Type systems for programming languages

Didier Rémy

Academic year 2020-2021
Version of August 25, 2022
## Contents

1 **Introduction** .................................................................................................................. 7  
1.1 Overview of the course .................................................. 7  
1.2 Requirements .................................................. 9  
1.3 About Functional Programming .................................................. 9  
1.4 About Types .................................................. 9  
1.5 Acknowledgment .................................................. 11  

2 **The untyped \( \lambda \)-calculus** ......................................................................................... 13  
2.1 Syntax .................................................. 13  
2.2 Semantics .................................................. 15  
   2.2.1 Strong \( v.s. \) weak reduction strategies .................................................. 15  
   2.2.2 Call-by-value semantics .................................................. 16  
2.3 Answers to exercises .................................................. 18  

3 **Simply-typed lambda-calculus** ....................................................................................... 21  
3.1 Syntax .................................................. 21  
3.2 Dynamic semantics .................................................. 21  
3.3 Type system .................................................. 22  
3.4 Type soundness .................................................. 25  
   3.4.1 Proof of subject reduction .................................................. 26  
   3.4.2 Proof of progress .................................................. 28  
3.5 Simple extensions .................................................. 30  
   3.5.1 Unit .................................................. 30  
   3.5.2 Boolean .................................................. 30  
   3.5.3 Pairs .................................................. 31  
   3.5.4 Sums .................................................. 32  
   3.5.5 Modularity of extensions .................................................. 32  
   3.5.6 Recursive functions .................................................. 33  
   3.5.7 A derived construct: let-bindings .................................................. 33  
3.6 Exceptions .................................................. 35  
   3.6.1 Semantics .................................................. 35
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.6.2</td>
<td>Typing rules</td>
<td>36</td>
</tr>
<tr>
<td>3.6.3</td>
<td>Variations</td>
<td>37</td>
</tr>
<tr>
<td>3.7</td>
<td>References</td>
<td>39</td>
</tr>
<tr>
<td>3.7.1</td>
<td>Language definition</td>
<td>39</td>
</tr>
<tr>
<td>3.7.2</td>
<td>Type soundness</td>
<td>41</td>
</tr>
<tr>
<td>3.7.3</td>
<td>Tracing effects with a monad</td>
<td>42</td>
</tr>
<tr>
<td>3.7.4</td>
<td>Memory deallocation</td>
<td>43</td>
</tr>
<tr>
<td>3.8</td>
<td>Ommitted proofs and answers to exercises</td>
<td>44</td>
</tr>
<tr>
<td>4</td>
<td>Polymorphism and System $F$</td>
<td>49</td>
</tr>
<tr>
<td>4.1</td>
<td>Polymorphism</td>
<td>49</td>
</tr>
<tr>
<td>4.2</td>
<td>Polymorphic $\lambda$-calculus</td>
<td>51</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Types and typing rules</td>
<td>51</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Semantics</td>
<td>52</td>
</tr>
<tr>
<td>4.2.3</td>
<td>Extended System $F$ with datatypes</td>
<td>54</td>
</tr>
<tr>
<td>4.3</td>
<td>Type soundness</td>
<td>58</td>
</tr>
<tr>
<td>4.4</td>
<td>Type erasing semantics</td>
<td>62</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Implicitly-typed System $F$</td>
<td>62</td>
</tr>
<tr>
<td>4.4.2</td>
<td>Type instance</td>
<td>65</td>
</tr>
<tr>
<td>4.4.3</td>
<td>Type containment in System $F$</td>
<td>66</td>
</tr>
<tr>
<td>4.4.4</td>
<td>A definition of principal typings</td>
<td>68</td>
</tr>
<tr>
<td>4.4.5</td>
<td>Type soundness for implicitly-typed System $F$</td>
<td>69</td>
</tr>
<tr>
<td>4.5</td>
<td>References</td>
<td>73</td>
</tr>
<tr>
<td>4.5.1</td>
<td>A counter example</td>
<td>73</td>
</tr>
<tr>
<td>4.5.2</td>
<td>Internalizing configurations</td>
<td>75</td>
</tr>
<tr>
<td>4.6</td>
<td>Damas and Milner's type system</td>
<td>77</td>
</tr>
<tr>
<td>4.6.1</td>
<td>Definition</td>
<td>78</td>
</tr>
<tr>
<td>4.6.2</td>
<td>Syntax-directed presentation</td>
<td>80</td>
</tr>
<tr>
<td>4.6.3</td>
<td>Type soundness for ML</td>
<td>82</td>
</tr>
<tr>
<td>4.7</td>
<td>Ommitted proofs and answers to exercises</td>
<td>84</td>
</tr>
<tr>
<td>5</td>
<td>Existential types</td>
<td>91</td>
</tr>
<tr>
<td>5.1</td>
<td>Towards typed closure conversion</td>
<td>92</td>
</tr>
<tr>
<td>5.2</td>
<td>Existential types</td>
<td>94</td>
</tr>
<tr>
<td>5.2.1</td>
<td>Existential types in Church style (explicitly typed)</td>
<td>94</td>
</tr>
<tr>
<td>5.2.2</td>
<td>Implicitly-typed existential types</td>
<td>97</td>
</tr>
<tr>
<td>5.2.3</td>
<td>Existential types in ML</td>
<td>99</td>
</tr>
<tr>
<td>5.2.4</td>
<td>Existential types in OCaml</td>
<td>100</td>
</tr>
<tr>
<td>5.3</td>
<td>Typed closure conversion</td>
<td>101</td>
</tr>
<tr>
<td>5.3.1</td>
<td>Environment-passing closure conversion</td>
<td>101</td>
</tr>
<tr>
<td>5.3.2</td>
<td>Closure-passing closure conversion</td>
<td>103</td>
</tr>
<tr>
<td>CONTENTS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>5.3.3 Mutually recursive functions</td>
<td>105</td>
<td></td>
</tr>
<tr>
<td><strong>7 Fomega: higher-kinds and higher-order types</strong></td>
<td>145</td>
<td></td>
</tr>
<tr>
<td>7.1 Introduction</td>
<td>145</td>
<td></td>
</tr>
<tr>
<td>7.2 From System F to System $F^\omega$</td>
<td>146</td>
<td></td>
</tr>
<tr>
<td>7.2.1 Properties</td>
<td>147</td>
<td></td>
</tr>
<tr>
<td>7.3 Expressiveness</td>
<td>149</td>
<td></td>
</tr>
<tr>
<td>7.3.1 Distrib pair in System $F^\omega$</td>
<td>149</td>
<td></td>
</tr>
<tr>
<td>7.3.2 Abstracting over type operators</td>
<td>149</td>
<td></td>
</tr>
<tr>
<td>7.3.3 Encoding of existential types</td>
<td>150</td>
<td></td>
</tr>
<tr>
<td>7.3.4 Church encoding of non-regular ADT</td>
<td>151</td>
<td></td>
</tr>
<tr>
<td>7.3.5 Encoding GADT—with explicit coercions</td>
<td>153</td>
<td></td>
</tr>
<tr>
<td>7.4 Beyond $F^\omega$</td>
<td>154</td>
<td></td>
</tr>
<tr>
<td><strong>9 Logical Relations</strong></td>
<td>183</td>
<td></td>
</tr>
<tr>
<td>9.1 Introduction</td>
<td>183</td>
<td></td>
</tr>
<tr>
<td>9.1.1 Parametricity</td>
<td>183</td>
<td></td>
</tr>
<tr>
<td>9.2 Normalization of simply-typed $\lambda$-calculus</td>
<td>185</td>
<td></td>
</tr>
<tr>
<td>9.3 Observational equivalence</td>
<td>187</td>
<td></td>
</tr>
<tr>
<td>9.4 Logical rel in simply-typed $\lambda$-calculus</td>
<td>189</td>
<td></td>
</tr>
<tr>
<td>9.4.1 Logical equivalence for closed terms</td>
<td>189</td>
<td></td>
</tr>
<tr>
<td>9.4.2 Logical equivalence for open terms</td>
<td>190</td>
<td></td>
</tr>
<tr>
<td>9.5 Logical rel. in $F$</td>
<td>193</td>
<td></td>
</tr>
<tr>
<td>9.5.1 Logical equivalence for closed terms</td>
<td>194</td>
<td></td>
</tr>
<tr>
<td>9.6 Applications</td>
<td>199</td>
<td></td>
</tr>
<tr>
<td>9.7 Extensions</td>
<td>201</td>
<td></td>
</tr>
<tr>
<td>9.7.1 Natural numbers</td>
<td>201</td>
<td></td>
</tr>
<tr>
<td>9.7.2 Products</td>
<td>201</td>
<td></td>
</tr>
<tr>
<td>9.7.3 Sums</td>
<td>202</td>
<td></td>
</tr>
<tr>
<td>9.7.4 Lists</td>
<td>202</td>
<td></td>
</tr>
<tr>
<td>9.7.5 Existential types</td>
<td>202</td>
<td></td>
</tr>
<tr>
<td>9.7.6 Step-indexed logical relations</td>
<td>203</td>
<td></td>
</tr>
<tr>
<td><strong>5 Type reconstruction</strong></td>
<td>91</td>
<td></td>
</tr>
<tr>
<td>5.1 Introduction</td>
<td>91</td>
<td></td>
</tr>
<tr>
<td>5.2 Type inference for simply-typed $\lambda$-calculus</td>
<td>92</td>
<td></td>
</tr>
<tr>
<td>5.2.1 Constraints</td>
<td>93</td>
<td></td>
</tr>
<tr>
<td>5.2.2 A detailed example</td>
<td>94</td>
<td></td>
</tr>
<tr>
<td>5.2.3 Soundness and completeness of type inference</td>
<td>96</td>
<td></td>
</tr>
<tr>
<td>5.2.4 Constraint solving</td>
<td>96</td>
<td></td>
</tr>
<tr>
<td>5.3 Type inference for ML</td>
<td>98</td>
<td></td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>5.3.1 Milner's Algorithm</td>
<td>98</td>
<td></td>
</tr>
<tr>
<td>5.3.2 Constraints</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>5.3.3 Constraint solving by example</td>
<td>103</td>
<td></td>
</tr>
<tr>
<td>5.3.4 Type reconstruction</td>
<td>106</td>
<td></td>
</tr>
<tr>
<td>5.4 Type annotations</td>
<td>109</td>
<td></td>
</tr>
<tr>
<td>5.4.1 Explicit binding of type variables</td>
<td>110</td>
<td></td>
</tr>
<tr>
<td>5.4.2 Polymorphic recursion</td>
<td>113</td>
<td></td>
</tr>
<tr>
<td>5.4.3 mixed-prefix</td>
<td>114</td>
<td></td>
</tr>
<tr>
<td>5.5 Equi- and iso-recursive types</td>
<td>115</td>
<td></td>
</tr>
<tr>
<td>5.5.1 Equi-recursive types</td>
<td>115</td>
<td></td>
</tr>
<tr>
<td>5.5.2 Iso-recursive types</td>
<td>117</td>
<td></td>
</tr>
<tr>
<td>5.5.3 Algebraic data types</td>
<td>118</td>
<td></td>
</tr>
<tr>
<td>5.6 HM(X)</td>
<td>119</td>
<td></td>
</tr>
<tr>
<td>5.7 Type reconstruction in System F</td>
<td>121</td>
<td></td>
</tr>
<tr>
<td>5.7.1 Type inference based on Second-order unification</td>
<td>121</td>
<td></td>
</tr>
<tr>
<td>5.7.2 Bidirectional type inference</td>
<td>122</td>
<td></td>
</tr>
<tr>
<td>5.7.3 Partial type inference in MLF</td>
<td>124</td>
<td></td>
</tr>
<tr>
<td>5.8 Proofs and Solution to Exercises</td>
<td>124</td>
<td></td>
</tr>
<tr>
<td>8 Overloading</td>
<td>159</td>
<td></td>
</tr>
<tr>
<td>8.1 An overview</td>
<td>159</td>
<td></td>
</tr>
<tr>
<td>8.1.1 Why use overloading?</td>
<td>159</td>
<td></td>
</tr>
<tr>
<td>8.1.2 Different forms of overloading</td>
<td>160</td>
<td></td>
</tr>
<tr>
<td>8.1.3 Static overloading</td>
<td>161</td>
<td></td>
</tr>
<tr>
<td>8.1.4 Dynamic resolution with a type passing semantics</td>
<td>161</td>
<td></td>
</tr>
<tr>
<td>8.1.5 Dynamic overloading with a type erasing semantics</td>
<td>162</td>
<td></td>
</tr>
<tr>
<td>8.2 Mini Haskell</td>
<td>162</td>
<td></td>
</tr>
<tr>
<td>8.2.1 Examples in MH</td>
<td>163</td>
<td></td>
</tr>
<tr>
<td>8.2.2 The definition of Mini Haskell</td>
<td>164</td>
<td></td>
</tr>
<tr>
<td>8.2.3 Semantics of Mini Haskell</td>
<td>165</td>
<td></td>
</tr>
<tr>
<td>8.2.4 Elaboration of expressions</td>
<td>167</td>
<td></td>
</tr>
<tr>
<td>8.2.5 Summary of the elaboration</td>
<td>168</td>
<td></td>
</tr>
<tr>
<td>8.2.6 Elaboration of dictionaries</td>
<td>170</td>
<td></td>
</tr>
<tr>
<td>8.3 Implicitly-typed terms</td>
<td>172</td>
<td></td>
</tr>
<tr>
<td>8.4 Variations</td>
<td>177</td>
<td></td>
</tr>
<tr>
<td>8.5 Omitted proofs and answers to exercises</td>
<td>181</td>
<td></td>
</tr>
<tr>
<td>A Proofs and Answers to exercises</td>
<td>183</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 5

