Univalence without function extensionality

Evan Cavallo Jonas Höfer

University of Gothenburg and Chalmers University of Technology

MFPS XLII, Ljubljana, Slovenia
June 01, 2026

Introduction

- We study intensional Martin-Löf type theory (ITT):
  - Σ and Π with strict η,
  - basic type formers $=, +, \ldots$,
  - a universe $\mathcal{U}$.
- The identity type in ITT is, to some extent, underspecified:
  - e.g. identity type of Π cannot be characterized
  - often this gap is bridged axiomatically (e.g. function extensionality, univalence)
- Our goal is to understand the relation between these axioms better

Introduction

Function extensionality

A *homotopy* between functions $f, g : \prod_A B$ is a pointwise equality:

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

Function extensionality states that "equality of functions is homotopy"

Axiom (Function extensionality, FE)

For all types A and families $A \vdash B$, the following is a family of equivalences:

$$\prod_{f,g:\prod_A B} f = g \longrightarrow f \sim g, \qquad \mathsf{refl}_f \longmapsto \lambda a.\mathsf{refl}_{f(a)}.$$

Introduction

Univalence

An *equivalence* of types A, B is a (well-behaved) bijection

$$A \simeq B :\equiv \sum_{f:A\to B}\left(\sum_{g:B\to A} fg \sim \mathrm{id}_B\right) \times \left(\sum_{h:B\to A} hf \sim \mathrm{id}_A\right).$$

Univalence states that "equality of types is equivalence"

Axiom (Univalence, $\mathsf{UA}_{\mathcal{U}}$)

A universe $\mathcal{U}$ is univalent if the following is a family of equivalences:

$$\text{id-to-eq} : \prod_{A,B:\mathcal{U}} A =_{\mathcal{U}} B \longrightarrow A \simeq B, \qquad \mathsf{refl}_A \longmapsto \mathrm{id}_A.$$

Univalence and function extensionality

Theorem (Voevodsky)

In ITT, the univalence axiom $\mathrm{UA}_{\mathcal{U}}$ implies function extensionality for functions between types between types in $\mathcal{U}$.

One of the many amazing consequences of univalence.

This one is to some extent unexpected:

- FE is not a universe-dependent axiom
- UA does not seem to imply extensionality for other negative type formers

Can we separate this consequence from the core of univalence?

Voevodsky's proof

Voevodsky's proof factors through a universe-independent fact:

$$\mathsf{UA}_{\mathcal{U}} \xRightarrow{\text{in } \mathcal{U}} \frac{f : A \simeq B}{f_* : A^C \simeq B^C}(*) \Longleftrightarrow \mathsf{FE}$$

FE $\Longrightarrow$ $(*)$ Inverse is f_*^{-1}, check $f \circ f^{-1} \circ g = g$ and $f^{-1} \circ f \circ g = g$ by FE

$(*)$ $\Longrightarrow$ FE Key idea: the homotopies yield paths in function types

$\mathsf{UA}_{\mathcal{U}}$ $\Longrightarrow$ $(*)$ By equivalence induction; less abstractly:

1. By $\mathsf{UA}_{\mathcal{U}}$ *improve* equivalence $f : A \simeq B$ to one coming from a path $A =_{\mathcal{U}} B$
2. For these the statement holds (by path induction, only check the case of id_A)

Categorical equivalences

Definition

A *categorical equivalence* of types A, B is

$$A \cong B :\equiv \sum_{f : A \to B} \left(\sum_{g : B \to A} fg =_{B \to B} \mathrm{id}_B \right) \times \left(\sum_{h : B \to A} hf =_{A \to A} \mathrm{id}_A \right).$$

- For $A, B : \mathcal{U}$, these are isomorphisms in the (wild) category $(\mathcal{U}, \to)$.
- Non-trivial examples: definitional isomorphisms, paths in $\mathcal{U}$ induce them
- Standard properties of isomorphisms in a (wild) category $\mathbb{C}$:
  id is isomorphism, 2-out-of-3, $f_* : \mathbb{C}(C, A) \to \mathbb{C}(C, B)$ is equivalence, ...

It suffices to improve equivalences to categorical ones in the Voevodsky's proof!

