Universes and the limits of Martin-Löf type theory

Michael Rathjen

School of Mathematics
University of Leeds

Russell'08 Proof Theory meets Type Theory
Swansea, March 15, 2008

Two foundational programmes and their limits

- Finitism
- Predicativism
- Kreisel: Finitist functions = provable functions of **PA**.
- Tait: Finitist reasoning = primitive recursive reasoning in the sense of Skolem (**PRA**).
- Kreisel, Feferman: Predicativism is captured by autonomous progressions of theories.
- Feferman, Schütte: Γ_0 is the limit of the predicatively provable ordinals.

Martin-Löf type theory, MLTT

- Developed " **with the philosophical motive of clarifying the syntax and semantics of intuitionistic mathematics**" (Martin-Löf)
- Intended to be a full scale system for formalizing constructive mathematics.
- What are the limits of **MLTT**?
- It is perhaps not surprising that a study of this kind has not been undertaken within the community of constructive type theorists.

Investigate a system of thought

- **Aim**: Establish the **limits** of what could be achieved by a logician who uses certain concepts and principles together with **reflection** on these.
- *"... we now asked ourselves: what is implicit in the given concepts* **together** *with the concept of reflection on these concepts?* (Kreisel 1970)
- Switching back and forth between **two modes of thought**:
  1. To explore a system of thought, we think **within** the system.
  2. We think **about** the system.

Expanding MLTT from within

Strength and expressiveness are obtained through the use of

- **inductive data types** and
- **reflection**, i.e. **type universes**.

The concept of an **inductive data type** is central to Martin-Löf's constructivism. Gödel (1933) described constructive mathematics by the following characteristics:

(1) *The application of the notion "all" or "any" is to be restricted to those infinite totalities for which we can give a* **finite procedure for generating all their elements** *(as we can, e.g., for the totality of integers by the process of forming the next greater integer and as we cannot, e.g., for the totality of all properties of integers).*

(2) [. . .] *it follows that we are left with essentially only one method for proving general propositions, namely,* **complete induction applied to the generating process of our elements.** [. . .] *and so we may say that our system is based exclusively on the* **method of complete induction in its definitions as well as its proofs.**

Inductive Types

- The type $\mathbb{N}$ natural numbers.
- **W**-types

  E.g. Kleene's $\mathcal{O}$

$$\bar{0} : \mathcal{O} \qquad \frac{a : \mathcal{O}}{a' : \mathcal{O}} \qquad \frac{f : \mathbb{N} \to \mathcal{O}}{\mathrm{sup}(f) : \mathcal{O}}$$

Universe Types

- Universe types aren't simple inductive data types.
- Combination of defining inductively a type (the universe) together with a type-valued function by structural recursion.
- Example of a non-monotone inductive definition.
- Peter Dybjer, Anton Setzer: **Inductive-recursive definitions**.

($\mathbf{U}_{\mathcal{C}}$-formation) $\quad \mathbf{U}_{\mathcal{C}} : \mathit{type} \quad \dfrac{a : \mathbf{U}_{\mathcal{C}}}{\mathbf{T}_{\mathcal{C}}(a) : \mathit{type}}$

($\mathbf{U}_{\mathcal{C}}$-introduction) $\quad \hat{\mathbb{N}} : \mathbf{U}_{\mathcal{C}} \quad \mathbf{T}_{\mathcal{C}}(\hat{\mathbb{N}}) = \mathbb{N}$

$$\frac{a : \mathbf{U}_{\mathcal{C}} \quad b : \mathbf{U}_{\mathcal{C}}}{a \hat{+} b : \mathbf{U}_{\mathcal{C}}} \qquad \frac{a : \mathbf{U}_{\mathcal{C}} \quad b : \mathbf{U}_{\mathcal{C}}}{\mathbf{T}_{\mathcal{C}}(a \hat{+} b) = \mathbf{T}_{\mathcal{C}}(a) + \mathbf{T}_{\mathcal{C}}(b)}$$