Existential types

Compilation is type-preserving when each intermediate language is *explicitly typed*, and each compilation phase transforms a typed program into a typed program in the next intermediate language.

Type preserving compilation is interesting for several reasons: it can help debug the compiler; types can be used to drive optimizations; types can also be used to produce *proof-carrying code*; proving that types are preserved during compilation can be the first step towards proving that the *semantics* is preserved [Chlipala (2007)].

Besides, type-preserving compilation is quite challenging as it exhibits an encoding of programming constructs into programming language that usually requires richer type systems. Sometimes, an encoding later becomes a programming idiom that is used directly in the source language. There are several examples: closure conversion requires an extension of the language with existential types, which happens to very useful on their own. Closures are themselves a simple form of objects. Defunctionalization may be done manually on some particular programs, e.g. in web applications to monitor the computation.

A classic paper by Morrisett et al. [1999] shows how to go from System F to “Typed Assembly Language”, while preserving types along the way. Its main passes are:

1. *CPS conversion* fixes the order of evaluation, names intermediate computations, and makes all function calls tail calls;

2. *closure conversion* makes environments and closures explicit, and produces a program where all functions are closed;

3. allocation and initialization of tuples is made explicit;

4. the calling convention is made explicit, and variables are replaced with (an unbounded number of) machine registers.
In general, a type-preserving compilation phase involves not only a translation of terms, mapping \( M \) to \( \llbracket M \rrbracket \), but also a translation of types, mapping \( \tau \) to \( \llbracket \tau \rrbracket \), with the property:

\[
\Gamma \vdash M : \tau \implies \llbracket \Gamma \rrbracket \vdash \llbracket M \rrbracket : \llbracket \tau \rrbracket
\]

The translation of types carries a lot of information: examining it is often enough to guess what the translation of terms will be.

### 5.1 Towards typed closure conversion

First-class functions may appear in the body of other functions. Hence, their own body may contain free variables that will be bound to values during the evaluation in the execution environment. Because they can be returned as values, and thus used outside of their definition environment, they must store their execution environment in their value. A closure is the packaging of the code of a first-class function with its runtime environment, so that it becomes closed, i.e. independent of the runtime environment and can be passed to another function and applied in another runtime environment. Closures can also be used to represent recursive functions and objects in the object-as-record-of-methods paradigm.

In the following, the source calculus has *unary* \( \lambda \)-abstractions, which can have free variables, while the target calculus has *binary* \( \lambda \)-abstractions, which must be closed. In the target language, we also use pattern matching over tuples. The translation will be naive, insofar as it will not handle functions of multiple arguments in a special way. One could argue that this is a feature, not a limitation, and that “uncurrying” (if desired) should be a separate type-preserving pass anyway. But closure conversion can also be easily extended to n-ary functions.

There are at least two variants of closure conversion: In the closure-passing variant, the closure and the environment are a single memory block; In the environment-passing variant, the environment is a separate block, to which the closure points. The impact of this choice on the term translations is minor. Closure-passing better supports simple recursive functions; but this is less obvious with mutually recursive ones. Closure-passing optimizes the case of closed functions: they do not need to create a closure—the code pointer can be passed directly [Steckler and Wand, 1997]. However, its impact on the type translations is more important: the closure-passing variant requires more type-theoretic machinery (recursive types and rows).

The closure-passing variant is as follows:

\[
\llbracket \lambda x. a \rrbracket = \text{let code} = \lambda (\text{clo}, x). \text{let} (\_ , x_1, \ldots, x_n) = \text{clo} \text{ in } \llbracket a \rrbracket \text{ in } (\text{code}, x_1, \ldots, x_n)
\]

\[
\llbracket a_1 \ a_2 \rrbracket = \text{let } \text{clo} = \llbracket a_1 \rrbracket \text{ in } \text{let code} = \text{proj}_0 \text{ clo in } \text{code} (\text{clo}, \llbracket a_2 \rrbracket)
\]
where \( \{x_1, \ldots, x_n\} \) is \( \text{fv}(\lambda x. a) \) (the variables \textit{code} and \textit{clo} must be suitably fresh). Note that the layout of the environment must be known only at the closure allocation site, not at the call site. In particular, \( \text{proj}_0 \text{clo} \) need not know the size of \textit{clo}.

The environment-passing variant is as follows:

\[
\begin{align*}
\llbracket \lambda x. a \rrbracket & = \text{let } \text{code} = \lambda(env, x).\text{let } (x_1, \ldots, x_n) = env \text{ in } \llbracket a \rrbracket \text{ in } \\
& \text{(code, } (x_1, \ldots, x_n)) \\
\llbracket a_1 a_2 \rrbracket & = \text{let } (\text{code}, env) = \llbracket a_1 \rrbracket \text{ in } \\
& \text{code (env, } \llbracket a_2 \rrbracket) \\
\end{align*}
\]

where \( \{x_1, \ldots, x_n\} = \text{fv}(\lambda x. a) \).

