Definitional Univalence in Cubical Type Theory

(work in progress)

Hugo Herbelin

Travail en commun avec Hugo Moeneclaey

Journées Logique, Homotopie, Catégories 2019

Lyon, 16 Octobre 2019

Two approaches to computing with equality

Parametric polymorphic equality

- Equality as smallest reflexive relation (Martin-Löf's type theory)
- $J_P\, \mathsf{refl}\, t \equiv t$

Ad hoc polymorphic equality

- Each type has its own equality (observational type theory, cubical type theory)
- Transport/substitutivity defined by (meta-level) induction on the type structure
  example: $J_{\lambda x.\lambda q.(A_1 \times A_2)}\, p\, (t_1, t_2) \equiv (J_{\lambda x.\lambda q.A_1}\, p\, t_1, J_{\lambda x.\lambda q.A_2}\, p\, t_2)$
- Supports extensionality
- Requires some techniques to deal with dependency and contravariance; better done using *equality over*

Equality as path

Cubical Type Theory reinterprets equality as a path over a formal interval

- Postulate a formal interval $\mathbb{I} \triangleq$ "$[0;1]''$ and treat equality as if characterised by

$$t =_A u \;\triangleq\; \{f : \mathbb{I} \to A \mid f0 \equiv t \wedge f1 \equiv u\}$$

- This notion of equality generalises into a (cubical) "equality over": $t =_\epsilon u$ depends on a proof $\epsilon : A = B$ (i.e. itself $\epsilon : \mathbb{I} \to \mathsf{U}$) stating that the type A of t is equal to the type B of u

$$t =_\epsilon u \;\triangleq\; \{f : (\Pi i : \mathbb{I}.\epsilon i) \mid f0 \equiv t \wedge f1 \equiv u\}$$

Our analysis of the contributions of Cubical Type Theory

- It decomposes equality as a path: abstraction/application allows to enter or conceal dimensions and reason within these dimensions.

  This provides functoriality (at all dimensions) and function extensionality which otherwise would have to be expressed by proper combinators.

  This can (a posteriori) be seen as *iterated parametricity* in *direct style*.

- It introduces *equality over* as a "consistent" heterogeneous equality (compare to Observational Type Theory which uses John Major equality).

  This allows to internalise a *cubical* geometrical shape in type theory (which otherwise is globular)[1].

- It provides a (Kan) box composition/filling structure which extends transport/substitutivity (together with specific definitional rules).

- An extra "gluing" operation provides *univalence*.

[1]This cubical structure can natively be equipped with algebraic structure echoing to logical structural rules: contraction (cartesian structure with diagonals), exchange (symmetric group of permutation), as well as symmetry (providing inverses called reversals), connections (for oblique commutative diagrams); this structure can be given either by term combinators or by interval combinators (e.g. one gets inverse either by adding a term operation p^{-1} or by adding an interval operation $-i$). There is also a room of manoeuvre about which properties of this structure is definitional (for instance, one would like $(p^{-1})^{-1} \equiv p$, resp. $--i = i$).

Our own approach of Cubical Type Theory

- Equality on types is *defined* to be equivalence
- Equivalence is enough to provide the substitutivity/transport/composition/filling structure

  This structure is "minimalistic" and we believe it is definitionally compatible with the rule $J_P\,\textsf{refl}\,t \equiv t$
- It is aimed to be iterated univalent parametricity in direct style and we inherit definitional rules from it
- In particular, abstraction/application over a variable in the formal interval are seen as operations

Core equality structure in Cubical Type Theory

Syntax

$$t, A, p, \epsilon \ ::= \ \dots \mid t =_\epsilon u \mid \lambda i.t \mid p\,i$$

Typing rules

$$\frac{\Gamma \vdash \epsilon : A =_{\lambda i.\mathsf{U}_n} B \qquad \Gamma \vdash t : A \qquad \Gamma \vdash u : B}{\Gamma \vdash t =_\epsilon u \ : \ \mathsf{U}_n}$$

$$\frac{\Gamma, i \vdash t : A}{\Gamma \vdash \lambda i.t : t[0/i] =_{\lambda i.A} t[1/i]} \qquad \frac{\Gamma \vdash v : t =_\epsilon u \qquad i \in \Gamma}{\Gamma \vdash v\,i : \epsilon\,i}$$

where the key steps of $t[0/i]$ and $t[1/i]$ are $(pi)[0/i] \triangleq t$ and $(pi)[1/i] \triangleq u$ whenever $p : t =_\epsilon u$.

