MATH 721: HOMOTOPY TYPE THEORY

EMILY RIEHL

Contents

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

Date: Fall 2021.

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) $\Gamma \vdash A\ \texttt{type}$, 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) $\Gamma \vdash A \doteq B\ \texttt{type}$, 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

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

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:

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

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.

[1] Indeed, 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.

[2] We'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:

$$\frac{\Gamma \vdash A\ \mathtt{type}}{\Gamma \vdash A \doteq A\ \mathtt{type}} \qquad \frac{\Gamma \vdash A \doteq B\ \mathtt{type}}{\Gamma \vdash B \doteq A\ \mathtt{type}} \qquad \frac{\Gamma \vdash A \doteq B\ \mathtt{type} \qquad \Gamma \vdash B \doteq C\ \mathtt{type}}{\Gamma \vdash A \doteq C\ \mathtt{type}}$$

and similarly for judgmental equality between terms.

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

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

(iii) Substitution rules:

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

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:

$$\frac{\Gamma \vdash A\ \mathtt{type} \qquad \Gamma, \Delta \vdash \mathcal{J}}{\Gamma, x : A, \Delta \vdash \mathcal{J}}$$

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:

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

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

$$\dfrac{\dfrac{\dfrac{\Gamma \vdash B\ \mathtt{type}}{\Gamma, y : B \vdash y : B}}{\Gamma, y : B, x : A \vdash y : B} \qquad \dfrac{\Gamma \vdash B\ \mathtt{type} \qquad \dfrac{\Gamma, x : A, y : B, \Delta \vdash \mathcal{J}}{\Gamma, x : A, z : B, \Delta[z/y] \vdash \mathcal{J}[z/y]}}{\Gamma, y : B, x : A, z : B, \Delta[z/y] \vdash \mathcal{J}[z/y]}}{\Gamma, y : B, x : A, \Delta \vdash \mathcal{J}}$$

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**

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

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 [R] for more details.

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

$$\frac{\Gamma, x : A \vdash B(x)\ \mathtt{type}}{\Gamma \vdash \Pi_{x:A} B(x)\ \mathtt{type}}$$

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

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

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 the evaluation function:

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

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:

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

Ordinary function types.

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

$$\frac{\dfrac{\Gamma \vdash A\ \mathtt{type} \qquad \Gamma \vdash B\ \mathtt{type}}{\Gamma, x : A \vdash B\ \mathtt{type}}}{\Gamma \vdash \Pi_{x:A} B\ \mathtt{type}}$$

We adopt the notation

$$A \to B \coloneqq \Pi_{x:A} B$$

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

$$\frac{\Gamma \vdash B\ \mathtt{type} \qquad \Gamma, x : A \vdash b(x) : B}{\Gamma \vdash \lambda x.b(x) : A \to B} \qquad \frac{\Gamma \vdash f : A \to B}{\Gamma, x : A \vdash f(x) : B}$$

plus the two computation rules:

$$\frac{\Gamma \vdash B\ \mathtt{type} \qquad \Gamma, x : A \vdash b(x) : B}{\Gamma, x : A \vdash (\lambda y.b(y))(x) \doteq b(x) : B} \qquad \frac{\Gamma \vdash f : A \to B}{\Gamma \vdash \lambda x.f(x) \doteq f : A \to B}$$

defn. Identity functions are defined as follows:

$$\frac{\dfrac{\Gamma \vdash A\ \mathtt{type}}{\Gamma, x : A \vdash x : A}}{\Gamma \vdash \lambda x.x : A \to A}$$

which is traditionally denoted by $\mathtt{id}_A \coloneqq \lambda x.x$.

The idea of composition is that given a function $f\colon A \to B$ and $g\colon B \to C$ you should get a function $g \circ f\colon A \to C$. 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]

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

defn. Composition has the form:

$$\frac{\Gamma \vdash A\ \mathtt{type} \qquad \Gamma \vdash B\ \mathtt{type} \qquad \Gamma \vdash C\ \mathtt{type}}{\Gamma \vdash \_ \circ \_ : (B \to C) \to ((A \to B) \to (A \to C))}$$

It is defined by

$$\_ \circ \_ \coloneqq \lambda g.\lambda f.\lambda x.g(f(x))$$

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 $f : A \to B$, $\mathtt{id}_B \circ f \doteq f : A \to B$ and $f \circ \mathtt{id}_A \doteq f : A \to B$.

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 $\mathtt{succ}_{\mathbb{N}} : \mathbb{N} \to \mathbb{N}$ and it satisfies the induction principle.

The $\mathbb{N}$-formation rule is

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

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

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

$$\frac{}{\vdash 0_{\mathbb{N}} : \mathbb{N}} \qquad \frac{}{\vdash \mathtt{succ}_{\mathbb{N}} : \mathbb{N} \to \mathbb{N}}$$

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 $P\colon \mathbb{N} \to \{\top, \bot\}$. 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 $P(n) = \varnothing$ (corresponding to $P(n)$ being false) or $P(n) = *$ (corresponding to $P(n)$ being true).

The induction principle then says

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

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:

$$\frac{\Gamma, n : \mathbb{N} \vdash P(n)\ \mathtt{type} \qquad \Gamma \vdash p_0 : P(0_{\mathbb{N}}) \qquad \Gamma \vdash p_S : \Pi_{n:\mathbb{N}}(P(n) \to P(\mathtt{succ}_{\mathbb{N}}(n)))}{\Gamma \vdash \mathtt{ind}_{\mathbb{N}}(p_0, p_S) : \Pi_{n:\mathbb{N}} P(n)}$$

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

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

$$\frac{\Gamma, n : \mathbb{N} \vdash P(n)\ \mathtt{type} \qquad \Gamma \vdash p_0 : P(0_{\mathbb{N}}) \qquad \Gamma \vdash p_S : \Pi_{n:\mathbb{N}}(P(n) \to P(\mathtt{succ}_{\mathbb{N}}(n)))}{\Gamma \vdash \mathtt{ind}_{\mathbb{N}}(p_0, p_S)(0_{\mathbb{N}}) \doteq p_0 : P(0_{\mathbb{N}})}$$

and under the same premises

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

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, $\mathtt{ind}_{\mathbb{N}}(p_0, p_S)$ 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) $\mathtt{succ}_{\mathbb{N}} : \mathbb{N} \to \mathbb{N}$

(iii) $\forall n, \mathsf{succ}_{\mathbb{N}}(n) \neq 0_{\mathbb{N}}$
(iv) $\forall n, m, \mathsf{succ}_{\mathbb{N}}(n) = \mathsf{succ}_{\mathbb{N}}(m) \to n = m$
(v) induction

We'll be able to *prove* the 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

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

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 $f : \Pi_{n:\mathbb{N}} P(n)$ by induction on n it suffices, by the elimination rule for the natural numbers type, to provide two terms:

$$p_0 : P(0_{\mathbb{N}}) \qquad p_S : \Pi_{n:\mathbb{N}} P(n) \to P(\mathsf{succ}_{\mathbb{N}}(n)).$$

Thus the definition of f may be presented by writing

$$f(0_{\mathbb{N}}) :\equiv p_0 \qquad f(\mathsf{succ}_{\mathbb{N}}(n)) :\equiv p_S(n, f(n)).$$

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 $\Pi_{n:\mathbb{N}} P(n)$ 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:

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

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

$$\mathtt{ind}_{\mathbb{1}} : P(\star) \to \prod_{x:1} P(x)$$

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

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 $\varnothing$ satisfying the induction principle that for any family of types $x : \varnothing \vdash P(x)$ there is a term

$$\mathtt{ind}_{\varnothing} : \prod_{x:\varnothing} P(x).$$

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

$$\frac{\vdash A\ \mathtt{type}}{\mathtt{ex\text{-}falso} \coloneqq \mathtt{ind}_{\varnothing} : \varnothing \to A}$$

By the elimination rule for function types it follows that if we had a term $x : \varnothing$ 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 its 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 $\neg A \coloneqq A \to \varnothing$ 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 $\varnothing$. 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

$$\neg\neg A \coloneqq (A \to \varnothing) \to \varnothing)$$

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*

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

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

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

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

$$\frac{\Gamma \vdash A\ \mathtt{type} \qquad \Gamma \vdash B\ \mathtt{type}}{\Gamma \vdash A + B\ \mathtt{type}}$$