To understand type-preserving closure conversion, let us first focus on the environment-passing variant. How can closure conversion be made \textit{type-preserving}? The key issue is to find a sensible definition of the type translation. In particular, what is the translation of a function type, \( \llbracket \tau_1 \rightarrow \tau_2 \rrbracket \)? Let us examine the closure allocation code again. Suppose \( \Gamma \vdash \lambda x. a : \tau_1 \rightarrow \tau_2 \). Suppose, without loss of generality (see Remark 5), that \( \text{dom}(\Gamma) \) is exactly \( \text{fv}(\lambda x. a) \), i.e. \( \{x_1, \ldots, x_n\} \). If \( \Gamma \) is \( x_1 : \tau'_1; \ldots; x_n : \tau'_n \), we write \( \llbracket \Gamma \rrbracket \) for \( x_1 : \llbracket \tau'_1 \rrbracket; \ldots; x_n : \llbracket \tau'_n \rrbracket \).

By abuse of notation, we also use \( \llbracket \Gamma \rrbracket \) in a type position to mean the tuple type \( \llbracket \tau'_1 \rrbracket \times \ldots \times \llbracket \tau'_n \rrbracket \).

By hypothesis, we have \( \llbracket \Gamma \rrbracket, x : \llbracket \tau_1 \rrbracket \vdash \llbracket a \rrbracket : \llbracket \tau_2 \rrbracket \), so \textit{env} has type \( \llbracket \Gamma \rrbracket \), \textit{code} has type \( (\llbracket \Gamma \rrbracket \times \llbracket \tau_1 \rrbracket) \rightarrow \llbracket \tau_2 \rrbracket \), and the entire closure has type \( ((\llbracket \Gamma \rrbracket \times \llbracket \tau_1 \rrbracket) \rightarrow \llbracket \tau_2 \rrbracket) \times \llbracket \Gamma \rrbracket \). So, can we adopt \( ((\llbracket \Gamma \rrbracket \times \llbracket \tau_1 \rrbracket) \rightarrow \llbracket \tau_2 \rrbracket) \times \llbracket \Gamma \rrbracket \) as a definition of \( \llbracket \tau_1 \rightarrow \tau_2 \rrbracket \)?

Naturally not. This definition is mathematically ill-formed, as we cannot use \( \Gamma \) out of the blue! That is, we cannot have a translation of \( \llbracket \tau_1 \rightarrow \tau_2 \rrbracket \) that depends on the type of free variables of \( a \)!

Indeed, we need a uniform translation of types, not just because it is nice to have one, but because it describes a uniform calling convention. If closures with distinct environment sizes or layouts receive distinct types, then we will be unable to translate well-typed code: if \( \ldots \text{then } \lambda x. x + y \text{ else } \lambda x. x \)\ldots \ then \( \lambda x. x + y \) \ else \( \lambda x. x \). Furthermore, we want function invocations to be translated uniformly, without knowledge of the size and layout of the closure’s environment.

The only sensible solution is: \( \exists \alpha. ((\alpha \times \llbracket \tau_1 \rrbracket) \rightarrow \llbracket \tau_2 \rrbracket) \times \alpha \). An existential quantification over the type of the environment abstracts away the differences in size and layout. Enough information is retained to ensure that the application of the code to the environment is valid: this is expressed by letting the variable \( \alpha \) occur twice on the right-hand side.

The existential quantification also provides a form of security. The caller cannot do anything with the environment except pass it as an argument to the code. In particular, it cannot inspect or modify the environment. For instance, in the source language, the following coding style guarantees that \( x \) remains even, no matter how \( f \) is used:

\[
\text{let } f = \text{let } x = \text{ref } 0 \text{ in } \lambda(). \text{ } x := (\! x + 2); \! x
\]

After closure conversion, the reference \( x \) is reachable via the closure of \( f \). A malicious, untyped client could write an odd value to \( x \). However, a \textit{well-typed} client is unable to do so. This encoding is not just type-preserving, but also \textit{fully abstract}: it preserves (a typed
version of) observational equivalence (Ahmed and Blume, 2008).

**Remark 5** In order to support the hypothesis \( \text{dom}(\Gamma) = \text{fv}(\lambda x. a) \) at every \( \lambda \)-abstraction, it is possible to introduce an (admissible) *weakening* rule:

\[
\text{Weakening}
\]

\[\frac{\Gamma_1; \Gamma_2 \vdash a : \tau \quad x \not\equiv a}{\Gamma_1; x : \tau'; \Gamma_2 \vdash a : \tau}\]

If the weakening rule is applied eagerly at every \( \lambda \)-abstraction, then the hypothesis is met, and closures have *minimal environments*. (In some cases, one may not use minimal environments, *e.g.* to allow sharing of environments between several closures.)

### 5.2 Existential types

One can extend System F with *existential types*, in addition to universals:

\[\tau ::= \ldots \mid \exists \alpha. \tau\]

As in the case of universals, there are *type-passing* and *type-erasing* interpretations of the terms and typing rules and, in the latter interpretation, there are *explicit* and *implicit* versions. Let us first look at the type-erasing interpretation with an explicit notation for introducing and eliminating existential types.

#### 5.2.1 Existential types in Church style (explicitly typed)

The existential quantifier are introduced and eliminated as follows:

\[
\begin{align*}
\text{Pack} & \quad \frac{\Gamma \vdash M : [\alpha \mapsto \tau']\tau}{\Gamma \vdash \text{pack } \tau', M \text{ as } \exists \alpha. \tau : \exists \alpha. \tau} \\
\text{Unpack} & \quad \frac{\Gamma \vdash M_1 : \exists \alpha. \tau_1 \quad \Gamma, \alpha, x : \tau_1 \vdash M_2 : \tau_2 \quad \alpha \not\equiv \tau_2}{\Gamma \vdash \text{let } \alpha, x = \text{unpack } M_1 \text{ in } M_2 : \tau_2}
\end{align*}
\]

The side condition \( \alpha \not\equiv \tau_2 \) is *mandatory* here to ensure well-formedness of the conclusion. If well-formedness conditions were explicit in judgments, this could be equivalently defined as \( \Gamma \vdash \tau_2 \), as it would imply \( \alpha \not\equiv \tau_2 \) since the last premise implies \( \alpha \not\equiv \Gamma \).

Notice the *imperfect* duality between existential and universals, reminded below:

\[
\begin{align*}
\text{Tabs} & \quad \frac{\Gamma, \alpha \vdash M : \tau}{\Gamma \vdash \Lambda \alpha. M : \forall \alpha. \tau} \\
\text{TApp} & \quad \frac{\Gamma \vdash M : \forall \alpha. \tau}{\Gamma \vdash M : \tau' : [\alpha \mapsto \tau']\tau}
\end{align*}
\]

This suggests a simpler elimination form, perhaps like this:

\[
\frac{\Gamma \vdash M : \exists \alpha. \tau}{\Gamma, \alpha \vdash \text{unpack } M : \tau} \quad \text{*Broken!*}
\]
Informally, this could mean that, if $M$ has type $\tau$ for some unknown $\alpha$, then it has type $\tau$, where $\alpha$ is “fresh”. Unfortunately, this is a broken rule, as we could immediately universally quantify over $\alpha$ and conclude that $\Gamma \vdash M : \forall \alpha. \tau$. This is nonsense! Replacing the premise $\Gamma, \alpha \vdash M : \exists \alpha. \tau$ by the conjunction $\Gamma \vdash M : \exists \alpha. \tau$ and $\alpha \in \text{dom}(\Gamma)$ would make the rule even more permissive, so it wouldn’t help.

A correct elimination rule must force the existential package to be used in a way that does not rely on the value of $\alpha$. Hence, the elimination rule must have control over the user or continuation of the package—that is, over the term $M_2$. The restriction $\alpha \not\in \tau_2$ prevents writing “let $\alpha, x = \text{unpack } M_1$ in $x$”, which would be equivalent to the unsound “unpack $M$” discussed above. The fact that $\alpha$ is bound within $M_2$ forces it to be treated abstractly. In fact, $M_2$ must be polymorphic in $\alpha$. The rule could be written:

$$
\begin{align*}
\Gamma &\vdash M_1 : \exists \alpha. \tau_1 \\
\Gamma &\vdash \Lambda \alpha. \lambda x. M_2 : \forall \alpha. \tau_1 \rightarrow \tau_2 \\
\Gamma &\vdash \text{let } \alpha, x = \text{unpack } M_1 \text{ in } M_2 : \tau_2 \\
\alpha &\not\in \tau_2
\end{align*}
$$

Or, more economically:

$$
\begin{align*}
\Gamma &\vdash M_1 : \exists \alpha. \tau_1 \\
\Gamma &\vdash M_0 : \forall \alpha. \tau_1 \rightarrow \tau_2 \\
\Gamma &\vdash \text{unpack } M_1 M_0 : \tau_2 \\
\alpha &\not\in \tau_2
\end{align*}
$$

where $M_0$ would evaluate to a value of the form $\Lambda \alpha. \lambda x. M_2$.

