

MATH 721: HOMOTOPY TYPE THEORY

EMILY RIEHL

CONTENTS

Part 1. Martin-Löf's Dependent Type Theory	1
August 30: Dependent Type Theory	1
September 1: Dependent function types & the natural numbers	3
September 8: The formal proof assistant <code>agda</code>	6
September 13: Inductive types	6
September 15: Identity types	9
September 20: More identity types	12
September 22: Universes	15
September 27: Modular arithmetic	18
September 29: Elementary number theory	18
Part 2. The Univalent Foundations of Mathematics	18
October 4: Equivalences	18
October 6: Contractibility	18
October 11: The fundamental theorem of identity types	18
October 13: Propositions, sets, and general truncation levels	18
October 18: Function extensionality	18
October 20: Propositional truncation	18
October 25: The image of a map	18
October 27: Finite types	18
November 1: The univalence axiom	18
November 3: Set quotients	18
November 8: Groups	18
November 10: Algebra	18
November 15: The real numbers	18
Part 3. Synthetic Homotopy Theory	18
November 17: The circle	18
November 29: The universal cover of the circle	18
December 1: Homotopy groups of types	18
December 6: Classifying types of groups	18

Part 1. Martin-Löf's Dependent Type Theory

AUGUST 30: DEPENDENT TYPE THEORY

Martin-Löf's dependent type theory is a formal language for writing mathematics: both constructions of mathematical objects and proofs of mathematical propositions. As we shall discover, these two things are treated in parallel (in contrast to classical Set theory plus first-order logic, where the latter supplies the proof calculus and the former gives the language which you use to state things to prove).

Judgments and contexts. I find it helpful to imagine I'm teaching a computer to do mathematics. It's also helpful to forget that you know other ways of doing mathematics.¹

defn. There are four kinds of **judgments** in dependent type theory, which you can think of as the “grammatically correct” expressions:

- (i) $\Gamma \vdash A \text{ type}$, meaning that A is a well-formed type in **context** Γ (more about this soon).
- (ii) $\Gamma \vdash a : A$, meaning that a is a well-formed term of type A in context Γ .
- (iii) $\Gamma \vdash A \doteq B \text{ type}$, meaning that A and B are **judgmentally** or **definitionally** equal types in context Γ .
- (iv) $\Gamma \vdash a \doteq b : A$, meaning that a and b are judgmentally equal terms of type A in context Γ .

These might be collectively abbreviated by $\Gamma \vdash \mathcal{J}$.

The statement of a mathematical theorem, often begins with an expression like “Let n and m be positive integers, with $n < m$, and let $\vec{v}_1, \dots, \vec{v}_m$ be vectors in \mathbb{R}^n . Then ...” This statement of the hypotheses defines a **context**, a finite list of types and hypothetical terms (called **variables**²) satisfying an inductive condition that that each type can be derived in the context of the previous types and terms using the inference rules of type theory.

defn. A **context** is a finite list of variable declarations:

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

satisfying the condition that for each $1 \leq k \leq n$ we can derive the judgment

$$x_1 : A_1, \dots, x_{k-1} : A_{k-1}(x_1, \dots, x_{k-2}) \vdash A_k(x_1, \dots, x_{k-1}) \text{ type}$$

using the inference rules of type theory.

We'll introduce the inference rules shortly but the idea is that it needs to be possible to form the type $A_k(x_1, \dots, x_{k-1})$ given terms x_1, \dots, x_{k-1} of the previously-formed types.

ex. For example, there is a unique context of length zero: the empty context.

ex. $n : \mathbb{N}, m : \mathbb{N}, p : n < m, \vec{v} : (\mathbb{R}^n)^m$ is a context. Here $n : \mathbb{N}, m : \mathbb{N} \vdash n < m$ is a dependent type that corresponds to the relation $\{n < m \mid n, m \in \mathbb{N}\} \subset \mathbb{N} \times \mathbb{N}$ and the variable p is a witness that $n < m$ is true (more about this later).

Type families. Absolutely everything in dependent type theory is context dependent so we always assume we're working in a background context Γ . Let's focus on the primary two judgment forms.

defn. Given a type A in context Γ a **family** of types over A in context Γ is a type $B(x)$ in context $\Gamma, x : A$, as represented by the judgment:

$$\Gamma, x : A \vdash B(x) \text{ type}$$

We also say that $B(x)$ is a type **indexed** by $x : A$, in context Γ .

ex. \mathbb{R}^n is a type indexed by $n \in \mathbb{N}$.

defn. Consider a type family B over A in context Γ . A **section** of the family B over A in context Γ is a term of type $B(x)$ in context $\Gamma, x : A$, as represented by the judgment:

$$\Gamma, x : A \vdash b(x) : B(x)$$

We say that b is a **section** of the family B over A in context Γ or that $b(x)$ is a term of type $B(x)$ indexed by $x : A$ in context Γ .

ex. $\vec{0}_n : \mathbb{R}^n$ is a term dependent on $n \in \mathbb{N}$.

Exercise. If you've heard the word “section” before you should think about what it is being used here.

¹Indeed, there are very deep theorems that describe how to interpret dependent type theory into classical set-based mathematics. You're welcome to investigate these for your final project but they are beyond the scope of this course.

²We're not going to say anything about proper syntax for variables and instead rely on instinct to recognize proper and improper usage.

Inference rules. There are five types of inference rules that collectively describe the structural rules of dependent type theory. They are

- (i) Rules postulating that judgmental equality is an equivalence relation:

$$\frac{\Gamma \vdash A \text{ type}}{\Gamma \vdash A \doteq A \text{ type}} \quad \frac{\Gamma \vdash A \doteq B \text{ type}}{\Gamma \vdash B \doteq A \text{ type}} \quad \frac{\Gamma \vdash A \doteq B \text{ type} \quad \Gamma \vdash B \doteq C \text{ type}}{\Gamma \vdash A \doteq C \text{ type}}$$

and similarly for judgmental equality between terms.

- (ii) Variable conversion rules for judgmental equality between types:

$$\frac{\Gamma \vdash A \doteq A' \text{ type} \quad \Gamma, x : A, \Delta \vdash \mathcal{J}}{\Gamma, x : A', \Delta \vdash \mathcal{J}}$$

- (iii) Substitution rules:

$$\frac{\Gamma \vdash a : A \quad \Gamma, x : A, \Delta \vdash \mathcal{J}}{\Gamma, \Delta[a/x] \vdash \mathcal{J}[a/x]}$$

If Δ is the context $y_1 : B_1(x), \dots, y_n : B_n(x, y_1, \dots, y_{n-1})$ then $\Delta[a/x]$ is the context $y_1 : B(a), \dots, y_n : B_n(a, y_1, \dots, y_{n-1})$. A similar substitution is performed in the judgment $\mathcal{J}[a/x]$. Further rules indicate that substitution by judgmentally equal terms gives judgmentally equal results.

- (iv) Weakening rules:

$$\frac{\Gamma \vdash A \text{ type} \quad \Gamma, \Delta \vdash \mathcal{J}}{\Gamma, x : A, \Delta \vdash \mathcal{J}}$$

Eg if A and B are types in context Γ , then B is also a type in context $\Gamma, x : A$.

- (v) The generic term:

$$\frac{\Gamma \vdash A \text{ type}}{\Gamma, x : A \vdash x : A}$$

This will be used to define the identity function of any type.

Derivations. A derivation in type theory is a finite rooted tree where each node is a valid rule of inference. The root is the conclusion.

ex. The interchange rule is derived as follows

$$\frac{\frac{\frac{\Gamma \vdash B \text{ type}}{\Gamma, y : B \vdash y : B}}{\Gamma, y : B, x : A \vdash y : B} \quad \frac{\Gamma \vdash B \text{ type} \quad \frac{\Gamma, x : A, y : B, \Delta \vdash \mathcal{J}}{\Gamma, x : A, z : B, \Delta[z/y] \vdash \mathcal{J}[z/y]}}{\Gamma, y : B, x : A, z : B, \Delta[z/y] \vdash \mathcal{J}[z/y]}}{\Gamma, y : B, x : A, \Delta \vdash \mathcal{J}}$$

SEPTEMBER 1: DEPENDENT FUNCTION TYPES & THE NATURAL NUMBERS

The rules for dependent function types. Consider a section b of a family B over A in context Γ , as encoded by a judgment:

$$\Gamma, x : A \vdash b(x) : B(x).$$

We think of the section b as a function that takes as input $x : A$ and produces a term $b(x) : B(x)$. Since the type of the output is allowed to depend on the term being input, this isn't quite an ordinary function but a **dependent function**. The type of all dependent functions is the **dependent function type**

$$\prod_{x:A} B(x)$$

What is a thing in mathematics? Structuralism says the ontology of a thing is determined by its behavior. In dependent type theory, we define dependent function types by stating their rules, which have the following forms:

- (i) **formation rules** tell us how a type may be formed
- (ii) **introduction rules** tell us how to introduce new terms of the type
- (iii) **elimination rules** tell us how the terms of a type may be used
- (iv) **computation rules** tell us how the introduction and elimination rules interact

There are also **congruence rules** that tell us that all constructions respect judgmental equality. See your book for more details.

defn (dependent function types). The Π -**formation rule** has the form:

$$\frac{\Gamma, x : A \vdash B(x) \text{ type}}{\Gamma \vdash \prod_{x:A} B(x) \text{ type}}$$

The Π -**introduction rule** has the form:

$$\frac{\Gamma, x : A \vdash b(x) : B(x)}{\Gamma \vdash \lambda x. b(x) : \prod_{x:A} B(x)}$$

The λ -**abstraction** $\lambda x. b(x)$ can be thought of as notation for $x \mapsto b(x)$.