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

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

$$\mathtt{inl} : A \to A + B \qquad \mathtt{inr} : B \to A + B$$

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

$$\mathtt{ind}_+ : (\Pi_{x:A} P(\mathtt{inl}(x))) \to \left(\Pi_{y:B} P(\mathtt{inr}(y))\right) \to \Pi_{z:A+B} P(z)$$

satisfying the computation rules

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

Not as a special case we have

$$\mathtt{ind}_+ : (A \to X) \to (B \to X) \to (A + B \to X)$$

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 $\mathbb{Z} \coloneqq \mathbb{N} + (\mathbb{1} + \mathbb{N})$ which comes equipped with inclusions:

$$\mathtt{in\text{-}pos} \coloneqq \mathtt{inr} \circ \mathtt{inr} : \mathbb{N} \to \mathbb{Z} \qquad \mathtt{in\text{-}neg} \coloneqq \mathtt{inl} : \mathbb{N} \to \mathbb{Z}$$

and constants

$$-1_{\mathbb{Z}} \coloneqq \mathtt{in\text{-}neg}(0_{\mathbb{N}}) \qquad 0_{\mathbb{Z}} \coloneqq \mathtt{inr}(\mathtt{inl}(\star)) \qquad 1_{\mathbb{Z}} \coloneqq \mathtt{in\text{-}pos}(0_{\mathbb{N}}).$$

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** $\lambda x.f(x) : \Pi_{x:A} B(x)$ is like an ordinary function except the output type is allowed to vary with the input term. Similarly, a **dependent pair** $(a, b) : \Sigma_{x:A} B(x)$ 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 $x : A \vdash B(x)\ \mathtt{type}$. The **dependent pair type** or **Σ-type** $\Sigma_{x:A} B(x)$ is the inductive type equipped with the function

$$\mathtt{pair} : \Pi_{x:A} \left(B(x) \to \Pi_{y:A} B(y)\right).$$

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

$$\mathtt{ind}_\Sigma : \left(\Pi_{x:A} \Pi_{y:B} P(\mathtt{pair}(x, y))\right) \to \left(\Pi_{z:\Sigma_{x:A} B(x)} P(z)\right)$$

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

It is common to write "(x, y)" as shorthand for "$\mathtt{pair}(x, y)$."

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

$$\mathtt{pr}_1 : \Sigma_{x:A} B(x) \to A$$

defined by $\mathtt{pr}_1(x, y) \coloneqq x$ and a dependent function

$$\mathtt{pr}_2 : \Pi_{p:\Sigma_{x:A} B(x)} B(\mathtt{pr}_1(p))$$

defined by $\mathtt{pr}_2(x, y) \coloneqq y$.

When B is a constant type family over A, the type $\Sigma_{x:A} B$ 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 $A \times B \coloneqq \Sigma_{x:A} B$. It comes with a pairing function

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

and satisfies an induction principle:

$$\mathtt{ind}_\times : \Pi_{x:A} \Pi_{y:B} P(x,y) \to \Pi_{z:A\times B} P(z)$$

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

As a special case, we have

$$\mathtt{ind}_\times : (A \to B \to C) \to ((A \times B) \to C).$$

This is the inverse of the **currying function**. Thus $\mathtt{ind}_\times$ and $\mathtt{ind}_\Sigma$ 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 $\varnothing$ to represent the proposition $\bot$. 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

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

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 $x : A \vdash a =_A x\ \mathtt{type}$ with a single constructor $\mathtt{refl}_a : a =_A a$. The induction principle is postulates that for any type family $x : A, p : a =_A x \vdash P(x,p)\ \mathtt{type}$ there is a function

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

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

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 $\mathtt{refl}_a$ and prove $P(a, \mathtt{refl}_a)$.

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

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

$$\frac{\Gamma \vdash a : A \qquad \Gamma, x : A, p : a =_A x \vdash P(x,p)\ \mathtt{type}}{\Gamma \vdash \mathtt{path\text{-}ind}_a : P(a, \mathtt{refl}_a) \to \Pi_{x:A} \Pi_{p:a=_A x} P(x,p)} \qquad \frac{\Gamma \vdash a : A \qquad \Gamma, x : A, p : a =_A x \vdash P(x,p)\ \mathtt{type}}{\Gamma \vdash \mathtt{path\text{-}ind}_a(q, a, \mathtt{refl}_a) \doteq q : P(a, \mathtt{refl}_a)}$$

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 $x : A, y : A \vdash x =_A y \ \mathtt{type}$ with a single constructor $x : A \vdash \mathtt{refl}_x : x =_A x$. The induction principle is postulates that for any type family $x : A, y : A, p : x =_A y \vdash P(x, y, p) \ \mathtt{type}$ there is a function

$$\mathtt{path\text{-}ind} : \Pi_{a:A} P(a, a, \mathtt{refl}_a) \to \Pi_{x:A} \Pi_{y:A} \Pi_{p:x=_A y} P(x, y, p)$$

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

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

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

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

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, three is an **inverse operation***

$$\mathtt{inv} : \Pi_{x,y:A} x = y \to y = x.$$

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

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

□

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

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

$$\mathtt{concat} : \Pi_{x,y,z:A} x = y \to y = z \to x = z.$$

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

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

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

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

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 $r : z = w$, there is an associator*

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

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

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

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

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

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

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

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

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

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

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

By path induction, it suffices to assume y is x and p is $\mathtt{refl}_x$, in which case we require terms

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

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

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

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

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

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

By path induction, it suffices to assume y is x and p is $\mathtt{refl}_x$, in which case we require terms

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

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

September 20: More identity types

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 $p =_{x=_A y} q$ 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 $p : x \to y$ 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 $p =_{x=_A y} q$. 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 $x, y : A \vdash x =_A y\ \mathtt{type}$ 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 $p, q : x =_A y \vdash p =_{x=_A y} q\ \mathtt{type}$ give $x =_A y$ its own groupoid structure. And the higher identity types, $\alpha, \beta : p =_{x=_A y} q \vdash \alpha = \beta\ \mathtt{type}$ give $p =_{x=_A y} q$ 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*.

[4] The 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.

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 $\mathtt{refl}_a : a = a$. 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 $p = \mathtt{refl}_a$ 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 $a : A$, $(a, \mathtt{refl}_a)$ is the unique term of the type $\Sigma_{x:A} a = x$. That is, for any $z : \Sigma_{x:A} a = x$, there is an identification $(a, \mathtt{refl}_a) = z$.*

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

In terminology to be introduced later, this result says that the type $\Sigma_{x:A} a = x$ is **contractible** with the term $(a, \mathtt{refl}_a)$ 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 $f : A \to B$. There is an operation that defines the **action on paths** of f*

$$\mathtt{ap}_f : \Pi_{x,y:A}(x = y) \to (f(x) = f(y))$$

that satisfies the coherence conditions

$$\mathtt{ap\text{-}id}_A : \Pi_{x,y:A} \Pi_{p:x=y} p = \mathtt{ap}_{\mathtt{id}_A}(p)$$

$$\mathtt{ap\text{-}comp}(f, g) : \Pi_{x,y:A} \Pi_{p:x=y} \mathtt{ap}_g(\mathtt{ap}_f(p)) = \mathtt{ap}_{g \circ f}(p).$$

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

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

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

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

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

$$\mathtt{ap\text{-}refl}(f, x) : \mathtt{ap}_f(\mathtt{refl}_x) = \mathtt{refl}_{f(x)}$$

$$\mathtt{ap\text{-}inv}(f, p) : \mathtt{ap}_f(p^{-1}) = \mathtt{ap}_f(p)^{-1}$$

$$\mathtt{ap\text{-}concat}(f, p, q) : \mathtt{ap}_f(p \cdot q) = \mathtt{ap}_f(p) \cdot \mathtt{ap}_f(q)$$

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

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

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

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

Transport. The term $\mathtt{ap}_f$ defines the action of a non-dependent function $f\colon A \to B$ on paths in A. It's natural to ask whether a dependent function $f : \Pi_{z:A} B(z)$ 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***

$$\mathtt{tr}_B : \Pi_{x,y:A}(x = y) \to (B(x) \to B(y)).$$

Proof. By path induction it suffices to define $\mathtt{tr}_B(\mathtt{refl}_x)) \coloneqq \mathtt{id}_{B(x)}$. □

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

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

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

Proof. The function

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

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

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

Since $\mathtt{tr}_B(\mathtt{refl}_x), f(x)) \doteq f(x)$ we may defined $\mathtt{apd}_f(\mathtt{refl}_x) \coloneqq \mathtt{refl}_{f(x)}$. □

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

$$m + 0 \doteq m \qquad m + \mathtt{succ}_{\mathbb{N}}(n) \doteq \mathtt{succ}_{\mathbb{N}}(m + n)$$

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*

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

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

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

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

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 $\mathtt{succ}_{\mathbb{N}} : \mathbb{N} \to \mathbb{N}$ to obtain a term $\mathtt{ap}_{\mathtt{succ}_{\mathbb{N}}}(p) : \mathtt{succ}_{\mathbb{N}}(0 + n) = \mathtt{succ}_{\mathbb{N}}(n)$. □

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

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

Proof. Again the second identification holds judgmentally so we define

$$\mathtt{right\text{-}sucessor\text{-}law\text{-}add}_{\mathbb{N}}(m, n) \coloneqq \mathtt{refl}_{\mathtt{succ}_{\mathbb{N}}(m+n)}.$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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 $\texttt{succ}_{\mathbb{N}}(m) + n = n + \texttt{succ}_{\mathbb{N}}(m)$. We have

