Proof Theory of Martin-Löf Type Theory – An Overview

Anton Setzer[*]

February 22, 2004

Abstract

We give an overview over the historic development of proof theory and the main techniques used in ordinal theoretic proof theory. We argue, that in a revised Hilbert's programme, ordinal theoretic proof theory has to be supplemented by a second step, namely the development of strong equiconsistent constructive theories. Then we show, how, as part of such a programme, the proof theoretic analysis of Martin-Löf type theory with W-type and one microscopic universe containing only two finite sets is carried out. Then we look at the analysis of Martin-Löf type theory with W-type and a universe closed under the W-type, and consider the extension of type theory by one Mahlo universe and its proof-theoretic analysis. Finally we repeat the concept of inductive-recursive definitions, which extends the notion of inductive definitions substantially. We introduce a closed formalisation, which can be used in generic programming, and explain, what is known about its strength.

Keywords: Martin-Löf type theory, proof theory, Kripke-Platek set theory, W-type, well-founded trees, Kleene's O, Mahlo universe, inductive-recursive definitions, generic programming.

1 Introduction

The goal of this article is to introduce the reader, who is not necessarily an expert in proof theory, into the current state of the art of proof theory of Martin-Löf type theory and the techniques used there. We start by giving a brief overview over the contents of this article.

In Sect. 2, we first reconsider the original form of Hilbert's programme: to prove the consistency of theories for carrying out mathematical proofs using finitary methods. We then discuss the theory PRA, which is usually taken as the formalisation of finitary methods. Next we look at Gödel's second incompleteness theorem and the fall of Hilbert's programme. Then we discuss Gentzen's proof theoretic analysis of Peano arithmetic. We will look at the techniques used there – the notion of an ordinal notation system and cut elimination. Then we introduce the two main (usually equivalent) notions of proof theoretic strength.

In the short Sect. 3, we discuss, why a traditional proof theoretic analysis should be supplemented by a second step, namely the development of equiconsistent constructive theories.

In Sect. 4, we consider a relatively small variant of type theory, Martin-Löf type theory with W-type and a microscopic universe atom. This is used in order

[*] Supported by Nuffield Foundation, grant ref. NAL/00303/G and EPSRC grant GR/S30450/01. Department of Computer Science, University of Wales Swansea, Singleton Park, Swansea SA2 8PP, UK. Email: a.g.setzer@swan.ac.uk, Fax: +44 1792 295651.

to introduce the basic techniques of proof theory, the relationship of these type theories to so called admissible ordinals, and to variants of a weak version of set theory called Kripke-Platek set theory.

In Sect. 5 we will look at the first major example, Martin-Löf type theory with W-type and one universe. We will see that, in order to obtain an upper bound, one can model type theory in Kripke-Platek set theory extended by one recursively inaccessible and finitely many admissibles above it, which can be analysed easily.

In Sect. 6, we first develop a formalisation of a Mahlo universe and discuss its constructive validity. We then discuss a suitable extension of Kripke-Platek set theory of the same strength, and give some hints about how to model type theory in this set theory.

In Sect. 7, we look at one application of the results obtained in the area of generic programming. We consider P. Dybjer's concept of inductive-recursive definitions, which is a substantial generalisation of the concept of inductive definitions and includes standard (non-Mahlo-) universe constructions. Then we will develop a theory in which one can introduce all inductive-recursive sets. This theory will have a data type of inductive-recursive definitions, and allows therefore to introduce functions, which take a data type, analyse it and create another data type from it – this is a very general form of generic programming. This type theory makes use of the ideas contained in the definition of the Mahlo universe, and subsumes a slightly weakened form of the Mahlo universe.

In appendix A, we give a direct well-ordering proof for an ordinal notation system of strength ϵ_0. This forms the basis for developing well-ordering proofs, the main tool for determining lower bounds for the strength of type theories.

In appendix B we give details about the how to model Martin-Löf type theory with W-type and one universe in a corresponding variant of Kripke-Platek set theory, which can be analysed easily. We will as well show how to obtain a lower bound by carrying out a direct well-ordering proof for a corresponding ordinal notation system.