The Π -**elimination rule** has the form of an evaluation rule:

$$\frac{\Gamma \vdash f : \prod_{x:A} B(x)}{\Gamma, x : A \vdash f(x) : B(x)}$$

Finally, there are two computation rules: the β -**rule**

$$\frac{\Gamma, x : A \vdash b(x) : B(x)}{\Gamma, x : A \vdash (\lambda y. b(y))(x) \doteq b(x) : B(x)}$$

and the η -**rule**, which says that all elements of a Π -type are dependent functions:

$$\frac{\Gamma \vdash f : \prod_{x:A} B(x)}{\Gamma \vdash \lambda x. f(x) \doteq f : \prod_{x:A} B(x)}$$

Ordinary function types.

defn (function types). The formation rule is derived from the formation rule for Π -types together with weakening:

$$\frac{\frac{\Gamma \vdash A \text{ type} \quad \Gamma \vdash B \text{ type}}{\Gamma, x : A \vdash B \text{ type}}}{\Gamma \vdash \prod_{x:A} B \text{ type}}$$

We adopt the notation

$$A \rightarrow B := \prod_{x:A} B$$

for the dependent function type in the case where the type family B is constant over $x : A$.

The introduction, evaluation, and computation rules are instances of term conversion: eg

$$\frac{\Gamma \vdash B \text{ type} \quad \Gamma, x : A \vdash b(x) : B}{\Gamma \vdash \lambda x. b(x) : A \rightarrow B} \quad \frac{\Gamma \vdash f : A \rightarrow B}{\Gamma, x : A \vdash f(x) : B}$$

plus the two computation rules:

$$\frac{\Gamma \vdash B \text{ type} \quad \Gamma, x : A \vdash b(x) : B}{\Gamma, x : A \vdash (\lambda y. b(y))(x) \doteq b(x) : B} \quad \frac{\Gamma \vdash f : A \rightarrow B}{\Gamma \vdash \lambda x. f(x) \doteq f : A \rightarrow B}$$

defn. Identity functions are defined as follows:

$$\frac{\frac{\Gamma \vdash A \text{ type}}{\Gamma, x : A \vdash x : A}}{\Gamma \vdash \lambda x. x : A \rightarrow A}$$

which is traditionally denoted by $\text{id}_A := \lambda x. x$.

The idea of composition is that given a function $f: A \rightarrow B$ and $g: B \rightarrow C$ you should get a function $g \circ f: A \rightarrow C$. Using infix notation you might denote this function by $_ \circ _$.

Q. $_ \circ _$ is itself a function, so it's a term of some type. What type?³

defn. Composition has the form:

$$\frac{\Gamma \vdash A \text{ type} \quad \Gamma \vdash B \text{ type} \quad \Gamma \vdash C \text{ type}}{\Gamma \vdash _ \circ _ : (B \rightarrow C) \rightarrow ((A \rightarrow B) \rightarrow (A \rightarrow C))}$$

It is defined by

$$_ \circ _ := \lambda g. \lambda f. \lambda x. g(f(x))$$

which can be understood as the term constructed by three applications of the Π -introduction rule followed by two applications of the Π -elimination rule.

Composition is associative essentially because both $(h \circ g) \circ f$ and $h \circ (g \circ f)$ are defined by $\lambda x. h(g(f(x)))$. We'll think about this more formally when we come back to identity types.

Similarly, you can compute that for all $f: A \rightarrow B$, $\text{id}_B \circ f \doteq f: A \rightarrow B$ and $f \circ \text{id}_A \doteq f: A \rightarrow B$.

The type of natural numbers. The type \mathbb{N} of natural numbers is the archetypical example of an **inductive type** about more which soon. It is given by rules which say that it has a term $0_{\mathbb{N}}: \mathbb{N}$, it has a successor function $\text{succ}_{\mathbb{N}}: \mathbb{N} \rightarrow \mathbb{N}$ and it satisfies the induction principle.

The \mathbb{N} -formation rule is

$$\frac{}{\vdash \mathbb{N} \text{ type}}$$

In other words, \mathbb{N} is a type in the empty context.

There are two \mathbb{N} -introduction rules:

$$\frac{}{\vdash 0_{\mathbb{N}}: \mathbb{N}} \quad \frac{}{\vdash \text{succ}_{\mathbb{N}}: \mathbb{N} \rightarrow \mathbb{N}}$$

Digression (traditional induction). In traditional first-order logic, the principle of \mathbb{N} -induction is stated in terms of a **predicate** P over \mathbb{N} . One way to think about P is as a function $P: \mathbb{N} \rightarrow \{\top, \perp\}$. That is, for each $n \in \mathbb{N}$, $P(n)$ is either true or false. We could also think of P as an indexed family of sets $(P(n))_{n \in \mathbb{N}}$ where for each n either $P(n) = \emptyset$ (corresponding to $P(n)$ being false) or $P(n) = *$ (corresponding to $P(n)$ being true).

The induction principle then says

$$\forall P: \{0, 1\}^{\mathbb{N}}, (P(0) \wedge (\forall n, P(n) \rightarrow P(n+1)) \rightarrow \forall n, P(n)).$$

In dependent type theory it is most natural to let P be an arbitrary type family over \mathbb{N} . This is a stronger assumption, as we'll see.

Q. What then corresponds to a proof that $\forall n, P(n)$?

The induction principle is encoded by the following rule:

$$\frac{\Gamma, n: \mathbb{N} \vdash P(n) \text{ type} \quad \Gamma \vdash p_0: P(0_{\mathbb{N}}) \quad \Gamma \vdash p_S: \prod_{n: \mathbb{N}} (P(n) \rightarrow P(\text{succ}_{\mathbb{N}}(n)))}{\Gamma \vdash \text{ind}_{\mathbb{N}}(p_0, p_S): \prod_{n: \mathbb{N}} P(n)}$$

Remark. There are other forms this rule might take that are interderivable with this one.

The computation rules say that the function $\text{ind}_{\mathbb{N}}(p_0, p_S): \prod_{n: \mathbb{N}} P(n)$ behaves like it should on $0_{\mathbb{N}}$ and successors:

$$\frac{\Gamma, n: \mathbb{N} \vdash P(n) \text{ type} \quad \Gamma \vdash p_0: P(0_{\mathbb{N}}) \quad \Gamma \vdash p_S: \prod_{n: \mathbb{N}} (P(n) \rightarrow P(\text{succ}_{\mathbb{N}}(n)))}{\Gamma \vdash \text{ind}_{\mathbb{N}}(p_0, p_S)(0_{\mathbb{N}}) \doteq p_0: P(0_{\mathbb{N}})}$$

and under the same premises

$$\Gamma, n: \mathbb{N} \vdash \text{ind}_{\mathbb{N}}(p_0, p_S)(\text{succ}_{\mathbb{N}}(n)) \doteq p_S(n, \text{ind}_{\mathbb{N}}(p_0, p_S, n)): P(\text{succ}_{\mathbb{N}}(n)).$$

³Really the type should involve three universe variables but let's save this for next week.

These computation rules don't matter so much if the type family $n : \mathbb{N} \vdash P(n)$ is really a predicate — $P(n)$ is either true or false and that's the end of the story — but they do matter if $P(n)$ is more like an indexed family of sets. In the latter case, $\text{ind}_{\mathbb{N}}(p_0, p_S)$ is the recursive function defined from p_0 and p_S and these are the computation rules for that recursion.

Remark. Recall Peano's axioms for the natural numbers:

- (i) $0_{\mathbb{N}} \in \mathbb{N}$
- (ii) $\text{succ}_{\mathbb{N}} : \mathbb{N} \rightarrow \mathbb{N}$
- (iii) $\forall n, \text{succ}_{\mathbb{N}}(n) \neq 0_{\mathbb{N}}$
- (iv) $\forall n, m, \text{succ}_{\mathbb{N}}(n) = \text{succ}_{\mathbb{N}}(m) \rightarrow n = m$
- (v) induction

We'll be able to *prove* this missing two axioms from the induction principle we've assumed once we have identity types and universes. We'll come back to this in a few weeks.

Addition on the natural numbers.

Remark. When addition is defined by recursion on the second variable, from the computation rules associated to function types and the natural numbers type you can derive judgmental equalities

$$m + 0 \doteq m \quad \text{and} \quad m + \text{succ}_{\mathbb{N}}(n) = \text{succ}_{\mathbb{N}}(m + n).$$

But you can't derive the symmetric judgmental equalities.