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

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

September 22: Universes

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 $\texttt{succ}_{\mathbb{N}} : \mathbb{N} \to \mathbb{N}$. 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 $\texttt{succ}_{\mathbb{N}}$ is injective—are missing. Using our type forming operations, we can define the types that assert those axioms:

$$\Pi_{n:\mathbb{N}}(n = 0_{\mathbb{N}}) \to \varnothing \qquad \Pi_{n,m:\mathbb{N}}(\texttt{succ}_{\mathbb{N}}(n) = \texttt{succ}_{\mathbb{N}}(m)) \to (n = m)$$

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 $x : A \vdash B(x)\ \texttt{type}$ over a type A can be encoded by a function $B\colon A \to \mathcal{U}$ 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 $X : \mathcal{U} \vdash \mathcal{T}(X)\ \texttt{type}$ 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 $\check{\varnothing}$, $\check{\mathbb{1}}$, $\check{\mathbb{N}}$ that satisfy the judgmental equalities

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

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

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

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

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

$$\check{\Pi} : \Pi_{X:\mathcal{U}}(\mathcal{T}(X) \to \mathcal{U}) \to \mathcal{U}$$

satisfying

$$\mathcal{T}(\check{\Pi}(X, P)) \doteq \Pi_{x:\mathcal{T}(X)}\mathcal{T}(P(x))$$

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

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

$$\check{\Sigma} : \Pi_{X:\mathcal{U}}(\mathcal{T}(X) \to \mathcal{U}) \to \mathcal{U}$$

satisfying

$$\mathcal{T}(\check{\Sigma}(X, P)) \doteq \Sigma_{x:\mathcal{T}(X)}\mathcal{T}(P(x))$$

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

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

$$\check{\mathrm{Id}} : \Pi_{X:\mathcal{U}}\mathcal{T}(X) \to \mathcal{T}(X) \to \mathcal{U}$$

satisfying

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

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

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

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

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

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 $\Gamma_i \vdash \mathcal{T}(\check{A}_i) \doteq A_i \ \mathtt{type}$ 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

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

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

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

[6] In `agda`, this structure is formalized in the file Agda.Primitive.

Observational equality on $\mathbb{N}$. To illustrate what universes are for, we define a type family $m : \mathbb{N}, n : \mathbb{N} \vdash \mathrm{Eq}_{\mathbb{N}}(m,n)\ \mathtt{type}$ that we call **observational equality** on $\mathbb{N}$. Because type families can now be thought of as functions $\mathrm{Eq}_{\mathbb{N}} : \mathbb{N} \to \mathbb{N} \to \mathcal{U}$ we can use the induction principle of $\mathbb{N}$ to define this type family. We'll then prove that $\mathrm{Eq}_{\mathbb{N}}$ 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 $\mathrm{Eq}_{\mathbb{N}}$ 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 $\mathrm{Eq}_{\mathbb{N}} : \mathbb{N} \to \mathbb{N} \to \mathcal{U}$ satisfying

$$\mathrm{Eq}_{\mathbb{N}}(0_{\mathbb{N}}, 0_{\mathbb{N}}) \doteq \mathbb{1} \quad \mathrm{Eq}_{\mathbb{N}}(\mathtt{succ}_{\mathbb{N}}(n), 0_{\mathbb{N}}) \doteq \varnothing \quad \mathrm{Eq}_{\mathbb{N}}(0, \mathtt{succ}_{\mathbb{N}}(n)) \doteq \varnothing \quad \mathrm{Eq}_{\mathbb{N}}(\mathtt{succ}_{\mathbb{N}}(m), \mathtt{succ}_{\mathbb{N}}(n)) \doteq \mathrm{Eq}_{\mathbb{N}}(m,n).$$

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

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

Proof. We define $\mathtt{refl} - \mathrm{Eq}_{\mathbb{N}}$ by induction by $\mathtt{refl} - \mathrm{Eq}_{\mathbb{N}}(0_{\mathbb{N}}) :\equiv \star$ and $\mathtt{refl} - \mathrm{Eq}_{\mathbb{N}}(\mathtt{succ}_{\mathbb{N}}(n)) :\equiv \mathtt{refl} - \mathrm{Eq}_{\mathbb{N}}(n)$. □

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

$$(m = n) \to \mathrm{Eq}_{\mathbb{N}}(m,n) \quad \textit{and} \quad \mathrm{Eq}_{\mathbb{N}}(m,n) \to (m = n).$$

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

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

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

□

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

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

Thus the logical equivalence defines a term in the type

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

Peano's axioms.

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

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

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

$$\mathtt{ap}_{\mathtt{succ}_{\mathbb{N}}} : (m = n) \to (\mathtt{succ}_{\mathbb{N}}(m) = \mathtt{succ}_{\mathbb{N}}(n))$$

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

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

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

□

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

Proof. We have a family of maps

$$\lambda n, \texttt{id-to-eq}(0_{\mathbb{N}}, n) : \Pi_{n:\mathbb{N}}(0_{\mathbb{N}} = n) \to \mathrm{Eq}_{\mathbb{N}}(0_{\mathbb{N}}, n).$$

Since $\mathrm{Eq}_{\mathbb{N}}(0_{\mathbb{N}}, \mathsf{succ}_{\mathbb{N}}(n)) \doteq \varnothing$ we have

$$\texttt{id-to-eq}(0_{\mathbb{N}}, \mathsf{succ}_{\mathbb{N}}(n)) : (0_{\mathbb{N}} = \mathsf{succ}_{\mathbb{N}}(n)) \to \varnothing$$

which is precisely the claim. □

September 27: Modular arithmetic

Having fully described Martin-Löf's dependent type theory, we may now start developing some mathematics in it. The fundamental idea used to develop mathematics is something we've already previewed: the Curry-Howard interpretation.

The Curry-Howard interpretation. The Curry-Howard interpretation is an interpretation of logic into type theory. In type theory, there is no separation between the logical framework and the general theory of collections of mathematical objects the way there is in the more traditional setup with Zermelo-Fraenkel set theory, which is postulated by axioms in first order logic. The idea is that propositions may be expressed as types with proofs of those propositions expressed as terms in those types. For example:

defn. We say that a natural number d divides a natural number n if there is a term in the type

$$d \mid n \coloneqq \Sigma_{k:\mathbb{N}} d \cdot k = n$$

defined using the multiplication $\cdot$ on $\mathbb{N}$, the identity type of $\mathbb{N}$, and the dependent sum of the type family $k : \mathbb{N} \vdash d \cdot k = n$ `type` .

Just as existential quantification ($\exists$) is expressed using Σ-types, universal quantification ($\forall$) is expressed using Π-types. For example, the type

$$\Pi_{n:\mathbb{N}} 1 \mid n$$

asserts that every natural number is divisible by 1. The term

$$\lambda n.(n, \texttt{left-unit}(n)) : \Pi_{n:\mathbb{N}} 1 \mid n$$

proves this result.

Proposition. *Let $d, m, n : \mathbb{N}$. If d divides any two of m, n, and $m + n$, then d divides the third.*

Proof. We prove only that if $d \mid m$ and $d \mid n$ then $d \mid m + n$. By hypothesis we have terms:

$$H : \Sigma_{k:\mathbb{N}} d \cdot k = m \quad \text{and} \quad K : \Sigma_{k:\mathbb{N}} d \cdot k = n.$$

By Σ-induction, we may assume that H is given by a pair $(h : \mathbb{N}, p : d \cdot h = m)$ and K is given by a pair $(k : \mathbb{N}, q : d \cdot k = n)$. To get a term in $\Sigma_{x:\mathbb{N}} d \cdot x = m + n$ we may use $x \coloneqq h + k$. Our goal is then to define an identification $d \cdot (h + k) = m + n$ which we obtain as a concatenation

$$d \cdot (h + k) \overset{\texttt{dist}}{=\!=} d \cdot h + d \cdot k \overset{\mathsf{ap}_{+d \cdot k} p}{=\!=} m + d \cdot k \overset{\mathsf{ap}_{m+} q}{=\!=} m + n$$

□

We have observed many similarities between the rules of various type constructors and tautologies from logic. For instance, the elimination rule for the non-dependent function type supplies a function

$$\texttt{modus-ponens} : A \times (A \to B) \to B.$$

One important difference is that general types may contain multiple terms that cannot be identified: i.e., for which it is possible to prove that $x =_A y \to \varnothing$. Later we'll study the following predicate on types:

$$\texttt{is-prop}(A) \coloneqq \Pi_{x,y:A} x =_A y$$

which asserts that *if* A has multiple terms (which it may not) those terms can always be identified. This will be the $n = -1$ level of a hierarchy of n-types for $n \geq -2$.

