Functional Programming Language Design

Christian J. Rudder

January 2025

Contents

This page is left intentionally blank.

Big thanks to **Professor Nathan Mull**
for teaching CS320: Concepts of Programming Languages
at Boston University [2].
Content in this document is based on content provided by Mull.

— 1 —

Functional Programming

1.1 Introduction

Programming Languages (PL) from the perspective of a programmer can be thought of as:

- A tool for programming
- A text-based way of interacting with hardware/a computer
- A way of organizing and working with data

However **This text concerns the design of PLs**, not the sole use of them. It's the difference between knowing how to fly an aircraft vs. designing one. A pilot may know how to steer, but may not know how to design one or, vise versa for the designer. We instead **think in terms of mathematics**, describing and defining the specifications of our language. Our program some mathematical object, a function with strict inputs and outputs.

Definition 1.1: Well-formed Expression

An expression (sequence of symbols) that is constructed according to established rules (syntax), ensuring clear and unambiguous meaning.

Definition 1.2: Programming Language

A PL is made up of three main components:

- **Syntax:** Defines well-formed expressions or programs.
- **Type System:** Delineates the characteristics of possible data.
- **Semantics:** Directs the evaluation of programs or expressions.

Example 1.1: Syntax:

If e_1 is a well-formed expression and e_2 is a well-formed expression, then $e_1 + e_2$ is a well-formed expression. Where <expr> is a placeholder for some arbitrary expression we **Denote:**

<expr> ::= <expr> + <expr>

■

Programmers may have some intuition about what a **variable** is, often thinking of it as a container for data. However, within this context, variables can represent entire expressions and are, in a sense, immutable.

Definition 1.3: Meta-variables

Meta-variables are placeholders that represent arbitrary expressions in a formal syntax. They are used to generalize the structure of expressions or programs within a language.

Example 1.2: Meta-variables:

An expression e could be represented as 3 (a literal) or $3 + 4$ (a compound expression). In this context, variables serve as shorthand for expressions rather than as containers for mutable data. ■

Before talking about types we must understand "**context**" when working with PLs.

Definition 1.4: Context and Typing Environment

In type theory, a context defines an environment which establishes data types for variables. In particular, an environment Γ is a set of ordered list of pairs $\langle x : \tau \rangle$, usually written as $x : \tau$, where x is a variable and τ is its type. We now write a **judgment**, a formal assertion about an expression or program within a given context. We denote:

$$\Gamma \vdash e : \tau$$

which reads "in the context Γ, the expression e has type τ". We may also write judgments for functions, denoting the type of the function and its arguments.

$$f : \tau_1, \tau_2, \ldots, \tau_n \to \tau$$

where f is a function taking n arguments $(\tau_1, \tau_2, \ldots, \tau_n)$, outputting the type τ.

[1]

Tip: Symbol names and command in LaTeX used above are as follows:

- Γ reads as "Gamma" (\Gamma).
- $\vdash$ reads as "turnstile" (\vdash).
- τ reads as "tau" (\tau).

Definition 1.5: Rule of Inference

In formal logic and type theory, an **inference rule** provides a formal structure for deriving conclusions from premises. Rules of inference are usually presented in a **standard form**:

$$\frac{\text{Premise}_1, \quad \text{Premise}_2, \quad \ldots, \quad \text{Premise}_n}{\text{Conclusion}} \text{ (Name)}$$

- **Premises (Numerator):** The conditions that must be met for the rule to apply.
- **Conclusion (Denominator):** The judgment derived when the premises are satisfied.
- **Name (Parentheses):** A label for referencing the rule. [3]

Now we may begin to create a type system for our language, starting with some basic rules.

Example 1.3: Typing Rule for Integer Addition

Consider the typing rule for integer addition for which the inference rule is written as:

$$\frac{\Gamma \vdash e_1 : \text{int} \quad \Gamma \vdash e_2 : \text{int}}{\Gamma \vdash e_1 + e_2 : \text{int}} \text{ (addInt)}$$

This reads as, "If e_1 is an **int** (in the context Γ) and e_2 is an **int** (in the context Γ), then $e_1 + e_2$ is an **int** (in the same context Γ)".

Therefore: let $\Gamma = \{x : \text{int}, y : \text{int}\}$. Then the expression $x + y$ is well-typed as an **int**, since both x and y are integers in the context Γ. ■

Example 1.4: Typing Rule for Function Application

If f is a function of type $\tau_1 \to \tau_2$ and e is of type τ_1, then $f(e)$ is of type τ_2.

$$\frac{\Gamma \vdash f : \tau_1 \to \tau_2 \quad \Gamma \vdash e : \tau_1}{\Gamma \vdash f(e) : \tau_2} \text{ (appFunc)}$$

This reads as, "If f is a function of type $\tau_1 \to \tau_2$ (in the context Γ) and e is of type τ_1 (in the context Γ), then $f(e)$ is of type τ_2 (in the same context Γ)".

Therefore: let $\Gamma = \{f : \text{int} \to \text{bool}, x : \text{int}\}$. Then the expression, $f(x)$, is well-typed as a **bool**, since f is a function that takes an integer and returns a boolean, and x is an integer in the context Γ. ■

Bibliography

[1] Wikipedia contributors. Typing environment — wikipedia, the free encyclopedia, 2023. Accessed: 2023-10-01.

[2] Nathan Mull. Cs320: Concepts of programming languages. Lecture notes, Boston University, Spring Semester, 2025. Boston University, CS Department.

[3] Wikipedia contributors. Rule of inference — Wikipedia, The Free Encyclopedia, 2025. [Online; accessed 22-January-2025].