Appendix C will describe some details about how to obtain an upper bound for the strength of type theory with W-type and one Mahlo universe by modelling it in Kripke-Platek set theory with one recursively Mahlo ordinal. We will not go into details w.r.t. the well-ordering proof for this type theory.

In appendix D we will show that type theory introduced in Sect. 7 reaches the strength of Kripke-Platek set theory plus recursive Mahloness of the universe.

2 The Notion of Proof-Theoretic Strength

Hilbert's programme. Proof theory was established as a science by D. Hilbert. In his famous list of mathematical problems [Hilbert, 1900], he posed as second problem to show the consistency of an axiomatisation of the real numbers developed by him. He argued that, if this axiomatisation is shown to be consistent, this would prove the mathematical existence of the concept of real numbers and of the continuum: consistency implies existence. He stated as well the main problem, namely that, if one shows the consistency of a theory for formalising mathematics in the same theory, one has not achieved anything: if the original theory is inconsistent, it proves everything, even its own consistency. So in order to achieve something, one has to do more: namely show the consistency using methods which are considered to be safe. According to Hilbert, finitary methods were to be considered to be safe. By finitary methods he considered finitary calculations, as we can carry them out on a piece of paper.

Later his problem was generalised to what is now known as Hilbert's programme: to prove the consistency of axiom systems, in which certain parts of mathematics can be carried out, by finitary means.

There are two main approaches for carrying out consistency proofs. One is to introduce a model of the system in question in the Meta-theory. However, it seems to be implausible to assume that one can prove this way the consistency of a theory by using finitary methods, since such methods do not allow the use of sets. Hilbert realized this and suggested therefore that one should instead analyse proofs and show this way directly that it is not possible to derive in the formal system in question a contradiction. He called the mathematical discipline, in which such investigations are carried out, proof theory.

The first step was to establish a precise formalisation of what is meant by mathematical theories. A theory for formalising basic logic had to be introduced. One of these formalisations, the Hilbert-calculus, is due to Hilbert. Later many other equivalent ones were developed, and the probably currently in proof theory most popular one is the Tait calculus [Tait, 1968]. By adding axioms about mathematical entities to such logic calculi, one obtains theories in which mathematical proofs can be formalised.

Hilbert developed one technique for carrying out consistency proofs in his sense, the epsilon-substitution method ([Hilbert and Bernays, 1939], see as well recent work by Mints and Tupailo, e.g. [Mints and Tupailo, 1999, Mints et al., 1996]).

Primitive-Recursive Arithmetic was introduced by Skolem in his article [Skolem, 1923]. There he reasoned informally in a system, which was later formalised and called primitive-recursive arithmetic (PRA). This system is nowadays generally regarded as being a formalisation of what Hilbert meant by finitary methods.

The only objects in PRA are natural numbers. The basic notion is that of a primitive recursive function. The primitive recursive functions, which are functions $\mathbb{N}^n \to \mathbb{N}$ for arbitrary n, are those, which can be constructed from the constant zero function ($\lambda x.0$), projection functions (i.e. $\lambda(x_0, \ldots, x_{n-1}).x_i$, which denotes the function $f : \mathbb{N}^n \to \mathbb{N}$ s.t. $f(x_0, \ldots, x_{n-1}) = x_i$), and successor function ($\lambda x.x + 1$) by using composition (i.e. if f, g_i are primitive recursive, so is $\lambda(x_0, \ldots, x_{n-1}).f(g_1(x_0, \ldots, x_{n-1}), \ldots, g_k(x_0, \ldots, x_{n-1}))$ and the schema of primitive recursion: if f and g are primitive recursive, so is the function h defined by $h(\vec{x}, 0) = f(\vec{x})$, $h(\vec{x}, y+1) = g(\vec{x}, y, h(\vec{x}, y))$. Addition, multiplication and exponentiation can easily be defined using primitive recursion. The defining equations for the primitive recursive functions provide a schema for calculating the result of these functions in finite amount of time. Therefore the primitive recursive functions can be regarded as finitary operations.