Conversion rules

$$\frac{\Gamma \vdash p : t =_\epsilon u \qquad i \text{ fresh}}{\Gamma \vdash \lambda i.(p\,i) \equiv p : t =_\epsilon u} \qquad \frac{\Gamma, i \vdash t : A \qquad j \in \Gamma}{\Gamma \vdash (\lambda i.t)\,j \equiv t[j/i] : A[j/i]}$$

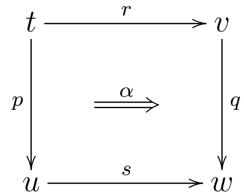
This is considered on top of an ambiant type theory with U_n, $\Sigma a : A.B$, $\Pi a : A.B$, ...

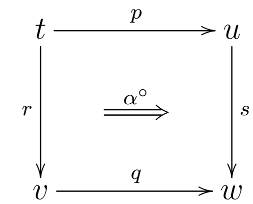
Examples

- *reflexivity*: $\widehat{t} \triangleq \lambda i.t$, for i fresh and t of type A, shall represent a proof of $t =_{\widehat{A}} t$
- *functoriality*: if $f : A \to B$ and $p : t =_{\widehat{A}} u$ then $\lambda i.f(pi)$ is a proof of $ft =_{\widehat{B}} fu$
- *dependent functoriality*: if $f : \Pi a : A.B$ and $p : t =_{\widehat{A}} u$ then $\lambda i.f(pi)$ is a proof of $ft =_{\lambda i.B[pi/a]} fu$
- *functional extensionality* trivially provable (swap term variable with direction variable): if $p : \Pi a\!:\!A.\,(f_0\,a =_{\lambda i.B} f_1\,a)$ then $\lambda ia.pa\,i : f_0 =_{\lambda i.\Pi a:A.\,B} f_1$

Further examples

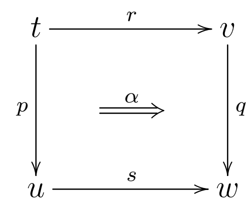
- commutation of sum with equality: if $p : t =_{\lambda i.\Sigma a:A.B} u$ then $\lambda i.\mathsf{snd}(pi)$ proves $\mathsf{snd}(t) =_{\lambda i.B[\mathsf{fst}(pi)/a]} \mathsf{snd}(u)$.
- nestings of equality have a *cubical* structure, stable by *permutation*

e.g. if $\alpha : p \underset{r \,\widetilde{=}_E\, s}{=} q$ (geometrically 

), then $\alpha^\circ \triangleq \lambda ij.\alpha\, j i : r \underset{p \,\widetilde{=}_{E^\circ}\, q}{=} s$ (

)

where we used the abbreviation $v \,\widetilde{=}_\xi\, w \triangleq \lambda i.(v\, i =_{\xi\, i} w\, i)$.

- *diagonals*: if $\alpha : p \underset{r \,\widetilde{=}_E\, s}{=} q$ (geometrically 

) then $\Delta\alpha \triangleq \lambda i.\alpha i i$ proves $t =_{\Delta E} w$
- supports reasoning with equality *over* an equality without breaking the symmetry

$$v_1 =_{\lambda i.vect\ (p\ i)} v_2 \quad \text{whenever} \quad p : n_1 =_{\widehat{\mathbb{N}}} n_2$$

- appropriate to compute with *Higher Inductive Types* (HITs)

$$case\ t\ of\ \mathsf{base} \Rightarrow b \mid \mathsf{loop}\ i \Rightarrow l\ i\ end$$

Cubical equality encourages to reason by pointwise transport

Let $f : A \to A$ and $p : \forall a\, f(a) =_{\widehat{A}} a$. For $a : A$, let us prove that $f(pa) = p(fa)$ where $f(pa)$ is functorial application of f, i.e. $\lambda i. f(pai)$.

