

An Implementation of Homotopy Type Theory in Isabelle/Pure

JOSHUA CHEN

Born 20th September 1991 in Kuching, Malaysia

12th September 2018

A thesis for the degree of Masters in Mathematics

Advisor: Prof. Dr. Peter Koepke

Second Examiner: P.D. Dr. Philipp Lücke

MATHEMATISCHES INSTITUT

MATHEMATISCH-NATURWISSENSCHAFTLICHE FAKULTÄT DER

RHEINISCHEN FRIEDRICH-WILHELMS-UNIVERSITÄT BONN

Contents

Introduction	v
Notation and conventions	vii
0.1 Orthography	vii
0.2 Notation	vii
0.3 Operator precedences	vii
1 Homotopy type theory	1
1.1 Terms, types, and contexts	1
1.2 Judgments	2
1.3 Universes	2
1.4 Rules for <i>ITT</i>	3
1.5 Univalence and <i>HoTT</i>	5
1.6 Metatheory	7
2 The Isabelle/Pure framework	13
2.1 Formal deductive systems	13
2.2 The Pure logic	17
2.3 The Pure system	21
2.4 Reasoning and proof in Pure	23
3 Implementing <i>HoTT</i> in Pure	31
3.1 Foundational setup	31
3.2 Design issues	34
3.3 Logical rules	36
3.4 Metatheory	40
4 <i>Isabelle/HoTT</i>: an invitation	43
Bibliography	47

Introduction

This thesis presents an implementation of an initial segment of a homotopy type theory in the proof assistant Isabelle.

To elaborate, a *homotopy type theory* is a type theory typically based on some version of Martin-Löf’s intuitionistic type theory [Mar75; Mar82; MS84] with universes, distinguished crucially by its use of the intensional equality type and the univalence axiom, as well as by the presence of higher inductive types. Isabelle [Uni18] is an interactive computer proof assistant, which allows for the formalization of various object logics in a natural deduction style within a “meta-logical framework” called Isabelle/Pure, or simply *Pure*. The logic of Pure is itself a fragment of simple type theory [Chu40], with the primitive logical connectives \implies for entailment/implication and \equiv for equality, and universal quantification \bigwedge over meta-level typed variables. Finally, the goal and major contribution of this thesis is to formalize within the Pure logic a version of homotopy type theory that corresponds suitably to that developed in the Homotopy Type Theory book [Uni13], and also to implement semi-automated routines for constructing and verifying proofs within this theory. We implement an “initial segment” of the theory consisting of the universe hierarchy, the basic types, and the univalence axiom. *W*-types and higher inductive types are left for future work.

In formalizing one formal system \mathcal{L} (the *object logic*) within another \mathcal{F} (the *metalogic/logical framework*), there are often significant differences in the structure and expressive ability¹ of the object- and meta- logics, which creates issues when one comes to implement one inside the other. For example, the standard formalism of the type theory presented in [Uni13] requires the explicit derivation of contexts to be used in hypothetical typing judgments, but there is no simple way of handling object-level contexts in Pure. This results in the impossibility of directly translating the type rules in a way which preserves the formal logical properties discussed above, and we are forced to make design decisions as to good alternative formulations, which must still deviate in some way from the original intended logic. Much of the work of this thesis consists

¹By which is meant the *convenience* with which one can express certain logical constructions, as opposed to the *proof-theoretic strength* of the two logics; it may be the case that both \mathcal{L} and \mathcal{F} can prove the same statements, but one in a much simpler and clearer way than the other. An analogy is the contrast between Conway’s Game of Life and the C++ programming language; both are Turing complete, but there is a reason software is not typically written using cellular automata.

in explaining and justifying these design decisions, and showing that the resulting modified implementation still corresponds well to the original theory.

The guiding goal of the implementation described in this thesis has been to serve as a formalization of some of the content of [Uni13], and an aid in exploring the theory described therein. It is however also hoped that this initial attempt at bringing homotopy type theory to the Isabelle prover will lead to the development of a more powerful and robust library for the logic, perhaps one that will ultimately be comparable to those of systems like Coq and Agda.

Notation and conventions

The following notation and conventions will be introduced throughout the course of this thesis. We gather them here for ease of reference.

0.1 Orthography

- Theory names are written in italics, e.g. *ITT*, *HoTT*.
- Names of axioms and inference rules are given in caps, e.g. `UA` or `UA` for the univalence axiom.
- Names of entities of the Isabelle/Pure framework are written in monospace font, e.g. `prop` for the type of Pure formulas, or `Prod_intro` for the Pure implementation of the dependent function introduction rule.
- Names of ML entities are denoted with underlined italic text, e.g. *thm* for the ML type of Pure theorems, *Var* for the ML term constructor of Pure schematic variables.

0.2 Notation

$\mathcal{P}_f(A)$ denotes the finite powerset of A .

$\mathcal{D}_{\mathcal{N}}$ denotes the deductive core of a natural deduction system \mathcal{N} .

$\mathfrak{t}[\sigma]$ is the instantiation of the term \mathfrak{t} by the higher-order unifier σ .

0.3 Operator precedences

In descending order; entries in the same item have equal precedence and must be disambiguated with parentheses:

1. \cdot (path composition)
2. \circ (function composition)

3. = (equality), \sim (homotopy), \simeq (equivalence)

4. \times (nondependent sum)

5. \rightarrow (nondependent product)

6. \equiv (definitional equality)

Juxtaposition of terms has higher precedence than definitional equality, so

$$\mathbf{f} \mathbf{x} \equiv \mathbf{g} \mathbf{x}$$

means

$$(\mathbf{f} \mathbf{x}) \equiv (\mathbf{g} \mathbf{x}).$$

Binders, e.g. \prod , \sum , λ , are greedy and continue to the end of their innermost enclosing scope. Arrows \implies , \rightarrow and function composition \circ associate to the right. The dependent sum constructor \times and path composition \cdot associate to the left.

Chapter 1

Homotopy type theory

We begin with a brief discussion of the homotopy type theory *HoTT* we aim to implement, and then provide a formal description in Sections 1.4 and 1.5.1. *HoTT* is a fragment of the type theory considered in the Homotopy Type Theory (HoTT) book [Uni13], consisting of the dependent product, dependent sum, empty, unit, coproduct, natural number and equality types, together with a Russell-style universe hierarchy and the univalence axiom.

This chapter elaborates on the presentation given in Appendix A.2 of [Uni13]. Although based on Per Martin-Löf’s intuitionistic type theories [Mar75; Mar82; MS84], an additional judgment is introduced for the well-formedness of contexts, which were handled more implicitly in the original development.

It should be noted that there is not one single “homotopy type theory”—we have chosen a particular system, but there are other ways to formalize and extend the homotopy interpretation of intensional type theory.¹

1.1 Terms, types, and contexts

As usual for a type theory, we have two syntactic classes of expressions: *terms* and *types*.

Types are introduced via *formation rules*. The terms are generated from countably many variable and constant symbols, augmented with additional constructors associated with particular types via *introduction rules*.

Terms a may be assigned types A , and as usual we write $a : A$ to express that a is assigned (“has”) type A .

Definition 1.1.1. A **context** is a list of variable symbol typing assignments

$$x_1 : A_1, \dots, x_n : A_n$$

for some $n \geq 0$ (i.e. it is allowed to be empty).

¹Cubical type theory [CCHM16] is a particularly promising approach in which univalence is a theorem, as opposed to our theory, in which it has to be taken as an axiom.

Contexts are used as “type assumptions” in the judgments of the theory. Each type A_i appearing in the context is allowed to depend on the variables x_1, \dots, x_{i-1} that appear before it, but not on any variables appearing after.

1.2 Judgments

Judgments are the “propositions” of our logical theory—they are the statements which the logic is allowed to prove.

There are three judgment forms

$$\Gamma \text{ ctx}, \quad \Gamma \vdash a : A, \quad \Gamma \vdash a \equiv b : A,$$

where Γ is a context, a and b are terms, and A is a type. They express, respectively, that Γ is a well-formed context, that term a has type A under the type assumptions in Γ , and that a and b are definitionally equal terms of type A under the type assumptions in Γ .

Definitional equality can be seen as a meta-level notion of “symmetric reduction”. It subsumes β - and η -conversion but not α -conversion, which is handled on the meta-level.²

Remark. Martin-Löf also considered the judgment forms

$$A \text{ type}, \quad A \equiv B \text{ type},$$

judging A to be a type and A, B to be definitionally equal types. We do not need these in our theory due to our use of universe types.

1.3 Universes

Intuitively, a universe type is a type whose inhabitants are themselves types. There are two ways to interpret this statement.

The first is to consider the inhabitants A of a universe type U to be *names* for actual types, and to introduce an interpretation operator El taking every name $A : U$ to a type $\text{El}(A)$. This is the *Tarski-style* formulation. It does not conflate the syntactic categories of terms and types and has nice meta-theoretic properties, but is more cumbersome to work with, especially when we come to introduce a hierarchy of universes.

²To be more explicit, standard formulations of type theory typically take the changing of bound variable names in a term expression as a meta-level operation preserving the syntactic identity of the term (subject to side conditions forbidding variable capture). That is, technically speaking, the class of terms is considered modulo the equivalence relation generated by changing bound variable symbols uniformly throughout a term, subject to no-variable-capture. This meta-level syntactic identity may be applied at any time, and is not handled by any of the rules governing definitional equality.

The second approach and the one we take is the *Russell-style* formulation. Here a term $A : U$ is also an actual type that may be used in typing assignments, so that there may be terms $a : A$. Note that in this formulation, the types are a proper subset of the terms.

In *HoTT* we consider not just one universe type, but introduce a countable hierarchy

$$U_0, U_1, \dots$$

such that each term U_i is an inhabitant of the next universe, i.e. we assign

$$U_i : U_{i+1}$$

for all natural numbers i . Note that the indices of the universes are *numerals* i.e. meta-level natural numbers. In addition, we require the hierarchy to be *cumulative*: if $A : U_i$ for some i , then also $A : U_j$ for all $j > i$.

Remark. It is known that while Russell-style universes seem more straightforward to use in practice, they can suffer from meta-theoretic issues, particularly in relation to canonicity and subject reduction [Luo12]. We discuss these issues further in Section 1.6.

1.4 Rules for ITT

We first formally state the axioms and inference rules for *ITT* (intuitionistic type theory), the fragment of *HoTT* without the univalence rule. These hold for all numerals i .

1.4.1 Structural rules

Contexts:

$$\frac{}{\cdot \text{ ctx}} \text{ ctx-EMP} \qquad \frac{\Gamma \vdash A : U_i}{\Gamma, x : A \text{ ctx}} \text{ ctx-EXT}$$

where x is a variable symbol that does not appear in Γ .

$$\frac{x_1 : A_1, \dots, x_n : A_n \text{ ctx}}{x_1 : A_1, \dots, x_n : A_n \vdash x_j : A_j} \text{ VBL}$$

for any $1 \leq j \leq n$. As discussed in Section 1.1, a type A_i appearing in a well-formed context may depend on the variables x_j for $1 \leq j < i$, and never on the variables x_j for $j \geq i$.

Substitution and weakening:

$$\frac{\Gamma \vdash a : A \quad \Gamma, x : A, \Delta \vdash b : B}{\Gamma, \Delta[a/x] \vdash b[a/x] : B[a/x]} \text{ SUBST}_1$$

$$\frac{\Gamma \vdash a : A \quad \Gamma, x : A, \Delta \vdash b \equiv c : B}{\Gamma, \Delta[a/x] \vdash b[a/x] \equiv c[a/x] : B[a/x]} \text{ SUBST}_2$$

$$\frac{\Gamma \vdash A : U_i \quad \Gamma, \Delta \vdash b : B}{\Gamma, x : A, \Delta \vdash b : B} \text{WKG}_1 \qquad \frac{\Gamma \vdash A : U_i \quad \Gamma, \Delta \vdash b \equiv c : B}{\Gamma, x : A, \Delta \vdash b \equiv c : B} \text{WKG}_2$$

Equivalence and structural congruence rules:

$$\frac{\Gamma \vdash a : A}{\Gamma \vdash a \equiv a : A} \equiv\text{-REFL} \qquad \frac{\Gamma \vdash a \equiv b : A}{\Gamma \vdash b \equiv a : A} \equiv\text{-SYM}$$

$$\frac{\Gamma \vdash a \equiv b : A \quad \Gamma \vdash b \equiv c : A}{\Gamma \vdash a \equiv c : A} \equiv\text{-TRANS}$$

$$\frac{\Gamma \vdash a : A \quad \Gamma \vdash A \equiv B : U_i}{\Gamma \vdash a : B} \text{TYPE-CONG} \qquad \frac{\Gamma \vdash a \equiv b : A \quad \Gamma \vdash A \equiv B : U_i}{\Gamma \vdash a \equiv b : B} \text{DEF-CONG}$$

1.4.2 Universes

$$\frac{\Gamma \text{ ctx}}{\Gamma \vdash U_i : U_{i+1}} U_i\text{-INTRO} \qquad \frac{\Gamma \vdash A : U_i}{\Gamma \vdash A : U_{i+1}} U_i\text{-CUMUL}$$

1.4.3 Type rules

These are the formation, introduction, elimination, computation and uniqueness rules as given in Sections A.2.4–A.2.10 of [Uni13].