$$\frac{a : \mathbf{U}_{\mathcal{C}} \quad \begin{array}{c}[x : \mathbf{T}_{\mathcal{C}}(a)] \\ t(x) : \mathbf{U}_{\mathcal{C}}\end{array}}{\hat{\Pi}(a, (\lambda x)t(x)) : \mathbf{U}_{\mathcal{C}}} \qquad \frac{a : \mathbf{U}_{\mathcal{C}} \quad \begin{array}{c}[x : \mathbf{T}_{\mathcal{C}}(a)] \\ t(x) : \mathbf{U}_{\mathcal{C}}\end{array}}{\mathbf{T}_{\mathcal{C}}(\hat{\Pi}(a, (\lambda x)t(x))) = (\Pi x : \mathbf{T}_{\mathcal{C}}(a))\mathbf{T}_{\mathcal{C}}(t(x))}$$

Expanding the realm of MLTT

- Palmgren: Universe operator and superuniverse
- Palmgren: Higher order universes
- R: Superjump universes

- Setzer: Mahlo and Π_3 reflecting universes

Should universes have elimination rules?

- In the absence of **elimination rules** and without closure under **W**-types and also no **W**-types in the ambient theory, the systems are rather weak.
- Hancock's conjecture Aczel, Feferman, Hancock: $|\bigcup_n \mathbf{ML}_n| = |\mathbf{MLU}| = \Gamma_0$.
- Crosilla, R (2002): $|\mathbf{CZF}^- + \forall x \exists y\, [x \in y \ \wedge\ y \text{ inaccessible}]| = \Gamma_0$.
- R (2000): $|\mathbf{MLS}| = |\mathbf{MLU} + \text{Superuniverse}| = \varphi\Gamma_0 00$.
- Gibbons, R (2002): $|\mathbf{ML} + \Pi_3\text{-reflection universe operator}|$ $= |\mathbf{CZF}^- + \Delta_0\text{-RDC} + \forall x \exists y\, [x \in y \ \wedge\ y \text{ is super-Mahlo}]|$ = **Big Veblen number** $= \theta\Omega^\Omega 0$

W-*types are essential*

- Aczel (1978) **CZF** $\hookrightarrow$ **MLV**,
  MLV has one universe **U** without elim rules and one **W**-type **V** on top of **U**.
- R (1992) $|\mathbf{MLV}| =$ Bachmann-Howard ordinal.
  Adding elim rules for **U** doesn't add any strength.
- $\mathbf{ML_{1W}}$ has one universe **U** closed under **W**-types but no elim rules and no **W**-types on top of **U**.
- R (1992): $\mathbf{ML_{1W}} \equiv \Delta^1_2$-CA $+$ BI. Adding **V** or elim rules for **U** doesn't increase the strength.

Stronger universe constructions

W-types always assumed

- Setzer (1993): Strength of $\mathbf{ML_1W}$. $\mathbf{ML_1W} > \Delta^1_2$-CA $+$ BI.
- Superjump universes: if $\mathbf{F} : \mathbf{Fam} \to \mathbf{Fam}$ then there exists a universe closed under **F**.
- R (2000):
  $\mathbf{MLF} = \mathbf{ML} +$ Superjump universes $\equiv \mathbf{CZF} + \mathbf{M} \equiv \mathbf{KPM} \restriction$ $= \mathbf{CZF} + \mathbf{M}$, where **M** is the rule

$$\frac{\phi}{\forall x\, \exists y \left[x \subseteq y \ \wedge \ y \ \texttt{is set-inaccessible} \ \wedge \ \phi^y\right]}$$

where ϕ is arbitrary sentence of **CZF** and ϕ^y is the result of restricting all quantifiers to y.

Even stronger universe constructions

- Palmgren's theory of higher universe operators $\leq$ **KPM**
- Setzer (2000): **ML** $+$ Mahlo universe $>$ **KPM**.
- Setzer (200?): **ML** $+$ Π_3-universe.

Classical theory of inductive definitions: the monotone case

- If A is set and

  $$\Phi : \mathcal{P}(A) \rightarrow \mathcal{P}(A)$$

  is a **monotone operator** then the the set-theoretic definition of the set inductively defined by Φ is given by

  $$\begin{aligned} \Phi^{\infty} &:= \bigcup_{\alpha} \Phi^{\alpha}, \\ \Phi^{\alpha} &:= \Phi\big(\bigcup_{\beta<\alpha} \Phi^{\beta}\big) \end{aligned}$$

  where α ranges over the ordinals.