The congruence relations on $\mathbb{N}$. The family of identity types can be understood as a type-valued binary relation on a type.

defn. For a type A, a **typal binary relation** on A is a family of types $x, y : A \vdash R(x,y)$ `type` . A binary relation R is

- **reflexive** if it comes with a term $\rho : \prod_{x:A} R(x,x)$,
- **symmetric** if it comes with a term $\sigma : \prod_{x,y:A} R(x,y) \to R(y,x)$,
- **transitive** if it comes with a term $\tau : \prod_{x,y,z:A} \to R(x,y) \to R(y,z) \to R(x,z)$

A **typal equivalence relation** on A is a reflexive, symmetric, and transitive, typal binary relation.

For instance, for each $k : \mathbb{N}$ we can define the relation of congruence modulo k by defining a type

$$x \equiv y \mod k$$

for each $x, y : \mathbb{N}$ comprised of proofs that x is equivalent to y modulo k. Following Gauss, we say that x is equivalent to y mod k if k divides the symmetric difference $\mathtt{dist}_{\mathbb{N}}(x,y)$ defined recursively by

$$\mathtt{dist}_{\mathbb{N}}(0,0) \coloneqq 0 \quad \mathtt{dist}_{\mathbb{N}}(0,y+1) \coloneqq y+1 \quad \mathtt{dist}_{\mathbb{N}}(x+1,0) \coloneqq x+1, \quad \mathtt{dist}_{\mathbb{N}}(x+1,y+1) \coloneqq \mathtt{dist}_{\mathbb{N}}(x,y).$$

defn. For $k, x, y : \mathbb{N}$ define

$$x \equiv y \mod k \coloneqq k \mid \mathtt{dist}_{\mathbb{N}}(x,y).$$

Note this defines the *type* $x \equiv y \mod k$. A term is then a pair comprised of an $\ell : \mathbb{N}$ together with an identification $k \cdot \ell = \mathtt{dist}_{\mathbb{N}}(x,y)$.

We leave the following to the course text:

Proposition. *For each k, the typal relation $\equiv \mod k$ is an equivalence relation.*

There are other important relations on $\mathbb{N}$ that are not-equivalence relations.

defn. The binary relation $\leq$ on $\mathbb{N}$ is defined by induction by

$$0 \leq 0 \coloneqq \mathbb{1} \quad 0 \leq n+1 \coloneqq \mathbb{1} \quad n+1 \leq 0 \coloneqq \varnothing \quad m+1 \leq n+1 \coloneqq m \leq n.$$

Similarly, the binary relation $<$ is defined by

$$0 < 0 \coloneqq \varnothing \quad 0 < n+1 \coloneqq \mathbb{1} \quad n+1 < 0 \coloneqq \varnothing \quad m+1 < n+1 \coloneqq m < n.$$

The standard finite types. The standard finite sets are classically defined as the sets $\{n \in \mathbb{N} \mid n < k\}$, so how do we interpret a subset $\{x \in A \mid P(x)\}$ characterized by a predicate in type theory?

In the Curry-Howard interpretation, the predicate $P(x)$ is interpreted as a type family and the type of terms x in A for which $P(x)$ is true is interpreted by the Σ-type $\Sigma_{x:A} P(x)$. Note for a general type family $P(x)$ it won't necessarily be the case that the map $\mathtt{pr}_1 : \Sigma_{x:A} P(x) \to A$ is a monomorphism[7] so this construction operates a bit differently than in set theory.

Through this mechanism it is possible to define the classical finite sets as

$$\text{Classical-Fin}_k \coloneqq \Sigma_{n:\mathbb{N}} n < k$$

though the standard definition is as follows:

defn. We define the type family Fin of **standard finite types** inductively (using the induction principle of $\mathbb{N}$ and the universe $\mathcal{U}$) as follows:

$$\text{Fin}_0 \coloneqq \varnothing, \quad \text{Fin}_{k+1} \coloneqq \text{Fin}_k + \mathbb{1}.$$

Write i for $\mathtt{inl} : \text{Fin}_k \to \text{Fin}_{k+1}$ and $\star$ for the point $\mathtt{inr}(\star) : \text{Fin}_{k+1}$.

By induction we can define functions $\iota_k : \text{Fin}_k \to \mathbb{N}$ for each k. When $k = 0$ there is nothing to show. To define $\iota_{k+1} : \text{Fin}_{k+1} \to \mathbb{N}$ we can use ι_k and define $\iota_{k+1}(i(x)) \coloneqq \iota_k(x)$ and $\iota_{k+1}(\star) \coloneqq k$.

[7] Though this will be the case if each type $P(x)$ is a proposition in the sense alluded to above.

The natural numbers modulo $k+1$. Given an equivalence relation $\sim$ on a set A the quotient $A_{/\sim}$ comes equipped with a quotient map $q\colon A \to A_{/\sim}$ that satisfies two important properties:

(i) q is **effective**: $q(x) = q(y)$ if and only if $x \sim y$
(ii) q is surjective: for all $[z] \in A_{/\sim}$ there is some $z \in A$ so that $q(z) = [z]$.

Both properties can be expressed in type theory, though there are some subtleties.

defn. In the context of types A and B there is a type family is-surj$\colon (A \to B) \to \mathcal{U}$ defined by

$$\text{is-surj}(f) :\equiv \Pi_{b:B}\Sigma_{a:A} f(a) =_B b.$$

The subtlety is that this really defines a *split* notion of surjectivity. A term p in is-surj(f) defines a function that for each term $b : B$ produces a term of $\Sigma_{a:A} f(a) =_B b$. By compositing p with $\mathtt{pr}_1 : \Sigma_{a:A} f(a) =_B b \to A$, we obtain a function $s\colon B \to A$. By composing p with $\mathtt{pr}_2 : \Sigma_{a:A} f(s(b)) =_B b$ we also obtain a proof that s is a **section** of f. Thus surjective functions in homotopy type theory are really **split** surjective functions.

Our next challenge is to define the quotient maps $[-]_k\colon \mathbb{N} \to \text{Fin}_k$ that compute the remainder modulo k. Our strategy will be to define this function by induction on $n : \mathbb{N}$. The idea is that the term $0_{\mathbb{N}} : \mathbb{N}$ should get sent to some 0 while successors in $\mathbb{N}$ should be sent to successors in Fin_k, taken in the cyclic order. We define these auxiliary structures first.

defn. We define $\mathtt{zero}_k : \text{Fin}_{k+1}$ recursively by

$$\mathtt{zero}_0 :\equiv \star \quad \mathtt{zero}_{k+1} :\equiv i(\mathtt{zero})_k.$$

We then define $\mathtt{skip\text{-}zero}_k : \text{Fin}_k \to \text{Fin}_{k+1}$ recursively by

$$\mathtt{skip\text{-}zero}_{k+1}(i(x)) :\equiv i(\mathtt{skip\text{-}zero}_k(x)) \quad \mathtt{skip\text{-}zero}_{k+1}(\star) :\equiv \star.$$

Finally, we define $\mathtt{succ}_k : \text{Fin}_k \to \text{Fin}_k$ recursively by

$$\mathtt{succ}_{k+1}(i(x)) :\equiv \mathtt{skip\text{-}zero}_k(x) \qquad \mathtt{succ}_{k+1}(\star) :\equiv \mathtt{zero}_k.$$

defn. For any $k : \mathbb{N}$ define $[-]_{k+1}\colon \mathbb{N} \to \text{Fin}_{k+1}$ by

$$[0]_{k+1} :\equiv 0 \quad \text{and} \quad [n+1]_{k+1} :\equiv \mathtt{succ}_{k+1}[n]_{k+1}.$$

The text goes on to show that

- $n \equiv \iota[n]_k \mod k$ for all n and k,
- $[n]_k = [m]_k$ if and only of $n \equiv m \mod k$,
- and the map $[-]_k\colon \mathbb{N} \to \text{Fin}_k$ is split surjective.

Then it is possible to use this quotient map to define the cyclic group structure on Fin_k.

September 29: Decidability in elementary number theory

In constructive mathematics it is not possible to prove the law of excluded middle: namely that $P \vee \neg P$ for an arbitrary proposition P. Similarly in type theory, it is not possible to construct a term of type $A + \neg A$ for arbitrary A. But certain types do come with such terms.

Decidability.

defn. A type A is **decidable** if it comes equipped with an element of type

$$\text{is-decidable}(A) :\equiv A + \neg A.$$

A type family $P\colon A \to \mathcal{U}$ is **decidable** if $P(a)$ is decidable for every $a : A$.

ex. The primary way to show that A is decidable is either to provide a term $a : A$ or provide a function $na : A \to \varnothing$. In particular $\mathbb{1}$ is decidable since we have $\mathtt{inl}(\star)$: is-decidable$(\mathbb{1})$. Similarly $\varnothing$ is decidable since we have $\mathtt{inr}(\mathtt{id})$: is-decidable$(\varnothing)$.

ex. Since the type families $n, m : \mathbb{N} \vdash n \leq m$ `type` and $n, m : \mathbb{N} \vdash n < m$ `type` were defined by induction from the types $\varnothing$ and $\mathbb{1}$, it follows that these type families are decidable.