1.4.4 Term congruence rules

The type rules introduce constants which are built from term constructors taking some number of arguments. A congruence rule expresses the principle that a term constructor applied to definitionally equal arguments yields definitionally equal terms. We state the rules for dependent sum and product here; the rules for the other types are analogous.

Dependent product:

$$\frac{\Gamma \vdash A \equiv A' : U_i \quad \Gamma, x : A \vdash B \equiv B' : U_i}{\Gamma \vdash \prod_{x : A} B \equiv \prod_{x : A'} B' : U_i} \Pi\text{-FORM-CONG}$$

$$\frac{\Gamma \vdash A \equiv A' : U_i \quad \Gamma, x : A \vdash b \equiv b' : B}{\lambda(x : A).b \equiv \lambda(x : A').b' : \prod_{x : A} B} \Pi\text{-INTRO-CONG}$$

$$\frac{\Gamma \vdash f \equiv f' : \prod_{x : A} B \quad \Gamma \vdash a \equiv a' : A}{\Gamma \vdash f(a) \equiv f'(a') : B[a/x]} \Pi\text{-ELIM-CONG}$$

Dependent sum:

$$\frac{\Gamma \vdash A \equiv A' : U_i \quad \Gamma, x : A \vdash B \equiv B' : U_i}{\Gamma \vdash \sum_{x : A} B \equiv \sum_{x : A'} B' : U_i} \Sigma\text{-FORM-CONG}$$

$$\frac{\Gamma, x : A \vdash B : U_i \quad \Gamma \vdash a \equiv a' : U_i \quad \Gamma \vdash b \equiv b' : B[a/x]}{\Gamma \vdash (a, b) \equiv (a', b') : \sum_{x : A} B} \Sigma\text{-INTRO-CONG}$$

$$\frac{\Gamma, z : \sum_{x : A} B \vdash C \equiv C' : U_i \quad \Gamma, x : A, y : B \vdash g \equiv g' : C[(x, y)/z] \quad \Gamma \vdash p \equiv p' : \sum_{x : A} B}{\Gamma \vdash \text{ind}_{\sum_{x : A} B}(z.C, x.y.g, p) \equiv \text{ind}_{\sum_{x : A} B}(z.C', x.y.g', p') : C[p/z]} \Sigma\text{-ELIM-CONG}$$

1.5 Univalence and *HOTT*

Thus far the theory has simply been that of intuitionistic Martin-Löf type theory. Here we introduce the final ingredient of *HOTT*—the univalence axiom.

Univalence is the first major homotopy-type-theoretic idea, so we pause to reiterate the foundational concepts and results. We assume familiarity with the basic terminology and definitions of [Uni13], particularly concerning type families, the equality type, and path induction.

Definition 1.5.1. Let $f, g : \prod_{x : A} B$. Let

$$f \sim g := \prod_{x : A} f(x) =_{B(x)} g(x)$$

be the type of **homotopies between f and g** .

Definition 1.5.2. Let $f : A \rightarrow B$. Define

$$\text{isequiv}(f) := \left(\sum_{g : B \rightarrow A} g \circ f \sim \text{id}_A \right) \times \left(\sum_{g : B \rightarrow A} f \circ g \sim \text{id}_B \right)$$

where $g \circ f := \lambda(x : A). g(f(x))$ and $\text{id}_A := \lambda(x : A). x$. If $\text{isequiv}(f)$ is inhabited we call f an **equivalence**.

Definition 1.5.3. Let $A, B : U_i$. Let

$$A \simeq B := \sum_{f : A \rightarrow B} \text{isequiv}(f)$$

be the type of **equivalences** from A to B .

That is, a homotopy between dependent functions $f \sim g$ expresses *pointwise equality* of f and g ; a function f is an equivalence if it has a pointwise left inverse and a pointwise right inverse, not necessarily identical; and two types A and B are equivalent if there is an equivalence $f: A \rightarrow B$.

Equivalence of types is meant to express the notion of *isomorphism* between mathematical structures, encoded as types. With this in mind, the following result is fitting:

Theorem 1.5.4 (Equal types are equivalent). *Let $A, B: U_i$. There is a function*

$$\text{idtoeqv}_{A,B}: A =_{U_i} B \rightarrow A \simeq B.$$

The proof of this theorem is straightforward and given in [Uni13]; we reproduce it more explicitly here in order to facilitate the formal statement of the univalence axiom later. First we need the following lemma.

Lemma 1.5.5 (Transport lemma). *Given a type $A: U_i$, a type family $P: A \rightarrow U_i$, $x, y: A$ and $p: x =_A y$, we can define a function $\text{transport}_{x,y,p}^P: P(x) \rightarrow P(y)$ satisfying*

$$\text{transport}_{x,x,\text{refl}(x)}^P \equiv \text{id}_{P(x)}$$

for all $x: A$.

Proof. By path induction on x, y, p . Let $C(\tilde{x}, \tilde{y}, \tilde{p}) := P(\tilde{x}) \rightarrow P(\tilde{y})$. It suffices to exhibit $c(\tilde{x}): C(\tilde{x}, \tilde{x}, \text{refl}(\tilde{x})) \equiv P(\tilde{x}) \rightarrow P(\tilde{x})$ for every $\tilde{x}: A$, from which it will follow that

$$\text{ind}_=(\tilde{x}.\tilde{y}.\tilde{p}.C, \tilde{x}.c, x, y, p): C(x, y, p) \equiv P(x) \rightarrow P(y)$$

for all $x, y: A$ and $p: x =_A y$. Defining $c(\tilde{x}) := \text{id}_{P(\tilde{x})}$ for all \tilde{x} gives such a function.

Define

$$\text{transport}_{x,y,p}^P := \text{ind}_=(\tilde{x}.\tilde{y}.\tilde{p}.C, \tilde{x}.c, x, y, p),$$

then

$$\begin{aligned} \text{transport}_{x,x,\text{refl}(x)}^P &\equiv \text{ind}_=(\tilde{x}.\tilde{y}.\tilde{p}.C, \tilde{x}.c, x, x, \text{refl}(x)) \\ &\equiv c(x) \\ &\equiv \text{id}_{P(x)}. \end{aligned}$$

□

Proof of Theorem 1.5.4. Let $p: A =_{U_i} B$ be given, and consider

$$\text{transport}_{A,B,p}^{\text{id}_{U_i}}: A \rightarrow B.$$

We claim that this is an equivalence, so that the desired function is

$$\text{idtoeqv}_{A,B} := \lambda(p: A =_{U_i} B). (\text{transport}_{A,B,p}^{\text{id}_{U_i}}, \mathbf{p}), \quad (1.1)$$

where $\mathfrak{p} : \text{isequiv}(\text{transport}_{A,B,p}^{\text{id}_{U_i}})$.

To construct such a term \mathfrak{p} we use path induction on A, B, p . Let

$$C(\tilde{A}, \tilde{B}, \tilde{p}) := \text{isequiv}(\text{transport}_{\tilde{A}, \tilde{B}, \tilde{p}}^{\text{id}_{U_i}}).$$

We have

$$\text{isequiv}(\text{transport}_{\tilde{A}, \tilde{B}, \tilde{p}}^{\text{id}_{U_i}}) \equiv \text{isequiv}(\text{id}_{\tilde{A}})$$

and

$$c(\tilde{A}) := ((\text{id}_{\tilde{A}}, \lambda(x : \tilde{A}). \text{refl}(x)), (\text{id}_{\tilde{A}}, \lambda(x : \tilde{A}). \text{refl}(x))) : \text{isequiv}(\text{id}_{\tilde{A}})$$

for all $\tilde{A} : U_i$. Thus

$$\mathfrak{p} := \text{ind}_{=}(\tilde{A}. \tilde{B}. \tilde{p}. C, \tilde{A}. c, A, B, p) : \text{isequiv}(\text{transport}_{A,B,p}^{\text{id}_{U_i}}).$$

□

1.5.1 The univalence axiom

In the previous paragraphs we proved that equal types are equivalent, or isomorphic. The converse, that equivalent types are equal, is however not provable. It is instead given by the *univalence axiom* which in fact states slightly more: postulating that the function idtoeqv is itself an equivalence, and hence “invertible”. As a formal inference rule, it states:

$$\frac{\Gamma \vdash A : U_i \quad \Gamma \vdash B : U_i}{\Gamma \vdash \text{univalence}(A, B) : \text{isequiv}(\text{idtoeqv}_{A,B})} \text{UA}$$

where isequiv and idtoeqv are as given in Definition 1.5.2 and (1.1).

Define

$$\text{HoTT} = \text{ITT} + \text{UA},$$

that is, the homotopy type theory we consider is ITT plus the univalence axiom schema.

1.6 Metatheory

When considering a type theory, we often ask if the following problems are decidable. In the following, Γ is some context.

1. **Definitional equality:** given two terms a and b , determine if $\Gamma \vdash a \equiv b : A$ for some type A .
2. **Type checking:** given a term a and a type A , determine if $\Gamma \vdash a : A$.
3. **Type inference:** given a term a , find a type A such that $\Gamma \vdash a : A$.

In [Mar75] a type theory which we will call $MLTT_{73}$ is presented and shown to have a normalization theorem, which then implies decidability of all three problems above *for well-typed terms*. The proof there is given for the empty context, but it is noted that it can be modified to accommodate open terms as well.

1.6.1 Metatheory of ITT

In this section we consider only ITT . ITT is very similar to $MLTT_{73}$, albeit different in a few ways. Its universes are cumulative, so that not every term has a unique type up to definitional equivalence. The rules for Π and Σ type formation (called reflection in [Mar75]) are slightly different to account for the different universe formulations. Also, the definitional equality judgment in ITT is typed while that of $MLTT_{73}$ is not. Despite these differences, we can carry the proof of normalization over to ITT with a small modification of the definitions.

Definition 1.6.1. Define the **reduction relation**

$$\Gamma \vdash a \triangleright b$$

on terms a, b to be the partial order generated by the rules governing definitional equality $\Gamma \vdash a \equiv b : A$, with the exception of the symmetry rule \equiv -SYM.

That is to say, the rules governing $\Gamma \vdash a \triangleright b$ are the reflexive, transitive, substitution, structural congruence, type computation, and term congruence rules, but replacing \equiv with \triangleright .

Note that the reduction relation does not specify *a priori* that a and b have the same type. In particular, it needs to be proved that $\Gamma \vdash a \triangleright b$ and $\Gamma \vdash a : A$ implies $\Gamma \vdash b : A$ (and the analogous statement for definitional equality). In fact this will be a consequence of the normalization theorem.

Theorem 1.6.2 (Normalization theorem for ITT). *There is a normal form $a^* : A^*$ defined on well-typed terms $a : A$, together with an algorithm that gives, for every well-typed closed term $a : A$, a normal term $a^* : A^*$ such that $\vdash a \triangleright a^*$.*

Proof sketch. The normal form on closed well-typed terms is defined inductively:

- Any nullary constructor $c : C$ given by a type formation or introduction rule is a normal term with corresponding *normal type* C . Formation and introduction rules are those rules ending in -FORM or -INTRO, respectively.
- If $c(x_1, \dots, x_n) : A(x_1, \dots, x_n)$ is a constructor with free variables

$$x_1 : A_1, x_2 : A_2(x_1), \dots, x_n : A_n(x_1, \dots, x_{n-1})$$

given by the conclusion of a formation or introduction rule,

$$c_1 : C_1, c_2 : C_2, \dots, c_n : C_n$$

are normal terms with corresponding normal types, and for each $1 \leq i \leq n$ either

- (i) $\vdash A_i[c_1, \dots, c_{i-1}/x_1, \dots, x_{i-1}] \triangleright C_i$, or
- (ii) there are numerals $k \leq l$ such C_i is syntactically identical to U_k and

$$\vdash A_i[c_1, \dots, c_{i-1}/x_1, \dots, x_{i-1}] \triangleright U_l,$$

then

$$c[c_1, \dots, c_n/x_1, \dots, x_n]: A_n[c_1, \dots, c_n/x_1, \dots, x_n]$$

is a normal term with corresponding normal type.

Observe by induction that the normal terms as we have defined them are indeed *normal*, that is, they do not reduce nontrivially.

Every normal term has a unique corresponding normal type. The proof will show that the normal types A^* corresponding to normal terms a^* as given by the inductive definition agree with the normal form of the type $A: U_i$ considered as a term, where U_i is the smallest universe containing A . That is to say, if $a: A$ has normal form $a^*: A^*$, then $A: U_i$ has normal form $A^*: U_i$.