Classical theory of inductive definitions: the general case

- If A is set and

$$\Phi : \mathcal{P}(A) \to \mathcal{P}(A)$$

  is an **arbitrary operator** then the the set-theoretic definition of the set inductively defined by Φ is given by

$$\begin{aligned} \Phi^\infty &:= \bigcup_\alpha \Phi^\alpha, \\ \Phi^\alpha &:= \Phi\Big(\bigcup_{\beta<\alpha} \Phi^\beta\Big) \cup \bigcup_{\beta<\alpha} \Phi^\beta, \end{aligned}$$

  where α ranges over the ordinals.
- $|\Phi| =$ least α s.t. $\Phi^\infty = \Phi^\alpha$

Some closure ordinals

- $|\mathfrak{X}| := \sup\{|\Phi| : \Phi \in \mathfrak{X}\}$
- $|\Pi^0_1| = |\Pi^1_1\text{mon}| = \omega_1^{CK}$ (Spector)
- $[\Phi_0, \Phi_1](X) = \begin{cases} \Phi_0(X) & \text{if } \Phi_0(X) \not\subseteq X \\ \Phi_1(X) & \text{if } \Phi_0(X) \subseteq X \end{cases}$
- $|\Pi^1_1, \Pi^0_0|$ = least recursively inaccessible (Richter)
- $|\Pi^0_1, \Pi^0_1|$ = least recursively Mahlo (Richter)
- $|\text{pos-}\Sigma^1_1| = |\text{mon-}\Sigma^1_1| = |\Sigma^1_1|$ (Grilliot)
- $n \geq 2$: $|\text{pos-}\Sigma^1_n| = |\text{mon-}\Sigma^1_n| = |\Sigma^1_n|$
  $|\text{pos-}\Pi^1_n| = |\text{mon-}\Pi^1_n| = |\Pi^1_n|$.

Coarse principles of Martin-Löf type theory

(A0) (Predicativism) The realm of types is built in stages (by the idealized type theorist). It is not a completed totality. In declaring what are the elements of a particular type it is disallowed to make reference to all types.

(A1) A type A is defined by describing how a canonical element of A is formed as well as the conditions under which two canonical elements of A are equal.

(A2) The canonical elements of a type must be namable, that is to say, they must allow for a symbolic representation, as a word in a language whose alphabet, in addition to countably many basic symbols, consists of the elements of previously introduced types. Here “previously" refers back to the stages of (A0).

Three kinds of types

- Explicitly defined types (e.g. the empty type and the type of Booleans N_1) as well types defined explicitly from given types or families of types (e.g. $A + B$, $(\Sigma x : A)B(x)$).
- Functions types: e.g. $A \rightarrow B$, $(\Pi x : A)B(x)$.
- Inductively defined types and universes.

Function types

- *"Since, in general, there are no exhaustive for generating all functions from one set to another, it follows that we cannot generate inductively all the elements of a set of the form* $(\Pi x \in A)B(x)$ *(or, in particular, of the form* B^A*, like* N^N*)."* (Martin-Löf)
- *"The reason that* B^A *can be constructed as a set is that we take the notion of function as* **primitive**, *instead of defining a function as a set of ordered pairs or a binary relation satisfying the usual existence and uniqueness conditions, which would make it a category (like* $\mathcal{P}(A)$*) instead of a set."* (Martin-Löf)
- *"The basic notion of function is an expression formed by abstraction."* (NPS)

- *“Power set seems especially nonconstructive and impredicative compared with the other axioms: it does not involve, as the others do, putting together or taking apart sets that one has already constructed but rather selecting, out of the totality of all sets those that stand in the relation of inclusion to a given set. One could make the same, admittedly vague, objection to the existence of the set $A \to B$ of mappings of A to B but I do not think the situation is parallel–a mapping or function is a* **rule***, a finite object which can actually be given;.."* (Myhill)
- *“An operation f defined over a domain D carries each element x of D into an element $f(x)$ of its range R. In intuitionistic mathematics the operation must be given as an* **effective means** *of determining the result $f(x)$ from the way in which x is given ... ."* (Dummett)

- The problem thus remains to delineate a class of functions that comprises all functions acceptable in Martin-Löf type theory. I will argue that all functions that deserve to be called effective must at least be definable in a way that is persistent with expansions of the universe of types.