Remark. If A and B are decidable then so are $A + B$, $A \times B$, and $A \to B$. Proofs use case analysis over the coproduct types $A + \neg A$ and $B + \neg B$.

defn. A type A **has decidable equality** if the identity type $x =_A y$ is decidable for every $x, y : A$. Thus

$$\text{has-decidable-eq}(A) \coloneqq \Pi_{x,y:A}\text{is-decidable}(x =_A y).$$

Lemma. *Suppose A and B are types so that $A \leftrightarrow B$. Then A is decidable if and only if B is decidable.*

Proof. A proof of $A \leftrightarrow B$ supplies functions $f\colon A \to B$ and $g\colon B \to A$. Using the contrapositive function we obtain $\texttt{contrapositive}(f) : \neg B \to \neg A$ and $\texttt{contrapositive}(g)\colon \neg A \to \neg B$. We therefore have functions

$$f + \texttt{contrapositive}(g)\colon A + \neg A \to B + \neg B \qquad g + \texttt{contrapositive}(f) : B + \neg B \to A + \neg A,$$

proving the logical equivalence of is-decidable(A) and is-decidable(B). In particular, if either type is inhabited, both must be. □

Corollary. $\mathbb{N}$ *has decidable equality.*

Proof. We have shown that the identity types of $\mathbb{N}$ are logically equivalent to the observational equality types, which were defined to be $\mathbb{1}$ or $\varnothing$. As both types are decidable, the identity types of $\mathbb{N}$ must be as well. □

Remark. We will prove later that if a type has decidable equality then it must be a **set** in a technical sense to be introduced. Even so, not all sets have decidable equality unless one assumes that the law of excluded middle is true for all propositions.

Case analysis. Suppose you'd like to define a function by case analysis such as $\texttt{collatz} : \mathbb{N} \to \mathbb{N}$

$$\texttt{collatz}(n) \coloneqq \begin{cases} n/2 & n \text{ is odd} \\ 2n+1 & n \text{ is even} \end{cases}$$

To justify this sort of case analysis we use a term

$$d : \Pi_{n:\mathbb{N}}\text{is-decidable}(2 \mid n)$$

whose construction we skipped. Note that $2 \mid n$ and $\neg(2 \mid n)$ cannot both hold because if so we could evaluate the function $\texttt{odd}(n) : \neg(2 \mid n)$ at the term $\texttt{even}(n) : 2 \mid n$ to get a contradiction. So this d can be thought of as a proof that for all $n : \mathbb{N}$, n is odd or n is even (but not both).

This puts us into the following abstract setup. Our goal is to define a function $c : \Pi_{x:A}C(x)$, namely the function $\texttt{collatz} : \mathbb{N} \to \mathbb{N}$. We already have a function $d : \Pi_{x:A}B(x)$, namely $d : \Pi_{n:\mathbb{N}}\text{is-decidable}(2 \mid n)$. So it suffices to define a function $h : \Pi_{x:A}B(x) \to C(x)$ because then we can define $c(x) \coloneqq h(x, d(x))$. In this case this means we need a function

$$h : \Pi_{n:\mathbb{N}}\text{is-decidable}(2 \mid n) \to \mathbb{N}$$

which we can now define by cases from

$$h\text{-}\texttt{even}(n) \coloneqq \lambda n.n/2\colon (2 \mid n) \to \mathbb{N} \quad \text{and} \quad h\text{-}\texttt{odd}(n) \coloneqq \lambda n.3n+1 : \neg(2 \mid n) \to \mathbb{N}.$$

There is something called the "with-abstraction" that gives a concise syntax for functions defined in this manner.

The well-ordering principle of $\mathbb{N}$. The traditional well-ordering principle is about subsets of $\mathbb{N}$, or equivalently, about predicates on $\mathbb{N}$. In type theory, we replace these by decidable type families over $\mathbb{N}$.

defn. Let $P\colon \mathbb{N} \to \mathcal{U}$. A number $n : \mathbb{N}$ is a **lower bound** for P if it comes equipped with a term in the type

$$\text{is-lower-bound}_P(n) \coloneqq \Pi_{k:\mathbb{N}}P(k) \to n \leq k$$

Similarly,

$$\text{is-uppwer-bound}_p(n) \coloneqq \Pi_{k:\mathbb{N}}P(k) \to k \leq n$$

Theorem (well-ordering principle). *Let P be a decidable family oer $\mathbb{N}$ with d a witness that P is decidable. Then there is a function*

$$w(P, d) : (\Sigma_{n:\mathbb{N}}P(n)) \to (\Sigma_{m:\mathbb{N}}P(m) \times \text{is-lower-bound}_P(m)).$$

In other words, if $P(n)$ is inhabited for some n then there is a smallest $m : \mathbb{N}$ so that $P(m)$ is inhabited.

Proof. We will show that for any decidable type family $Q : \mathbb{N} \to \mathcal{U}$ that there is a function

$$Q(n) \to \Sigma_{m:\mathbb{N}} Q(m) \times \text{is-lower-bound}_Q(m)$$

by induction on n. When $n = 0$ we can use $m = 0$ since 0 is always a lower bound. For the inductive step we may assume we have the displayed function for every type family Q and consider a decidable type family Q with a term $q : Q(n+1)$. Our goal is to construct a term in the type

$$\Sigma_{m:\mathbb{N}} Q(m) \times \text{is-lower-bound}_Q(m)$$

. Since $Q(0)$ is decidable it suffices to construct a function

$$Q(0) + \neg Q(0) \to \Sigma_{m:\mathbb{N}} Q(m) \times \text{is-lower-bound}_Q m$$

so we can do a case analysis. If we have $Q(0)$ then it follows immediately that $m = 0$ is minimal. If $\neq Q(0)$, then we can consider the decidable family $Q' : \mathbb{N} \to \mathcal{U}$ defined by $Q(n) :\equiv Q'(\mathsf{succ}_{\mathbb{N}}(n))$. Since $q : Q'(n)$ we get a minimal element m for Q' by the inductive hypothesis. But since $Q(0)$ is assumed to be false then $m + 1$ is the minimal element for Q. □

The infinitude of primes. For natural numbers d and n we say d is a **proper divisor** of n if it comes with a term in the type

$$\text{is-proper-divisor}(n, d) :\equiv (d \neq n) \times (d \mid n)$$

With this notation we can say a natural number n is **prime** if it comes with a term in the type

$$\text{is-prime}(n) :\equiv \Pi_{x:\mathbb{N}} \text{is-proper-divisor}(n, x) \leftrightarrow (x = 1)$$

The proof of the infinitude of primes proceeds by constructing a prime number larger than n for any $n : \mathbb{N}$. So we can consider the type family for $n, m : \mathbb{N}$

$$R(n, m) :\equiv (n < m) \times \Pi_{x:\mathbb{N}} (x \leq n) \to (x \mid m) \to (x = 1)$$

of pairs so that m is greater than n and m is relatively prime to all numbers $x \leq n$. Since $n! + 1$ satisfies these properties for any n, the type family $m \mapsto R(n, m)$ in context $n : \mathbb{N}$ is inhabited. Thus, by the well-ordering principle, it has a least element p and this p must be prime.

Using the results we skipped it's possible to prove:

Lemma. *The type $R(n, m)$ is decidable for each $n, m : \mathbb{N}$.*

We leave it to the reader to verify that $R(n, n! + 1)$ is inhabited. Using these ingredients, we prove the infinitude of primes in the following form:

Theorem. *For each n, there is a prime number $p : \mathbb{N}$ so that $n < p$.*

Proof. It suffices to show this for each non-zero n since the case $n = 0$ follows. So let n be a non-zero natural number.

Since the type $R(n, m)$ is decidable for each m and since $R(n, n! + 1)$ holds by the well-ordering principle there is a minimal $p : \mathbb{N}$ so that $R(n, p)$. We will show that p is prime by constructing a term in the type

$$\text{is-prime}(p) :\equiv (p \neq 1) \times \Pi_{x:\mathbb{N}} \text{is-proper-divisor}(p, x) \to (x = 1).$$

By construction, $n < p$ and n is non-zero so $p \neq 1$. So now let x be a proper divisor of p. Since $R(n, p)$ holds by construction we can show that $x = 1$ by proving that $x \leq n$. Since p is non-zero and $x \mid p$ we must have $x < p$. By minimality of p it follows that $\neg R(n, x)$ holds. However, any divisor of x must also divide p by transitivity of divisibility, so

$$\Pi_{y:\mathbb{N}} (y \leq n) \to (y \mid x) \to (y = 1).$$

Since

$$\neg R(n, x) \doteq \neg \Big((n < x) \times \Pi_{y:\mathbb{N}} (y \leq n) \to (y \mid x) \to (y = 1) \Big)$$

holds we conclude that $\neg(n < x)$. On account of the logical equivalence $\neg(n < x) \leftrightarrow (x \leq n)$, it follows that $x \leq n$. □

Part 2. The Univalent Foundations of Mathematics

The univalent foundations build's on Martin-Löf's dependent type theory by adding a few new axioms inspired by the interpretation of types as ∞-groupoids or spaces rather than sets. One of these axioms, Voevodsky's *univalence axiom*, is inconsistent with the interpretation of types as sets. When we meet the hierarchy of truncation levels, we'll see that some types behave like types and some types behave like sets while others have non-trivial higher-dimensional structures.