We *will* be able to prove such equalities using the identity types, to be introduced shortly.

Pattern matching. To define a dependent function $f : \prod_{n:\mathbb{N}} P(n)$ by induction on n it suffices, by the elimination rule for the natural numbers type, to provide two terms:

$$p_0 : P(0_{\mathbb{N}}) \quad p_S : \prod_{n:\mathbb{N}} P(n) \rightarrow P(\text{succ}_{\mathbb{N}}(n)).$$

Thus the definition of f may be presented by writing

$$f(0_{\mathbb{N}}) := p_0 \quad f(\text{succ}_{\mathbb{N}}(n)) := p_S(n, f(n)).$$

This defines the function f by **pattern matching** on the variable n . When a function is defined in this form, the judgmental equalities accompanying the definition are immediately displayed.

SEPTEMBER 8: THE FORMAL PROOF ASSISTANT `agda`

See <https://github.com/emilyriehl/721/blob/master/introduction.agda>

SEPTEMBER 13: INDUCTIVE TYPES

The rules for the natural numbers type \mathbb{N} tell us:

- (i) how to form terms in \mathbb{N} , and
- (ii) how to define dependent functions in $\prod_{n:\mathbb{N}} P(n)$ for any type family $n : \mathbb{N} \vdash P(n)$ type ,

while providing two computation rules for those dependent functions.

Many types can be specified by stating how to form their terms and how to define dependent functions out of them. Such types are called **inductive types**.

The idea of inductive types. Recall a type is specified by its formation rules, its introduction rules, its elimination rules, and its computation rules. For inductive types, the introduction rules specify the **constructors** of the inductive type, while the elimination rule provides the **induction principle**. The computation rules provide definitional equalities for the induction principle.

In more detail:

- (i) The constructors tell us what structure the identity type is given with.
- (ii) The induction principle defines sections of any type family over the inductive type by specifying the behavior at the constructors.
- (iii) The computation rules assert that the inductively defined section agrees on the constructors with the data used to define it. So there is one computation rule for each constructor.

The unit type. The formal definition of the **unit** type is as follows:

$$\vdash \mathbb{1} \text{ type} \quad \vdash \star : \mathbb{1} \quad \frac{x : \mathbb{1} \vdash P(x) \text{ type} \quad p : P(\star)}{x : \mathbb{1} \vdash \text{ind}_{\mathbb{1}}(p, x) : P(x)} \quad \frac{x : \mathbb{1} \vdash P(x) \text{ type} \quad p : P(\star)}{x : \mathbb{1} \vdash \text{ind}_1(p, \star) \doteq p : P(\star)}$$

As an inductive type, the definition is packaged as follows:

defn. The **unit** type is a type $\mathbb{1}$ equipped with a term $\star : \mathbb{1}$ satisfying the inductive principle that for any family $x : \mathbb{1} \vdash P(x)$ there is a function

$$\text{ind}_{\mathbb{1}} : P(\star) \rightarrow \prod_{x:\mathbb{1}} P(x)$$

with the computation rule $\text{ind}_1(p, \star) \doteq p$.

In agda, this definition has the form:

```
data unit : UU lzero where
  star : unit
```

Q. What does the induction rule look like for a constant type family A that does not depend on $\mathbb{1}$?

The empty type.

defn. The empty type is a type \emptyset satisfying the induction principle that for any family of types $x : \emptyset \vdash P(x)$ there is a term

$$\text{ind}_{\emptyset} : \prod_{x:\emptyset} P(x).$$

That is the empty type is the inductive type with no constructors. Thus there are no computation rules. In agda, this definition has the form:

```
data empty : UU lzero where
```

Remark. As a special case of the elimination rule for the empty type we have

$$\frac{\vdash A \text{ type}}{\text{ex-falso} := \text{ind}_{\emptyset} : \emptyset \rightarrow A}$$

By the elimination rule for function types it follows that if we had a term $x : \emptyset$ then we could get a term in any type. The name comes from latin *ex falso quodlibet*: “from falsehood, anything.”

We’ve already seen a few glimpses of logic in type theory, something we’ll discuss more formally soon. The basic idea is that we can interpret the formation of a type as akin to the process of formulating a mathematical statement that could be a sentence (if it’s a type in the empty context) or a predicate (if it’s a dependent type). The act of constructing a term in that type is then analogous to proving the proposition so-encoded. These ideas motivate the logically-inflected terms in what follows.

For instance, we can use the empty type to define a negation operation on types:

defn. For any type A , we define its **negation** by $\neg A := A \rightarrow \emptyset$ and say the type A is **empty** if there is a term in this type.

Remark. To construct a term of type $\neg A$, use the introduction rule for function types and assume given a term $a : A$. The task then is to derive a term of \emptyset . In other words, we prove $\neg A$ by assuming A and deriving a contradiction. This proof technique is called **proof of negation**.

This should be contrasted with **proof by contradiction**, which aims to prove a proposition P by assuming $\neg P$ and deriving a contradiction. This uses the logical step “ $\neg\neg P$ implies P .” In type theory, however, $\neg\neg A$ is the type of functions

$$\neg\neg A := (A \rightarrow \emptyset) \rightarrow \emptyset$$

and it is not possible in general to use a term in this type to construct a term of type A .

The law of contraposition does work, at least in one direction.

Proposition. For any types P and Q there is a function

$$(P \rightarrow Q) \rightarrow (\neg Q \rightarrow \neg P).$$

Proof. By λ -abstraction assume given $f : P \rightarrow Q$ and $\tilde{q} : Q \rightarrow \emptyset$. We seek a term in $P \rightarrow \emptyset$, which we obtain simply by composing: $\tilde{q} \circ f : P \rightarrow \emptyset$. Thus

$$\lambda f. \lambda \tilde{q}. \lambda p. \tilde{q}(f(p)) : (P \rightarrow Q) \rightarrow (\neg Q \rightarrow \neg P).$$

□

Coproducts. Inductive types can be defined outside the empty context. For instance, the formation and introduction rules for the coproduct type have the form:

$$\frac{\Gamma \vdash A \text{ type} \quad \Gamma \vdash B \text{ type}}{\Gamma \vdash A + B \text{ type}}$$

$$\frac{\Gamma \vdash A \text{ type} \quad \Gamma \vdash B \text{ type} \quad \Gamma \vdash a : A}{\Gamma \vdash \text{inl}a : A + B} \quad \frac{\Gamma \vdash A \text{ type} \quad \Gamma \vdash B \text{ type} \quad \Gamma \vdash b : B}{\Gamma \vdash \text{inr}b : A + B}$$

defn. Given types A and B the **coproduct type** is the type equipped with

$$\text{inl} : A \rightarrow A + B \quad \text{inr} : B \rightarrow A + B$$

satisfying the induction principle that says that for any family of types $x : A + B \vdash P(x) \text{ type}$ there is a term

$$\text{ind}_+ : \left(\prod_{x:A} P(\text{inl}(x)) \right) \rightarrow \left(\prod_{y:B} P(\text{inr}(y)) \right) \rightarrow \prod_{z:A+B} P(z)$$

satisfying the computation rules

$$\text{ind}_+(f, g, \text{inl}(x)) \doteq f(x) \quad \text{ind}_+(f, g, \text{inr}(y)) \doteq g(y).$$

Not as a special case we have

$$\text{ind}_+ : (A \rightarrow X) \rightarrow (B \rightarrow X) \rightarrow (A + B \rightarrow X)$$

which is similar to the elimination rule for disjunction in first order logic: if you've proven that A implies X and that B implies X then you can conclude that A or B implies X .

The type of integers. There are many ways to define the integers in Martin-Löf type theory, one of which is as follows:

defn. Define the **integers** to be the type $\mathbb{Z} := \mathbb{N} + (\mathbb{1} + \mathbb{N})$ which comes equipped with inclusions:

$$\text{in-pos} := \text{inr} \circ \text{inr} : \mathbb{N} \rightarrow \mathbb{Z} \quad \text{in-neg} := \text{inl} : \mathbb{N} \rightarrow \mathbb{Z}$$

and constants

$$-1_{\mathbb{Z}} := \text{in-neg}(0_{\mathbb{N}}) \quad 0_{\mathbb{Z}} := \text{inr}(\text{inl}(\star)) \quad 1_{\mathbb{Z}} := \text{in-pos}(0_{\mathbb{N}}).$$

Since \mathbb{Z} is built from inductive types it is then an inductive type given with its own induction principle.

Dependent pair types. Of all the inductive types we've introduced, the final one is perhaps the most important.