- To put flesh to this idea, I consider it fruitful to investigate a rigorous model of the principles underlying **MLTT** within set theory. In the following, let us adopt a classical Cantorian point of view and analyze the principles (A0),(A1),(A2) on this basis.

- Firstly, types are to be interpreted as sets. By Gödel numbering, (A2) hereditarily has the consequence that nothing will be lost by considering all types to be surjective images of subsets of $\mathbb{N}$. In combination with (A1), such an encoding yields that every inductive type A can be emulated by an inductive definition Φ over the natural numbers together with a decoding function D, where

$$\Phi : \mathcal{P}(\mathbb{N}) \to \mathcal{P}(\mathbb{N})$$

is a (class) function from the class of all subsets of $\mathbb{N}$, $\mathcal{P}(\mathbb{N})$, to $\mathcal{P}(\mathbb{N})$. Thus the type A will be identified with the set

$$\{D(x) \,:\, x \in \Phi^{\infty}\}.$$

A further step in delineating **MLTT** consists in describing the allowable operators Φ and decoding functions D_A. A common way of classifying inductive definitions proceeds by their syntactic complexity. To find such a syntactic bound it is in order to recall that the type theorists develop their universe of types in stages. Introducing a new type A consists in describing a method for generating its elements. Taking into account that the type-theoretic universe is always in a state of expansion it becomes clear that each time a new element of A is formed by the method of generation for A, this method can only refer to types that have been built up hitherto.

Furthermore, the method of generation of elements should also obey a persistency condition of the following form: If at a certain stage an object t is recognized as an element of A then an expansion of the type-theoretic universe should not nullify this fact, i.e. the method should remain to be applicable and yield $t : A$ in the expanded universe as well. And in the same vein, if A is a type of codes of types which comes endowed with a type-valued decoding function D (like in the case of type universes), then the validity of equations between types of the form $D(x) = B$ with $x : A$ should remain true under expansions of the universe of types.

- Framing the foregoing in set-theoretic terms amounts to saying that the truth of formulas describing $t \in \Phi(X)$ and $D(t) = b$, respectively, ought to be **persistent under adding more sets to a set-theoretic universe**.
  In more technical language this means that whenever $\mathbb{M}$ and $\mathbb{P}$ are transitive sets of sets such that $t, X \in \mathbb{M}$, $\mathbb{M} \subseteq \mathbb{P}$, then

$$(\mathbb{M}, \in_{\mathbb{M}}) \models t \in \Phi(X) \;\Rightarrow\; (\mathbb{P}, \in_{\mathbb{P}}) \models t \in \Phi(X).$$

  The same persistency property should hold for formulas of the form '$D(x) = b$'.
- The formulas which can be characterized by the latter property are known in set theory as the **Σ-formulas**.
  They are exactly the collection of set-theoretic formulas generated from the atomic and negated atomic formulas by closing off under $\wedge, \vee$, bounded quantifiers $(\forall x \in a), (\exists x \in a)$ and unbounded existential quantification $\exists x$.

- In view of the preceding, one thus is led to impose restrictions on the complexity of inductive definitions for generating types in **MLTT** as follows.

(A3) Every inductive definition $\Phi : \mathcal{P}(\mathbb{N}) \to \mathcal{P}(\mathbb{N})$ for generating the elements of an inductive type *A* in **MLTT** and its pertinent decoding function are definable by set-theoretic Σ-formulas. These formulas may contain further sets as parameters, corresponding to previously defined types.

- To avoid misunderstandings, I'd like to emphasize that (A3) is not meant to say that every such Σ inductive definition gives rise to a type acceptable in **MLTT**. (A3) is intended only as a delineation of an upper bound.