Recall the notion of **logical equivalence** between types A and B

$$A \leftrightarrow B \coloneqq (A \to B) \times (B \to A).$$

For instance, you proved on your homework that bool and $\mathbb{1} + \mathbb{1}$ are logically equivalent by constructing functions

$$\begin{array}{ccc} \texttt{bool} \xrightarrow{\texttt{bool-to-1+1}} \mathbb{1} + \mathbb{1} & & \mathbb{1} + \mathbb{1} \xrightarrow{\texttt{1+1-to-bool}} \texttt{bool} \\ \texttt{true} \longmapsto \texttt{inl}(\star) & & \texttt{inl}(\star) \longmapsto \texttt{true} \\ \texttt{false} \longmapsto \texttt{inr}(\star) & & \texttt{inr}(\star) \longmapsto \texttt{bool} \end{array}$$

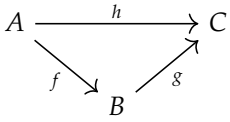
These inverse implications are obviously related. Our next goal is to develop language to explain how.

Homotopies. Surprisingly, we are not able, in dependent type theory, to construct identifications between functions $\texttt{1+1-to-bool} \circ \texttt{bool-to-1+1}$ and $\texttt{id}_{\text{bool}}$ or between $\texttt{succ}_{\mathbb{N}}$ and $\lambda n, n+1$. We can, however, construct *pointwise identifications* between these pairs of functions, which in homotopy type theory are commonly called **homotopies**:

defn. Let $f, g : \prod_{x:A} B(x)$. The type $f \sim g$ of **homotopies** from f to g is defined to be the type of pointwise identifications:

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

Remark. Given three functions as below



we say the triangle **commutes** if $h \sim g \circ f$.

Note if $H, K : f \sim g \coloneqq \prod_{x:A} f(x) = g(x)$ are homotopies we can treat them as dependent functions in their own right and consider **homotopies between homotopies**, i.e., terms of the type $H \sim K \coloneqq \prod_{x:A} H(x) = K(x)$. This continues all the way up.

Indeed, the ∞-groupoid structure of identity types extends to homotopies as follows:

defn. For any type family $B \colon A \to \mathcal{U}$ we have operations

$$\texttt{refl-htpy} : \prod_{f:\prod_{x:A} B(x)} f \sim f \quad \texttt{inv-htpy} : \prod_{f,g:\prod_{x:A} B(x)} (f \sim g) \to (g \sim f)$$

$$\texttt{concat-htpy} : \prod_{f,g,h:\prod_{x:A} B(x)} f \sim g \to g \sim h \to f \sim h$$

defined pointwise by

$$\texttt{refl-htpy}(f) = \lambda x.\texttt{refl}_{f(x)} \quad \texttt{inv-htpy}(H) \coloneqq \lambda x.H(x)^{-1} \quad \texttt{concat-htpy}(H, K) \coloneqq \lambda x.H(x) \cdot K(x)$$

We abbreviate these latter two terms by writing H^{-1} and $H \cdot K$.

Note we don't abbreviate $\texttt{refl-htpy}_f : f \sim f \coloneqq \prod_{x:A} f(x) =_{B(x)} f(x)$ as $\texttt{refl}_f : f =_{\prod_{x:A} B(x)} f$ as these terms live in different types.

Proposition. *Homotopies satisfy the groupoid laws up to homotopy:*

(i) *There is a homotopy* $\texttt{assoc-htpy}(H, K, L) : (H \cdot K) \cdot L \sim H \cdot (K \cdot L)$ *for any homotopies* $H : f \sim g$, $K : g \sim h$, $L \colon h \sim i$.
(ii) *There are homotopies* $\texttt{left-unit-htpy}(H) : \texttt{refl-htpy}_f \cdot H \sim H$ *and* $\texttt{right-unit-htpy}(H) : H \cdot \texttt{refl-htpy}_g \sim H$.
(iii) *There are homotopies* $\texttt{left-inv-htpy}(H) : H^{-1} \cdot H \sim \texttt{refl-htpy}_g$ *and* $\texttt{right-inv-htpy}(H) \colon H \cdot H^{-1} \sim \texttt{refl-htpy}_f$.

In addition to the groupoid operations we make use of the following **whiskering operations** on homotopies which are relevant in the setting of functions

$$A \xrightarrow{f} B \underset{h}{\overset{g}{\rightrightarrows}} C \xrightarrow{k} D$$

defn. Given the functions above and a homotopy $H\colon g \sim h$ define homopies

$$H \circ f \coloneqq \lambda a.H(f(a))\colon g \circ f \sim h \circ f$$

and

$$k \circ H \coloneqq \lambda_b.\mathsf{ap}_k(H(b))\colon k \circ g \sim k \circ h.$$

Bi-invertible maps. We now explain what it means for a function $f\colon A \to B$ between types to define an **equivalence**. On explanation is that an equivalence is like a homotopy equivalence in topology, but we actually define the type family is-equiv in a manner that makes equivalences look more like "bi-invertible maps" for reasons that we will explain.

defn. Let $f\colon A \to B$.

(i) The type of **sections** of f is defined to be the type

$$\mathrm{sec}(f) \coloneqq \Sigma_{g:B\to A} f \circ g \simeq \mathtt{id}_B.$$

(ii) The type of **retractions** of f is defined to be the type

$$\mathrm{retr}(f) \coloneqq \Sigma_{h:B\to A} h \circ f \simeq \mathtt{id}_A.$$

(iii) We say f is an **equivalence** if it has a section and a retraction:

$$\text{is-equiv}(f)\colon \mathrm{sec}(f) \times \mathrm{retr}(f).$$

We write $A \simeq B$ for the type $\Sigma_{f:A\to B}\text{is-equiv}(f)$ of all equivalences from A to B.

Explicitly the data of a term in is-equiv(f) involves two maps $g, h : B \to A$ and two homotopies $G : f \circ g \simeq \mathtt{id}_B$ and $H\colon h \circ f \simeq \mathtt{id}_A$. We have seen a bunch of examples:

ex.

- Identity functions are equivalences.
- `bool-to-1+1`$\colon$ bool $\to \mathbb{1} + \mathbb{1}$ and `1+1-to-bool`$\colon \mathbb{1} + \mathbb{1} \to$ bool are inverse equivalences.
- `neg-bool`$\colon$ bool $\to$ bool is an equivalence.
- `pred`$\colon \mathbb{Z} \to \mathbb{Z}$ is an equivalence, inverse to `succ`$\colon \mathbb{Z} \to \mathbb{Z}$.

We might also define

$$\text{has-inverse}(f) \coloneqq \Sigma_{g:B\to A}(f \circ g \simeq \mathtt{id}_B) \times (g \circ f \simeq \mathtt{id}_A).$$

The reason we did not define equivalences to be functions that have inverses is that we wanted being an equivalence to be a **property** associated to a map f rather than a **structure** on the map, in a sense we'll discuss soon. Essentially because of the uniqueness of reflexivity paths, if f is an equivalence then it only has one section and homotopy and only one retraction and homotopy, up to homotopy. We'll see, however, that the data of inverses to f can be more complicated.

But even though is-equiv(f) and has-inverse(f) may differ:

Proposition. *The types* is-equiv(f) *and* has-inverse(f) *are logically equivalent. In particular, if f is an equivalence it has a two-sided inverse.*

Proof. The data of an inverse obviously defines the data of an equivalence, so we have has-inverse$(f) \to$ is-equiv(f).