We need to find a "continuous" term q that evaluates into $p(fa)$ in 0 and in $f(p(a)$ in 1. To connect these terms, it is convenient to rephrase them into

$$\lambda i. f(p(\mathsf{id}\, a)i)$$

and

$$\lambda i.\mathsf{id}(p(fa)i)$$

(using η and β-expansions) so as to expose the similarity of structure. Then, for any t, the equation $ft \stackrel{?}{=} \mathsf{id}\, t$ unifies along the interval if we can find a term $?q'$ such that $ft \equiv ?q'0$ and $\mathsf{id}\, t \equiv ?q'1$. The solution is $?q' \triangleq pt$. Similarly, $\mathsf{id}\, t \stackrel{?}{=} ft$ unifies along the interval by setting $?q'' \triangleq \overline{pt}$ where $\overline{e}$ denotes a proof of $v = w$ whenever e proves $w = v$. It finally suffices to combine this into a unifier of the original problem:

$$\begin{array}{l}
?q\,0 \;=\; \lambda i.\; f\;\; (p\;\; (\mathsf{id}\; a\;\;\;)\;\; i) \\
?q\,1 \;=\; \lambda i.\; \mathsf{id}\; (p\;\; (f\;\; a\;\;\;)\;\; i) \\
\hline
?q\,j \;=\; \lambda i.\; p\;\; (p\;\; (\overline{p\; a}\;\; j)\;\; i)\;\; j
\end{array}$$

Hence $q \triangleq \lambda j.\lambda i.p(p(\overline{p\,a}\,j)\,i)\,j$

A symmetric definition of equivalence

(exercise 4.2 of the HoTT Book)

We extend the theory with a record type $A \simeq_n B$ (equivalence in universe U_n) defined as follows. If $\epsilon : A \simeq_n B$, the following projections are available:

$$
\begin{array}{lcl}
=_\epsilon & : & A \to B \to \mathsf{U}_n; \\
\overrightarrow{\epsilon} & : & A \to B; \\
\overset{\Rightarrow}{\epsilon} & : & \Pi a : A.\, a =_A \overleftarrow{\epsilon}\, \overrightarrow{\epsilon}\, a; \\
\overrightarrow{\mathsf{coe}}_\epsilon & : & \Pi a : A. \Pi b : B.\, a =_\epsilon b \to \overrightarrow{\epsilon}\, a =_B b \\
\overrightarrow{\overrightarrow{\mathsf{coe}}}_\epsilon & : & \Pi a : A. \Pi b : B. \Pi p : a =_\epsilon b.\, \overset{\Rightarrow}{\epsilon}\, a =_{\lambda i.(a =_\epsilon \overrightarrow{\mathsf{coe}}_\epsilon(p) i)} p \\
\overleftarrow{\epsilon} & : & B \to A; \\
\overset{\Leftarrow}{\epsilon} & : & \Pi b : B.\, \overrightarrow{\epsilon}\, \overleftarrow{\epsilon}(b) =_B b; \\
\overleftarrow{\mathsf{coe}}_\epsilon & : & \Pi a : A. \Pi b : B.\, a =_\epsilon b \to a =_A \overleftarrow{\epsilon}\, b \\
\overleftarrow{\overleftarrow{\mathsf{coe}}}_\epsilon & : & \Pi a : A. \Pi b : B. \Pi p : a =_\epsilon b.\, p =_{\lambda i.(\overleftarrow{\mathsf{coe}}_\epsilon(p) i =_\epsilon b)} \overset{\Leftarrow}{\epsilon}(b)
\end{array}
$$