- The case for the predicativity of function types rests on the requirement that functions be given by rules that enable one to compute their values effectively, it is plain that any such function must be **definable in an absolute way**. In view of the foregoing arguments for restrictions imposed on inductive types in conjunction with (A2) one is led to require the following:

(A4) All the functions figuring in **MLTT** belong to the set

$$\textbf{Func} := \{f \subseteq \mathbb{N} \times \mathbb{N} : f \text{ is a } \Sigma\text{-definable function}\}.$$

Note that the functions in **Func** are required to have a lightface Σ definition, that is to say definitions must not involve parameters (oracles).

- The functions in **Func** are known from generalized recursion theory on ordinals. **Func** consists all ∞-partial recursive functions from $\mathbb{N}$ to $\mathbb{N}$. In terms of the analytical hierarchy, **Func** can be characterized as the class of all (lightface) Σ^1_2-definable partial functions from $\mathbb{N}$ to $\mathbb{N}$.

- (A4) and previous considerations induce us to delineate the interpretation of product types as follows:

(A5) Every product type $(\Pi x : A)B(x)$ in **MLTT** is a set of functions from A to $\bigcup_{x:A} B(x)$ Σ-definable (with parameters) from previously defined types and the set **Func**. Moreover, $(\Pi x : A)B(x)$ is a subset of **Func**.

- The principles (A0)-(A5) will allow us to draw a limit to **MLTT** in the guise of a small fragment of **ZF**. This fragment, notated **T**, will be based on the ubiquitous **Kripke-Platek set theory**, **KP**. Kripke-Platek set theory is a truly remarkable subsystem of **ZF**. Though considerably weaker than **ZF**, a great deal of set theory requires only the axioms of this subsystem. **KP** arises from **ZF** by omitting the power set axiom and restricting separation and collection to bounded formulas, that is formulas without unbounded quantifiers. **KP** has been a major site of interaction between many branches of logic (for more information see the book by Barwise). The transitive models of **KP** are called **admissible sets**.

To describe **T**, we have to alter **KP** slightly. Among the axioms of **KP** is the foundation scheme which says that every non-empty definable class has an $\in$-least element. Let $\mathbf{KP}^r$ result from **KP** by restricting the foundation scheme to sets. In addition to $\mathbf{KP}^r$, **T** has an axiom asserting that every set is contained in a transitive set which is a Σ_1 elementary substructure of the set-theoretic universe V (written $M \prec_1 V$). To be more precise, let $M \prec_1 V$ stand for the scheme

$$\forall a{\in}M \left[\exists x\, \phi(x,a) \rightarrow \exists x \in M\, \phi(x,a)\right]$$

for all bounded formulas $\phi(x,y)$ with all free variables exhibited. Using a Σ_1 satisfaction predicate, $M \prec_1 V$ can actually be expressed via a single formula.

We take **T** to be the theory

$$\mathbf{KP}^r + \forall x\, \exists M \left(x{\in}M \ \wedge\ M \prec_1 V\right).$$

- The following theorems are provable in **T**.
- **Theorem 1: Func** is a set.
- **Theorem 2:** If $\Phi : \mathcal{P}(\mathbb{N}) \to \mathcal{P}(\mathbb{N})$ is definable by a Σ formula with parameters in M and $M \prec_1 V$, then $\Phi^\infty \in M$.
- The above theorem supports the claim that everything a Martin-Löf type theorist can ever develop can be emulated in **T** or, to put it more pictorially, that the boundaries of the type theorist world are to be drawn inside M, where M satisfies $M \prec_1 V$.

- Before elaborating further on this question, it might be interesting to give an equivalent characterization of **T** which is couched in terms of subsystems of second order arithmetic.

- **Theorem 3:** The theories $(\Pi^1_2\text{-}\mathbf{CA}) \restriction$ and **T** prove the same statements of second order arithmetic.

- Resuming the question of the type theorist's limit, I shall now argue on the basis of **T** to support the following **Claim**:
- **Theorem**: Every set $M \prec_1 V$ with **Func** $\in M$ is a model of **MLTT**, i.e. it contains all the types that may ever be constructed in **MLTT**.