For the converse, suppose we have $g, h\colon B \to A$ and homotopies $G : f \circ g \simeq \mathtt{id}_B$ and $H\colon h \circ f \simeq \mathtt{id}_A$. Then for any $b : B$ we have

$$g(b) \overset{H(g(b))^{-1}}{=\!=} h(f(g(b))) \overset{\mathsf{ap}_h(G(b))}{=\!=} h(b)$$

defining a homotopy $K\colon g \simeq h$. Using this we can see that g is also a retraction of f by the identification for any $a : A$

$$g(f(a)) \overset{K(f(a))}{=\!=} h(f(a)) \xrightarrow{H(a)} a$$

□

Corollary. *A section or retraction of an equivalence is an equivalence.*

Proof. We've just seen that the section of an equivalence is also a retraction and thus is an invertible map with inverse f. A dual construction applies to a retraction. □

Up to equivalence, types satisfy many familiar laws, such as

$$A + B \simeq B + A \qquad A \times B \simeq B \times A$$
$$A \times \mathbb{1} \simeq A \qquad A + \varnothing \simeq A$$
$$A \times (B + C) \simeq (A \times B) + (A \times C$$

where the required equivalences and homotopies can be constructed by induction. These equivalences extend to cases such as

$$\Sigma_{x:\varnothing} B(x) \simeq \varnothing \quad \Sigma_{x:\mathbb{1}} B(x) \simeq B(\star).$$

and

$$\Sigma_{z:\Sigma_{x:A} B(x)} C(z) \simeq \Sigma_{x:A} \Sigma_{y:B(x)} C(x,y) \quad \Sigma_{w:A+B} C(w) \simeq \Sigma_{x:A} C(\mathtt{inl}x) + \Sigma_{y:B} C(\mathtt{inr}y).$$

In the presence of an additional axiom, we'll be able to prove similar equivalences involving function types, but we can't deduce these just yet.

Characterizing the identity types of Σ-types. We return to the theme of characterizing identity types to characterize the identity type of a Σ-type, up to equivalence. We follow the same outline used to characterize the identity type family of the natural numbers above:

(i) We define a binary relation $R\colon A \to A \to \mathcal{U}$ that we suspect is equivalent to the identity type family.
(ii) Prove reflexivity by constructing a term in $\Pi_{x:A} R(x,x)$.
(iii) Use reflexivity and path induction to define a canonical map $\Pi_{x,y:A} x =_A y \to R(x,y)$.
(iv) Show that the map $x =_A y \to R(x,y)$ is an equivalence for each $x, y : A$.

The last step is usually the most difficult and we will refine our techniques for dealing with it soon.

In this section, we consider $B\colon A \to \mathcal{U}$ and the corresponding type $\Sigma_{x:A} B(x)$. Given dependent pairs (a,b) and (a',b') we might imagine that the data of an identification between them involves a pair of identifications comprised firstly of a path $p : a =_A a'$ and then of a path $q : \mathtt{tr}_B(p,b) = b'$. We first turn this idea into a binary relation.

defn. Define

$$\mathrm{Eq}_\Sigma\colon (\Sigma_{x:A} B(x)) \to (\Sigma_{x:A} B(x)) \to \mathcal{U}$$

by

$$\mathrm{Eq}_\Sigma(s,t) :\equiv \Sigma_{p:\mathtt{pr}_1(s)=\mathtt{pr}_1(t)} \mathtt{tr}_B(p, \mathtt{pr}_2(s)) = \mathtt{pr}_2(t).$$

Lemma. *The relation* Eq_Σ *is reflexive.*

Proof. By Σ-induction it suffices to let $s : \Sigma_{x:A} B(x)$ be a pair $(a : A, b : B(a))$ in which case we have

$$\lambda a, \lambda b.(\mathtt{refl}_a, \mathtt{refl}_b)\colon \Pi_{x:A} \Pi_{y:B(x)} \Sigma_{p:x=x} \mathtt{tr}_B(p,y) = y$$

since we have a definitional equality $\mathtt{tr}_B(\mathtt{refl}_a, b) \doteq b$. □

By path induction, we may define a map $\mathtt{pair\text{-}eq} : s = t \to \mathrm{Eq}_\Sigma(s,t)$ for any $s, t : \Sigma_{x:A} B(x)$ by sending $\mathtt{refl}_s$ to the corresponding reflexivity term.

Theorem. *For any type family* $B\colon A \to \mathcal{U}$*, the map*

$$\mathtt{pair\text{-}eq}\colon (s = t) \to \mathrm{Eq}_\Sigma(s,t)$$

is an equivalence for every $s, t : \Sigma_{x:A} B(x)$*.*

Proof. We define an inverse equivalence $\mathtt{eq\text{-}pair}\colon \mathrm{Eq}_\Sigma(s,t) \to (s = t)$ by repeated Σ-induction. For pairs (a,b) and (a',b') we must define

$$\mathtt{eq\text{-}pair}\colon \Sigma_{p:a=a'} \mathtt{tr}_B(p,b) = b' \to ((a,b) = (a',b'))$$

which we again do by Σ-induction. It suffices to define a function in the type

$$\Pi_{p:a=a'} (\mathtt{tr}_B(p,b) = b') \to ((a,b) = (a',b')).$$

By double path induction we may first assume b' is $\mathtt{tr}_B(p,b)$ and the second path is $\mathtt{refl}$. Then by path induction we may assume that a' is a and p is refl in which case we send this pair $(\mathtt{refl}_a, \mathtt{refl}_b)$ to $\mathtt{refl}_{(a,b)} : ((a,b) = (a,b))$.

To see that $\mathtt{eq\text{-}pair}$ is a section of $\mathtt{pair\text{-}eq}$ we require identifications

$$\mathtt{pair\text{-}eq}(\mathtt{eq\text{-}pair}(p,q)) = (p,q)$$

for each $(p,q) : \Sigma_{p:a=a'} \mathtt{tr}_B(p,b) = b'$. By a double path induction it suffices to identify

$$\mathtt{pair\text{-}eq}(\mathtt{eq\text{-}pair}(\mathtt{refl}_a, \mathtt{refl}_b)) = (\mathtt{refl}_a, \mathtt{refl}_b)$$

and the left-hand side is judgmentally equal to the right-hand side, so we may use $\mathtt{refl}_{(\mathtt{refl}_a, \mathtt{refl}_b)}$.

To see that `eq-pair` is a retraction of `pair-eq` we require identifications

$$\mathtt{eq\text{-}pair}(\mathtt{pair\text{-}eq}(p)) = p$$

for each $p : s = t$. By path induction we may take t to be s and p to be $\mathtt{refl}_s$. Then we require

$$\mathtt{eq\text{-}pair}(\mathtt{refl}_{\mathtt{pr}_1(s)}, \mathtt{refl}_{\mathtt{pr}_2(t)}) = \mathtt{refl}_s.$$

By Σ-induction we may consider the case of a pair (a,b) in which case we require

$$\mathtt{eq\text{-}pair}(\mathtt{refl}_a, \mathtt{refl}_b) = \mathtt{refl}_{(a,b)}.$$

Since the left-hand side we defined to be the right-hand side, we may use `refl`. □

October 6: Contractibility

Contractible types are singletons up to homotopy. In this section, we'll see several characterizations of contractibility and then relativize our discussion to include so-called "contractible maps."

Contractible types. The term "contractibility" is inspired by the homotopical interpretation of type theory but it can also be thought of as an expression of "uniqueness": we say that a type A is contractible just when it contains a unique element in a sense encoded by the Curry-Howard interpretation:

defn. A type A is **contractible** if it comes equipped with a term in the type

$$\text{is-contr}(A) \coloneqq \Sigma_{c:A} \Pi_{x:A} c = x.$$

Given $(c, C) : \text{is-contr}(A)$ we refer to $c : A$ as the **center of contraction** and $C : \Pi_{x:A} c = x$ as the **contraction** or **contracting homotopy**.

Q. This terminology suggests that C is a homotopy between some pair of functions. Which functions?

ex. The unit type is easily seen to be contractible. We take $\star : \mathbb{1}$ as the center of contraction and define $C : \Pi_{x:\mathbb{1}} \star = x$ by singleton induction: in the case where x is $\star$, we define $C(\star) = \mathtt{refl}_\star$.

We've met another example of a contractible type already.

Theorem. *For any $a : A$ the type*

$$\Sigma_{x:A} a = x$$

is contractible.

Proof. We take $(a, \mathtt{refl}_a) : \Sigma_{x:A} a = x$ to be the center of contraction. In our proof of the uniqueness of reflexivity we constructed the required contracting homotopy. □

Singleton induction. Contractible types satisfy an induction principle similar to singleton types.

defn. Suppose A comes with a term $a : A$. Then we say A satisfies **singleton induction** if for every type family B over A the map

$$\mathtt{ev\text{-}pt} : (\Pi_{x:A} B(x)) \to B(a)$$

defined by $\mathtt{ev\text{-}pt} f \coloneqq f(a)$ has a section. The data of this section is then given by a function and a homotopy

$$\mathtt{ind\text{-}sing}_a : B(a) \to \Pi_{x:A} B(x) \quad \text{and} \quad \mathtt{comp\text{-}sing}_a : \mathtt{ev\text{-}pt} \circ \mathtt{ind\text{-}sing}_a \simeq \mathtt{id}.$$

The singleton induction principle is almost the same as the induction principle for the unit type except for one point of difference: the "computation rule" for singleton induction is stated using an *identification* rather than a judgmental equality. Note in particular, that $\star : \mathbb{1}$ satisfies singleton induction with $\lambda x.\mathtt{refl}_x$ as the required homotopy.