Recall a **dependent function** $\lambda x.f(x) : \prod_{x:A} B(x)$ is like an ordinary function except the output type is allowed to vary with the input term. Similarly, a **dependent pair** $(a, b) : \sum_{x:A} B(x)$ is like an ordinary (ordered) pair except the type of the second term $b : B(a)$ is allowed to vary with the first term $a : A$.

defn. Consider a type family $x : A \vdash B(x) \text{ type}$. The **dependent pair type** or **Σ -type** $\sum_{x:A} B(x)$ is the inductive type equipped with the function

$$\text{pair} : \prod_{x:A} \left(B(x) \rightarrow \prod_{y:A} B(y) \right).$$

The induction principle asserts that for any family of types $p : \sum_{x:A} B(x) \vdash P(p) \text{ type}$ there is a function

$$\text{ind}_{\Sigma} : \left(\prod_{x:A} \prod_{y:B} P(\text{pair}(x, y)) \right) \rightarrow \left(\prod_{z:\sum_{x:A} B(x)} P(z) \right)$$

satisfying the computation rule $\text{ind}_{\Sigma}(g, \text{pair}(x, y)) \doteq g(x, y)$.

It is common to write “ (x, y) ” as shorthand for “ $\text{pair}(x, y)$.”

defn. Given a type family $x : A \vdash B(x)$ **type** by the induction principle for Σ -types, we have a function

$$\text{pr}_1 : \sum_{x:A} B(x) \rightarrow A$$

defined by $\text{pr}_1(x, y) := x$ and a dependent function

$$\text{pr}_2 : \prod_{p : \sum_{x:A} B(x)} B(\text{pr}_1(p))$$

defined by $\text{pr}_2(x, y) := y$.

When B is a constant type family over A , the type $\sum_{x:A} B$ is the type of ordinary pairs (x, y) where $x : A$ and $y : B$. Thus **product types** arise as special cases of Σ -types.

defn. Given types A and B their product type is the type $A \times B := \sum_{x:A} B$. It comes with a pairing function

$$(-, -) : A \rightarrow B \rightarrow A \times B$$

and satisfies an induction principle:

$$\text{ind}_\times : \prod_{x:A} \prod_{y:B} P(x, y) \rightarrow \prod_{z:A \times B} P(z)$$

satisfying the computation rule $\text{ind}_\times(g, (x, y)) \doteq g(x, y)$.

As a special case, we have

$$\text{ind}_\times : (A \rightarrow B \rightarrow C) \rightarrow ((A \times B) \rightarrow C).$$

This is the inverse of the **currying function**. Thus ind_\times and ind_Σ sometimes go by the name **uncurrying**.

SEPTEMBER 15: IDENTITY TYPES

We have started to develop an analogy in which types play the role of mathematical propositions and terms in a type play the role of proofs of that proposition. More exactly, we might think of a type as a “proof-relevant” proposition, the distinction being that the individual proofs of a given proposition—the terms of the type—are first class mathematical objects, which may be used as ingredients in future proofs, rather than mere witnesses to the truth of the particular proposition.

The various constructions on types that we have discussed are analogous to the logical operations “and,” “or,” “implies,” “not,” “there exists,” and “for all.” We also have the unit type $\mathbb{1}$ to represent the proposition \top and the empty type \emptyset to represent the proposition \perp . There is one further ingredient from first-order logic that is missing a counterpart in dependent type theory: the logical operation “=.”

Given a type A and two terms $x, y : A$ it is sensible to ask whether $x = y$. From the point of view of types as proof-relevant propositions, “ $x = y$ ” should be the name of a type, in fact a dependent type. The formation rule for **identity types** says

$$\frac{\Gamma \vdash A \text{ type}}{\Gamma, x : A, y : A \vdash x =_A y \text{ type}}$$

where “ $x = y$ ” is commonly used as an abbreviation for “ $x =_A y$ ” when the type of x and y is clear from context. A term $p : x = y$ of an identity type is called an **identification** of x and y or a **path** from x to y (more about this second term later). Identifications have a rich structure that follows from a very simple characterization of the identity type due to Per Martin-Löf: it is the inductive type family freely generated by the reflexivity terms.

The inductive definition of identity types. We can define identity types as inductive types in either a one-sided or two-sided fashion. The induction rule may be easier to understand from the one-sided point of view, so we present it first.

defn (one-sided identity types). Given a type A and a term $a : A$, the **identity type** of A at a is the inductive family of types $x : A \vdash a =_A x$ **type** with a single constructor $\text{refl}_a : a =_A a$. The induction principle is postulates that for any type family $x : A, p : a =_A x \vdash P(x, p)$ **type** there is a function

$$\text{path-ind}_a : P(a, \text{refl}_a) \rightarrow \prod_{x:A} \prod_{p:a=_A x} P(x, p)$$

satisfying $\text{path-ind}_a(q, a, \text{refl}_a) \doteq q$.

This is a very strong induction principle: it says that to prove a predicate $P(x, p)$ depending on any term $x : A$ and any identification $p : a =_A x$ it suffices to assume x is a and p is refl_a and proof $P(a, \text{refl}_a)$.

More formally, identity types are defined by the following rules:

$$\frac{\Gamma \vdash a : A}{\Gamma, x : A \vdash a =_A x \text{ type}} \quad \frac{\Gamma \vdash a : A}{\Gamma \vdash \text{refl}_a : a =_A a}$$

$$\frac{\Gamma \vdash a : A \quad \Gamma, x : A, p : a =_A x \vdash P(x, p) \text{ type}}{\Gamma \vdash \text{path-ind}_a : P(a, \text{refl}_a) \rightarrow \prod_{x:A} \prod_{p:a=A x} P(x, p)} \quad \frac{\Gamma \vdash a : A \quad \Gamma, x : A, p : a =_A x \vdash P(x, p) \text{ type}}{\Gamma \vdash \text{path-ind}_a(q, a, \text{refl}_a) \doteq q : P(a, \text{refl}_a)}$$

Equally, the identity type can be considered in a two-sided fashion:

defn (two-sided identity types). Given a type A , the **identity type** of A is the inductive family of types $x : A, y : A \vdash x =_A y \text{ type}$ with a single constructor $x : A \vdash \text{refl}_x : x =_A x$. The induction principle is postulates that for any type family $x : A, y : A, p : x =_A y \vdash P(x, y, p) \text{ type}$ there is a function

$$\text{path-ind} : \prod_{a:A} P(a, a, \text{refl}_a) \rightarrow \prod_{x:A} \prod_{y:A} \prod_{p:x=A y} P(x, y, p)$$

satisfying $\text{path-ind}(q, a, a, \text{refl}_a) \doteq q$.

In this form, the identity types are defined by the following rules:

$$\frac{\Gamma \vdash A}{\Gamma, x : A, y : A \vdash x =_A y \text{ type}} \quad \frac{\Gamma \vdash A}{\Gamma, x : A \vdash \text{refl}_x : x =_A x}$$

$$\frac{\Gamma \vdash A \quad \Gamma, x : A, y : A, p : x =_A y \vdash P(x, y, p) \text{ type}}{\Gamma \vdash \text{path-ind} : \prod_{a:A} P(a, a, \text{refl}_a) \rightarrow \prod_{x:A} \prod_{y:A} \prod_{p:x=A y} P(x, y, p)} \quad \frac{\Gamma \vdash A \quad \Gamma, x : A, y : A, p : x =_A y \vdash P(x, y, p) \text{ type}}{\Gamma, a : A \vdash \text{path-ind}(q, a, a, \text{refl}_a) \doteq q : P(a, a, \text{refl}_a)}$$

These presentations are interderivable.

The groupoid structure on types. Mathematical equality, as traditionally understood, is an equivalence relation: it's reflexive, symmetric, and transitive. But all we've asserted about identity types is that they are inductively generated by the reflexivity terms! As we'll now start to discover, considerable additional structure follows.

Proposition (symmetry). *For any type A , there is an inverse operation*

$$\text{inv} : \prod_{x, y : A} x = y \rightarrow y = x.$$

Proof. We define inv by path induction. By the introduction rule for function types it suffices to define $\text{inv}p : y = x$ for $p : x = y$. Consider the type family $x : A, y : A, p : x = y \vdash P(x, y, p) := y = x$. By path induction to inhabit $y = x$ it suffices to assume $x = y$ and p is refl_x in which case we may define $\text{invrefl}_x := \text{refl}_x : x = x$. Thus inv is

$$\text{path-ind}(\lambda x, \text{refl}_x) : \prod_{x:A} \prod_{y:A} \prod_{x=y} y = x.$$

□

Notation. Write p^{-1} for $\text{inv}(p)$.

Proposition (transitivity). *For any type A , there is a concatenation operation*

$$\text{concat} : \prod_{x, y, z : A} x = y \rightarrow y = z \rightarrow x = z.$$

Proof. We define concat by appealing to the path induction principle for identity types. By the introduction rule for dependent function types, to define concat you may assume given $p : x = y$. The task is then to define $\text{concat}(p) : \prod_{z:A} y = z \rightarrow x = z$. For this, consider the type family $x : A, y : A, p : x = y \vdash P(x, y, p)$ where $P(x, y, p) := \prod_{z:A} (y = z \rightarrow (x = z))$. By applying the function path-ind to get a term of this type it suffices to assume y is x and p is refl_x . So we need only define $\text{concat}(\text{refl}_x) : \prod_{z:A} x = z \rightarrow x = z$ and we define this to be the identity function $\text{id}_{x=z}$. Thus the function concat is

$$\text{path-ind}(\lambda x, \lambda z, \text{id}_{x=z}) : \prod_{x:A} \prod_{y:A} \prod_{p:x=y} \prod_{z:A} y = z \rightarrow x = z,$$

which can be regarded as a function in the type $\prod_{x,y,z:A} x = y \rightarrow y = z \rightarrow x = z$ by swapping the order of the arguments p and z . \square

Notation. Write $p \cdot q$ for $\text{concat}(p, q)$.