With the above definition of normal terms, the proof of the normalization theorem in Section 3.3 of [Mar75] goes through *mutatis mutandis* for *ITT*. The core idea is to show by induction on the length of a derivation of $\vdash a: A$ that there is a normal term $a^*: A^*$ such that $\vdash a \triangleright a^*$, and by induction on the length of a derivation of $\vdash a \equiv b: A$ that a^* and b^* are syntactically identical. \square

Corollary 1.6.3 (Syntactic uniqueness of normal forms). *If $a: A$ and $b: A$ are closed normal terms with $\vdash a \equiv b: A$, then a and b are syntactically identical.*

Proof. Since a and b are closed normal terms they are syntactically identical to their normal forms a^* and b^* respectively. The result follows from the last sentence of the proof sketch above and the derivation of $\vdash a \equiv b: A$. \square

This means that the closed normal terms form a collection of unique representatives for every \equiv -equivalence class of well-typed terms. We are now able to prove decidability of the problems considered at the beginning of this section.

Theorem 1.6.4. *Definitional equality for well-typed terms is decidable.*

Proof. We can reduce any two well-typed terms $a: A$ and $b: B$ to their normal forms $a^*: A^*$ and $b^*: B^*$. By syntactic uniqueness, $\vdash a \equiv b: A$ if and only if a^* is syntactically identical to b^* and A^* is syntactically identical to B^* , which we can clearly check. \square

We observe that this means that if $\Gamma \vdash a \triangleright b$ for well-typed terms a and b , then $\Gamma \vdash a \equiv b: A$ for some type A .

Theorem 1.6.5. *It is possible to check if a well-typed term is a type. Type checking and type inference for well-typed terms are decidable.*

Proof. Checking if a well-typed term A is a type is solved by reducing A to its normal form and observing its head constant.

Given a well-typed term a and a well-formed type A , we check if $a : A$ by reducing a to its normal form $a^* : B$, reducing A to its normal form $A^* : U_i$, and then checking if B and A^* are syntactically identical.

To infer a type for a well-typed term a , we simply reduce to normal form to obtain $a^* : A^*$. Then $a \equiv a^* : A^*$ and one can show that this implies $a : A^*$. \square

Finally *ITT* is consistent, in that there is no proof of contradiction from the empty context.

Theorem 1.6.6. *In ITT, there is no derivable judgment $\vdash a : \mathbf{0}$.*

Proof. Otherwise a would reduce to a normal term $a^* : \mathbf{0}$. But since $\mathbf{0}$ has no introduction rule, there is no associated constructor and hence no normal term of type $\mathbf{0}$. \square

The following notions are closely related to the normalization and decidability of definitional equality.

Definition 1.6.7. A **canonical term** $a : A$ is one generated only from constructors given by the introduction rules for A . Note that we view the type formation rules as introduction rules for the respective universe types. A type theory satisfies **canonicity** if for every type A and term $a : A$, a is definitionally equal to a canonical term in A .

Definition 1.6.8. A type theory with a reduction relation \rightarrow on terms satisfies **subject reduction** if whenever $a : A$ and $a \rightarrow b$, we also have $b : A$.

Using the normalization theorem, the following theorem is straightforward.

Theorem 1.6.9. *ITT satisfies canonicity. With the reduction relation \triangleright , it also satisfies subject reduction.*

Remark. In Section 1.3 we remarked that cumulative Russell-style universes may cause issues with canonicity and subject reduction. For example, in Section 3 of [Luo12] examples are given showing the failure of both properties for a particular type theory. However we note that both the examples given there use features that are either not present, or formulated differently in *ITT*.

Remark. The results we have stated in this section hold for well-typed terms. It is further known that there are algorithms for type checking arbitrary terms in type theories closely related to *MLTT*₇₃; see for example [Coq96; Mag95]. However it is unclear to the author if these algorithms adapt to the specific case of *ITT*.

1.6.2 Metatheory of $HoTT$ (and variants)

We have seen that the univalence axiom expresses a metamathematical property we would like our type theory to possess, namely that isomorphic structures are equal. We have also seen that the foundational theory ITT we have chosen has a normalization property, and as a consequence, is consistent and has decidable judgments for all well-typed terms. Unfortunately, it is for the moment unclear if all these desirable properties can be merged into $HoTT$.

The univalence axiom schema UA introduces a non-canonical term constructor $\text{univalence}(\cdot, \cdot)$ into specific Σ -types, which breaks canonicity and normalization as defined for ITT . It appears to still be open if there exists a modified normalization process for $HoTT$ (and the theory in [Uni13]).

Metatheoretic consequences of the univalence axiom together with Martin-Löf type theory in general have been considered in other works, for example in [KL12] where it is shown that a variant of homotopy type theory with one Tarski-style universe is consistent relative to

$$\text{ZFC} + \text{there exist two strong inaccessible cardinals } \alpha, \beta.$$

More recently, Coquand et al. [CCHM16] have constructed models of a *cubical type theory* which uses Russell-style universes and *path types* in place of the standard equality type, in which univalence is a theorem and canonicity for the natural numbers holds [Hub17]. In this theory the standard intensional equality type can be defined in terms of the path type.

Chapter 2

The Isabelle/Pure framework

We aim to implement a working version of *HoTT* inside the interactive computer proof assistant Isabelle [Uni18]. Isabelle is a *generic* proof assistant, in that it supports the usage of many different logics of which, loosely speaking, any one may be “loaded” for a given proof session. Currently the largest and most developed logic is *higher-order logic* or HOL, although there are also implementations of first-order logic and Zermelo-Fraenkel set theory, extensional Martin-Löf type theory, modal logic and various sequent calculi, etc. Each of these *object logics* is implemented on top of Isabelle’s base “meta-logical framework”, which is called Isabelle/Pure or simply *Pure*.

Pure provides both an extensible logic in which the terms and formulas of a given object logic can be formalized, as well as a set of procedures for manipulating and proving such formalized statements. In this chapter we give an introduction to both aspects of the framework. This introduction is not intended to be a complete description; we have omitted parts (e.g. type classes, locales etc.) that are not directly used in our implementation of *HoTT*.¹

2.1 Formal deductive systems

Before introducing Pure, it will help to make explicit the framework we have been implicitly using to formulate logical systems like *HoTT*. The definitions in this section are based on Ch. 6 of [HS08] and Ch. 4 of [Gra05].

2.1.1 Hilbert-style deductive systems

Definition 2.1.1. A (Hilbert-style) **deductive system** is a triple

$$\mathcal{D} = (\mathcal{L}, \mathcal{F}, \mathcal{R})$$

¹Readers interested in knowing more are referred to Section 2.1 of [Wen18b] Chapters 2, 4 and 5 of [Wen18a], and Part I of [Pau18] as starting points.

consisting of the *language* of the system, its *formulas*, and its *inference rules*. The language \mathcal{L} is a set with distinguished subsets called *syntactic categories*. The set of formulas \mathcal{F} is a set of well-formed expressions typically defined inductively over the elements of \mathcal{L} . An inference rule $R \in \mathcal{R}$ is given by a set $R = \{(\Phi_\alpha, \phi_\alpha) \mid \alpha \in I_R\}$ of *rule instances*, where each $\Phi_\alpha \subseteq \mathcal{F}$ is a finite set of formulas called the *premises* of the particular rule instance, and $\phi_\alpha \in \mathcal{F}$ is the *conclusion* of the rule instance.

In other words, an inference rule R for a deductive system is a binary relation between $\mathcal{P}_f(\mathcal{F})$ and \mathcal{F} .² We use the usual proof tree notation

$$\frac{\phi_\alpha^1 \quad \cdots \quad \phi_\alpha^n}{\phi_\alpha} R$$

where $\Phi_\alpha = \{\phi_\alpha^1, \dots, \phi_\alpha^n\}$, to denote instances of a rule R (the order of the premises does not matter). Defining a rule by specifying all its members is frequently infeasible, so we will instead define a rule by specifying a *schema* for its instances, in a similar way to how one defines axiom schemas for ZFC and other theories. Rule schemas frequently include *side conditions* further constraining the rule instances.

Definition 2.1.2. An inference rule of the form

$$R = \{(\emptyset, \phi_\alpha) \mid \alpha \in I_R\}$$

is known as an **axiom schema**. The conclusions ϕ_α of its individual instances are known as **axioms**.

Let us view the theory *HoTT* as a deductive system. Its language consists of the expressions of the typed λ -calculus, the term constructors given by the type formation and introduction rules, the univalence constructor, the context variables Γ , Δ , and the improper symbols ctx , \vdash , \equiv and \cdot . There are distinguished syntactic categories of *terms* and *types*. Its formulas are given by the judgment forms described in Section 1.2, and its inference rules are those described in Sections 1.4 and 1.5. There is only one axiom of *HoTT*, namely the conclusion of ctx-EMP ,

$$\cdot \text{ctx.}$$

Note that in the terminology of deductive systems, the univalence axiom schema is really a rule schema.

The following definitions are standard.

Definition 2.1.3. Let $\mathcal{D} = (\mathcal{L}, \mathcal{F}, \mathcal{R})$ be a deductive system, $\Gamma \subseteq \mathcal{F}$, and $\varphi \in \mathcal{F}$. A **derivation of φ from premises Γ in \mathcal{D}** is a labelled tree:

²An inference rule R is usually required to be a *partial function* from $\mathcal{P}_f(\mathcal{F})$ to \mathcal{F} , so that the result of applying R is fully determined by the premises to which it is applied. However we will not require this.

- (i) whose root is labelled φ ,
- (ii) whose set of leaf labels is Γ , and
- (iii) where every label of an internal node u is the conclusion of an instance of an inference rule whose premises are the labels of the children of u .

Leaf labels that are not axioms are the **assumptions** of the derivation. A derivation of φ only from axioms (i.e. without assumptions) is a **derivation of φ** .

We will write derivations using the usual proof tree notation, with conclusion on the bottom.

We write

$$\Gamma \vdash_{\mathcal{D}} \varphi$$

to express that there is a derivation of φ from premises Γ in \mathcal{D} . In such a case φ is said to be *derivable* from Γ . A derivation of φ without assumptions may be written

$$\vdash_{\mathcal{D}} \varphi$$

instead.

Definition 2.1.4. A **theorem** of a deductive system \mathcal{D} is a formula φ of \mathcal{D} such that $\vdash_{\mathcal{D}} \varphi$, i.e. φ is derivable from the axioms of \mathcal{D} alone.

The following concepts will be used when we later come to analyze our implementation of *HoTT*.

Definition 2.1.5. Let $\mathcal{D} = (\mathcal{L}, \mathcal{F}, \mathcal{R})$ be a deductive system, and let

$$R = \{(\Phi_{\alpha}, \phi_{\alpha}) \mid \alpha \in I_R\}$$

be a binary relation from $\mathcal{P}_f(\mathcal{F})$ to \mathcal{F} , not necessarily in \mathcal{R} . R is called:

- (i) a **derivable rule in \mathcal{D}** if $\Phi_{\alpha} \vdash_{\mathcal{D}} \phi_{\alpha}$ for every instance $(\Phi_{\alpha}, \phi_{\alpha})$ of R .
- (ii) an **admissible rule in \mathcal{D}** if adding R to \mathcal{R} does not increase the set of theorems of \mathcal{D} , i.e. if every theorem of $\mathcal{D} + R$ is a theorem of \mathcal{D} .
- (iii) a **correct rule in \mathcal{D}** if for every instance $(\Phi_{\alpha}, \phi_{\alpha})$ of R , whenever the formulas in Φ_{α} are all theorems of \mathcal{D} , then so is ϕ_{α} .

It is clear that a rule R derivable in \mathcal{D} is also admissible in \mathcal{D} : consider any derivation that uses R and replace the application of R with its derivation in \mathcal{D} . The converse is not true in general (see Remark 15.2 of [HS08]).

We also have equivalence of rule admissibility and correctness for deductive systems:

Proposition 2.1.6. *R is an admissible rule in \mathcal{D} if and only if it is correct in \mathcal{D} .*

Proof. Assume R is admissible in \mathcal{D} and let

$$\frac{\phi_\alpha^1 \quad \cdots \quad \phi_\alpha^n}{\phi_\alpha} R$$

be an instance of R whose premises are all theorems of \mathcal{D} . Then ϕ_α is a theorem of $\mathcal{D} + R$, and by admissibility of R is also a theorem of \mathcal{D} .

Conversely let R be correct in \mathcal{D} , and let D be a derivation of a theorem φ that uses R . For each application

$$\frac{\phi_\alpha^1 \quad \cdots \quad \phi_\alpha^n}{\phi_\alpha} R$$

of an instance of R in D , by correctness of R the subderivation of D with root ϕ_α may be replaced with a derivation of ϕ_α that uses only the rules of \mathcal{D} . Doing this iteratively for every occurrence of an application of R yields a derivation of φ using only rules of \mathcal{D} , so φ is a theorem of \mathcal{D} . (One proves this more rigorously using induction.) \square

2.1.2 Natural deduction systems

Here we give a brief indication of how one can define *natural deduction systems*, which differ from Hilbert-style deductive systems in that they incorporate the notions of *assumption* and *discharge*.