Theorem. *For a type A the following are logically equivalent.*

(i) A is contractible.
(ii) A comes equipped with a term $a : A$ satisfying singleton induction.

Proof. Suppose we have $(a, C) : \text{is-contr}(A) :\equiv \sum_{c:A} \prod_{x:A} c = x$. We may replace the homotopy $C : \mathtt{const}_a \sim \mathtt{id}_A$ by a new homotopy C' defined by $C'(x) = C(a)^{-1} \cdot C(x)$. We then have an identification $\mathtt{left\text{-}inv} : C'(a) = \mathtt{refl}_a$. So without loss of generality we suppose $(a, C) : \text{is-contr}(A)$ comes with a term $p : C(a) = \mathtt{refl}_a$.

Consider a type family B over A. For each $b : B(a)$ our goal is to define $\mathtt{ind\text{-}sing}_a(b) : \prod_{x:A} B(x)$, which we do by

$$\mathtt{ind\text{-}sing}_a(b) :\equiv \mathtt{tr}_B(C(x), b) : B(x).$$

It remains to construct an identification $\mathtt{ind\text{-}sing}_a(b, a) = b$. We have a function

$$\lambda(q : a = x).\mathtt{tr}_B(q, b)$$

that evaluates at the path $C(a)$ to give $\mathtt{ind\text{-}sing}_a(b)$ and evalutes at the path $\mathtt{refl}_a$ to give b. When we apply this function to the path $p : C(a) = \mathtt{refl}_a$ we get the desired identification.

Conversely, suppose $a : A$ satisfies singleton induction. To prove that A is contractible with center of contraction a we apply singleton induction to the type family $B(x) :\equiv a = x$ to obtain

$$\mathtt{ind\text{-}sing}_a : a = a \to \Pi_{x:A} a = x.$$

Now $\mathtt{ind\text{-}sing}_a(\mathtt{refl}_a)$ is a contracting homotopy. □

Fibers.

defn. Let $f : A \to B$ be a function between types and consider $b : B$. The **fiber** of f over b is the type

$$\text{fib}_f(b) :\equiv \Sigma_{a:A} f(a) = b.$$

That is the fiber is what in other contexts might get called the "homotopy fiber": it is comprised of those terms $a : A$ whose images $f(a) : B$ are identified with b.

It will be useful to identify the identity type of the fiber, which we do by following our usual strategy.

defn. Let $f : A \to B$ and $b : B$. Given two terms $(a, p), (a', p') : \text{fib}_f(b)$ define a type

$$\text{Eq-fib}_f((a, p), (a', p')) :\equiv \Sigma_{\alpha:a=a'} p = \mathtt{ap}_f(\alpha) \cdot p'.$$

Note $\text{Eq-fib}_f : \text{fib}_f(b) \to \text{fib}_f(b) \to \mathcal{U}$ is reflexive since we have

$$\lambda(a, p).(\mathtt{refl}_a, \mathtt{refl}_p) : \Pi_{(a,p):\text{fib}_f(b)} \text{Eq-fib}_f((a, p), (a, p)).$$

Proposition. *Consider $f : A \to B$ and $b : B$. The canonical map*

$$(a, p) =_{\text{fib}_f(b)} (a', p') \to \text{Eq-fib}_f((a, p), (a', p'))$$

induced by the reflexivity term is an equivalence for any pair of points $(a, p), (a', p') : \text{fib}_f(b)$.

Proof. For the inverse equivalence, we need a function

$$\left(\Sigma_{\alpha:a=a'} p = \mathtt{ap}_f(\alpha) \cdot p'\right) \to (a, p) =_{\text{fib}_f(b)} (a', p').$$

By Σ-induction it suffices to consider the image of a pair $\alpha : a = a'$ and $\beta : p = \mathtt{ap}_f(\alpha) \cdot p'$. By path induction we may take p to be $\mathtt{ap}_f(\alpha) \cdot p'$ and β to be $\mathtt{refl}$. By path induction again, we may take a to be a' and α to be $\mathtt{refl}$. Since $\mathtt{ap}_f(\mathtt{refl}) \cdot p'$ is judgmentally equal to p' we may use $\mathtt{refl}$ as our identification between $(a', \mathtt{ap}_f(\mathtt{refl}) \cdot p')$ and (a', p'). The construction of the homotopies witnessing the left and right inverses are similar. □

Contractible maps and equivalences.

defn. We say that a function $f : A \to B$ is **contractible** if it comes equipped with a term of type

$$\text{is-contr}(f) :\equiv \Pi_{b:B} \text{is-contr}(\text{fib}_f(b)).$$

Theorem. *Any contractible map is an equivalence.*

Proof. Suppose $f : A \to B$ is contractible. The center of contraction $(g(b), G(B))$ in each fiber gives rise to a dependent function

$$\lambda b.(g(b), G(B)) : \Pi_{b:B} \mathrm{fib}_f(b).$$

In particular $g : B \to A$ defines a map and $G : f \circ g \sim \mathtt{id}_B$ defines a homotopy. So g is a section of f.

It remains to show that g is also a retraction of f by defining a term in the type $g \circ f \sim \mathtt{id}_A$. For each $a : A$ we have an identification $p : f(g(f(a))) = f(a)$ since g is a section of f. So $(g(f(a)), p) : \mathrm{fib}_f(f(a))$. Since this type is contractible there is an identification $\beta : (g(f(a)), p) = (a, \mathtt{refl}_{f(a)})$. The base path $\mathtt{ap}_{\mathtt{pr}_1}(\beta) : g(f(a)) = a$ gives the desired identification. □

In fact, equivalences necessarily also define contractible maps, which we show in a few steps. Recall an **invertible map** is a map $f : A \to B$ equipped with $g : B \to A$ and homotopies $G : f \circ g \sim \mathtt{id}$ and $H : g \circ f \sim \mathtt{id}$. Then the whiskered homotopies $G \circ f$ and $f \circ H$ both have the $f \circ g \circ f \simeq f$.

defn. A map $f : A \to B$ is **coherently invertible** if it comes with

$$g : B \to A, \quad G : f \circ g \sim \mathtt{id}, \quad H : g \circ f \sim \mathtt{id}, \quad K : G \circ f \sim f \circ H.$$

Proposition. *Any coherently invertible map has contractible fibers.*

Proof. Given $f : A \to B$ together with

$$g : B \to A, \quad G : f \circ g \sim \mathtt{id}, \quad H : g \circ f \sim \mathtt{id}, \quad K : G \circ f \sim f \circ H.$$

we wish to show that $\mathrm{fib}_f(b)$ is contractible for any $b : B$. We take $(g(b), G(b)) : \mathrm{fib}_f(b)$ as the center of contraction.

For the contracting homotopy it suffices to define a term in the equivalent type

$$\prod_{a:A} \prod_{p:f(a)=b} \text{Eq-fib}_f((g(b), G(B)), (a, p).$$

By path induction on p it suffices to define a term in the type

$$\prod_{a:A} \text{Eq-fib}_f((g(f(a)), G(f(a))), (a, \mathtt{refl}_{f(a)})).$$

By the definition of Eq-fib_f this means that we need, for each $a : A$, a path $H(a) : g(f(a)) = a$ and a further identification

$$G(f(a)) = \mathtt{ap}_f(H(a)) \cdot \mathtt{refl}_{f(a)}.$$

We have $K(a) : G(f(a)) = \mathtt{ap}_f(H(a))$ so we get the identification we want by composing with `right-unit-htpy`. □

Our next goal is to show that any invertible map $f : A \to B$ equipped with $g : B \to A$ and homotopies $G : f \circ g \sim \mathtt{id}$ and $H : g \circ f \sim \mathtt{id}$ can be improved to a coherently invertible map at the cost of replacing G with a new homotopy $G' : f \circ g \sim \mathtt{id}$ satisfying the coherence $K : f \circ H \sim G' \circ f$. The upshot is that we have a map

$$\text{has-inverse}(f) \to \text{is-coh-invertible}(f).$$

The construction is by messy path algebra that you can read about in [R, §10.4].

Corollary. *For any type A and term $a : A$ the type*

$$\Sigma_{x:A} x = a$$

is contractible.

Proof. This type is the fiber of the identity function $\mathtt{id}_A : A \to A$ over $a : A$. Since $\mathtt{id}_A$ is an equivalence, this type must be contractible. □

Soon we'll have a second proof: we'll be able to use the equivalence $\mathtt{inv} : (a = x) \to (x = a)$ to define an equivalence $\lambda x.\mathtt{inv} : \Sigma_{x:A} a = x \to \Sigma_{x:A} x = a$. It follows easily that any type equivalent to a contractible type is contractible.

References

[R] Egbert Rijke, *Introduction to Homotopy Type Theory*, available from https://hott.zulipchat.com

Dept. of Mathematics, Johns Hopkins University, 3400 N Charles St, Baltimore, MD 21218
E-mail address: eriehl@math.jhu.edu