One could even view “unpack” as a constant with all the types $(\exists \alpha. \tau_1) \rightarrow (\forall \alpha. (\tau_1 \rightarrow \tau_2)) \rightarrow \tau_2$. or, letting $\beta$ range over $\tau_2$, all types $\forall \beta. (\exists \alpha. \tau) \rightarrow (\forall \alpha. (\tau \rightarrow \beta)) \rightarrow \beta$ or even better, $\exists \alpha. \tau \rightarrow \forall \beta. ((\forall \alpha. (\tau \rightarrow \beta)) \rightarrow \beta)$, since $\beta$ should not occur free in $\tau$. We thus introduce a family of constants “$\text{unpack}_{\exists \alpha. \tau}$” with type $\exists \alpha. \tau \rightarrow \forall \beta. ((\forall \alpha. (\tau \rightarrow \beta)) \rightarrow \beta)$. Notice that the variable $\beta$, which stands for $\tau_2$, is bound prior to $\alpha$, so it naturally cannot be instantiated to a type that refers to $\alpha$. This reflects the side condition $\alpha \not\in \tau_2$. If desired, “$\text{pack}_{\exists \alpha. \tau}$” could also be viewed as a constant of type $\forall \alpha. (\tau \rightarrow \exists \alpha. \tau)$. Similarly, we may introduce a constant $\text{pack}$ with all the types $[\alpha \mapsto \tau'] \tau \rightarrow \exists \alpha. \tau$, which we may factor as the following types $\forall \alpha. (\tau \rightarrow \exists \alpha. \tau)$.

In summary, System $F$ with existential types can also be presented by introducing two families of constants with the following types:

$$
\text{pack}_{\exists \alpha. \tau} : \forall \alpha. (\tau \rightarrow \exists \alpha. \tau) \quad \text{unpack}_{\exists \alpha. \tau} : \exists \alpha. \tau \rightarrow \forall \beta. ((\forall \alpha. (\tau \rightarrow \beta)) \rightarrow \beta)
$$

These can be read as follows: for any $\alpha$, if you have a $\tau$, then, for some $\alpha$, you have a $\tau$; conversely, if, for some $\alpha$, you have a $\tau$, then, (for any $\beta$,) if you wish to obtain a $\beta$ out of $\exists \alpha. \tau$, you must present a function which, for any $\alpha$, obtains a $\beta$ out of a $\tau$. This is somewhat reminiscent of ordinary first-order logic: $\exists x. F$ is equivalent to, and can be defined as, $\neg(\forall x. \neg F)$.

One can go one step further and entirely encode existential types into universal types. This encoding is actually a small example of type-preserving translation! The type transla-
CHAPTER 5. EXISTENTIAL TYPES

The term translation is double negation:

\[
\llbracket \exists \alpha. \tau \rrbracket = \forall \beta. ((\forall \alpha. (\llbracket \tau \rrbracket \rightarrow \beta)) \rightarrow \beta)
\]

if \( \beta \neq \tau \)

There is actually little choice for the term translation, if the translation is to be type-preserving:

\[
\llbracket \text{pack}_{\exists \alpha. \tau} \rrbracket : \forall \alpha. (\llbracket \tau \rrbracket \rightarrow \llbracket \exists \alpha. \tau \rrbracket) = \Lambda \alpha. \lambda x : \llbracket \tau \rrbracket. \Lambda \beta. \lambda k : \forall \alpha. (\llbracket \tau \rrbracket \rightarrow \beta). k \alpha x
\]

\[
\llbracket \text{unpack}_{\exists \alpha. \tau} \rrbracket : \llbracket \exists \alpha. \tau \rrbracket \rightarrow (\forall \beta. ((\forall \alpha. (\llbracket \tau \rrbracket \rightarrow \beta)) \rightarrow \beta)) = \lambda x : \llbracket \exists \alpha. \tau \rrbracket. x
\]

This encoding is a continuation-passing transform. This encoding is due to Reynolds [1983], although it has more ancient roots in logic.

When existential are presented as constrants, their semantics is defined by seeing \( \text{pack}_{\exists \alpha. \tau} \) as a unary constructor and \( \text{unpack}_{\exists \alpha. \tau} \) as a unary destructor with the following reduction rule:

\[
\text{unpack}_{\exists \alpha. \tau} (\text{pack}_{\exists \alpha. \tau} \tau' V) \rightarrow \Lambda \beta. \lambda y : \forall \alpha. \tau \rightarrow \beta. y \tau' V \quad (\delta_3)
\]

Exercise 35 Show that this \( \delta \)-rule satisfies the progress and subject reduction assumptions for constants with the types in \( \Delta_3 \). (You may assume that the standard lemmas still hold.)

(Solution p. 183)

Exercise 36 The \( \delta_3 \) reduction for existential is permissive it allows reducing of ill-typed terms. Give a more restrictive version of the rule. What will need to be changed in the proof of subject reduction and progress for the \( \delta \)-rule (Exercise 35)?

(Solution p. 183)

Notice that our \( \delta_3 \)-reduction reduces an “unpack of a pack” to a polymorphic function that applies its argument to the packed value. This is still a form of continuation-passing-style encoding. It seems more natural to treat \( \text{unpack}_{\exists \alpha. \tau} \) as a binary destructor to avoid this intermediate step and have the more intuitive reduction rule:

\[
\text{unpack}_{\exists \alpha. \tau} (\text{pack}_{\exists \alpha. \tau} \tau' V) \tau_1 (\Lambda \alpha. \lambda x : \tau. M) \rightarrow [x \mapsto V][\alpha \mapsto \tau'] M \quad (\delta_3)
\]

However, this does not fit in our framework and notion of arity for constants where all type arguments must be passed first and not interleaved with value arguments. Our framework could be extended to the above \( \delta \)-rules for existentials, but the presentation would become cumbersome.

Alternatively, if existential are primitive, their semantics is defined by extending values and evaluation contexts as follows:

\[
V ::= \ldots \mid \text{pack } \tau', V \text{ as } \tau \quad E ::= \ldots \mid \text{pack } \tau', [] \text{ as } \tau \mid \text{let } \alpha, x = \text{unpack } [] \text{ in } M
\]
and by adding the following reduction rule:

\[
let \alpha, x = \text{unpack} (\text{pack} \tau', V \text{ as } \tau) \text{ in } M \rightarrow [\alpha \mapsto \tau'][x \mapsto V]M
\]

**Exercise 37** Check that the proofs of subject reduction and progress for System F extend to existential types. (Just check the new cases, assuming that the standard lemmas still hold.)

The reduction rule for existential destructs its arguments. Hence, let \( \alpha, x = \text{unpack} M_1 \) in \( M_2 \) cannot be reduced unless \( M_1 \) is itself a packed expression, which is indeed the case when \( M_1 \) is a value (or in head normal form). This contrasts with let \( x : \tau = M_1 \) in \( M_2 \) where \( M_1 \) need not be evaluated and may be an application (e.g. in call-by-name or with strong reduction).

**Exercise 38** The reduction of let \( \alpha, x = \text{unpack} M_1 \) in \( M_2 \) could be problematic when \( M_1 \) is not a value. Illustrate this on an example (You may use the following hint if needed: *lanoitidnocaesu.*)

One may wonder whether the pack construct is not too verbose: isn’t the type witness type annotation \( \tau' \) in rule **Pack** superfluous? The type \( \tau_0 \) of \( M \) is fully determined by \( M \) and the given type \( \exists \alpha. \tau \) of the packed value. Checking that \( \tau_0 \) is of the form \( [\alpha \mapsto \tau'] \tau \) is the matching problem for second-order types, which is simple. However, the reduction rule need the witness type \( \tau' \). If it were not available, it would have to be computed during reduction. The reduction rule would then not be pure rewriting. The explicitly-typed language need the witness type for simplicity, while in the surface language, it could be omitted and reconstructed by second-order matching.

### 5.2.2 Implicitly-typed existential types

Intuitively, pack and unpack are just type information that can be dropped by type erasure. More precisely, the erasure of pack \( \tau', M \) as \( \exists \alpha. \tau \exists \alpha. \tau \) is \( M \) and the erasure of let \( \alpha, x = \text{unpack} M_1 \) in \( M_2 \) is a let-binding let \( x = M_1 \) in \( M_2 \). After type-erasure, the following typing rules for existential types in implicit-typed System F:

\[
\begin{align*}
\text{if-Unpack} & \quad \Gamma \vdash a_1 : \exists \alpha . \tau_1 & \Gamma, \alpha, x : \tau_1 \vdash a_2 : \tau_2 & \alpha \# \tau_2 \\
\Gamma \vdash \text{let } x = a_1 \text{ in } a_2 : \tau_2 \\
\text{if-Pack} & \quad \Gamma \vdash a : [\alpha \mapsto \tau'] \tau \\
\Gamma \vdash a : \exists \alpha . \tau
\end{align*}
\]

Notice, that the let-binding is not typechecked as syntactic sugar for an immediate application. Its semantics remains the same.

\[
E ::= \ldots \text{let } x = [] \text{ in } M \quad \text{let } x = V \text{ in } M \rightarrow [x \mapsto V]M
\]

Is the semantics still type-erasing? Yes, it is, but there is a subtlety! This is only true in call-by-value. In a call-by-name semantics, a let-bound expression is not reduced prior to
substitution of the argument, that is, the rule would be:

$$\text{let } x = a_1 \text{ in } a_2 \rightarrow [x \mapsto a_1]a_2$$

With existential types, this breaks subject reduction! This was first noticed by Sørensen and Urzyczyn (2006). See also Fujita and Schubert (2009, §9).