Definition 2.1.7. A **natural deduction system** is a triple

$$\mathcal{N} = (\mathcal{L}, \mathcal{F}, \mathcal{R})$$

whose language \mathcal{L} and formulas \mathcal{F} are defined as for deductive systems. A rule $R \in \mathcal{R}$ is a set of the form

$$R = \{(\Phi_\alpha, \phi_\alpha) \mid \alpha \in I_R\}$$

where each $\phi_\alpha \in \mathcal{F}$ and each Φ_α is a finite set whose elements are now either formulas $\phi_\alpha^i \in \mathcal{F}$ or *pairs* $(\psi_\alpha^i, \phi_\alpha^i) \in \mathcal{F} \times \mathcal{F}$. As before, the elements $(\Phi_\alpha, \phi_\alpha)$ of R are known as the *instances* of the rule R .

Instances $(\Phi_\alpha, \phi_\alpha) = (\{(\psi_\alpha^{i_1}, \phi_\alpha^{i_1}), \dots, (\psi_\alpha^{i_m}, \phi_\alpha^{i_m}), \phi_\alpha^{j_1}, \dots, \phi_\alpha^{j_n}\}, \phi_\alpha)$ of a rule R are written

$$\frac{\begin{array}{c} [\psi_\alpha^{i_1}] \\ \vdots \\ \phi_\alpha^{i_1} \end{array} \quad \cdots \quad \begin{array}{c} [\psi_\alpha^{i_m}] \\ \vdots \\ \phi_\alpha^{i_m} \end{array} \quad \phi_\alpha^{j_1} \quad \cdots \quad \phi_\alpha^{j_n}}{\phi_\alpha} R$$

That is, the first coordinate of a pair $(\psi_\alpha^i, \phi_\alpha^i)$ in Φ_α expresses an assumption that is discharged by application of the rule. Again, the order of the premises does not matter.

The definitions of axiom schema and axiom for deductive systems carry over to the case of natural deduction systems.

We can define the notion of a derivation in a natural deduction system by modifying the definition for deductive systems to incorporate discharge of assumptions, essentially by requiring every internal node u of a proof tree D to be labelled by the conclusion of some rule instance $(\Phi_\alpha, \phi_\alpha)$ where

- (i) every $\phi_\alpha^i \in \Phi_\alpha$ is the label of a child of u in D , and
- (ii) for every $(\psi_\alpha^i, \phi_\alpha^i) \in \Phi_\alpha$, there is a child v of u labelled ϕ_α^i and a branch ending at v whose leaf is labelled ψ_α^i .

As we will not actually use such a definition in this thesis, we will not say any more beyond this.

Note that every natural deduction system $\mathcal{N} = (\mathcal{L}, \mathcal{F}, \mathcal{R})$ has a (possibly trivial) Hilbert-style **deductive core** $\mathcal{D}_\mathcal{N} = (\mathcal{L}, \mathcal{F}, \mathcal{R}')$ where $\mathcal{R}' \subseteq \mathcal{R}$ is the collection of rules of \mathcal{N} whose premises are all formulas, i.e. they do not involve discharge of assumptions. In $\mathcal{D}_\mathcal{N}$, all the definitions and results of Section 2.1.1 hold. We will formulate the Pure logic as a natural deduction system, but will thereafter mainly work in its deductive core.

2.2 The Pure logic

The logic of Pure [Pau00] is based on Church's simple type theory, also known as higher-order logic [Chu40; And02]. It is called *simple* because types do not depend on any other values, in contrast to dependent type theory, where types may depend on terms, or polymorphic type theory where they may depend on other types (although as we will see, the Pure logic implements type polymorphism in practice).

2.2.1 Types and terms

The types of Pure are generated inductively by the basic types, and function types of the form

$$\mathbf{T} \Rightarrow \mathbf{S}$$

where \mathbf{T} and \mathbf{S} are types. The single basic type declared by Pure is **prop**, the type of logical propositions; additional basic types may be declared by object logics.

The terms of Pure are the usual terms of the typed lambda-calculus, classified into three major syntactic classes: constants, variables, and lambda expressions. Isabelle annotates terms with the syntactic class to which they belong. The terms are further augmented with the three constants \bigwedge , \implies , and \equiv , which we will define shortly. Object logics may also define additional constants. Terms of type **prop** are called *formulas*; they are the judgments of the Pure logic. They are also the formulas of the theory, seen as a natural deduction system.

The application of a term \mathbf{t} to argument terms $\mathbf{t}_1, \dots, \mathbf{t}_n$ is written in higher-order style, i.e. as

$$(\mathbf{t} \ \mathbf{t}_1 \ \dots \ \mathbf{t}_n)$$

instead of $\mathbf{t}(\mathbf{t}_1, \dots, \mathbf{t}_n)$. In a term \mathbf{t} obtained by the successive application of terms to arguments, the outermost leading term is known as the *head* of \mathbf{t} .

Term expressions in Pure may contain free variables. As usual we will write

$$\mathbf{t}[\mathbf{a}_1, \dots, \mathbf{a}_n/\mathbf{x}_1, \dots, \mathbf{x}_n]$$

to denote the simultaneous replacement of the free variables \mathbf{x}_i with the terms \mathbf{a}_i in the term \mathbf{t} .³

We write

$$\mathbf{t} :: \mathbf{T}$$

to express that term \mathbf{t} has type \mathbf{T} . Note that this expression is not a formula but simply a type annotation, and only appears as a fragment of a larger formula. Isabelle performs automatic inference of the most general type of entities in the system, so type annotations on terms are optional, except in cases where a more specific type is needed.

Pure's type implementation has parametric polymorphism. Type variables are prefixed with apostrophes, as in $'\mathbf{a}$. Note that the original formulation of simple type theory is non-parametric, replacing every instance of a polymorphic constant $\mathbf{t} :: '\mathbf{a}$ with a family of constants $\mathbf{t}_\mathbf{T} :: \mathbf{T}$ for every type \mathbf{T} .

2.2.2 Logical constants

Pure provides three constants for constructing formulas,

$$\begin{aligned} \equiv &:: '\mathbf{a} \Rightarrow '\mathbf{a} \Rightarrow \mathbf{prop} \\ \Longrightarrow &:: \mathbf{prop} \Rightarrow \mathbf{prop} \Rightarrow \mathbf{prop} \\ \bigwedge &:: (' \mathbf{a} \Rightarrow \mathbf{prop}) \Rightarrow \mathbf{prop} \end{aligned}$$

expressing definitional equality, implication, and universal quantification respectively. Note the difference between the Pure arrows \Rightarrow and \Longrightarrow : the first is the function type constructor, which takes type arguments; the second is the implication operator, which takes formula arguments.

We write, as usual, $(\mathbf{a} \equiv \mathbf{b})$ and $(\mathbf{p} \Longrightarrow \mathbf{q})$ in place of $(\equiv \ \mathbf{a} \ \mathbf{b})$ and $(\Longrightarrow \ \mathbf{p} \ \mathbf{q})$. Universal quantification is written as

³This is a notation of theory, as opposed to actual Pure syntax. We will see later that although Pure has a notion of "free variable", it has no native way of expressing the substitution of terms in place of these free variables. We will instead have to encode substitutions $\mathbf{t}[\mathbf{a}/\mathbf{x}]$ as applications of function terms \mathbf{t} of type $\mathbf{T} \Rightarrow \mathbf{S}$ to terms \mathbf{a} .

$$\bigwedge \mathbf{x}. \mathbf{p}$$

instead of $\bigwedge(\lambda \mathbf{x}. \mathbf{p})$. By convention, the scope of the binder \bigwedge extends to the end of the line. We further abbreviate

$$\bigwedge \mathbf{x}^1. \dots \bigwedge \mathbf{x}^k. \mathbf{p}$$

by

$$\bigwedge \mathbf{x}^1 \dots \mathbf{x}^k. \mathbf{p}.$$

2.2.3 Logical rules

Here we state the theory of the inference rules governing the Pure logical constants. We will elaborate on their implementation and usage in practice later. In the following statements, \mathbf{p} and \mathbf{q} are terms of type **prop**, \mathbf{x} and \mathbf{y} are variables, and $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{f}, \mathbf{g}$ stand for arbitrary terms of the appropriate type.

Implication:

$$\frac{\begin{array}{c} [\mathbf{p}] \\ \vdots \\ \mathbf{q} \end{array}}{\mathbf{p} \implies \mathbf{q}} \implies \text{I} \qquad \frac{\mathbf{p} \implies \mathbf{q} \quad \mathbf{p}}{\mathbf{q}} \implies \text{E}$$

where \mathbf{p} is discharged from the premises of $\implies \text{I}$.

Universal quantification:

$$\frac{\mathbf{p}}{\bigwedge \mathbf{x}. \mathbf{p}} \bigwedge \text{I} \qquad \frac{\bigwedge \mathbf{x}. \mathbf{p}}{\mathbf{p}[\mathbf{t}/\mathbf{x}]} \bigwedge \text{E}$$

In $\bigwedge \text{I}$, \mathbf{x} must not be free in any of the assumptions of the derivation of \mathbf{p} . In $\bigwedge \text{E}$, \mathbf{t} is any term of appropriate type.

Definitional equality and lambda expressions:

$$\frac{}{\mathbf{a} \equiv \mathbf{a}} \equiv \text{R} \qquad \frac{\mathbf{a} \equiv \mathbf{b}}{\mathbf{b} \equiv \mathbf{a}} \equiv \text{S} \qquad \frac{\mathbf{a} \equiv \mathbf{b} \quad \mathbf{b} \equiv \mathbf{c}}{\mathbf{a} \equiv \mathbf{c}} \equiv \text{T}$$

$$\frac{}{(\lambda \mathbf{x}. \mathbf{b}) \equiv (\lambda \mathbf{y}. \mathbf{b}[\mathbf{y}/\mathbf{x}])} \alpha\text{-CONV} \qquad \frac{}{(\lambda \mathbf{x}. \mathbf{b}) \mathbf{a} \equiv \mathbf{b}[\mathbf{a}/\mathbf{x}]} \beta\text{-CONV}$$

$$\frac{}{(\lambda \mathbf{x}. \mathbf{f} \mathbf{x}) \equiv \mathbf{f}} \eta\text{-CONV}$$

In $\alpha\text{-CONV}$, \mathbf{y} must not be free in \mathbf{b} . In $\eta\text{-CONV}$, \mathbf{x} must not be free in \mathbf{f} . Note that in contrast

to *HoTT*, the Pure definitional equality has α -conversion built in.

Abstraction and combination rules:

$$\frac{a \equiv b}{(\lambda x. a) \equiv (\lambda x. b)} \text{ ABS} \qquad \frac{f \equiv g \quad a \equiv b}{f a \equiv g b} \text{ COMB}$$

In ABS, x must not be free in any of the assumptions of the derivation of $a \equiv b$.

Function extensionality:

$$\frac{f x \equiv g x}{f \equiv g} \text{ EXT}$$

where x must not be free in f , g , or the assumptions of the derivation of $f x \equiv g x$.

In fact, η -conversion and function extensionality are equivalent:

Proposition 2.2.1 (Theorem 7.4, Hindley and Seldin [HS08]). *Let Pure_η be the theory consisting of the rules given above without the extensionality rule EXT, and let Pure_{EXT} be the theory consisting of the rules given above without η -conversion $\eta\text{-CONV}$. Then $\eta\text{-CONV}$ is derivable in Pure_{EXT} and EXT is derivable in Pure_η .*

Proof. Assume x is not free in f . Then we have the following derivation:

$$\frac{\frac{}{(\lambda x. f x) x \equiv f x} \beta\text{-CONV}}{(\lambda x. f x) \equiv f} \text{ EXT}$$

Conversely, assume x is not free in f , g , or the assumptions of the derivation of

$$f x \equiv g x.$$

Then we have

$$\frac{\frac{\frac{}{(\lambda x. f x) \equiv f} \eta\text{-CONV}}{f \equiv (\lambda x. f x)} \equiv S \quad \frac{f x \equiv g x}{(\lambda x. f x) \equiv (\lambda x. g x)} \text{ ABS}}{f \equiv (\lambda x. g x)} \equiv T \quad \frac{}{(\lambda x. g x) \equiv g} \eta\text{-CONV}}{f \equiv g} \equiv T$$

□

Pure implements EXT as a rule derived from $\eta\text{-CONV}$.

2.3 The Pure system

We have so far mainly kept the presentation of Pure theoretical, however it is unavoidable that we will eventually have to consider details of implementation. In this section we begin to explain how the theory of the previous sections is implemented and used in the Isabelle system. We have attempted, as best as possible, to be clear in our explanations, but there is no substitute for seeing a concrete example for many of the issues discussed here. Hence it may be helpful to return to this section after having read Chapter 3.

2.3.1 Pure and ML

The Isabelle system is written in the statically-typed programming language ML, so the terms and types of Pure are themselves implemented as typed ML terms. Hence there is a distinction between the types T and terms t of Pure on the object-level and the types \underline{T} and terms \underline{t} of ML on the meta-level, which is separate from the object/meta distinction between object logics and Pure.