While the elimination rule for identity types is quite strong the corresponding computation rule is relatively weak. It's not strong enough to show that $(p \cdot q) \cdot r$ and $p \cdot (q \cdot r)$ are judgmentally equal for any $p : x = y$, $q : y = z$, and $r : z = q$. In fact there are countermodels that show that this is false in general. However, since both $(p \cdot q) \cdot r$ and $p \cdot (q \cdot r)$ are terms of type $x = w$ we can ask whether there is an identification between them and it turns out this is always true.

Proposition (associativity). *Given $x, y, z, w : A$ and identifications $p : x = y$, $q : y = z$, and $r : z = q$, there is an associator*

$$\text{assoc}(p, q, r) : (p \cdot q) \cdot r = p \cdot (q \cdot r)$$

Proof. We define $\text{assoc}(p, q, r)$ by path induction.

Consider the type family $x : A, y : A, p : x = y \vdash \prod_{z:A} \prod_{q:y=z} \prod_{w:A} \prod_{r:z=w} (p \cdot q) \cdot r = p \cdot (q \cdot r)$. To define a term $\text{assoc}(p, q, r)$ in here it suffices to assume y is x and p is refl_x and define

$$\lambda z. \lambda q. \lambda w. \lambda r. \text{assoc}(\text{refl}_x, q, r) : \prod_{z:A} \prod_{q:x=z} \prod_{w:A} \prod_{r:z=w} (\text{refl}_x \cdot q) \cdot r = \text{refl}_x \cdot (q \cdot r).$$

By the definition of concatenation, $\text{refl}_x \cdot q \doteq q$ and $\text{refl}_x \cdot (q \cdot r) \doteq q \cdot r$. So we must define

$$\text{assoc}(\text{refl}_x, q, r) : q \cdot r = q \cdot r$$

and we can take this term to be $\text{refl}_{q \cdot r}$. \square

Proposition (units). *For any type A , there are left and right unit laws*

$$\lambda x. \lambda y. \lambda p. \text{left-unit}(p) : \lambda x, y : A \prod_{p:x=y} \text{refl}_x \cdot p = p \quad \lambda x. \lambda y. \lambda p. \text{right-unit}(p) : \prod_{x,y:A} \prod_{p:x=y} p \cdot \text{refl}_y = p.$$

Proof. We are asked to define dependent functions that takes $x, y : A$ and $p : x = y$ and produce terms

$$\text{left-unit}(p) : \text{refl}_x \cdot p = p \quad \text{right-unit}(p) : p \cdot \text{refl}_y = p.$$

By path induction, it suffices to assume y is x and p is refl_x , in which case we require terms

$$\text{left-unit}(\text{refl}_x) : \text{refl}_x \cdot \text{refl}_x = \text{refl}_x \quad \text{right-unit}(\text{refl}_x) : \text{refl}_x \cdot \text{refl}_x = \text{refl}_x.$$

By the definition of concatenation $\text{refl}_x \cdot \text{refl}_x \doteq \text{refl}_x$ so we can take $\text{refl}_{\text{refl}_x}$ as both $\text{left-unit}(\text{refl}_x)$ and $\text{right-unit}(\text{refl}_x)$. \square

Proposition (inverses). *For any type A , there are left and right inverse laws*

$$\lambda x. \lambda y. \lambda p. \text{left-inv}(p) : \prod_{x,y:A} \prod_{p:x=y} p^{-1} \cdot p = \text{refl}_y \quad \lambda x. \lambda y. \lambda p. \text{right-inv}(p) : \prod_{x,y:A} \prod_{p:x=y} p \cdot p^{-1} = \text{refl}_x.$$

Proof. We are asked to define dependent functions that takes $x, y : A$ and $p : x = y$ and produce terms

$$\text{left-inv}(p) : p^{-1} \cdot p = \text{refl}_y \quad \text{right-inv}(p) : p \cdot p^{-1} = \text{refl}_x.$$

By path induction, it suffices to assume y is x and p is refl_x , in which case we require terms

$$\text{left-inv}(\text{refl}_x) : \text{refl}_x^{-1} \cdot \text{refl}_x = \text{refl}_x \quad \text{right-inv}(\text{refl}_x) : \text{refl}_x \cdot \text{refl}_x^{-1} = \text{refl}_x.$$

By the definitions of concatenation and inverses, again both left-hand and right-hand sides are judgementally equal so we take $\text{left-inv}(\text{refl}_x)$ and $\text{right-inv}(\text{refl}_x)$ to be $\text{refl}_{\text{refl}_x}$. \square

Types as ∞ -groupoids. Martin-Löf’s rules for the identity types date from a 1975 paper “An Intuitionistic Theory of Types.” In the following two decades, there was a conjecture that went by the name “uniqueness of identity proofs” that for any $x, y : A$, $p, q : x =_A y$, the type $p =_{x=Ay} q$ is inhabited, meaning that it’s possible to construct an identification between p and q . In 1994, Martin Hofmann and Thomas Streicher constructed a model of Martin-Löf’s dependent type theory in the category of groupoids that refutes uniqueness of identity proofs.⁴

In the Hofmann-Streicher model, types A correspond to *groupoids* and terms $x, y : A$ correspond to *objects* in the groupoid. An identification $p : x = y$ corresponds to a(n iso)morphism $p : x \rightarrow y$ in the groupoid, while an identification between identifications exists if and only if p and q define the same morphism. Since there are groupoids with multiple distinct morphisms between a fixed pair of objects, we see that it is not always the case that $p =_{x=Ay} q$. Following Hofmann-Streicher, it made sense to start viewing types as more akin to groupoids than to sets. The proofs of symmetry and transitivity for identity types are more accurately described as inverses and concatenation operations in a groupoid. As we’ve seen, these satisfy various associativity, unit, and inverse laws—up to identification at least—as required by a groupoid.

But that last caveat is important. We’ve shown that for any type A , its identity types $x, y : A \vdash x =_A y$ type give it something like the structure of a groupoid. But for each $x, y : A$, $x =_A y$ is also a type, so its identity types $p, q : x =_A y \vdash p =_{x=Ay} q$ type give $x =_A y$ its own groupoid structure. And the higher identity types, $\alpha, \beta : p =_{x=Ay} q \vdash \alpha = \beta$ type give $p =_{x=Ay} q$ its own groupoid structure and so on. So a modern point of view is that the types in Martin-Löf’s dependent type theory should be thought of as ∞ -groupoids.

If A is an ∞ -groupoid, its terms $x : A$ might be called **points** and its identifications $p : x =_A y$ might be called **paths**. This explains the modern name “path induction” for the induction principle for identity types. These ideas are at the heart of the homotopical interpretation of type theory, about more which later.

The uniqueness of refl . The definition of the identity types says that the family of types $a = x$ indexed by $x : A$ is inductively generated by the term $\text{refl}_a : a = a$. It does *not* say that the type $a = a$ is inductively generated by $a : A$. In particular, we cannot apply path induction to prove that $p = \text{refl}_a$ for any $p : a = a$ because in this case neither endpoint of the identity type is free.

There is a sense however in which the reflexivity term is unique:

Proposition. *For any type A and $a : A$, (a, refl_a) is the unique term of the type $\sum_{x:A} a = x$. That is, for any $z : \sum_{x:A} a = x$, there is an identification $(a, \text{refl}_a) = z$.*

Proof. We’re trying to define a dependent function that takes $z : \sum_{x:A} a = x$ and gives a term in the identity type $(a, \text{refl}_a) =_{\sum_{x:A} a=x} z$. By Σ -induction it suffices to assume z is a pair (x, p) where $x : A$ and $p : a = x$ and construct an identification $(a, \text{refl}_a) =_{\sum_{x:A} a=x} (x, p)$. So now we’re trying to define a dependent function that takes $x : A$ and $p : a = x$ and constructs an identification $(a, \text{refl}_a) =_{\sum_{x:A} a=x} (x, p)$. By path induction, it suffices to assume x is a and p is refl_a . But now we can use reflexivity to show that $(a, \text{refl}_a) = (a, \text{refl}_a)$. \square

In terminology to be introduced later, this result says that the type $\sum_{x:A} a = x$ is **contractible** with the term (a, refl_a) serving as its **center of contraction**.

The action of paths on functions. The structural rules of type theory guarantee that any function (and indeed any construction in type theory) preserve definitional equality. We now show that in addition every function preserves identifications.