In particular, setting $(A =_{\widehat{\mathsf{U}_n}} B) \triangleq (A \simeq_n B)$, substitutivity shall become a consequence of $t =_\xi u \to P(t) =_{\widehat{\mathsf{U}_n}} P(u)$

Excerpt of rules defining $\lambda i.A$ as a proof of equivalence

Excerpt of the semantics of $\widehat{\mathsf{U}_n}$:

$$(A =_{\widehat{\mathsf{U}_n}} B) \;\equiv\; A \simeq_n B \qquad \begin{array}{l} \overrightarrow{\widehat{\mathsf{U}_n}}\, A \;\equiv\; A \\ \overleftarrow{\widehat{\mathsf{U}_n}}\, B \;\equiv\; B \end{array} \qquad \begin{array}{l} \overset{\Longrightarrow}{\widehat{\mathsf{U}_n}}\, A \;\equiv\; \widehat{A} \\ \overset{\Longleftarrow}{\widehat{\mathsf{U}_n}}\, A \;\equiv\; \widehat{A} \end{array}$$

Excerpt of the semantics of $\lambda i.\Sigma a : A.B$:

$$\begin{array}{lcl}
(t =_{\lambda i.\Sigma a:A.B} u) & \equiv & \Sigma a : (\mathsf{fst}\, t =_{\lambda i.A} \mathsf{fst}\, u).(\mathsf{snd}\, t =_{\lambda i.B[\mathsf{fst}\,(ai)/a]} \mathsf{snd}\, u) \\
\overrightarrow{\lambda i.\Sigma a : A.B}\, t & \equiv & (\overrightarrow{\lambda i.A}(\mathsf{fst}\, t), \overrightarrow{\lambda i.B[\overset{\Longrightarrow}{\lambda i.A}\, a\, i/a]}(\mathsf{snd}\, t)) \\
\overleftarrow{\lambda i.\Sigma a : A.B}\, t & \equiv & (\overleftarrow{\lambda i.A}(\mathsf{fst}\, t), \overleftarrow{\lambda i.B[\overset{\Longleftarrow}{\lambda i.A}\, a\, i/a]}(\mathsf{snd}\, t)) \\
\overset{\Longrightarrow}{\lambda i.\Sigma a : A.B}\, t & \equiv & (\overset{\Longrightarrow}{\lambda i.A}\,(\mathsf{fst}\, t), \overset{\Longrightarrow}{\lambda i.B[\overset{\Longrightarrow}{\lambda i.A}\, a\, i/a]}\,(\mathsf{snd}\, t)) \\
\overset{\Longleftarrow}{\lambda i.\Sigma a : A.B}\, t & \equiv & (\overset{\Longleftarrow}{\lambda i.A}\,(\mathsf{fst}\, t), \overset{\Longleftarrow}{\lambda i.B[\overset{\Longleftarrow}{\lambda i.A}\, a\, i/a]}\,(\mathsf{snd}\, t))
\end{array}$$

And similar other rules, including for $A \simeq_n B$ and $\Pi a : A.B$ (though the design for the latter is not yet stabilised)

Excerpt of related works

Takeuti (1953), Gandy (1956): setoid interpretation in Church's simple type theory

Hofmann (1995), Altenkirch (1999): setoid interpretation in type theory

Altenkirch-McBride-Swiestra (2007): setoid interpretation in direct style

Licata-Harper (2012): two-dimensional type theory

Barras-Coquand-Huber (2015): semi-simplicial interpretation

Bernardy-Coquand-Moulin (2015): iterated parametricity in direct style

Altenkirch-Kaposi (2015): towards univalent parametricity

Tabareau-Tanter-Sozeau (2018): univalent parametricity at dimension 1

More generally, a motto is that we should eventually have a "polysemy" between some type theory in direct style, a corresponding indirect interpretation type theory by translation, a corresponding higher-dimensional presheaf interpretation.

In particular, we generalise Bernardy-Coquand-Moulin into an iterated *univalent* parametricity translation (in progress).