2.3.2 Axioms and theorems

Pure terms of type **prop** encompass all formulas, including those that are not necessarily true or provable. To distinguish those formulas that should be considered true judgments, Isabelle defines ML terms of a special type *thm*, which wrap formulas p with additional information telling the system to consider p to be proven. These *ML theorems* are introduced into an Isabelle proof session in one of two ways: either axiomatically with the Pure command **axiomatization**, or constructed from other ML theorems using Pure's proof procedures.

Most of the logical rules in the previous section are defined directly as ML theorems in the Isabelle source files [18a; 18b], object logics introduce additional axioms.

2.3.3 Free versus schematic variables

Before explaining how logical rules are encoded and applied in Isabelle, we need to clarify the distinction between two different kinds of variable in the system.

Recall from Section 2.2.1 that the Pure terms are classified syntactically into constants, lambda expressions, and variables. In the actual Pure implementation, a further distinction is made between *free* and *schematic* variables. Logically, these are the same, but internally they are represented differently: free variables are represented by ML terms with head *Free*; schematic variables by ML terms with head *Var*. Functionally, during the proof process the Isabelle system is allowed to instantiate instances of schematic variables, but not free variables. That is, schematic variables, written

$$?x :: T,$$

indicate “placeholders” in formulas which may be instantiated by the system with any Pure term of type T . This instantiation occurs when the system performs unification. On the other hand, a free variable $x :: T$ indicates an arbitrary but *fixed* term of type T , which is not modified by unification.

2.3.4 Unification

Definition 2.3.1. Let t and s be Pure terms containing schematic variables $?x_1, \dots, ?x_n$ and $?y_1, \dots, ?y_m$ respectively. A **higher-order unifier** of t and s is an assignment σ of terms to schematic variables

$$\begin{aligned} ?x_1 &:= a_1, \dots, ?x_n := a_n \\ ?y_1 &:= b_1, \dots, ?y_m := b_m \end{aligned}$$

that respects types, such that

$$t[\sigma] \equiv t[a_1, \dots, a_n / ?x_1, \dots, ?x_n]$$

and

$$s[\sigma] \equiv s[b_1, \dots, b_m / ?y_1, \dots, ?y_m]$$

are syntactically identical modulo β -reduction.

For example, the terms $f (?x_1 \ a)$ and $?y_1 \ (g \ ?y_2)$ have a unifier

$$?x_1 := g, \ ?y_1 := f, \ ?y_2 := a.$$

They also admit more complicated unifiers such as

$$?x_1 := (\lambda x. \ g \ b), \ ?y_1 := f, \ ?y_2 := b.$$

A unifier may leave schematic variables in terms uninstantiated, and may also occasionally introduce new schematic variables.

Isabelle performs higher-order unification using a modification of Huet’s procedure [Hue75], and may occasionally fail to return a unifier even if one exists. This incompleteness is due to the higher-order nature of the problem, in particular the pair of terms

$$?f \ x \ \text{and} \ ?g \ x$$

admit infinitely many unifiers, e.g. $?f = ?g := \lambda x. \ ?h$. Such a pair is known as a *flex-flex pair*. If Isabelle encounters a flex-flex pair, it does not make a choice of unifier and instead separately records the constraint to be handled manually. We will not discuss the details of the unification process any further, interested readers may refer to Sections 3.1 and 7.4 of [Pau18] for more information.

2.4 Reasoning and proof in Pure

In this section we continue considering implementation details specifically pertaining to the proof process in Isabelle.

2.4.1 Inference rules

Much of Pure's power and flexibility comes from its ability to encode natural deduction rules as formulas. Hence it can implement the inference rules of many object logics.

The basic pattern is the following. The Pure implication \Longrightarrow expresses the notion of *entailment* or *derivability*. An inference rule schema

$$\frac{\begin{array}{ccccccc} [\psi^{i_1}] & & [\psi^{i_m}] & & & & \\ \vdots & & \vdots & & & & \\ \phi^{i_1} & \dots & \phi^{i_m} & \phi^{j_1} & \dots & \phi^{j_n} & \\ & & & & & & R \end{array}}{\phi}$$

can be encoded by the formula

$$\llbracket \bigwedge \mathbf{x}_{i_1}^1 \dots \mathbf{x}_{i_1}^{k_{i_1}}. \mathbf{q}_{i_1} \Longrightarrow \mathbf{p}_{i_1}; \dots; \bigwedge \mathbf{x}_{i_m}^1 \dots \mathbf{x}_{i_m}^{k_{i_m}}. \mathbf{q}_{i_m} \Longrightarrow \mathbf{p}_{i_m}; \bigwedge \mathbf{x}_{j_1}^1 \dots \mathbf{x}_{j_1}^{l_{j_1}}. \mathbf{p}_{j_1}; \dots; \bigwedge \mathbf{x}_{j_n}^1 \dots \mathbf{x}_{j_n}^{l_{j_n}}. \mathbf{p}_{j_n} \rrbracket \Longrightarrow \mathbf{p}, \quad (2.1)$$

where \mathbf{p} is the encoding of ϕ , the \mathbf{q}_* and \mathbf{p}_* are encodings of the respective formula schemas ψ^* and ϕ^* as Pure formulas, and each universal quantifier \bigwedge optionally binds occurrences of free variables \mathbf{x}_* in its scope. Here we have used the Pure notation

$$\llbracket \mathbf{p}_1; \dots; \mathbf{p}_n \rrbracket \Longrightarrow \mathbf{p}$$

which abbreviates

$$\mathbf{p}_1 \Longrightarrow \dots \Longrightarrow \mathbf{p}_n \Longrightarrow \mathbf{p},$$

where \Longrightarrow is right-associative.

That is, a formula such as (2.1) encodes an inference rule in which the rightmost implication represents the final “line” of the rule, and the implications inside the double brackets $\llbracket \rrbracket$ represent subderivations, whose premises are to be discharged. This structure may furthermore be nested. In the jargon of proof theory, we say that an inference rule is represented by a Horn clause of hereditary Harrop formulas in Pure.

Inference rules having no premises simply have their conclusion encoded as axioms, so

$$\frac{}{\phi} R$$

is simply given by an encoding \mathbf{p} of ϕ .

From now on we will usually use the term “inference rule” to refer both to abstract inference rules as defined in Definition 2.1.1, as well as to their encoding as Pure formulas, and trust the context to disambiguate which sense we mean. In cases where the distinction matters we will be very explicit.

2.4.2 Goals and backward proof

Consider a deductive system \mathcal{D} . In a derivation D of φ from Γ in \mathcal{D} , one starts with the formulas of Γ and iteratively applies instances of appropriate inference rules in order to finally obtain φ , in this way constructing a proof tree from the top down. In Isabelle, this is known as *forward proof*, where one works forward from assumptions, applying inference rule transformations to finally obtain the conclusion.

We can also read D from the bottom up, starting with the root φ which is to be shown. Now we read the application of inference rule instances *backwards*: a derivation step

$$\frac{\phi^1 \quad \dots \quad \phi^n}{\phi} R$$

in D expresses that in order to derive the *goal* ϕ , it suffices to derive the *subgoals* ϕ^1, \dots, ϕ^n . This forms the basis of Isabelle’s notion of *backward proof*, which we now illustrate with an example.

Consider the derivation given in the proof of Proposition 2.2.1,

$$\frac{\frac{\frac{}{(\lambda x. f x) \equiv f} \eta\text{-CONV}}{f \equiv (\lambda x. f x)} \equiv S \quad \frac{f x \equiv g x}{(\lambda x. f x) \equiv (\lambda x. g x)} \text{ABS}}{f \equiv (\lambda x. g x)} \equiv T \quad \frac{}{(\lambda x. g x) \equiv g} \eta\text{-CONV}}{f \equiv g} \equiv T$$

where the aim is to show that $(f x \equiv g x) \vdash_{\text{Pure}_\eta} (f \equiv g)$. Starting from the goal

$$f \equiv g \tag{2.2}$$

and applying an appropriate instance of the rule $\equiv T$, we obtain the two subgoals

$$f \equiv (\lambda x. f x), \tag{2.3}$$

$$(\lambda x. g x) \equiv g, \tag{2.4}$$

which we can consider as replacing the original goal, in the sense that if we can derive both (2.3) and (2.4), then we can also derive (2.2).

We then repeat this process for each of these subgoals. The second subgoal $(\lambda x. g x) \equiv g$ is the conclusion of an instance of the rule $\eta\text{-CONV}$, which has no premises. Hence applying

η -CONV yields no new subgoals in place of (2.4), that is to say, in order to derive (2.4) it suffices to derive the empty set. Thus (2.4) is trivially proved, befitting its status as an axiom.

This leaves subgoal (2.3). Again we apply an appropriate instance of $\equiv T$ to replace (2.3) by new subgoals

$$\mathbf{f} \equiv (\lambda \mathbf{x}. \mathbf{f} \ \mathbf{x}), \quad (2.5)$$

$$(\lambda \mathbf{x}. \mathbf{f} \ \mathbf{x}) \equiv (\lambda \mathbf{x}. \mathbf{g} \ \mathbf{x}), \quad (2.6)$$

the first of which is then transformed to the trivial goal set by applying $\equiv S$ and then η -CONV. To the second new subgoal (2.6) we apply rule ABS, yielding the subgoal

$$\mathbf{f} \ \mathbf{x} \equiv \mathbf{g} \ \mathbf{x}.$$

This is the single assumption of the derivation we aim to show, and hence we are done.

Remark. The example above illustrates a backward proof that

$$(\mathbf{f} \ \mathbf{x} \equiv \mathbf{g} \ \mathbf{x}) \vdash_{\text{Pure}_\eta} (\mathbf{f} \equiv \mathbf{g}),$$

subject to the relevant side conditions. In practice, we would normally perform an initial application of the Pure rule $\implies I$ in order to show, equivalently, that

$$\vdash_{\text{Pure}_\eta} (\mathbf{f} \ \mathbf{x} \equiv \mathbf{g} \ \mathbf{x}) \implies (\mathbf{f} \equiv \mathbf{g}).$$

2.4.3 Proof states, lifting, and resolution

Backward proof is a basic way Pure constructs new ML theorems from existing ones. We now describe its implementation.

Definition 2.4.1. A **proof state** is a Pure formula of the form

$$\llbracket \mathbf{p}_1; \dots; \mathbf{p}_n \rrbracket \implies \mathbf{p} \quad (2.7)$$

where the \mathbf{p}_i are formulas known as the **subgoals** of the proof state, and \mathbf{p} is a formula known as the **goal**.

A proof state (2.8) expresses that in order to prove the goal \mathbf{p} , it suffices to prove the subgoals $\mathbf{p}_1, \dots, \mathbf{p}_n$, where n may be zero. The goal \mathbf{p} may itself be an implication. Note that proof states and inference rules have the same syntactic form and dual interpretations.

Proof states are transformed using inference rules by a method called *lifted resolution*. In the following definitions, let

$$\mathbf{R}: \llbracket \mathbf{r}_1; \dots; \mathbf{r}_m \rrbracket \implies \mathbf{r}$$

be an inference rule encoded in the form (2.1).

Definition 2.4.2. Let q_1, \dots, q_k be formulas. The **lift of R over formulas** q_1, \dots, q_k is the formula

$$\begin{aligned} \llbracket q_1; \dots; q_k \rrbracket \Longrightarrow r_1; \dots; \llbracket q_1; \dots; q_k \rrbracket \Longrightarrow r_m \Longrightarrow \\ (\llbracket q_1; \dots; q_k \rrbracket \Longrightarrow r). \quad (\text{LIFT}_{\text{FORMS}}(\mathbf{R})) \end{aligned}$$

It is easily shown that $\text{LIFT}_{\text{FORMS}}(\mathbf{R})$ is in fact a derivable inference rule in the theory $\text{Pure}+\mathbf{R}$ consisting of the inference rules of Section 2.2.3 together with \mathbf{R} .

Definition 2.4.3. Let x^1, \dots, x^k be terms. The **lift of R over parameters** x^1, \dots, x^k is the formula

$$\llbracket \bigwedge x^1 \dots x^k. r_1^*; \dots; \bigwedge x^1 \dots x^k. r_m^* \rrbracket \Longrightarrow \bigwedge x^1 \dots x^k. r^*, \quad (\text{LIFT}_{\text{PARAMS}}(\mathbf{R}))$$

where s^* is the formula obtained from s by replacing every instance of a schematic variable $?y$ in s by the schematic term

$$(?y \ x^1 \dots x^k).$$

Again, one can show (Section 6.1, [Pau00]) that $\text{LIFT}_{\text{PARAMS}}(\mathbf{R})$ is a derivable rule in $\text{Pure}+\mathbf{R}$.