The argument

The argument may run in this way: Types are interpreted as sets. At a certain stage the idealized type theorist, called *ITT*, has a certain repertoire of type forming operations, say $\mathcal{C}$. The operations correspond to a collection $\mathcal{C}_{Set}$ of Σ-definable operations on sets. Further, assume that *ITT* introduces a new type A by utilizing $\mathcal{C}$. Inductively we may assume that any set M with $M \prec_1 V$ and **Func** $\in M$ is a model of *ITT*'s reasoning as developed up to this point. Thus any such M is closed under $\mathcal{C}_{Set}$. According to (A3), the generation of the elements of A gives rise to an operator $\Phi_M : \mathcal{P}(\mathbb{N}) \cap M \to \mathcal{P}(\mathbb{N}) \cap M$ and a decoding function D_M which are both Σ-definable on M whenever $M \prec_1 V$. Moreover, Φ_M and D_M are uniformly definable on all $M \prec_1 V$, that is to say, there are Σ-formulas $\psi(x, y)$ and $\delta(u, v)$ such that $\Phi_M(X) = Y$ iff $(M, \in_M) \models \psi(X, Y)$ and $D_M(u, v)$ iff $(M, \in_M) \models \delta(u, v)$ whenever $X, Y \in \mathcal{P}(\mathbb{N}) \cap M$, $u, v \in M$, and $M \prec_1 V$.

Now define

$$\Phi : \mathcal{P}(\mathbb{N}) \to \mathcal{P}(\mathbb{N})$$

by letting $\Phi(X) = \Phi_M(X)$, where $X \in M$ and $M \prec_1 V$. Φ defines a function since the Φ_M are Σ definable and for every $X \subseteq \mathbb{N}$ there exists $M \prec_1 V$ such that $X \in M$. Thus **T** proves that Φ is a Σ-definable operator, i.e.,

$$\mathbf{T} \vdash \forall X \subseteq \mathbb{N}\, \exists Y\, \Phi(X) = Y.$$

Employing Theorem 2, one can deduce that Φ^∞ is a set. Moreover, as Φ^∞ is Σ definable too, one can infer that $\Phi^\infty \in M$ and thus

$$A = \{D(u) : u \in \Phi^\infty\} = \{D_M(u) : u \in \Phi^\infty_M\} \in M$$

for every $M \prec_1 V$.

Pushing the boundaries

- Feferman (1975): **Explicit Mathematics**, $\mathbf{T}_0$.
- $\mathbf{T}_0$ is a formal framework serving many purposes. It is suitable for representing Bishop-style constructive mathematics as well as generalized recursion, including direct expression of structural concepts which admit self-application.
- Let **MID** be the axiom asserting the existence of a least fixed point for any monotone operation f on classifications (the notion of set in explicit mathematics), and let **UMID** be its uniform rendering, where a least solution $\mathbf{clfp}(f)$ is presented as a function of the operation by adjoining a new constant **clfp** to the language of $\mathbf{T}_0$.

- Feferman (1982) *What is the strength of* $\mathbf{T}_0 + \mathbf{MID}$*? [...] I have tried, but did not succeed, to extend my interpretation of* $\mathbf{T_0}$ *in* $\Sigma^1_2 - AC + BI$ *to include the statement* **MID**. *The theory* $\mathbf{T}_0 + \mathbf{MID}$ *includes all constructive formulations of iteration of monotone inductive definitions of which I am aware, while* $\mathbf{T}_0$ *(in its* **IG** *axiom) is based squarely on the general iteration of accessibility inductive definitions. Thus it would be of great interest for the present subject to settle the relationship between these theories.* (p. 88)
- R (1996-1998)
  $\mathbf{T}_0 \restriction + \mathbf{UMID} \equiv \Pi^1_2\text{-CA}_0$
  $\mathbf{T}_0 \restriction +\text{full induction} + \mathbf{UMID} \equiv (\Pi^1_2\text{-CA})$.
- Tupailo (2005)
  $\mathbf{T}^i_0 \restriction + \mathbf{UMID} \equiv \Pi^1_2\text{-CA}_0$
- Tupailo, R
  $\mathbf{T}^i_0 \restriction +\text{full induction} + \mathbf{UMID} \equiv (\Pi^1_2\text{-CA})$

- R (2005) **CZF** $+$ **GID** $>$ (Π^1_2-CA)