To see this, let $$\tau_0$$ be $$\exists \alpha. (\alpha \rightarrow \alpha) \rightarrow (\alpha \rightarrow \alpha)$$ and let $$v_0$$ be a value of type $$\text{bool}$$. Then, let $$v_1$$ and $$v_2$$ two values of type $$\tau_0$$ with incompatible witness types, taking for instance, $$\lambda f. \lambda x. 1 + (f \ (1 + x))$$ and $$\lambda f. \lambda x. \text{not} \ (f \ (\text{not} \ x))$$. Let $$v$$ be the function $$\lambda b. \text{if } b \text{ then } v_1 \text{ else } v_2$$ of type $$\text{bool} \rightarrow \tau_0$$, which returns either one of $$V_1$$ or $$V_2$$ depending on its argument $$b$$. We then have the reduction

$$a_1 = \text{let } x = v \ v_0 \text{ in } x \ (\lambda y. y) \rightarrow v \ v_0 \ (v \ v_0 \ (\lambda y. y)) = a_2$$

The typing judgment $$\emptyset \vdash a_1 : \exists \alpha. \alpha \rightarrow \alpha$$ holds, while $$\emptyset \vdash a_2 : \tau$$ does not hold for any $$\tau$$. Indeed, the term $$a_1$$ is well-typed since $$v \ v_0$$ has type $$\tau_0$$, hence $$x$$ can be assumed of type $$(\beta \rightarrow \beta) \rightarrow (\beta \rightarrow \beta)$$ for some unknown type $$\beta$$ and $$\lambda y. y$$ is of type $$\beta \rightarrow \beta$$. However, without the outer existential type $$v \ v_0$$ can only be typed with $$(\forall \alpha. \alpha \rightarrow \alpha) \rightarrow \exists \alpha. (\alpha \rightarrow \alpha)$$, because the value returned by the function need different witnesses for $$\alpha$$. This is demanding too much on its argument and the outer application is ill-typed.

One may wonder whether the syntax should not allow the implicit introduction of unpacking instead. For instance, one could argue that if some expression is the expansion of a well-typed let-binding, then it should also be well-typed:

$$\frac{\Gamma \vdash a_1 : \exists \alpha. \tau_1 \quad \Gamma, \alpha, x : \tau_1 \vdash a_2 : \tau_2 \quad \alpha \neq \tau_2}{\Gamma \vdash [x \mapsto a_1]a_2 : \tau_2}$$

However, this rule is not quite satisfactory as it does not have a logical flavor. Moreover, it fixes the previous example, but does not help with the general case: Pick $$a_1$$ that is not yet a value after one reduction step. Then, after let-expansion reduce one of the two occurrences of $$a_1$$. The result is no longer of the form $$[x \mapsto a_1]a_2$$.

In summary, existential types are tricky: The subject reduction property breaks if reduction is not restricted to expressions in head-normal forms. Unrestricted reduction is still safe because well-typedness may eventually be recovered by further reduction steps—so that progress will never break.

Interestingly, the CPS encoding of existential types (1) enforces the evaluation of the packed value (2) before it can be unpacked (3) and substituted (4):

\[
\begin{align*}
\text{unpack } a_1 (\lambda x. a_2) & = [a_1] (\lambda x. [a_2]) & (1) \\
& \rightarrow (\lambda k. [a_2]) (\lambda x. [a_2]) & (2) \\
& \rightarrow (\lambda x. [a_2]) [a] & (3) \\
& \rightarrow [x \mapsto [a]] [a_2] & (4)
\end{align*}
\]

In the call-by-value setting, $$\lambda k. [a] \ k$$ would come from the reduction of $$\text{pack } a$$, i.e. is $$(\lambda k. \lambda x. k \ x) \ [a]$$, so that $$a$$ is always a value $$v$$. However, $$a$$ need not be a value. What is
essential is again that $a_1$ be reduced to some head normal form $\lambda k. [a] k$.

### 5.2. EXISTENTIAL TYPES

What if one wished to extend ML with existential types? Full type inference for existential types is undecidable, just like type inference for universals. However, introducing existential types in ML is easy if one is willing to rely on user-supplied annotations that indicate where to pack and unpack.

This iso-existential approach was suggested by Läufer and Odersky (1994). Iso-existential types are explicitly declared, much as datatypes:

$$D \overset{\alpha}{=} \exists \beta. \tau$$

if $\text{ftv}(\tau) \subseteq \overset{\alpha}{\alpha} \cup \overset{\beta}{\beta}$ and $\overset{\alpha}{\alpha} \# \overset{\beta}{\beta}$

This introduces two constants, with the following type schemes:

$$\text{pack}_D : \forall \overset{\alpha}{\alpha} \overset{\beta}{\beta}. \tau \rightarrow D \overset{\alpha}{=}$$
$$\text{unpack}_D : \forall \overset{\alpha}{\alpha} \overset{\beta}{\beta}. D \overset{\alpha}{=} \rightarrow (\forall \overset{\beta}{\beta}. (\tau \rightarrow \gamma)) \rightarrow \gamma$$

(Compare with basic iso-recursive types, where $\overset{\beta}{\beta} = \emptyset$.)

Unfortunately, the “type scheme” of $\text{unpack}_D$ is not an ML type scheme. A solution is to make $\text{unpack}_D$ a binary primitive construct, rather than a constant, with an ad hoc typing rule:

$$\Gamma \vdash M_1 : D \overset{\tau}{=} \quad \Gamma \vdash M_2 : \forall \overset{\beta}{\beta}. ([\overset{\alpha}{\alpha} \mapsto \overset{\beta}{\tau}] \tau \rightarrow \tau_2) \quad \overset{\beta}{\beta} \# \overset{\tau}{\tau}, \tau_2$$

where $D \overset{\alpha}{=} \exists \overset{\beta}{\beta}. \tau$

We have seen a version of this rule in System F earlier; this in an ML version. The term $M_2$ must be polymorphic, which can prove.

Iso-existential types are perfectly compatible with ML type inference. The constant $\text{pack}_D$ admits an ML type scheme, so it is not problematic. The construct $\text{unpack}_D$ leads to this constraint generation rule (cf. §3):

$$\text{unpack}_D : \Gamma \vdash M_1 : D \overset{\tau}{=} \quad \Gamma \vdash M_2 : \forall \overset{\beta}{\beta}. ([\overset{\alpha}{\alpha} \mapsto \overset{\beta}{\tau}] \tau \rightarrow \tau_2) \quad \overset{\beta}{\beta} \# \overset{\tau}{\tau}, \tau_2$$

where $D \overset{\alpha}{=} \exists \overset{\beta}{\beta}. \tau$ and, w.l.o.g., $\overset{\beta}{\alpha} \overset{\beta}{\beta} \# M_1, M_2, \tau_2$. Note that a universally quantified constraint appears where polymorphism is required.

In practice, Läufer and Odersky suggest fusing iso-existential types with algebraic data types. The somewhat bizarre Haskell syntax for this is:

```haskell
data D \overset{\alpha}{=} \forall \overset{\beta}{\beta}. \ell \tau
```

where $\ell$ is a data constructor. The elimination construct $\text{case } M_1 \text{ of } \ell x \rightarrow M_2 : \tau_2$ and is typed as follows:

$$\text{case } M_1 \text{ of } \ell x \rightarrow M_2 : \tau_2 = \exists \overset{\alpha}{\alpha}. (\text{def } x : \tau \text{ in } \text{case } M_2 : \tau_2)$$

where, w.l.o.g., $\overset{\alpha}{\alpha} \overset{\beta}{\beta} \# M_1, M_2, \tau_2$. 

Examples  Define \( \text{Any} \approx \exists \beta. \beta \). The following code that attempts to extract the raw content of a package fails:

\[
\text{unpack}_{\text{Any}} M_1 (\lambda x. x) : \tau_2 = \text{unpack}_{\text{Any}} M_1 \wedge \forall \beta. (\lambda x. x : \beta \rightarrow \tau_2) \implies \forall \beta. \beta = \tau_2 \equiv \text{false}
\]

Now, define \( D \alpha \approx \exists \beta. (\beta \rightarrow \alpha) \times \beta \). A client that regards \( \beta \) as abstract succeeds:

\[
\text{unpack}_{D} M_1 (\lambda (f, y). f y) : \tau = \exists \alpha. (\text{unpack}_{D} M_1 : D \alpha) \wedge \forall \beta. (\beta \rightarrow \alpha \times \beta) \rightarrow \tau
\]

Remark 6  We reuse the type \( D \alpha \approx \exists \beta. (\beta \rightarrow \alpha) \times \beta \) of frozen computations, defined above. Assume given a list \( l \) of elements of type \( D \tau_1 \). Assume given a function \( g \) of type \( \tau_1 \rightarrow \tau_2 \). We may transform the list into a new list \( l' \) of frozen computations of type \( D \tau_2 \) (without actually running any computation).

\[
\text{List.map} (\lambda (z) \text{ let } (f, y) = z \text{ in } D((\lambda (z) g (f z)), y))
\]

We may generalize the code into a functional that receives \( g \) and and \( l \) as arguments and returns \( l' \). Unfortunately, the following code does not typecheck:

\[
\text{let lift } g \ l = \text{List.map} (\lambda (z) \text{ let } (f, y) = z \text{ in } D((\lambda (z) g (f z)), y))
\]

The problem is that, in expression \( \text{let } \alpha, x = \text{unpack } M_1 \text{ in } M_2 \), occurrences of \( x \) can only be passed to polymorphic functions so that the type \( \alpha \) of \( x \) does not escape from its scope. That is first-class existential types calls for first-class universal types as well!