Categorical univalence

Lemma

FE *is valid exactly if any equivalence can be improved to a categorical equivalence.*

$$
\begin{array}{ccccc}
 & & (A \cong B) & & \\
 & \overset{\text{id-to-ceq}}{\nearrow} & & \overset{\text{ceq-to-eq}}{\searrow} & \\
(A =_{\mathcal{U}} B) & & \xrightarrow[\text{id-to-eq}]{\hspace{6em}} & & (A \simeq B)
\end{array}
$$

Univalence ($\mathsf{UA}_{\mathcal{U}}$) $\text{id-to-eq}_{A,B}$ is equivalence for $A, B : \mathcal{U}$

Function extensionality (FE) $\text{ceq-to-eq}_{A,B}$ has section ($\Leftrightarrow$ is equivalence) for all A, B

Definition (Categorical univalence, $\mathsf{CUA}_{\mathcal{U}}$)

$\mathcal{U}$ is *categorically univalent* if $\text{id-to-ceq}_{A,B} : A =_{\mathcal{U}} B \to A \cong B$ is an equivalence for $A, B : \mathcal{U}$. Equivalently, if the (wild) category $\mathcal{U}$ is univalent.

This axiom was proposed in a MathOverflow question by Dorais (2013)

Counter models for function extensionality

Categorical univalence is defined such that $\mathsf{UA}_{\mathcal{U}} \iff \mathsf{CUA}_{\mathcal{U}} \land \mathsf{FE}_{\mathcal{U}}$
Is the decomposition proper? Need model $\mathbb{M}$ of ITT with $\mathbb{M} \vDash \mathsf{CUA}_{\mathcal{U}}$ and $\mathbb{M} \nvDash \mathsf{FE}_{\mathcal{U}}$

Getting models refuting FE but validating η for Π types ($f \doteq \lambda x.f(x)$) is non-trivial.

Realizability e.g. Streicher (1993)

- Hard to combine with homotopical features without validating FE

Projective model structure on $[\mathbf{B}\mathbb{Z}_2, \mathsf{Gpd}]$ Bordg (2015)

The exceptional translation Pédrot and Tabareau (2018)

- Refute $\mathsf{CUA}_{\mathcal{U}}$

Polynomial models Von Glehn (2015)

von Glehn's polynomial model

Theorem (von Glehn)

For a nice[1] *model* $\mathbb{M}$ *of* ITT, *there is a model* $\mathbf{Poly}(\mathbb{M})$ *of* ITT *s.t.* $\mathbf{Poly}(\mathbb{M}) \vDash \neg\mathrm{FE}$.

Categorical univalence survives this construction:

Theorem (Cavallo, H.)

Given a nice model $\mathbb{M}$ *with* $\mathbb{M} \vDash \mathrm{UA}$ *then* $\mathbf{Poly}(\mathbb{M}) \vDash \mathrm{CUA}$.

In fact, we get a more uniform statement:

Theorem (Cavallo, H.)

Given a nice model $\mathbb{M}$ *with* $\mathbb{M} \vDash \mathrm{CUA}^{\bullet}$ *then* $\mathbf{Poly}(\mathbb{M}) \vDash \mathrm{CUA}^{\bullet}$.

Here, $\mathrm{CUA}^{\bullet}$ is categorical univalence for the category $\mathcal{U}^I$ for all types $I : \mathcal{U}$

[1] modeling coproducts and empty types with η

von Glehn's polynomial model

Basics

in **Poly**($\mathbb{M}$)	in $\mathbb{M}$	in $\mathbb{M}$
Context $\vdash \Gamma$ is	$\vdash \Gamma : \mathcal{U}$	$\gamma : \Gamma \vdash \underline{\Gamma}(\gamma) : \mathcal{U}$
Type $\Gamma \vdash A$ is	$\gamma : \Gamma \vdash A(\gamma) : \mathcal{U}$	$\gamma : \Gamma, a : A(\gamma) \vdash \underline{A}(\gamma, a) : \mathcal{U}$
Context extension $\Gamma.A$	$\vdash \Sigma_\Gamma A : \mathcal{U}$	$(\gamma, a) : \Sigma_\Gamma A \vdash \underline{\Gamma}(\gamma) + \underline{A}(\gamma, a) : \mathcal{U}$
Elements $\Gamma \vdash u : A$	$\gamma : \Gamma \vdash u(\gamma) : A(\gamma)$	$\gamma : \Gamma, a : \underline{A}(\gamma, u(\gamma)) \vdash \underline{u}(\gamma, a) : \underline{\Gamma}(\gamma)$
Id types $\Gamma \vdash u =_A v$	$\gamma : \Gamma \vdash u(\gamma) =_A v(\gamma)$	$\gamma : \Gamma, \_ : u(\gamma) =_A v(\gamma) \vdash 0.$

