - 1, x : k\rangle \qquad (x > 0) \end{aligned}$$

The configuration $[\texttt{Nil}, r]$ is terminal. We can see this as a mode of pushing and popping from a stack, alongside arithmetic.

The **CEK abstract machine** is a machine that deals with the syntactic values which are either closures or integers. That is, we have `SValue` is defined by

$$v ::= (\texttt{lambda}(x)e, env) \mid n$$

A control stack `KStack` is a set of 'things to do next' as

$$k ::= Show \mid (v(\Box)) : k \mid (\Box(e), env) : k$$

where the $\Box$ is used to represent what kind of element we are adding. The configurations are

$$\langle e, env, k\rangle \qquad [k, v]$$

The set of all such configurations are written `Conf`.

The translation relations are as follows:

$$\begin{aligned}
&\langle e(e'), env, k\rangle \to \langle e, env, (\Box(e'), env) : k\rangle \\
&\langle \texttt{lambda}(x)e, env, k\rangle \to [k, (\texttt{lambda}(x)e, env)] \\
&\langle n, env, k\rangle \to [k, n] \\
&\langle x, env, k\rangle \to [k, env(x)] \\
&[(\Box(e), env) : k, v] \to \langle e, env, (v(\Box)) : k\rangle \\
&[((\texttt{lambda}(x)e, env)(\Box)) : k, v] \to \langle e, (x, v) : env, k\rangle
\end{aligned}$$

The $\langle\rangle$ talks about evaluating a certain expression. The stack keeps track of promises to be evaluated, where the $\Box$ represents what is currently being evaluated. Note that when evaluating function application, we evaluate the function body first and then the body. Once the evaluation is complete, we are given a form $[k, n]$ (or $[k, f]$), which based on the promise we pass onto the top layer stack, continuing the evaluation.

4 Typing

In this section we consider giving a simple set of programs types in the usual sense. In particular, we explore a relation $R \subseteq \texttt{Ctx} \times \texttt{Expr} \times \texttt{Type}$ (writing $\Gamma \vdash e : A$ for $(\Gamma, e, A) \in R$).

First, Types are given by the grammar

$$A, B ::= \texttt{int} \mid (A \Rightarrow B)$$

We define typing rules as follows

$$\frac{}{\Gamma, x_i : A_i, \Gamma' \vdash x_i : A_i} \qquad \frac{}{\Gamma \vdash n : \texttt{int}}\ n \in \mathbb{Z}$$

$$\frac{\Gamma, x : A \vdash e : B}{\Gamma \vdash \texttt{lambda}(x)e : A \Rightarrow B} \qquad \frac{\Gamma \vdash e : A \Rightarrow B \qquad \Gamma \vdash e' : A}{\Gamma \vdash e(e') : B}$$

For each type A, define a subset $[\![A]\!] \subseteq \texttt{Values}$ to be

$$[\![int]\!] := \mathbb{Z} \qquad [\![A \Rightarrow B]\!] := ([\![A]\!] \to [\![B]\!])$$

In a similar way, as the environment is a partial map from identifiers to values,

5 Simple Domain Theory

We can define an inductive set by simply taking the least set that satisfies certain rule-based properties (constructors). Then, principle of induction says that if this is a monotonic function F, given any set S and $F\ S \subseteq S$, S contains at least the least fixed point.

Some things to note: - Tarski's Fixed Point Theorem (for chain cpo) - factorial function as example - continuity

Construction of cpos

- adding $\perp$
- $[X \to Y]$
- $X \times Y$

6 Monads in Functional Programming

Definition 6.0.1. *A **monad**, written M is a type constructor equipped with*

```
return :: A → M a
▷ :: M a → (a → M b) → M b
```

such that

```
- (xm ▷ f) ▷ g = xm ▷ (λ x → f x ▷ g)
- (result x) ▷ f = f x
- xm ▷ result = xm
```

where the three rules are called the monad laws, representing associativity and identity rules.

Example 6.0.2. We give some examples of monads:

- The identity monad

  ```
  M a = a
  return x = x
  xm ▷ f = f xm
  ```

- Maybe Monad

  ```
  M a = Maybe a
  return x = Just x
  xm ▷ f = case xm of
      Just x  → f x
      Nothing → Nothing
  ```

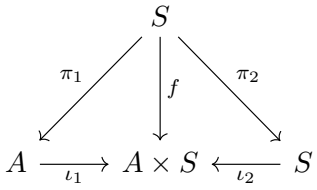
- List Monad
- Print Monad
- Memory Monad
- Continuation Monad

6.0.1 State Monad Decomposition

The state Monad over s is

```
M a = s → (a, s)
return x = (λ y → (x, y))
(xm ▷ f) s =
    case xm s of
        (x, y) → f x y
```

We can view the monad as a function from $s \to a \times s$, thus decompose it based on the components. So, we have



Now consider expressions

```
read :: M s
read x = (x, x)
write :: s → M ()
write x _ = ((), x)
```

Now, we claim that given any `m :: M a`, we can decompose this as

```
m = read ▷ (λ x → write (m2 x) ▷ λ () → result (m1 x))
```

where `m1` and `m2` are projections onto each coordinate.

Indeed, we have

```
m x = (read ▷ (λ x → write (m2 x) ▷ λ () → result (m1 x))) x
    = (write (m2 x) ▷ λ () → result (m1 x)) x
    = result (m1 x) (m2 x)
    = (m1 x, m2 x)
```

6.0.2 Monad Morphism

Definition 6.0.3. *Given monads M and N, a monad morphism is a polymorphic function*

```
h :: M a → N a
- h(result_M x) = (result_N x)
- h (m ▷_M f) = (h m) ▷_N (h · f)
```

Intuitively this represents

7 Other

7.1 On elab with abstract

- some sort of closure, lasyness

- how to think of let, where, how far things are captured

- Env keeps track of variables and which values they map to. This doesn't change, unless some 'let' overrides the previous environment. Mem keeps track of a mapping from addresses to values, which are mutable. Thus, environments map variables to addresses, which map to different values instructed by Mem. Mem only changes when assignments happen.