Mitchell and Plotkin (1988) note that existential types offer a means of explaining abstract types. For instance, the type:

\[
\exists \text{stack.} \{ \text{empty : stack; push : int } \times \text{stack } \rightarrow \text{stack; pop : stack } \rightarrow \text{option (int } \times \text{stack) } \}
\]

specifies an abstract implementation of integer stacks.

Unfortunately, it was soon noticed that the elimination rule is too awkward, and that existential types alone do not allow designing module systems Harper and Pierce (2005). Montagu and Rémy (2009) make existential types more flexible in several important ways, and argue that they might explain modules after all.

5.2.4 Existential types in OCaml

Amusingly, existential types were first available in OCaml via abstract types and first-class modules. There are now also available as a degenerate case of Generalized Algebraic DataTypes (GADT) which coincides with the appraoached described above.

For example, one may defined the previous datatype of frozen computations:
5.3. Typed closure conversion

Equipped with existential types, we may now revisit type closure conversion.

5.3.1 Environment-passing closure conversion

Remember that we came to the conclusion that the translation of arrow types \([\tau_1 \rightarrow \tau_2]\) must be \(\exists \alpha.(\alpha \times [\tau_1]) \rightarrow [\tau_2] \times \alpha\). Let us show that we may translate expressions so as to preserve well-typedness, i.e. so that \(\Gamma \vdash M : \tau\) implies \([\Gamma] \vdash [M] : [\tau]\). Assume \(\Gamma \vdash \lambda x. M : \tau_1 \rightarrow \tau_2\) and \(\text{dom}(\Gamma) = \{x_1, \ldots, x_n\} = \text{fv}(\lambda x : \tau_1. M)\). We may now hide the dependence on \(\Gamma\) using an existential type:

\[
[\lambda x : \tau_1. M] = \text{let code} : ([\Gamma] \times [\tau_1]) \rightarrow [\tau_2] = \lambda (env : [\Gamma], x : [\tau_1]). \text{let} (x_1, \ldots, x_n : [\Gamma]) = \text{env} \text{ in } [M] \text{ in } \text{pack } [\Gamma], (\text{code}, (x_1, \ldots, x_n)) \text{ as } \exists \alpha.((\alpha \times [\tau_1]) \rightarrow [\tau_2]) \times \alpha = [\tau_1 \rightarrow \tau_2].
\]

In the case of application, assume \(\Gamma \vdash M : \tau_1 \rightarrow \tau_2\) and \(\Gamma \vdash M_1 : \tau_1\) and take:

\[
[M \ M_1] = \text{let } \alpha, (\text{code} : (\alpha \times [\tau_1]) \rightarrow \tau_2, \text{ env } : \alpha) = \text{unpack } [M] \text{ in } \text{code } \text{ (env, } [M_1]) \text{ : } [\tau_2].
\]

For recursive functions we may use the “fix-code” variant (Morrisett and Harper, 1998):

\[
[\mu f. \lambda x. a] = \text{let } \text{rec code } (\text{env, } x) = \text{let } f = \text{pack } \text{ (code, env) in let } (x_1, \ldots, x_n) = \text{env} \text{ in } [a] \text{ in } \text{pack } \text{ (code, } (x_1, \ldots, x_n)).
\]

where \(\{x_1, \ldots, x_n\} = \text{fv}(\mu f. \lambda x. a)\). The translation of applications is unchanged as recursive and non-recursive functions have an identical calling convention. This translation builds recursive code, avoiding a recursive closure, hence the code is easy to type. Unfortunately, as a counterpart, a new closure is allocated at every call, which is the weak point of this variant.
Instead, the “fix-pack” variant \cite{Morisset and Harper, 1998} uses an extra field in the environment to store a back pointer to the closure:

$$\begin{align*}
\llbracket \mu f. \lambda x. a \rrbracket &= \text{let code} = \lambda (env, x). \text{let } (f, x_1, \ldots, x_n) = \text{env in } \llbracket a \rrbracket \text{ in } \\
&\quad \text{let rec } clo = (\text{code}, (clo, x_1, \ldots, x_n)) \text{ in } clo
\end{align*}$$

where \(\{x_1, \ldots, x_n\} = \text{fv}(\mu f. \lambda x. a)\). Hence, we avoid rebuilding the closure at every call by creating a recursive closure. However, this requires, in general, recursively-defined values and closures are now cyclic data structures.

Here is how the “fix-pack” variant is type-checked. Assume \(\Gamma \vdash \mu f : \tau_1 \rightarrow \tau_2. \lambda x. M : \tau_1 \rightarrow \tau_2\) and \(\text{dom}(\Gamma) = \{x_1, \ldots, x_n\} = \text{fv}(\mu f. \lambda x. M)\).

$$\begin{align*}
\llbracket \mu f : \tau_1 \rightarrow \tau_2. \lambda x. M \rrbracket &= \\
&\quad \text{let code} : ((\llbracket f : \tau_1 \rightarrow \tau_2; \Gamma \rrbracket \times \llbracket \tau_1 \rrbracket) \rightarrow \llbracket \tau_2 \rrbracket) = \\
&\quad \lambda (\text{env} : \llbracket f : \tau_1 \rightarrow \tau_2, \Gamma \rrbracket, x : \llbracket \tau_1 \rrbracket). \text{let } (f, x_1, \ldots, x_n) : \llbracket f : \tau_1 \rightarrow \tau_2, \Gamma \rrbracket = \text{env in } \llbracket M \rrbracket \text{ in } \\
&\quad \text{let rec } clo : \llbracket \tau_1 \rightarrow \tau_2 \rrbracket = \\
&\quad \text{pack } \llbracket f : \tau_1 \rightarrow \tau_2, \Gamma \rrbracket, (\text{code}, (clo, x_1, \ldots, x_n)) \text{ as } \exists \alpha((\alpha \times \llbracket \tau_1 \rrbracket) \rightarrow \llbracket \tau_2 \rrbracket) \times \alpha) \text{ in } clo
\end{align*}$$

This implements monomorphic recursion, as by default in ML. To allow the recursive function to be polymorphic, we can generalize the encoding afterwards:

$$\begin{align*}
\llbracket \Lambda \bar{\beta}. \mu f : \tau_1 \rightarrow \tau_2. \lambda x. M \rrbracket &= \Lambda \bar{\beta}. \llbracket \mu f : \tau_1 \rightarrow \tau_2. \lambda x. M \rrbracket
\end{align*}$$

whenever the right-hand side is well-defined. This allows the \textit{indirect} compilation of polymorphic recursive functions as long as the recursion is monomorphic.

Fortunately, the encoding can be straightforwardly adapted to \textit{directly} compile polymorphically recursive functions into polymorphic closure.

$$\begin{align*}
\llbracket \mu f : \forall \bar{\beta}. \tau_1 \rightarrow \tau_2, \lambda x. M \rrbracket &= \\
&\quad \text{let code} : \forall \bar{\beta}. ((\llbracket f : \forall \bar{\beta}. \tau_1 \rightarrow \tau_2; \Gamma \rrbracket \times \llbracket \tau_1 \rrbracket) \rightarrow \llbracket \tau_2 \rrbracket) = \\
&\quad \Lambda \bar{\beta}. \lambda (\text{env} : \llbracket f : \forall \bar{\beta}. \tau_1 \rightarrow \tau_2, \Gamma \rrbracket, x : \llbracket \tau_1 \rrbracket). \\
&\quad \text{let } (f, x_1, \ldots, x_n) : \llbracket f : \forall \bar{\beta}. \tau_1 \rightarrow \tau_2, \Gamma \rrbracket = \text{env in } \llbracket M \rrbracket \text{ in } \\
&\quad \text{let rec } clo : \llbracket \forall \bar{\beta}. \tau_1 \rightarrow \tau_2 \rrbracket = \Lambda \bar{\beta}. \\
&\quad \text{pack } \llbracket f : \forall \bar{\beta}. \tau_1 \rightarrow \tau_2, \Gamma \rrbracket, (\text{code} \ \bar{\beta}, (clo, x_1, \ldots, x_n)) \text{ as } \exists \alpha((\alpha \times \llbracket \tau_1 \rrbracket) \rightarrow \llbracket \tau_2 \rrbracket) \times \alpha) \text{ in } clo
\end{align*}$$

In summary, the environment-passing closure conversion is simple, but it requires the introduction of recursive non-functional values \texttt{let rec } \texttt{x = V in M}. While this is a useful construct, it really alters the operational semantics and requires updating the type soundness proof (as recursive non-functional values were not permitted so far).
5.3.2 Closure-passing closure conversion

Recall the closure-passing variant:

\[
\begin{align*}
\llbracket \lambda x. a \rrbracket &= \text{let } code = \lambda (\text{clo}, x_1, \ldots, x_n) = \text{clo in } \llbracket a \rrbracket \text{ in } \\
\llbracket a_1 \ a_2 \rrbracket &= \text{let } \text{clo} = \llbracket a_1 \rrbracket \text{ in } \text{let } \text{code} = \text{proj}_0 \text{clo in } \text{code} (\text{clo}, \llbracket a_2 \rrbracket)
\end{align*}
\]

where \(\{x_1, \ldots, x_n\} = \text{fv}(\lambda x. a)\).

There are two difficulties to typecheck this: first, a closure is a tuple, whose first field—the code pointer—should be exposed, while the number and types of the remaining fields—the environment—should be abstract; second, the first field of the closure contains a function that expects the closure itself as its first argument.