von Glehn's polynomial model

Equivalences and categorical equivalences

Ordinary equivalences An equivalence $\Gamma \vdash e \doteq (f, g, G, h, H) \colon A \simeq B$ is given by

$$\Gamma \vdash e \doteq (f, g, G, h, H) \colon A \simeq B$$
$$\Gamma, a \colon A \vdash \underline{f}_a \colon \underline{B}(f(a)) \longrightarrow \underline{\Gamma} + \underline{A}(a)$$
$$\Gamma, b \colon B \vdash \underline{g}_b \colon \underline{A}(g(b)) \longrightarrow \underline{\Gamma} + \underline{B}(b)$$
$$\Gamma, b \colon B \vdash \underline{h}_b \colon \underline{A}(h(b)) \longrightarrow \underline{\Gamma} + \underline{B}(b)$$

No condition on $\underline{f}, \underline{g}, \underline{h}$!

Categorical equivalences For $\Gamma \vdash e \doteq (f, g, G, h, H) \colon A \cong B$ analogous and additionally

$$\Gamma, b \colon B \vdash \_ \colon \underline{g}_b \circ \underline{f}_{g(b)} \underset{G}{=} \mathrm{inr} \qquad \Gamma, a \colon A \vdash \_ \colon \underline{f}_a \circ \underline{h}_{f(a)} \underset{H}{=} \mathrm{inr}$$

Hence, $\underline{f}, \underline{g}, \underline{h}$ factor over inr and induce equivalences on $\underline{2^{\text{nd}}}$ components.

By $\mathrm{UA}_{\mathcal{U}}$ in base model, get $A =_{\mathcal{U}} B$ and $\underline{A} =_{\mathcal{U}^A} \underline{B}e$.

Bibliography I

Bordg, Anthony (2015). "On lifting univalence to the equivariant setting". PhD thesis. Université Nice Sophia Antipolis. URL: https://arxiv.org/abs/1512.04083.

Dorais, François G. (2013). *Equivalent form of the Univalence Axiom*. MathOverflow. version: 2013-06-22. URL: https://mathoverflow.net/q/134449.

Pédrot, Pierre-Marie and Nicolas Tabareau (2018). "Failure is Not an Option: An Exceptional Type Theory". In: *Programming Languages and Systems - 27th European Symposium on Programming (ESOP 2018), Held as Part of the European Joint Conferences on Theory and Practice of Software (ETAPS 2018).* Ed. by Amal Ahmed. Vol. 10801. Lecture Notes in Computer Science. Thessaloniki, Greece: Springer, pp. 245–271. ISBN: 978-3-319-89883-4. DOI: 10.1007/978-3-319-89884-1_9. URL: https://doi.org/10.1007/978-3-319-89884-1%5C_9.

Bibliography II

Streicher, Thomas (1993). "Investigations into Intensional Type Theory". Habilitation thesis. Ludwig-Maximilians-Universität München. URL: https://www2.mathematik.tu-darmstadt.de/~streicher/HabilStreicher.pdf.

Von Glehn, Tamara (2015). "Polynomials and models of type theory". PhD thesis. University of Cambridge. DOI: 10.17863/CAM.16245. URL: https://www.repository.cam.ac.uk/handle/1810/254394.

Some details on Voevodsky's proof

The key step in the UA $\implies$ FE proof is as follows:

1. For $A \vdash P$ contractible, the map $\pi \colon \sum_A P \to A$ is an equivalence
2. By univalence $\pi_* \colon (\sum_A P)^A \to A^A$ is an equivalence
3. $\mathrm{fib}_{\pi_*}(\mathrm{id}_A) \simeq \prod_A P$, i.e, the following is a homotopy pullback

$$\begin{array}{ccc}
\prod_{a:A} B(a) & \xrightarrow{\langle \mathrm{id}_A, - \rangle} & \left(\sum_{a:A} B(a)\right)^A \\
\downarrow \wr & & \pi_* \downarrow \wr \\
1 & \xrightarrow[\mathrm{id}_A]{} & A^A.
\end{array}$$

4. Hence, $\Pi_A P \simeq 1$ and thus contractible