Proposition. *Let $f : A \rightarrow B$. There is an operation that defines the **action on paths** of f*

$$\text{ap}_f : \prod_{x,y:A} (x = y) \rightarrow (f(x) = f(y))$$

that satisfies the coherence conditions

$$\begin{aligned} \text{ap-id}_A : \prod_{x,y:A} \prod_{p:x=y} p &= \text{ap-id}_A(p) \\ \text{ap-comp}(f, g) : \prod_{x,y:A} \prod_{p:x=y} \text{ap}_g(\text{ap}_f(p)) &= \text{ap}_{g \circ f}(p). \end{aligned}$$

⁴The technical details of what exactly it means to “construct a model of type theory” are quite elaborate and would be interesting to explore as a final project.

Proof. By path induction to define $\text{ap}_f(p) : f(x) = f(y)$ it suffices to assume y is x and p is refl_x . We may then define $\text{ap}_f(\text{refl}_x) := \text{refl}_{f(x)} : f(x) = f(x)$.

Next to define ap-id_A it similarly suffices to suppose y is x and p is refl_x . Since $\text{ap}_{\text{id}_A}(\text{refl}_x) \doteq \text{refl}_x$, we may define $\text{ap-id}_A(\text{refl}_x) := \text{refl}_{\text{refl}_x} : \text{refl}_x = \text{refl}_x$.

Finally, to define $\text{ap-comp}(f, g)$, by path induction we may again assume y is x and p is refl_x . Since both $\text{ap}_g(\text{ap}_f(\text{refl}_x))$ and $\text{ap}_{g \circ f}(\text{refl}_x)$ are defined to be $\text{refl}_{g(f(x))}$ we may define $\text{ap-comp}(f, g)(\text{refl}_x)$ to be $\text{refl}_{\text{refl}_{g(f(x))}}$. \square

If the types A and B are thought of as ∞ -groupoids, then $f : A \rightarrow B$ can be thought of as a functor of ∞ -groupoids in a sense hinted at by the following lemma.

Lemma. *For $f : A \rightarrow B$ there are identifications*

$$\begin{aligned} \text{ap-refl}(f, x) : \text{ap}_f(\text{refl}_x) &= \text{refl}_{f(x)} \\ \text{ap-inv}(f, p) : \text{ap}_f(p^{-1}) &= \text{ap}_f(p)^{-1} \\ \text{ap-concat}(f, p, q) : \text{ap}_f(p \cdot q) &= \text{ap}_f(p) \cdot \text{ap}_f(q) \end{aligned}$$

for every $p : x = y$ and $q : y = z$.

Proof. For the first coherence, there is a definitional equality $\text{ap}_f(\text{refl}_x) \doteq \text{refl}_{f(x)}$ so we take $\text{ap-refl}(f, x) := \text{refl}_{\text{refl}_{f(x)}}$.

We define $\text{ap-inv}(f, p)$ by path induction on p by defining $\text{ap-inv}(f, \text{refl}_x) := \text{refl}_{\text{refl}_{f(x)}}$.

Similarly, we define $\text{ap-concat}(f, p, q)$ by path induction on p (since concat was defined by path induction on p) by defining $\text{ap-concat}(f, \text{refl}_x, q)$ to be $\text{refl}_{\text{ap}_f(q)}$. \square

Transport. The term ap_f defines the action of a non-dependent function $f : A \rightarrow B$ on paths in A . It's natural to ask whether a dependent function $f : \prod_{z:A} B(z)$ also induces an action on paths. There's a challenge here, though. If $x, y : A$ are terms belonging to the base type, then we can form the type $x =_A y$ to ask whether they are identifiable. But the terms $f(x) : B(x)$ and $f(y) : B(y)$ belong to different types and are not identifiable. But nevertheless if there is path $p : x = y$ identifying y with x intuition suggests there should be some way to compare $f(y)$ to $f(x)$.

To achieve this, we must construct a different sort of action of paths function first. This is called the **transport** function for dependent types $x : A \vdash B(x)$ type that, given an identification $p : x = y$ in the base type, can be used to transport any term in $B(x)$ to a term in $B(y)$.

Proposition. *For any type family $x : A \vdash B(x)$ type, there is a transport operation*

$$\text{tr}_B : \prod_{x,y:A} (x = y) \rightarrow (B(x) \rightarrow B(y)).$$

Proof. By path induction it suffices to define $\text{tr}_B(\text{refl}_x) := \text{id}_{B(x)}$. \square

As an application of transport we can now defined the action on paths of a dependent function.

Proposition. *For any dependent function $f : \prod_{z:A} B(z)$ and identification $p : x =_A y$ there is a path*

$$\text{apd}_f(p) : \text{tr}_B(p, f(x)) =_{B(y)} f(y).$$

Proof. The function

$$\lambda x. \lambda y. \lambda p. \text{apd}_f(p) : \prod_{x,y:A} \prod_{p:x=y} \text{tr}_B(p, f(x)) =_{B(y)} f(y)$$

may be defined by path induction on p . It suffices to construct a path

$$\lambda x. \text{apd}_f(\text{refl}_x) : \prod_{x:A} \text{tr}_B(\text{refl}_x, f(x)) =_{B(x)} f(x).$$

Since $\text{tr}_B(\text{refl}_x, f(x)) \doteq f(x)$ we may defined $\text{apd}_f(\text{refl}_x) := \text{refl}_{f(x)}$. \square

The laws of addition on \mathbb{N} . Recall that we defined the addition of natural numbers in such a way that

$$m + 0 \doteq m \quad m + \text{succ}_{\mathbb{N}}(n) \doteq \text{succ}_{\mathbb{N}}(m + n)$$

by induction on the second variable. With this definition, these are the only definitional equalities. However, it is possible to produce identifications proving the other commutative monoid axioms.

Lemma. *For any $n : \mathbb{N}$ there are identifications*

$$\text{left-unit-law-add}_{\mathbb{N}}(n) : 0 + n = n \quad \text{right-unit-law-add}_{\mathbb{N}}(n) : n + 0 = n.$$

Proof. The second of these can be taken to be refl_n but the first is more complicated. We define $\text{left-unit-law-add}_{\mathbb{N}}(n)$ by induction on $n : \mathbb{N}$. When $n = 0$, $0 + 0 = 0$ holds by reflexivity.

Our final goal is to show $0 + \text{succ}_{\mathbb{N}}(n) = \text{succ}_{\mathbb{N}}(n)$, for which it suffices to construct an identification

$$\text{succ}_{\mathbb{N}}(0 + n) = \text{succ}_{\mathbb{N}}(n)$$

by the definition of addition. We may assume we have an identification $p : 0 + n = n$. Thus, we can use the action on paths of $\text{succ}_{\mathbb{N}} : \mathbb{N} \rightarrow \mathbb{N}$ to obtain a term $\text{ap}_{\text{succ}_{\mathbb{N}}}(p) : \text{succ}_{\mathbb{N}}(0 + n) = \text{succ}_{\mathbb{N}}(n)$. \square

Proposition. *For any $m, n : \mathbb{N}$ there are identifications*

$$\text{left-successor-law-add}_{\mathbb{N}}(m, n) : \text{succ}_{\mathbb{N}}(m) + n = \text{succ}_{\mathbb{N}}(m + n)$$

$$\text{right-successor-law-add}_{\mathbb{N}}(m, n) = m + \text{succ}_{\mathbb{N}}(n) = \text{succ}_{\mathbb{N}}(m + n)$$

Proof. Again the second identification holds judgmentally so we define

$$\text{right-successor-law-add}_{\mathbb{N}}(m, n) := \text{refl}_{\text{succ}_{\mathbb{N}}(m+n)}.$$

We construct the former using induction on $n \in \mathbb{N}$. The base case $\text{succ}_{\mathbb{N}}(m) + 0 = \text{succ}_{\mathbb{N}}(m + 0)$ holds by $\text{refl}_{\text{succ}_{\mathbb{N}}(m)}$. For the inductive step we assume we have an identification $p : \text{succ}_{\mathbb{N}}(m) + n = \text{succ}_{\mathbb{N}}(m + n)$. Our goal is to show that $\text{succ}_{\mathbb{N}}(m) + \text{succ}_{\mathbb{N}}(n) = \text{succ}_{\mathbb{N}}(m + \text{succ}_{\mathbb{N}}(n))$. By action of paths of $\text{succ}_{\mathbb{N}} : \mathbb{N} \rightarrow \mathbb{N}$ we obtain a term

$$\text{ap}_{\text{succ}_{\mathbb{N}}}(p) : \text{succ}_{\mathbb{N}}(\text{succ}_{\mathbb{N}}(m) + n) = \text{succ}_{\mathbb{N}}(\text{succ}_{\mathbb{N}}(m + n))$$

but here the left hand side is judgmentally equal to $\text{succ}_{\mathbb{N}}(m) + \text{succ}_{\mathbb{N}}(n)$ while the right hand side is judgmentally equal to $\text{succ}_{\mathbb{N}}(m + \text{succ}_{\mathbb{N}}(n))$. \square

Proposition (associativity). *For all $k, m, n : \mathbb{N}$,*

$$\text{associative-add}_{\mathbb{N}}(k, m, n) : (m + n) + k = m + (n + k).$$

Proof. We construct $\text{associative-add}_{\mathbb{N}}(k, m, n)$ by induction on n . In the base case we have

$$(k + m) + 0 \doteq k + m \doteq k + (m + 0),$$

so we define $\text{associative-add}_{\mathbb{N}}(k, m, 0) := \text{refl}_{m+n}$.