To describe this, we use two type-theoretic mechanisms; first existential quantification over the tail of a tuple (a.k.a. a row) to allow the environment to remain abstract; and recursive types to allow the closure to point to itself.

**Tuples, rows, row variables** Let us first introduce extensible tuples. The standard tuple types that we have used so far are:

\[
\begin{align*}
\tau &::= \ldots | \Pi R & \quad \text{— types} \\
R &::= \epsilon | (\tau; R) & \quad \text{— rows}
\end{align*}
\]

The notation \((\tau_1 \times \ldots \times \tau_n)\) was sugar for \(\Pi (\tau_1; \ldots; \tau_n; \epsilon)\). Let us introduce row variables and allow quantification over them:

\[
\begin{align*}
\tau &::= \ldots | \Pi R | \forall \rho. \tau | \exists \rho. \tau & \quad \text{— types} \\
R &::= \rho | \epsilon | (\tau; R) & \quad \text{— rows}
\end{align*}
\]

This allows reasoning about the first few fields of a tuple whose length is not known. The typing rules for tuple construction and deconstruction are:

\[
\begin{align*}
\text{Tuple} \\
\forall i. \epsilon [1, n] & \quad \Gamma \vdash M_i : \tau_i \\
\therefore & \quad \Gamma \vdash (M_1, \ldots, M_n) : \Pi (\tau_1; \ldots; \tau_n; \epsilon) \\
\text{Proj} \\
\Gamma \vdash M : \Pi (\tau_1; \ldots; \tau_i; R) & \quad \Gamma \vdash \text{proj}_i M : \tau_i
\end{align*}
\]

These rules make sense with or without row variables. Projection does not care about the fields beyond \(i\). Thanks to row variables, this can be expressed in terms of parametric polymorphism: \(\text{proj}_i : \forall \alpha_1 \ldots \alpha_i \rho. \Pi (\alpha_1; \ldots; \alpha_i; \rho) \rightarrow \alpha_i\).

**Remark 7** Rows were invented by Wand (1988) and improved by Rémy (1994b) in order to ascribe precise types to operations on records. The case of tuples, presented here, is simpler. Rows are used to describe objects in OCaml (Rémy and Vouillon, 1998). Rows are explained in depth by Pottier and Rémy (2005).
Back to closure-passing closure conversion  Rows and recursive types allow to define the translation of types in the closure-passing variant:

\[ [\tau_1 \rightarrow \tau_2] = \exists \rho. \mu \alpha. \Pi (((\alpha \times [\tau_1]) \rightarrow [\tau_2]) ; \rho) \]

\( \rho \) describes the environment represented as a row of fields, which is abstract; \( \alpha \) is the concrete type of the closure that is to refer to recursively; \( \Pi (((\alpha \times [\tau_1]) \rightarrow [\tau_2]) ; \rho) \) is a tuple that begins with a code pointer of type \( (\alpha \times [\tau_1]) \rightarrow [\tau_2] \) and continues with the environment \( \rho \). See the “fix-type” encoding proposed by Morrisett and Harper (1998).

Notice that the type is \( \exists \rho. \mu \alpha. \tau \) and not \( \mu \alpha. \exists \rho. \tau \): The type of the environment is fixed once for all and does not change at each recursive call. Notice that \( \rho \) appears only once, which may seem surprising. Usually, an existential type variable appears both at positive and negative occurrences. Here, the variable \( \alpha \) appear only at a negative occurrence, but in a recursive part of the type that can be unfolded.

To help checking well-typedness of the encoding, let \( \text{Clo}(R) \) abbreviate the concrete type of a closure of row \( R \) and \( \text{UClo}(R) \) its unfolded version:

\[
\text{Clo}(R) \triangleq \mu \alpha. \Pi (((\alpha \times [\tau_1]) \rightarrow [\tau_2]) ; R)
\]

\[
\text{UClo}(R) \triangleq \Pi (((\text{Clo}(R) \times [\tau_1]) \rightarrow [\tau_2]) ; R)
\]

The encoding of arrow types \( [\tau_1 \rightarrow \tau_2] \) is \( \exists \rho. \text{Clo}(\rho) \). The encoding of abstractions and applications is:

\[
[\lambda x: \tau_1. M] = \text{let } code : (\text{Clo}([\Gamma]) \times [\tau_1]) \rightarrow [\tau_2] = \\
\quad \lambda (clo : \text{Clo}([\Gamma]), x : [\tau_1]). \\
\quad \text{let } (\_ \, x_1, \ldots, x_n) : \text{UClo}[\Gamma] = \text{unfold clo in } [M] \text{ in} \\
\quad \text{pack } [\Gamma], (\text{fold } (\text{code}, x_1, \ldots, x_n)) \text{ as } \exists \rho. \text{Clo}(\rho)
\]

\[
[M_1 M_2] = \text{let } \rho, clo = \text{unpack } [M_1] \text{ in} \\
\quad \text{let } code : (\text{Clo}(\rho) \times [\tau_1]) \rightarrow [\tau_2] = \text{proj}_0 (\text{unfold clo}) \text{ in} \\
\quad \text{code } (clo, [M_2])
\]

where \( \{x_1, \ldots, x_n\} = \text{fv}(\lambda x : \tau_1. M) \).

In the closure-passing variant, recursive functions can be translated as follows:

\[
[\mu f. \lambda x. a] = \text{let } code = \lambda (clo, x). \\
\quad \text{let } f = \text{clo in } (\lambda x_1, \ldots, x_n) = \text{clo in } [a] \text{ in} \\
\quad (\text{code}, x_1, \ldots, x_n)
\]

where \( \{x_1, \ldots, x_n\} = \text{fv}(\mu f. \lambda x. a) \). No extra field or extra work is required to store or construct a representation of the free variable \( f \): the closure itself plays this role. However, this untyped code can only be typechecked when recursion is monomorphic.

**Exercise 39** Carefully check well-typedness of the above translation with monomorphic recursion.  \( \Box \)
5.3. TYPED CLOSURE CONVERSION

To adapt this encoding to polymorphic recursion, the problem is that recursive occurrences of \( f \) are rebuilt from the current invocation of the closure, this with the same type since the closure is invoked after type specialization.

By contrast, in the environment passing encoding, the environment contained a polymorphic binding for the recursive calls that was filled with the closure before its invocation, \( i.e. \) with a polymorphic type.

Fortunately, we may slightly change the encoding, using a recursive closure as in the type-passing version, to allow typechecking in System \( \text{F} \).

Remark 8 One could think of changing the encoding of closure types \([\tau_1 \rightarrow \tau_2]\) to make the encoding work. However, although this should be possible in some more expressive type systems, there seems to be no easy way to do so and certainly not within System \( \text{F} \).

Let \( \tau \) be \( \forall \bar{\alpha}. \tau_1 \rightarrow \tau_2 \) and \( \Gamma_f \) be \( f : \tau, \Gamma \) where \( \tilde{\beta} \neq \Gamma \)

\[
[\mu f : \tau. \lambda x. M] = \text{let } code = \\
\Lambda \tilde{\beta}. \lambda (clo : \text{Clo}[\Gamma_f], x : [\tau_1]). \\
\text{let } (\_, code, f_1, \ldots, f_n) : \forall \bar{\beta}. \text{UClo}([\Gamma_f]) = \text{unfold } clo \text{ in } [M] \text{ in} \\
\text{let rec } clo : \forall \tilde{\beta}. \exists \rho. \text{Clo}(\rho) = \\
\Lambda \tilde{\beta}. \text{pack } [\Gamma], (\text{fold } (code \tilde{\beta}, clo, x_1, \ldots, x_n)) \text{ as } \exists \rho. \text{Clo}(\rho) \\
\text{in } clo
\]

Remind that \( \text{Clo}(R) \) abbreviates \( \mu \alpha. \Pi ((\alpha \times [\tau_1]) \rightarrow [\tau_2]; R) \). Hence, \( \tilde{\beta} \) are free variables of \( \text{Clo}(R) \). Here, a polymorphic recursive function is directly compiled into a polymorphic recursive closure. Notice that the type of closures is unchanged, so the encoding of applications is also unchanged.

Optimizing representations Closure-passing and environment-passing closure conversions cannot be mixed because the calling-convention \( i.e., \) the encoding of application) must be uniform. However, theirs is some flexibility in the representation of the closure. For instance, the following change is completely local:

\[
[\lambda x. a] = \text{let } code = \lambda (clo, x). \text{let } (\_, (x_1, \ldots, x_n)) = clo \text{ in } [a] \text{ in} \\
(code, (x_1, \ldots, x_n))
\]

This allows for sharing the closure (or part of it) may be shared when many definitions share the same closure.