As an illustration consider the following encoding of the \vee -introduction rule of intuitionistic propositional logic

$$\text{Or_intro: } ?p \Longrightarrow (?p \vee ?q),$$

which corresponds to

$$\frac{p}{p \vee q} \vee\text{I.}$$

Lifting Or_intro over the formula r and the parameter x yields

$$(\bigwedge x. r \Longrightarrow (?p \ x)) \Longrightarrow (\bigwedge x. r \Longrightarrow ((?p \ x) \vee (?q \ x))),$$

which expresses the derivable inference rule

$$\frac{\begin{array}{c} [r] \\ \vdots \\ p(x) \end{array}}{r \Longrightarrow (p(x) \vee q(x))}$$

of intuitionistic predicate logic, where x is arbitrary. Note that we lifted over the *free* formula r and not the *schematic* formula $?r$, so we do not replace r by $(r \ x)$.

The derivability of the two lifting rules means that lifting is sound:

Theorem 2.4.4. *The lift of a Pure theorem over arbitrary formulas and parameters is a Pure theorem. Applying a lifted Pure theorem to Pure theorems yields a Pure theorem.*

Proof. Let \mathbf{R} be a Pure theorem. Then the lifting rules $\text{LIFT}_{\text{FORMS}}(\mathbf{R})$ and $\text{LIFT}_{\text{PARAMS}}(\mathbf{R})$ are derivable as abstract inference rules in $\text{Pure} + \mathbf{R} = \text{Pure}$. By discharging the premises of their derivations using $\implies\text{I}$, we see that they are in fact derivable as concrete theorems of Pure. Hence the lift of a Pure theorem over arbitrary formulas and parameters is also a Pure theorem.

The lift of a Pure theorem is in fact derivable as an abstract inference rule in the deductive core $\mathcal{D}_{\text{Pure}}$ of Pure, i.e. it is derivable using only the inference rules that do not discharge assumptions. Hence it is admissible, and by Proposition 2.1.6 it is also correct. Thus applying it to theorems yields a theorem. \square

The Pure logical rules and any additional inference rules defined by object logics are theorems of Pure. Theorem 2.4.4 then says that applying lifted inference rules to Pure theorems yields a Pure theorem. To implement backward proof we need one final definition:

Definition 2.4.5. Let

$$\mathbf{P}: \llbracket \mathbf{p}_1; \dots; \mathbf{p}_n \rrbracket \implies \mathbf{p} \quad (2.8)$$

be a proof state, and

$$\mathbf{R}: \llbracket \mathbf{r}_1; \dots; \mathbf{r}_m \rrbracket \implies \mathbf{r} \quad (2.9)$$

an inference rule.

Suppose there is a subgoal \mathbf{p}_i of \mathbf{P} , a lift

$$\mathbf{R}^\uparrow: \llbracket \mathbf{r}_1^\uparrow; \dots; \mathbf{r}_m^\uparrow \rrbracket \implies \mathbf{r}^\uparrow \quad (2.10)$$

of \mathbf{R} over formulas $\mathbf{q}_1, \dots, \mathbf{q}_k$ and parameters $\mathbf{x}^1, \dots, \mathbf{x}^l$, and a unifier σ of \mathbf{r}^\uparrow and \mathbf{p}_i . Then the **resolvent of \mathbf{P} with \mathbf{R} with respect to subgoal \mathbf{p}_i** is the formula

$$(\llbracket \mathbf{p}_1; \dots; \mathbf{p}_{i-1}; \mathbf{r}_1^\uparrow; \dots; \mathbf{r}_m^\uparrow; \mathbf{p}_{i+1}; \dots; \mathbf{p}_n \rrbracket \implies \mathbf{p})[\sigma] \quad (2.11)$$

The resolvent of \mathbf{P} with \mathbf{R} is a new proof state. In words, it is formed by finding a subgoal \mathbf{p}_i with which the lifted conclusion of \mathbf{R} unifies, and then replacing \mathbf{p}_i with new subgoals given by the lifted hypotheses of \mathbf{R} and propagating the assignments given by the unifier throughout the new proof state.

Given a rule \mathbf{R} and a proof state \mathbf{P} as in (2.8) and (2.9), the process of finding a resolvent of \mathbf{P} with \mathbf{R} is known as **resolution**. It can be encapsulated by the following rule, where \mathbf{P}_{RES} is as given in (2.11):

$$\frac{\mathbf{R} \quad \mathbf{P}}{\mathbf{P}_{\text{RES}}} \text{ RES}$$

which holds whenever there is a unifier σ of some subgoal \mathbf{p}_i of \mathbf{P} with the conclusion of \mathbf{R}^\uparrow lifted over the premises and parameters of \mathbf{p}_i .

Theorem 2.4.6. *The inference rule RES is correct in Pure.*

Proof. Assume that \mathbf{R} and \mathbf{P} are theorems of Pure. By Theorem 2.4.4, \mathbf{R}^\uparrow is a theorem. Also, $\mathbf{R}^\uparrow[\sigma]$ and $\mathbf{P}[\sigma]$ are theorems for any instantiation σ of schematic variables.

Consider \mathbf{P}_{RES} as an abstract inference rule

$$\frac{\mathbf{p}_1[\sigma] \quad \cdots \quad \mathbf{p}_{i-1}[\sigma] \quad \mathbf{r}_1^\uparrow[\sigma] \quad \cdots \quad \mathbf{r}_m^\uparrow[\sigma] \quad \mathbf{p}_{i+1}[\sigma] \quad \cdots \quad \mathbf{p}_n[\sigma]}{\mathbf{p}[\sigma]}$$

This rule is derivable in $\text{Pure} + \mathbf{R}^\uparrow[\sigma] + \mathbf{P}[\sigma]$ since from $\mathbf{r}_1^\uparrow[\sigma], \dots, \mathbf{r}_m^\uparrow[\sigma]$ we can derive $\mathbf{r}^\uparrow[\sigma]$, which is syntactically identical to $\mathbf{p}_i[\sigma]$. In turn, together with the other premises $\mathbf{p}_j[\sigma]$ we can then derive $\mathbf{p}[\sigma]$. Finally, by applying $\implies\text{I}$ we obtain a derivation

$$\vdash_{\text{Pure} + \mathbf{R}^\uparrow[\sigma] + \mathbf{P}[\sigma]} \mathbf{P}_{\text{RES}}.$$

But $\text{Pure} + \mathbf{R}^\uparrow[\sigma] + \mathbf{P}[\sigma] = \text{Pure}$ since $\mathbf{R}^\uparrow[\sigma]$ and $\mathbf{P}[\sigma]$ are theorems. Hence \mathbf{P}_{RES} is a Pure theorem. \square

2.4.4 Proof by resolution

The results of the last section give us a rigorous theoretical basis for backward proof in Pure.

Assume we are working in an extension

$$\text{Pure} + \mathcal{R}$$

of *Pure*, with additional inference rules \mathcal{R} encoded as Pure formulas by some object logic. To prove a formula \mathbf{p} , we begin with the proof state

$$\mathbf{p} \implies \mathbf{p},$$

which is trivially a Pure theorem. We transform the proof state by applying instances of RES with appropriate inference rules: this top-down proof using resolution corresponds to building a proof tree using the inference rules from the bottom up. Searching for a unifier in each resolution step corresponds to choosing a suitable inference rule instance to apply.

If an instance of an axiom \mathbf{R} occurs as a subgoal \mathbf{p}_i of a proof state, unifying with \mathbf{R} with respect to \mathbf{p}_i will result in \mathbf{p}_i being replaced with no new subgoals, effectively removing a subgoal from the proof state. Our aim is thus to transform a proof state in order to eliminate subgoals in this way. Once we have succeeded in removing all subgoals, the proof state will read

$$\implies \mathbf{p}[\sigma]$$

or trivially, $\mathbf{p}[\sigma]$, where there may be instantiations of schematic variables in \mathbf{p} from the unification performed during resolution. From the results of the previous section, we then have that $\mathbf{p}[\sigma]$ is a theorem of $\text{Pure} + \mathcal{R}$.

2.4.5 Proof by assumption

Resolution is one of Isabelle's many methods of proof, which are known as *tactics*. We briefly mention another tactic here, called *assumption*. It is trivial but often necessary, and proves formulas of the form

$$\llbracket \mathbf{p}_1; \dots; \mathbf{p}_n \rrbracket \Longrightarrow \mathbf{p}$$

where one of the premises \mathbf{p}_i unifies with the conclusion \mathbf{p} . It is an easy exercise to show that assumption can be formulated as a derivable inference rule in Pure.

There are numerous examples of lifting and proof by resolution and assumption in Sections 5.3 and 5.4 of [Pau18], as well as in [Pau00].

Chapter 3

Implementing *HoTT* in Pure

Now that we have a basic understanding of the Isabelle/Pure framework, we are ready to implement *HoTT* as an object logic inside Pure. We will call the theory of our implementation *HoTT_P*.¹ Currently the implementation itself is provided as an Isabelle object logic library *Isabelle/HoTT* [Che18]. In this chapter we provide a high-level overview of the theory and implementation of *HoTT_P*, leaving the details to the description text contained within the library files.

3.1 Foundational setup

3.1.1 Object types and terms

We begin by defining a new basic Pure type

$$\mathbf{Term},$$

whose terms will be the object-level types and terms of *HoTT_P*. In formulations of type theory without universes or with Tarski-style universes, one would usually define distinct metatypes \mathbf{i} and \mathbf{t} for the object terms and object types respectively. This is done for example in *CTT*, the Pure implementation of Martin-Löf constructive type theory. However, when we use Russell-style universes, the universe types may act as terms in some judgments and as types in others. This forces us to treat object terms and types uniformly, as instances of the same sort of metaterm.

Remark. One can view this setup as a triply-sorted pure type system (PTS) [nLa18b], where the sorts are the sort of object-level terms *term*, the sort of object-level types *type*, and \mathbf{Term} . The axioms are then

$$term : \mathbf{Term} \quad \text{and} \quad type : \mathbf{Term}.$$

¹Other names that were considered but rejected were *PHo* (for being too culinary) and *PHTT* (too dismissive). A personal favorite of the author's was *HoTTiP*, aka “Homotopy Type Theory in Pure”, which in his admittedly biased opinion reflects the promise of this area of research.

We will however not use the PTS framework in this thesis.

3.1.2 Judgments

Next we need to implement the judgments of *HoTT*. Judgments of object logics are implemented in Pure by defining a function term that takes arguments of judgment forms and wraps them to return a Pure formula. In an Isabelle theory session we define a new Pure constant

$$\text{has_type} :: [\text{Term}, \text{Term}] \Rightarrow \text{prop}$$

that takes two arguments of type `Term` and returns a Pure formula. In this way, formulas

$$(\text{has_type } a \ A)$$

now encode the *HoTT* typing assignment $a : A$ and are able to be reasoned about using the proof methods discussed in Chapter 2. Instead of $(\text{has_type } a \ A)$, we will write

$$a : A.$$

Note that the notation

$$[t_1, \dots, t_n] \Rightarrow t$$

abbreviates

$$t_1 \Rightarrow \dots \Rightarrow t_n \Rightarrow t,$$

where \Rightarrow is right-associative.

Recall that there are three judgment forms of *HoTT*,

$$\Gamma \text{ ctx}, \quad \Gamma \vdash a : A, \quad \Gamma \vdash a \equiv b : A,$$

all of which involve contexts Γ . It does not appear to be straightforward to directly implement contexts in the Pure framework—a particular difficulty is the question of how variables should be assigned object-level types. Hence we handle variable typing assignments differently in Pure, and encode the *HoTT* judgments

$$x_1 : A_1, \dots, x_n : A_n \vdash a : A$$

and

$$x_1 : A_1, \dots, x_n : A_n \vdash a \equiv b : A$$

as the *HoTT_P* formulas

$$\bigwedge x_1 \dots x_n. \llbracket x_1 : A_1; x_2 : (A_1 \ x_1); \dots; x_n : (A_n \ x_1 \dots x_{n-1}) \rrbracket \implies (a \ x_1 \dots x_n) : (A \ x_1 \dots x_n) \quad (3.1)$$

and

$$\bigwedge_{\mathbf{x}_1 \dots \mathbf{x}_n} \llbracket \mathbf{x}_1 : \mathbf{A}_1; \mathbf{x}_2 : (\mathbf{A}_1 \ \mathbf{x}_1); \dots; \mathbf{x}_n : (\mathbf{A}_n \ \mathbf{x}_1 \dots \mathbf{x}_{n-1}) \rrbracket \implies (\mathbf{a} \ \mathbf{x}_1 \dots \mathbf{x}_n) \equiv (\mathbf{b} \ \mathbf{x}_1 \dots \mathbf{x}_n) \quad (3.2)$$

respectively. Observe that whereas in $HoTT$ we treat the terms a , A , etc. as *term expressions* which may contain *free variables* x_i , in $HoTT_P$ we must consider them as *function terms* \mathbf{a} , \mathbf{A} , etc. of the Pure framework, and explicitly express their dependency on the *parameters* \mathbf{x}_i (though this dependency is optional).