For the inductive step let $p : (k + m) + n = k + (m + n)$. We then have

$$\text{ap}_{\text{succ}_{\mathbb{N}}}(p) : \text{succ}_{\mathbb{N}}((k + m) + n) = \text{succ}_{\mathbb{N}}(k + (m + n)).$$

We have $\text{succ}_{\mathbb{N}}((k + m) + n) \doteq (k + m) + \text{succ}_{\mathbb{N}}(n)$ and $\text{succ}_{\mathbb{N}}(k + (m + n)) \doteq k + \text{succ}_{\mathbb{N}}(m + n) \doteq k + (m + \text{succ}_{\mathbb{N}}(n))$ so this term is the term we wanted. \square

Proposition (commutativity). *For all $m, n : \mathbb{N}$,*

$$\text{commutative-add}_{\mathbb{N}}(m, n) : m + n = n + m.$$

Proof. By induction on m we have to show $0 + n = n + 0$, which holds by the unit laws for n . Then we may assume $p : m + n = n + m$ and must show $\text{succ}_{\mathbb{N}}(m) + n = n + \text{succ}_{\mathbb{N}}(m)$. We have

$$\text{ap}_{\text{succ}_{\mathbb{N}}}(p) : \text{succ}_{\mathbb{N}}(m + n) = \text{succ}_{\mathbb{N}}(n + m).$$

We then concatenate this path with the paths $\text{left-successor-law-add}_{\mathbb{N}}(m, n)$ and $\text{right-successor-law-add}_{\mathbb{N}}(n, m)$ to obtain the identification we want. \square

Recall that in Martin-Löf's dependent type theory, \mathbb{N} was defined as the inductive type freely generated by a term $0_{\mathbb{N}} : \mathbb{N}$ and a function $\text{succ}_{\mathbb{N}} : \mathbb{N} \rightarrow \mathbb{N}$. The corresponding induction principle gives a strengthened version of the Dedekind-Peano principle of mathematical induction, but two of the traditional axioms—namely that $0_{\mathbb{N}}$ is not a successor and $\text{succ}_{\mathbb{N}}$ is injective—are missing. Using our type forming operations, we can define the types that assert those axioms:

$$\prod_{n:\mathbb{N}} (n = 0_{\mathbb{N}}) \rightarrow \emptyset \quad \prod_{n,m:\mathbb{N}} (\text{succ}_{\mathbb{N}}(n) = \text{succ}_{\mathbb{N}}(m)) \rightarrow (n = m)$$

but we don't yet have the tools needed to construct terms in those types. Type theoretic *universes* will enable us to construct terms in these types and prove many other things besides.

Informally, a universe \mathcal{U} can be thought of as a “type whose terms are types.” More precisely, a universe is a type \mathcal{U} together with a type family $X : \mathcal{U} \vdash \mathcal{T}(X)$ called the *universal type family*. We think of the term X as an *encoding* of the type $\mathcal{T}(X)$ though its common to conflate these notions notationally, writing “ X ” for both the encoding and the type.

Universes are assumed to be closed under all the type constructors in a sense to be made precise below. To avoid a famous inconsistency, however, we do not assume that the universe is contained in itself. One way to think about this is that \mathcal{U} is the type of “small” types, but \mathcal{U} itself is not “small.”

In the presence of a universe \mathcal{U} , a family of small types $x : A \vdash B(x)$ type over a type A can be encoded by a function $B : A \rightarrow \mathcal{U}$ defined by sending the term x to the encoding of the type $B(x)$.⁵ In particular, if A is an inductive type, freely generated by some finite list of constructors, then *type families* over A —not just dependent functions over A —can be defined inductively by specifying types for each of the constructors. We will see examples of this soon.

Type theoretic universes.

defn. A **universe** is a type \mathcal{U} in the empty context equipped with a type family $X : \mathcal{U} \vdash \mathcal{T}(X)$ type over \mathcal{U} called the **universal family of types** that is closed under the type forming operations in the sense that it is equipped with the following structure:

- (i) \mathcal{U} contains terms $\emptyset, \mathbb{1}, \mathbb{N}$ that satisfy the judgmental equalities

$$\mathcal{T}(\emptyset) \doteq \emptyset, \quad \mathcal{T}(\mathbb{1}) \doteq \mathbb{1}, \quad \mathcal{T}(\mathbb{N}) \doteq \mathbb{N}.$$

- (ii) \mathcal{U} is closed under coproducts in the sense that it comes equipped with a function

$$\dot{+} : \mathcal{U} \rightarrow \mathcal{U} \rightarrow \mathcal{U}$$

that satisfies $\mathcal{T}(X \dot{+} Y) \doteq \mathcal{T}(X) + \mathcal{T}(Y)$.

- (iii) \mathcal{U} is closed under Π -types in the sense that it comes equipped with a function

$$\dot{\Pi} : \prod_{X:\mathcal{U}} (\mathcal{T}(X) \rightarrow \mathcal{U}) \rightarrow \mathcal{U}$$

satisfying

$$\mathcal{T}(\dot{\Pi}(X, P)) \doteq \prod_{x:\mathcal{T}(X)} \mathcal{T}(P(x))$$

for all $X : \mathcal{U}$ and $P : \mathcal{T}(X) \rightarrow \mathcal{U}$.

- (iv) \mathcal{U} is closed under Σ -types in the sense that it comes equipped with a function

$$\dot{\Sigma} : \prod_{X:\mathcal{U}} (\mathcal{T}(X) \rightarrow \mathcal{U}) \rightarrow \mathcal{U}$$

satisfying

$$\mathcal{T}(\dot{\Sigma}(X, P)) \doteq \sum_{x:\mathcal{T}(X)} \mathcal{T}(P(x))$$

for all $X : \mathcal{U}$ and $P : \mathcal{T}(X) \rightarrow \mathcal{U}$.

⁵This is already how we have been defining type families in `agda`.

(v) \mathcal{U} is closed under identity types in the sense that it comes equipped with a function

$$\text{Id} : \prod_{X:\mathcal{U}} \mathcal{T}(X) \rightarrow \mathcal{T}(X) \rightarrow \mathcal{U}$$

satisfying

$$\mathcal{T}(\text{Id}(X, x, y)) \doteq (x = y)$$

for all $X : \mathcal{U}$ and $x, y : \mathcal{T}(X)$.

defn. Given a universe \mathcal{U} , we say a type A in context Γ is **small** if it occurs in the universe: i.e., if it comes equipped with a term $\check{A} : \mathcal{U}$ in context Γ for which the judgment

$$\Gamma \vdash \mathcal{T}(\check{A}) \doteq A \text{ type}$$

holds.

When A is a small type, it's common to write A for both \check{A} and $\mathcal{T}(A)$. So by $A : \mathcal{U}$ we mean that A is a small type.

Assuming enough universes. Most of the time it's sufficient to assume just one universe \mathcal{U} . But on occasion, it is useful to assume that \mathcal{U} itself is a type in some universe.

Postulate. We assume that there are **enough universes**, i.e., that for every finite list of types in context

$$\Gamma_1 \vdash A_1 \text{ type} \quad \cdots \quad \Gamma_n \vdash A_n \text{ type}$$

there is a universe \mathcal{U} that contains each A_i in the sense that \mathcal{U} has terms

$$\Gamma_i \vdash \check{A}_i : \mathcal{U}$$

for which $\Gamma_i \vdash \mathcal{T}(\check{A}_i) \doteq A_i \text{ type}$ holds.

With this assumption it's rarely necessary to work with more than one universe at the same time.