5.3.3 Mutually recursive functions

Can we compile mutually recursive functions \( \mu(f_1, f_2). (\lambda x_1. a_1, \lambda x_2. a_2), \text{ say } a? \)
The environment passing encoding is as follows:

\[
[a] = \text{let } code_1 = \lambda(env, x). \text{let } (f_1, f_2, x_1, \ldots, x_n) = env \text{ in } [a_1] \text{ in }
\]

\[
\text{let rec } env = (clo_1, clo_2, x_1, \ldots, x_n)
\]

\[
\text{and } clo_1 = (code_1, env)
\]

\[
\text{and } clo_2 = (code_2, env) \text{ in }
\]

\[
clo_1, clo_2
\]

Notice that we can share the environment inside the two closures. The closure passing encoding is:

\[
[a] = \text{let } code_1 = \lambda(clo, x). \text{let } (\_, f_1, f_2, x_1, \ldots, x_n) = clo \text{ in } [a_1] \text{ in }
\]

\[
\text{let rec } clo_1 = (code_1, clo_1, clo_2, x_1, \ldots, x_n)
\]

\[
\text{and } clo_2 = (code_2, clo_1, clo_2, x_1, \ldots, x_n) \text{ in } clo_1, clo_2
\]

**Question:** Can we share the closures \(c_1\) and \(c_2\) in case \(n\) is large?

Here the environment cannot be shared between the two closures, since they belong to tuples of different size. Unless the runtime, in particular the garbage collector, supports such an operation as returning the tail of a tuple without allocating a new tuple. Then we could write:

\[
[a] = \text{let } code_1 = \lambda(clo, x). \text{let } (\_, \_, f_1, f_2, x_1, \ldots, x_n) = clo \text{ in } [a_1] \text{ in }
\]

\[
\text{let } code_2 = \lambda(clo, x). \text{let } (\_, \_, f_1, f_2, x_1, \ldots, x_n) = clo \text{ in } [a_2] \text{ in }
\]

\[
\text{let rec } clo_1 = (code_1, code_2, clo_1, clo_2, x_1, \ldots, x_n)
\]

\[
\text{and } clo_2 = clo_1.\text{tail}
\]

\[
in clo_1, clo_2
\]

Here \(clo_1.\text{tail}\) returns a pointer to the tail \((code_2, clo_1, clo_2, x_1, \ldots, x_n)\) of \(clo_1\) without allocating a new tuple.

**Encoding of objects**  The closure-passing representation of mutually recursive functions is similar to the representation of objects in the object-as-record-of-functions paradigm:

A class definition is an object generator:

\[
\text{class } c (x_1, \ldots x_q) \{ \text{meth } m_1 = a_i; \ldots \text{meth } m_q = a_i \}
\]

Given arguments for parameter \(x_1, \ldots x_n\), it builds recursive methods \(m_1, \ldots m_n\). A class can be compiled into an object closure:

\[
\text{let } m = \text{let } m = \lambda(m, x_1, \ldots x_q). [a_1] ;
\]

\[
\vdots
\]

\[
m_p = \lambda(m, x_1, \ldots x_q). [a_p] \text{ in }
\]

\[
\lambda x_1, \ldots x_q. (m, x_1, \ldots x_q)
\]
Each \( m_i \) is bound to the code for the corresponding method. All codes are combined into a record of codes. Then, calling method \( m_i \) of an object \( p \) is \((\text{proj}_0 \ p).m_i \ p\).

Let us write the typed version of this encoding. Let \( \tau_i \) be the type of \( M_i \) and row \( R \) describe the types of \((x_1, \ldots, x_q)\). Let \( \text{Clo}(R) \) be \( \mu \alpha.\Pi(\{(m_i : \alpha \rightarrow \tau_i)_{i \in 1..n}; R\}) \) and \( \text{UClo}(R) \) its unfolding.

Fields \( R \) are hidden in an existential type \( \mu \alpha.\Pi(\{(m_i : \alpha \rightarrow \tau_i)_{i \in 1..n}; \rho\}) \):

\[
\text{let } m = \\
\{ \begin{array}{l}
  m_1 = \lambda(m, x_1, \ldots x_q : \text{UClo}(R)).\ M_1; \\
  \vdots \\
  m_p = \lambda(m, x_1, \ldots x_q : \text{UClo}(R)).\ M_p \\
\end{array} \text{ in} \\
\lambda x_1. \ldots \lambda x_q. \text{pack } R, \text{fold } (m, x_1, \ldots x_q) \text{ as } \exists \rho. (M, \rho)
\]

Calling a method of an object \( p \) of type \( M \) is

\[
p \# m_i \triangleq \text{let } \rho, z = \text{unpack } p \text{ in } (\text{proj}_0 \text{ unfold } z). m_i \ z
\]

An object has a recursive type but it is \textit{not} a recursive value.

Typed encoding of objects were first studied in the 90’s to understand what objects really are in a type setting. These encodings are in fact type-preserving compilation of (primitive) objects. There are several variations on these encodings. See Bruce et al. (1999) for a comparison. See Rémy (1994a) for an encoding of objects in (a small extension of) ML with iso-existentials and universals. See Abadi and Cardelli (1996, 1995) for more details on primitive objects.

**Summary**

Type-preserving compilation is rather \textit{fun}. (Yes, really!) It forces compiler writers to make the structure of the compiled program \textit{fully explicit}, in type-theoretic terms. In practice, building explicit type derivations, ensuring that they remain small and can be efficiently typechecked, can be a lot of work.

Because we have focused on type preservation, we have studied only naive closure conversion algorithms. More ambitious versions of closure conversion require program analysis: see, for instance, Steckler and Wand (1997). These versions \textit{can} be made type-preserving.

Defunctionalization, an alternative to closure conversion, offers an interesting challenge, with a simple solution. See, for instance Pottier and Gauthier (2006). Designing an efficient, type-preserving compiler for an \textit{object-oriented language} is quite challenging. See, for instance, Chen and Tarditi (2005).

One may think that references in System F could be translated away by making the store explicit. In fact, this can be done, but not in System F, nor even in System \( F^\infty \): the translation is quite tricky and in order for the translation to be well-typed, the type system must be reach enough to express monotonicity of the store in a context where the store is itself recursively defined. See Pottier (2011) for details.
Exercise 40 (CPS conversion) Here is an untyped version of call-by-value CPS conversion:

\[
\begin{align*}
\llbracket V \rrbracket &= \lambda k. k \llbracket V \rrbracket \\
\llbracket M_1 \ M_2 \rrbracket &= \lambda k. \llbracket M_1 \rrbracket \ (\lambda x_1. \llbracket M_2 \rrbracket \ (\lambda x_2. x_1 \ x_2 \ k)) \\
\llparenthesis x \rrparenthesis &= x \\
\llparenthesis () \rrparenthesis &= () \\
\llparenthesis (V_1, \ \ V_2) \rrparenthesis &= (\llparenthesis V_1 \rrparenthesis, \ \llparenthesis V_2 \rrparenthesis) \\
\llparenthesis (\lambda x. \ M) \rrparenthesis &= \lambda x. \llbracket M \rrbracket
\end{align*}
\]

Is this a type-preserving transformation? (Solution p. 184)
Appendix A

Proofs and Answers to exercises

Solution of Exercise 35

We first need to show that the $\delta_3$ preserves typings. Assume that

$$ \Gamma \vdash \text{unpack}_{\exists \alpha. \tau} (\text{pack}_{\exists \alpha. \tau} \tau' V) : \tau_0 $$

By inversion of typing, $\tau_1$ and $\tau_0$ must be equal to $\tau$ and $\forall \beta. (\forall \alpha. \tau \rightarrow \beta) \rightarrow \beta$, respectively, and the judgment $\Gamma \vdash V : [\alpha \mapsto \tau'] \tau$ must hold. Let $\Gamma'$ be $\Gamma, \beta, y : \forall \alpha. \tau \rightarrow \beta$. By weakening, we have $\Gamma' \vdash V : [\alpha \mapsto \tau'] \tau$. We then have $\Gamma' \vdash y \tau' V : \beta$ and finally, we have

$$ \Gamma \vdash \Lambda \beta. \lambda y : \forall \alpha. \tau \rightarrow \beta. y \tau' V : \tau_0 $$

as expected.

We then need to show that $\delta_3$ satisfies progress, i.e., a full well typed application of $\text{unpack}_{\exists \alpha. \tau}$ can always be reduced. Assume that $\Gamma \vdash \text{unpack}_{\exists \alpha. \tau} V : \tau_0$. By inversion of typing, we must have $\Gamma \vdash V : \exists \alpha. \tau$. By the classification lemma (to be extended and rechecked), $V$ must be an existential value, i.e. of the form $\text{pack}_{\exists \alpha. \tau} \tau_0 V_0$. Hence, $\text{unpack}_{\exists \alpha. \tau} V$ reduces by $\delta_3$.

Solution of Exercise 36

We just force $\tau_1$ to coincide with $\tau$:

$$ \text{unpack}_{\exists \alpha. \tau} (\text{pack}_{\exists \alpha. \tau} \tau' V) \rightarrow \Lambda \beta. \lambda y : \forall \alpha. \tau \rightarrow \beta. y \tau' V $$

The proof of subject reduction will know by construction that $\tau_0$ is $\tau$ instead of learning it by inversion of typing. Conversely for progress, we will have to show that $\tau_1$ and $\tau$ are equal by inversion so that $\delta_3$ can be applied.
Solution of Exercise 38

Let $M_1$ be if $M$ then $V_1$ else $V_2$ where $V_i$ is of the form pack $\tau_i$, $V_i$ as $\exists \alpha \tau$ and the two witnesses $\tau_1$ and $\tau_2$ differ. There is no common type for the unpacking of the two possible results $V_1$ and $V_2$. The choice between those two possible results must be made, by evaluating $M_1$, before unpacking.

Solution of Exercise 40

The answer is in the 2007–2008 exam.
Bibliography

▷ A tour of scala: Implicit parameters. Part of scala documentation.