There is no equivalent in $HoTT_P$ of the context judgment

$$\Gamma \text{ ctx}$$

of $HoTT$. Its purpose in $HoTT$ is to ensure well-typedness of terms and well-formedness of dependent types when formulating the type formation and introduction rules; we will later accomplish the same effect using universal quantification and premises giving type constraints for parameters. In some cases, this means we have to add premises to the type rules in $HoTT_P$ in order to maintain certain properties present in $HoTT$ that would be lost if we simply translated the rules “naïvely”.

3.1.3 Definitional equality

We remark briefly on the implementation of definitional equality.

We use the Pure equality as the definitional equality in $HoTT_P$, and do not define a separate judgment function for it as we did for the typing judgment. Definitional equality is typed in $HoTT$, but the Pure equality is untyped, and when we translate the definitional equality rules over into $HoTT_P$ as in (3.2) we simply drop the type annotation. This means that in $HoTT_P$ it automatically holds that $\mathbf{a} : \mathbf{A}$ and $\mathbf{a} \equiv \mathbf{b}$ together imply $\mathbf{b} : \mathbf{A}$, hence subject reduction automatically holds.

It turns out that from the Pure rules we can derive most of the rules governing definitional equality in $HoTT_P$, and we only have to explicitly axiomatize term congruence rules for some of the constructors for the dependent types.

An issue to be aware of is that the definitional equality of $HoTT$ does not have α -conversion, which is instead handled on the meta-level. However, the Pure equality does have rules for α -conversion, thus our implementation conflates this aspect of the object- and meta-theory of $HoTT$. Functionally, however, there is no difference. Such issues are known and solutions have been considered before, we refer to [Tas93] for further reading.

3.1.4 Universes

To implement the universe hierarchy we first declare another basic Pure type \mathbf{Ord} representing meta-level natural numbers, along with new defined constants

$$\begin{aligned} 0 &:: \text{Ord}, \\ S &:: \text{Ord} \Rightarrow \text{Ord} \end{aligned}$$

and the strict and non-strict order relations

$$\begin{aligned} < &:: [\text{Ord}, \text{Ord}] \Rightarrow \text{prop}, \\ \leq &:: [\text{Ord}, \text{Ord}] \Rightarrow \text{prop}, \end{aligned}$$

satisfying the axioms

$$\begin{aligned} &\bigwedge n. 0 < (S n), \\ &\bigwedge n. 0 \leq n, \\ &\bigwedge n. n < (S n), \\ &\bigwedge n. n \leq (S n), \\ &\bigwedge m n. m < n \implies (S m) < (S n), \\ &\bigwedge m n. m \leq n \implies (S m) \leq (S n). \end{aligned} \tag{Ord}$$

This is all we need to define the terms of universe hierarchy as values of the function term

$$U :: \text{Ord} \Rightarrow \text{Term}.$$

We give the rules governing the universes in the next section.

3.2 Design issues

3.2.1 Analytic versus synthetic formulations

We define the object-level types and terms as terms of the Pure framework. In particular, the object-level typing judgment is implemented as a Pure term, and hence the entirety of $HoTTP$ is contained “separately” inside the Pure metalogic.

This is in contrast to systems like Coq and Agda, there the types of $HoTT$ may be *directly* encoded as types of the meta-framework. This is because those systems are themselves based on extensions of Martin-Löf type theory, whereas the simple type theory that Pure is based on does not have dependent types and hence cannot natively express the $HoTT$ types.

This illustrates the distinction between the *synthetic* and *analytic* formulations [nLa18a; Har12] of a logic \mathcal{L} inside a logical framework \mathcal{F} . In the analytic approach we interpret the types and judgments of \mathcal{L} directly using the semantics of \mathcal{F} , which usually entails using the types of \mathcal{F} as types of \mathcal{L} . This approach, however, requires the logics \mathcal{L} and \mathcal{F} to be similar. In the synthetic approach, one simply uses the syntax of \mathcal{F} to encode the entities of \mathcal{L} , as we do for $HoTT$ inside Pure, essentially because we have no other choice.

3.2.2 Monomorphism versus polymorphism

The theory of *HoTT* is *monomorphic with type annotations*, meaning that terms are decorated with the types of their arguments, and are not polymorphic—each of their arguments can only come from one specific type (these are sometimes called *Church-style* terms). For example, we annotate lambda terms $\lambda(x : A).x$ with the type A of their argument, and dependent eliminators with the type family C over which they are constructed, e.g. the first argument of

$$\text{ind}_{\Sigma}(z.C, x.y.f, p) : C[p/z].$$

Monomorphic theories with annotations have several advantages, first of which is that different terms are distinguished purely syntactically, so for example the lambda terms $\lambda(x : A).x$ and $\lambda(x : B).x$ are different identity functions for $A \neq B$. As another example, the projection functions for dependent pairs $\sum_{x : A} B$ are defined separately for each instance of A and B , so that we do not just write $\text{fst}(p)$ but

$$\text{fst}_{A,B}(p) := \text{ind}_{\Sigma}(z.A, x.y.x, p) : A,$$

one distinct constant for every A and B .

The second advantage of annotating terms t is that the type annotations contain all the information required to reconstruct the derivation of t . Furthermore, such annotations are usually necessary for a theory to have decidable type inference; type theories without annotations are known to suffer from undecidable typability (see for example [Dow93]).

Unfortunately, the monomorphic theory has some distinct *disadvantages* for us when we try to implement it in Pure. Firstly—and this is true of any implementation of a type theory in general—the annotations create extra notational baggage that must be carried around during the course of a proof, and which is yet irrelevant for most of it. For example, in a formal proof using path induction, it suffices to know the form of the type family C over which we are inducting at the moment we apply the equality elimination rule. Once we have done so and instantiated any relevant variables, the proof continues without having to explicitly use the form of C again.

Secondly, recording C as part of the proof term we are constructing creates an additional component of the expression that must be checked for well-formedness further on in the proof, often multiple times. This second issue is a particular problem for us when we work in Isabelle, due to the fact that in the Pure framework we cannot easily implement automatic type checking or inference, which is normally how the overhead is mitigated.

Thus the monomorphic version of the theory is useful if we want to be able to normalize terms or check for well-formedness of proofs, but it is less convenient if our main purpose is to derive proof terms of given types. Hence we choose to implement *HoTT_P* as a *polymorphic* theory without annotations. Here, for example, the object lambda expression $\lambda \mathbf{x}. \mathbf{x}$ has a different semantic interpretation depending on whether it is considered an inhabitant of

$$(U\ 0) \rightarrow (U\ 0)$$

or

$$\mathbf{Nat} \rightarrow \mathbf{Nat},$$

and in fact in our theory we will be able to prove for any type \mathbf{A} that

$$\lambda \mathbf{x}. \mathbf{x} : \mathbf{A} \rightarrow \mathbf{A}.$$

We also drop the type family annotations on the dependent eliminators. Terms in this formulation are sometimes called *Curry-style* terms. Formulating $HoTT_P$ as a polymorphic theory does not affect the form of the type rules other than to drop the type annotations.

3.3 Logical rules

We formalize the rules in Chapter 1 inside Pure according to the patterns indicated in Section 3.1.2. We present inference rules abstractly, but in the actual implementation these are encoded according to the pattern (2.1) discussed earlier.

3.3.1 Structural rules

As our theory does not have contexts, there are no equivalents of `ctx-EMP`, `ctx-EXT`, or `VBL` in $HoTT_P$. As the substitution and weakening rules are admissible (Section A.2.2, [Uni13]), we also do not need to axiomatize them. The $HoTT$ rules \equiv -REFL, \equiv -SYM, and \equiv -TRANS are effectively implemented as the Pure rules \equiv I, \equiv S, and \equiv T. `TYPE-CONG` is derivable using the Pure rule `COMB`, and `DEF-CONG` becomes trivial in $HoTT_P$.

3.3.2 Universes

The universe rules are formulated differently from those of $HoTT$ in order to make proofs in Isabelle easier. They are

$$\frac{i < j}{(U\ i) : (U\ j)} \text{U_hierarchy} \qquad \frac{\mathbf{A} : (U\ i) \quad i \leq j}{\mathbf{A} : (U\ j)} \text{U_cumulative}$$

where Isabelle/Pure's automatic meta-level type-checking ensures that i and j have Pure type `Ord` and \mathbf{A} has Pure type `Term`.

It is not hard to show that if we had instead translated the rules U_i -INTRO and U_i -CUMUL directly as

$$\frac{}{(U\ i) : (U\ (S\ i))} \text{U_intro} \qquad \frac{\mathbf{A} : (U\ i)}{\mathbf{A} : (U\ (S\ i))} \text{U_cumul}$$

then `U_hierarchy` and `U_cumulative` are derivable in $Pure + Ord + U_intro + U_cumul$, and `U_intro`, `U_cumul` are derivable in $Pure + Ord + U_hierarchy + U_cumulative$.

3.3.3 Type rules

The basic type rules are implemented in their respective `.thy` files in [Che18].

Dependent product:

$$\frac{A: \text{U } i \quad \bigwedge x. x: A \implies B \ x: \text{U } i}{(\prod x: A. B \ x): \text{U } i} \text{Prod_form}$$

$$\frac{A: \text{U } i \quad \bigwedge x. x: A \implies b \ x: B \ x}{(\lambda x. b \ x): \prod x: A. B \ x} \text{Prod_intro} \qquad \frac{f: \prod x: A. B \ x \quad a: A}{f \ 'a: B \ a} \text{Prod_elim}$$

$$\frac{\bigwedge x. x: A \implies b \ x: B \ x \quad a: A}{(\lambda x. b \ x) \ 'a \equiv b \ a} \text{Prod_appl} \qquad \frac{f: \prod x: A. B \ x}{\lambda x. (f \ 'x) \equiv f} \text{Prod_uniq}$$

Object-logic lambda terms are distinct from the lambda terms of Pure, and are written with boldface λ instead of the regular λ . We use the backtick `'` to denote object-level application of lambda terms.

Dependent sum:

$$\frac{A: \text{U } i \quad \bigwedge x. x: A \implies B \ x: \text{U } i}{(\sum x: A. B \ x): \text{U } i} \text{Sum_form}$$

$$\frac{\bigwedge x. x: A \implies B \ x: \text{U } i \quad a: A \quad b: B \ a}{\langle a, b \rangle: \sum x: A. B \ x} \text{Sum_intro}$$

$$\frac{\bigwedge p. p: \sum x: A. B \ x \implies C \ p: \text{U } i \quad \bigwedge x \ y. \llbracket x: A; y: B \ x \rrbracket \implies f \ x \ y: C \ \langle x, y \rangle \quad p: \sum x: A. B \ x}{\text{ind}_{\sum} f \ p: C \ p} \text{Sum_elim}$$

$$\frac{\bigwedge p. p: \sum x: A. B \ x \implies C \ p: \text{U } i \quad \bigwedge x \ y. \llbracket x: A; y: B \ x \rrbracket \implies f \ x \ y: C \ \langle x, y \rangle \quad \bigwedge x. x: A \implies B \ x: \text{U } i \quad a: A \quad b: B \ a}{\text{ind}_{\sum} f \ p: C \ p} \text{Sum_comp}$$

Equality:

$$\frac{A: \text{U } i \quad a: A \quad b: A}{a =_A b: \text{U } i} \text{Equal_form} \qquad \frac{a: A}{(\text{refl } a): a =_A a} \text{Equal_intro}$$

$$\begin{array}{c}
\wedge x y p. \llbracket x: A; y: A; p: x =_A y \rrbracket \implies C x y p: U i \\
\wedge x. x: A \implies f x: C x x (\text{refl } x) \\
\frac{x: A \quad y: A \quad p: x =_A y}{\text{ind}_= f p: C x y p} \text{Equal_elim} \\
\wedge x y p. \llbracket x: A; y: A; p: x =_A y \rrbracket \implies C x y p: U i \\
\frac{\wedge x. x: A \implies f x: C x x (\text{refl } x) \quad a: A}{\text{ind}_= f (\text{refl } a) \equiv f a} \text{Equal_comp}
\end{array}$$

Natural numbers:

$$\begin{array}{c}
\frac{}{\mathbb{N}: U 0} \text{Nat_form} \quad \frac{}{0: \mathbb{N}} \text{Nat_intro_0} \quad \frac{n: \mathbb{N}}{\text{succ } n: \mathbb{N}} \text{Nat_intro_succ} \\
\wedge n. n: \mathbb{N} \implies C n: U i \quad \wedge n c. \llbracket n: \mathbb{N}; c: C n \rrbracket \implies f n c: C (\text{succ } n) \\
\frac{a: C 0 \quad n: \mathbb{N}}{\text{ind}_{\mathbb{N}} f a n: C n} \text{Nat_elim} \\
\wedge n. n: \mathbb{N} \implies C n: U i \quad \wedge n c. \llbracket n: \mathbb{N}; c: C n \rrbracket \implies f n c: C (\text{succ } n) \quad a: C 0 \\
\frac{}{\text{ind}_{\mathbb{N}} f a 0 \equiv a} \text{Nat_comp_0} \\
\wedge n. n: \mathbb{N} \implies C n: U i \quad \wedge n c. \llbracket n: \mathbb{N}; c: C n \rrbracket \implies f n c: C (\text{succ } n) \\
\frac{a: C 0 \quad n: \mathbb{N}}{\text{ind}_{\mathbb{N}} f a (\text{succ } n) \equiv f n (\text{ind}_{\mathbb{N}} f a n)} \text{Nat_comp_succ}
\end{array}$$