As a consequence of our postulate that there exist enough universes, we obtain specific universes:⁶

defn. The **base universe** \mathcal{U}_0 is obtained by applying the postulate to the empty list of types in context.

defn. The **successor universe** of any universe \mathcal{U} is the universe \mathcal{U}^+ obtained from the finite list

$$\vdash \mathcal{U} \text{ type} \quad X : \mathcal{U} \vdash \mathcal{T}(X) \text{ type}$$

Thus the successor universe contains both \mathcal{U} and any type in \mathcal{U} .

defn. The **join** of two universes \mathcal{U} and \mathcal{V} is the universe $\mathcal{U} \sqcup \mathcal{V}$ obtained by applying the postulate to the type families

$$X : \mathcal{U} \vdash \mathcal{T}_{\mathcal{U}}(X) \text{ type} \quad Y : \mathcal{V} \vdash \mathcal{T}_{\mathcal{V}}(Y) \text{ type}$$

Observational equality on \mathbb{N} . To illustrate what universes are for, we define a type family $m : \mathbb{N}, n : \mathbb{N} \vdash \text{Eq}_{\mathbb{N}}(m, n) \text{ type}$ that we call **observational equality** on \mathbb{N} . Because type families can now be thought of as functions $\text{Eq}_{\mathbb{N}} : \mathbb{N} \rightarrow \mathbb{N} \rightarrow \mathcal{U}$ we can use the induction principle of \mathbb{N} to define this type family. We'll then prove that $\text{Eq}_{\mathbb{N}}$ is **logically equivalent** to the identity type family; in fact, we'll later see that these types are **equivalent**, once we know what that means. The advantage of the type family $\text{Eq}_{\mathbb{N}}$ is that it's characterized more explicitly, so this will help us prove theorems about the identity type family over the natural numbers.

defn. We define **observational equality** of \mathbb{N} as the type family $\text{Eq}_{\mathbb{N}} : \mathbb{N} \rightarrow \mathbb{N} \rightarrow \mathcal{U}$ satisfying

$$\text{Eq}_{\mathbb{N}}(0_{\mathbb{N}}, 0_{\mathbb{N}}) \doteq \mathbb{1} \quad \text{Eq}_{\mathbb{N}}(\text{succ}_{\mathbb{N}}(n), 0_{\mathbb{N}}) \doteq \emptyset \quad \text{Eq}_{\mathbb{N}}(0, \text{succ}_{\mathbb{N}}(n)) \doteq \emptyset \quad \text{Eq}_{\mathbb{N}}(\text{succ}_{\mathbb{N}}(m), \text{succ}_{\mathbb{N}}(n)) \doteq \text{Eq}_{\mathbb{N}}(m, n).$$

Lemma. *Observational equality on \mathbb{N} is reflexive:*

$$\text{refl} - \text{Eq}_{\mathbb{N}} : \prod_{n:\mathbb{N}} \text{Eq}_{\mathbb{N}}(n, n).$$

Proof. We define $\text{refl} - \text{Eq}_{\mathbb{N}}$ by induction by $\text{refl} - \text{Eq}_{\mathbb{N}}(0_{\mathbb{N}}) := \star$ and $\text{refl} - \text{Eq}_{\mathbb{N}}(\text{succ}_{\mathbb{N}}(n)) := \text{refl} - \text{Eq}_{\mathbb{N}}(n)$. \square

Proposition. *For any $m, n : \mathbb{N}$, the types $\text{Eq}_{\mathbb{N}}(m, n)$ and $(m = n)$ are **logically equivalent**: that is there are functions*

$$(m = n) \rightarrow \text{Eq}_{\mathbb{N}}(m, n) \quad \text{and} \quad \text{Eq}_{\mathbb{N}}(m, n) \rightarrow (m = n).$$

⁶In *agda*, this structure is formalized in the file *Agda.Primitive*.

Proof. By path induction, there is a function $\text{id-to-eq} : \prod_{m,n:\mathbb{N}} (m = n) \rightarrow \text{Eq}_{\mathbb{N}}(m, n)$ defined by $\text{id-to-eq}(n, \text{refl}_n) := \text{refl} - \text{Eq}_{\mathbb{N}}(n)$.

For the converse, we define a function $\text{eq-to-id} : \prod_{m,n:\mathbb{N}} \text{Eq}_{\mathbb{N}}(m, n) \rightarrow (m = n)$ by induction on m and n . We define $\text{eq-to-id}(0_{\mathbb{N}}, 0_{\mathbb{N}}) : \text{Eq}_{\mathbb{N}}(0_{\mathbb{N}}, 0_{\mathbb{N}}) \rightarrow (0_{\mathbb{N}} = 0_{\mathbb{N}})$, by induction on $\text{Eq}_{\mathbb{N}}(0_{\mathbb{N}}, 0_{\mathbb{N}}) \doteq \mathbb{1}$ to be the function that sends $\star : \mathbb{1}$ to $\text{refl}_{0_{\mathbb{N}}} : 0_{\mathbb{N}} = 0_{\mathbb{N}}$. We define the functions $\text{eq-to-id}(\text{succ}_{\mathbb{N}}(n), 0_{\mathbb{N}})$ and $\text{eq-to-id}(0_{\mathbb{N}}, \text{succ}_{\mathbb{N}}(n))$ using ex-falso , since both of these are maps out of the empty type. Finally, to define $\text{eq-to-id}(\text{succ}_{\mathbb{N}}(m), \text{succ}_{\mathbb{N}}(n))$ we may use a function $f : \text{Eq}_{\mathbb{N}}(m, n) \rightarrow (m = n)$, in which case, $\text{eq-to-id}(\text{succ}_{\mathbb{N}}(m), \text{succ}_{\mathbb{N}}(n))$ is defined to be the composite function

$$\text{Eq}_{\mathbb{N}}(\text{succ}_{\mathbb{N}}(m), \text{succ}_{\mathbb{N}}(n)) \xrightarrow{\text{id}} \text{Eq}(m, n) \xrightarrow{f} (m = n) \xrightarrow{\text{ap}_{\text{succ}_{\mathbb{N}}}} (\text{succ}_{\mathbb{N}}(m) = \text{succ}_{\mathbb{N}}(n)).$$

□

Notation. For types A and B , we write $A \leftrightarrow B$ as an abbreviation for the type

$$(A \rightarrow B) \times (B \rightarrow A).$$

Thus the logical equivalence defines a term in the type

$$\prod_{m,n:\mathbb{N}} \text{Eq}_{\mathbb{N}}(m, n) \leftrightarrow (m = n).$$

Peano's axioms.

Theorem. For any $m, n : \mathbb{N}$ we have

$$(m = n) \leftrightarrow (\text{succ}_{\mathbb{N}}(m) = \text{succ}_{\mathbb{N}}(n))$$

Proof. The action of paths of the successor function proves the forwards implication

$$\text{ap}_{\text{succ}_{\mathbb{N}}} : (m = n) \rightarrow (\text{succ}_{\mathbb{N}}(m) = \text{succ}_{\mathbb{N}}(n))$$

The direction of interest is the converse which proves that successor is injective.

Using the logical equivalences $(m = n) \leftrightarrow \text{Eq}_{\mathbb{N}}(m, n)$ we define the reverse implication to be the composite

$$(\text{succ}_{\mathbb{N}}(m) = \text{succ}_{\mathbb{N}}(n)) \xrightarrow{\text{id-to-eq}(\text{succ}_{\mathbb{N}}(m), \text{succ}_{\mathbb{N}}(n))} \text{Eq}_{\mathbb{N}}(\text{succ}_{\mathbb{N}}(m), \text{succ}_{\mathbb{N}}(n)) \xrightarrow{\text{id}} \text{Eq}_{\mathbb{N}}(m, n) \xrightarrow{\text{eq-to-id}(m, n)} (m = n).$$

□

Theorem. For any $n : \mathbb{N}$, $\neg(0_{\mathbb{N}} =_{\mathbb{N}} n)$.

Proof. We have a family of maps

$$\lambda n, \text{id-to-eq}(0_{\mathbb{N}}, n) : \prod_{n:\mathbb{N}} (0_{\mathbb{N}} = n) \rightarrow \text{Eq}_{\mathbb{N}}(0_{\mathbb{N}}, n).$$

Since $\text{Eq}_{\mathbb{N}}(0_{\mathbb{N}}, \text{succ}_{\mathbb{N}}(n)) \doteq \emptyset$ we have

$$\text{id-to-eq}(0_{\mathbb{N}}, \text{succ}_{\mathbb{N}}(n)) : (0_{\mathbb{N}} = \text{succ}_{\mathbb{N}}(n)) \rightarrow \emptyset$$

which is precisely the claim.

□

SEPTEMBER 27: MODULAR ARITHMETIC

SEPTEMBER 29: ELEMENTARY NUMBER THEORY

Part 2. The Univalent Foundations of Mathematics

OCTOBER 4: EQUIVALENCES

OCTOBER 6: CONTRACTIBILITY

OCTOBER 11: THE FUNDAMENTAL THEOREM OF IDENTITY TYPES

OCTOBER 13: PROPOSITIONS, SETS, AND GENERAL TRUNCATION LEVELS

OCTOBER 18: FUNCTION EXTENSIONALITY

OCTOBER 20: PROPOSITIONAL TRUNCATION

OCTOBER 25: THE IMAGE OF A MAP

OCTOBER 27: FINITE TYPES

NOVEMBER 1: THE UNIVALENCE AXIOM

NOVEMBER 3: SET QUOTIENTS

NOVEMBER 8: GROUPS

NOVEMBER 10: ALGEBRA

NOVEMBER 15: THE REAL NUMBERS

Part 3. Synthetic Homotopy Theory

NOVEMBER 17: THE CIRCLE

NOVEMBER 29: THE UNIVERSAL COVER OF THE CIRCLE

DECEMBER 1: HOMOTOPY GROUPS OF TYPES

DECEMBER 6: CLASSIFYING TYPES OF GROUPS

DEPT. OF MATHEMATICS, JOHNS HOPKINS UNIVERSITY, 3400 N CHARLES ST, BALTIMORE, MD 21218

E-mail address: eriehl@math.jhu.edu