Coproduct:

$$\begin{array}{c}
\frac{A: U i \quad B: U i}{A + B: U i} \text{Coproduct_form} \\
\frac{a: A \quad B: U i}{\text{inl } a: A + B} \text{Coproduct_intro_inl} \quad \frac{A: U i \quad b: B}{\text{inr } b: A + B} \text{Coproduct_intro_inr} \\
\wedge u. u: A + B \implies C u: U i \\
\wedge x. x: A \implies c x: C (\text{inl } x) \quad \wedge y. y: B \implies d y: C (\text{inr } y) \\
\frac{u: A + B}{\text{ind}_+ c d u: C u} \text{Coproduct_elim}
\end{array}$$

$$\begin{array}{c}
\wedge u. u: A + B \implies C u: U i \\
\wedge x. x: A \implies c x: C (\text{inl } x) \quad \wedge y. y: B \implies d y: C (\text{inr } y) \\
\hline
a: A \\
\text{ind}_+ c d (\text{inl } a) \equiv c a \quad \text{Coproduct_comp_inl}
\end{array}$$

$$\begin{array}{c}
\wedge u. u: A + B \implies C u: U i \\
\wedge x. x: A \implies c x: C (\text{inl } x) \quad \wedge y. y: B \implies d y: C (\text{inr } y) \\
\hline
b: B \\
\text{ind}_+ c d (\text{inr } b) \equiv d b \quad \text{Coproduct_comp_inl}
\end{array}$$

Empty and unit types:

$$\begin{array}{c}
\frac{}{0: U 0} \text{Empty_form} \quad \frac{\wedge x. x: 0 \implies C x: U i \quad z: 0}{\text{ind}_0 z: C z} \text{Empty_elim} \\
\\
\frac{}{1: U 0} \text{Unit_form} \quad \frac{}{*: 1} \text{Unit_intro} \\
\\
\frac{\wedge x. x: 1 \implies C x: U i \quad c: C * \quad a: 1}{\text{ind}_1 c a: C a} \text{Unit_elim} \\
\\
\frac{\wedge x. x: 1 \implies C x: U i \quad c: C *}{\text{ind}_1 c * \equiv c} \text{Unit_comp}
\end{array}$$

3.3.4 Additional rules

We axiomatize the additional congruence rule

$$\frac{A: U i \quad \wedge x. x: A \implies b x \equiv c x}{\lambda x. b x \equiv \lambda x. c x} \text{Prod_eq}$$

governing definitional equality of object lambda terms.

3.3.5 Univalence

The formulation of the inference rules of the previous sections is sufficient to encode the definitions and proofs needed to define univalence as presented in Section 1.5. These are implemented in the theory file `Univalence.thy` in [Che18], and its dependencies `EqualProps.thy`, `ProdProps.thy`, and `Sum.thy`.

Univalence:

$$\frac{}{\text{univalence: isequiv idtoeqv}} \text{UA}$$

where `isequiv` and `idtoeqv` are defined constants encoding `isequiv` and `idtoeqv` in Pure.

3.4 Metatheory

We discuss the metatheory of $HoTT_P$ as compared with that of $HoTT$.

As discussed earlier, formulating $HoTT_P$ as a polymorphic theory results in differences in the metatheory, in particular regard to normalization, and decidability of type checking and inference in the underlying implementation of Martin-Löf type theory

$$ITT_P = HoTT_P - \mathbf{UA}.$$

It is at the moment unclear to the author if ITT_P enjoys the same properties as ITT , and what exactly is lost when removing annotations and introducing polymorphism.

There are, however, properties of ITT that we have taken care to preserve. For example, it is an important consequence of the normalization theorem in ITT that if

$$\vdash a : A,$$

then

$$\vdash A : U_i$$

for some i (Theorem 3.12, [Mar75]), i.e. well-typedness of terms implies well-formedness of their types. The type rules of $HoTT_P$ have been deliberately formulated in order to maintain this property, which is the reason some of the rules in Section 3.3 have additional premises not present in the corresponding $HoTT$ rules. That is to say, by a structural induction on the depth of proof terms, we can show the following:

Theorem 3.4.1. *If*

$$\Gamma \vdash_{ITT_P} a : A$$

then

$$\Gamma \vdash_{ITT_P} A : U_i$$

for some numeral i .

Using a similar technique, one should be able to show that if the judgment

$$\prod x : A. B x : U_i$$

is derivable in ITT_P , then so are the judgments

$$A : U_i$$

and

$$\bigwedge x. x : A \implies B x : U_i,$$

essentially because there is no other way the type $\prod x : A. B x$ can be formed.

Thus we conjecture

Conjecture 3.4.2. *The inference rules*

$$\frac{\prod x: A. B \quad x: U \quad i}{A: U \quad i} \text{Prod_wellform1} \qquad \frac{\prod x: A. B \quad x: U \quad i}{\wedge x. x: A \implies B \quad x: U \quad i} \text{Prod_wellform2}$$

are correct in ITT_P , as well as similar well-formedness rules for the dependent sum, equality, and coproduct types.

These type well-formedness rules are axiomatized in their respective theory files, and form a component of one of the automated proof methods implemented as part of *Isabelle/HoTT*.

Chapter 4

Isabelle/HoTT: an invitation

We now come to the end of this thesis, which is really only just the beginning, as we now point the reader in the direction of the work that we have actually been doing all along—the implementation in code of the theory *HoTT* as a working object logic in the Isabelle system.

There is much in the code that we have not been able to describe here: a collection of semi-automated proof methods that can be used to assist the derivation process, numerous examples illustrating its use to prove nontrivial statements and search for proof terms, formalizations of proofs from [Uni13], along with descriptive text explaining the details of the implementation in theory and practice.

As hinted at in the closing paragraphs of the previous chapter, there is still much work left to be done, both theoretically and practically. It is however hoped that the work in this thesis will ultimately be of use in laying the foundations for bringing homotopy type theory to the Isabelle prover.

Code, issues, and documentation updated at <https://github.com/jaycech3n/Isabelle-HoTT>.

Bibliography

- [18a] *src/Pure/drule.ML*, ML file in the Isabelle2018 distribution source directory, Aug. 2018 (cit. on p. 21).
- [18b] *src/Pure/thm.ML*, ML file in the Isabelle2018 distribution source directory, Aug. 2018 (cit. on p. 21).
- [And02] P. B. Andrews, *An Introduction to Mathematical Logic and Type Theory: To Truth Through Proof*, 2nd ed. Springer, 2002 (cit. on p. 17).
- [CCHM16] C. Cohen, T. Coquand, S. Huber, and A. Mörtberg. (Nov. 7, 2016). Cubical type theory: A constructive interpretation of the univalence axiom. arXiv: **1611.02108** [**cs.LG**] (cit. on pp. 1, 11).
- [Che18] J. Chen, *jaycech3n/Isabelle-HoTT: Initial release*, Isabelle proof assistant object logic library, Sep. 2018. DOI: **10.5281/zenodo.1413984**. [Online]. Available: <https://doi.org/10.5281/zenodo.1413984> (cit. on pp. 31, 37, 39).
- [Chu40] A. Church, “A formulation of the simple theory of types,” *The Journal of Symbolic Logic*, vol. 5, no. 2, pp. 56–68, Jun. 1940. DOI: **10.2307/2266170**. [Online]. Available: <https://www.jstor.org/stable/2266170> (cit. on pp. v, 17).
- [Coq96] T. Coquand, “An algorithm for type-checking dependent types,” *Science of Computer Programming*, vol. 26, no. 1, pp. 167–177, 1996, ISSN: 0167-6423. DOI: **10.1016/0167-6423(95)00021-6** (cit. on p. 10).
- [Dow93] G. Dowek, “The undecidability of typability in the lambda-pi-calculus,” in *Typed Lambda Calculi and Applications*, M. Bezem and J. F. Groote, Eds., Springer Berlin Heidelberg, 1993, pp. 139–145, ISBN: 978-3-540-47586-6 (cit. on p. 35).
- [Gra05] C. A. Grabmayer, “Relating proof systems for recursive types,” PhD thesis, Vrije Universiteit Amsterdam, 2005 (cit. on p. 13).
- [Har12] R. Harper, “Notes on logical frameworks,” Lecture notes, Institute for Advanced Study, Nov. 29, 2012, [Online]. Available: <http://www.cs.cmu.edu/~rwh/papers/lfias/lf.pdf> (cit. on p. 34).
- [HS08] J. R. Hindley and J. P. Seldin, *Lambda-Calculus and Combinators, an Introduction*. Cambridge University Press, 2008 (cit. on pp. 13, 15, 20).

- [Hub17] S. Huber. (Oct. 30, 2017). Canonicity for cubical type theory. arXiv: 1607.04156 [cs.LO] (cit. on p. 11).
- [Hue75] G. P. Huet, “A unification algorithm for typed λ -calculus,” *Theoretical Computer Science*, vol. 1, no. 1, pp. 27–57, 1975, ISSN: 0304-3975. DOI: 10.1016/0304-3975(75)90011-0 (cit. on p. 22).
- [KL12] K. Kapulkin and P. LeFanu Lumsdaine. (Nov. 2012). The simplicial model of Univalent Foundations (after Voevodsky). arXiv: 1211.2851 [math.LO] (cit. on p. 11).
- [Luo12] Z. Luo, “Notes on universes in type theory,” Talk notes, Institute for Advanced Study, Oct. 2012, [Online]. Available: <http://www.cs.rhul.ac.uk/home/zhaohui/universes.pdf> (cit. on pp. 3, 10).
- [Mag95] L. Magnusson, “The implementation of ALF - a proof editor based on Martin-Löf’s monomorphic type theory with explicit substitution,” PhD thesis, Chalmers University of Technology and University of Göteborg, Jan. 2, 1995 (cit. on p. 10).
- [Mar75] P. Martin-Löf, “An intuitionistic theory of types: Predicative part,” in *Logic Colloquium ’73*, ser. Studies in Logic and the Foundations of Mathematics, H. Rose and J. Shepherdson, Eds., vol. 80, Elsevier, 1975, pp. 73–118. DOI: 10.1016/S0049-237X(08)71945-1 (cit. on pp. v, 1, 8, 9, 40).
- [Mar82] —, “Constructive mathematics and computer programming,” in *Logic, Methodology and Philosophy of Science VI*, ser. Studies in Logic and the Foundations of Mathematics, vol. 104, Elsevier, 1982, pp. 153–175. DOI: 10.1016/S0049-237X(09)70189-2 (cit. on pp. v, 1).
- [MS84] P. Martin-Löf and G. Sambin, *Intuitionistic type theory*, ser. Studies in proof theory. Bibliopolis, 1984, vol. 1 (cit. on pp. v, 1).
- [nLa18a] nLab authors, *Logical framework*, <https://ncatlab.org/nlab/show/logical+framework>, Revision 24, Sep. 2018 (cit. on p. 34).
- [nLa18b] —, *Pure type system*, <https://ncatlab.org/nlab/show/pure+type+system>, Revision 20, Sep. 2018 (cit. on p. 31).
- [Pau00] L. C. Paulson. (Oct. 2000). The foundation of a generic theorem prover. arXiv: cs/9301105 (cit. on pp. 17, 26, 29).
- [Pau18] —, *Old introduction to Isabelle*, Aug. 15, 2018. [Online]. Available: <https://isabelle.in.tum.de/website-Isabelle2018/dist/Isabelle2018/doc/intro.pdf> (cit. on pp. 13, 22, 29).
- [Tas93] A. Tasistro, “Formulation of Martin-Löf’s theory of types with explicit substitutions,” Master’s thesis, Chalmers University of Technology and University of Göteborg, May 14, 1993 (cit. on p. 33).

- [Uni13] Univalent Foundations Program, *Homotopy Type Theory: Univalent Foundations of Mathematics*. Institute for Advanced Study, 2013. [Online]. Available: <https://homotopytypetheory.org/book> (cit. on pp. v, vi, 1, 4–6, 11, 36, 43).
- [Uni18] University of Cambridge, Technische Universität München, and contributors, *Isabelle2018*, Aug. 2018. [Online]. Available: <https://isabelle.in.tum.de/website-Isabelle2018/> (cit. on pp. v, 13).
- [Wen18a] M. Wenzel, *The Isabelle/Isar implementation*, Aug. 15, 2018. [Online]. Available: <https://isabelle.in.tum.de/website-Isabelle2018/dist/Isabelle2018/doc/implementation.pdf> (cit. on p. 13).
- [Wen18b] —, *The Isabelle/Isar reference manual*, Aug. 15, 2018. [Online]. Available: <https://isabelle.in.tum.de/website-Isabelle2018/dist/Isabelle2018/doc/isar-ref.pdf> (cit. on p. 13).