The terms of PRA are now expressions which are constructed from variables and 0 by application of symbols for primitive recursive functions. If we substitute the free variables by numbers, they can be evaluated in finite amount of time. Therefore these terms can be regarded as finitary schemata.

The formulae in Skolem's system were all propositional formulae constructed from equations, i.e. the set of formulae is the least set containing the equations, and which is closed under negation, conjunction and disjunction. Define first $a \dot{-} b := \max\{0, a-b\}$, which can be defined primitive-recursively. Then we can encode $a = b$ as $(a \dot{-} b) + (b \dot{-} a) = 0$, $\neg(a = 0)$ as $(1 \dot{-} a) = 0$, $(a = 0) \wedge (b = 0)$ as $a + b = 0$ and $a = 0 \vee b = 0$ as $a \cdot b = 0$. By repetitively applying these operations we can encode all propositional formulae as equations of the form $a = 0$, and can restrict the set of formulae therefore to equations. These equations relate finitary schemata with each other.

Skolem used in his reasoning as basic laws the defining equations for primitive-recursive functions, the standard laws of = (reflexivity, symmetry, transitivity and substitution) and the classical laws for the propositional connectives. (One can show that restricted to quantifier free formulae with decidable prime formulae, classical and intuitionistic logic coincide). The only strong law he used is that of induction over primitive recursive formulae. Using the above encoding, the system can be restricted to a theory having as formulae equations only, and where we omit therefore the laws for the propositional connectives (a systematic development of this can be found in [Goodstein, 1964]). The resulting laws can now be regarded as a formulation of what Hilbert meant by finitary methods: Assume an equation $t = t'$ is derived this way, and let s, s' be the result of substituting all variables in t, t' respectively by numbers. By going through the derivation (a proof of a corresponding Meta-theorem requires induction on the derivation) one can easily see that s and s' reduce to the same number, and, when investigating it, one has never to refer to the set of natural numbers as an entity, but needs to refer only to finitely many numbers.

It can easily be shown that the following system is conservative over Skolem's version of PRA for equational formulae, i.e. that both theories derive the same propositional formulae: One takes as formulae arbitrary first-order formulae (i.e. all formulae with quantifiers ranging over natural numbers), built from equations of terms as given before. As rules one takes the basic rules of the predicate calculus (i.e. basic logic for formulae), and basic laws of equality (which involve the extended language). Furthermore, one adds the rule of induction over quantifier-free (i.e. propositional) formulae. The proof that we obtain a conservative extension can be carried out using proof theoretic techniques in PRA. In proof theory, by PRA one usually means the just mentioned theory, and we will follow in the rest of this article this convention.

Gödel's second incompleteness theorem and the failure of Hilbert's original programme. 1931 Gödel showed in his second incompleteness theorem [Gödel, 1931], that Hilbert's original programme cannot be carried out – assuming minimal conditions on a theory T, which hold for practically all theories with a natural encoding of the natural numbers (natural theories which have been considered and do not fulfil these conditions are weaker than PRA), he could show that a consistent theory T does not prove its own consistency. It follows that the consistency of theories T with a natural embedding of PRA and which fulfil Gödel's conditions cannot be shown by finitary means. Most natural theories except for extremely weak ones fulfil the premise of the last sentence – Hilbert's original programme had failed.

Gentzen's proof of the consistency of Peano Arithmetic. 1936 Gerhard Gentzen [Gentzen, 1936] showed the consistency of Peano Arithmetic (PA) using transfinite induction up to ϵ_0. This was the birth of ordinal theoretic proof theory.

PA is the extension of PRA by allowing induction over all (first-order) formulae. (one often restricts the set of functions to addition, multiplication, successor function and the constant zero however – the addition of all primitive-recursive functions is conservative over that theory). Gentzen considered a primitive recursive ordinal notation system up to ϵ_0:

In set theory, an ordinal α is a set which is transitive and the elements of which are transitive. Especially the elements of ordinals are ordinals. Let Ord be the class of all ordinals. As usual, Greek letters will in the following refer to ordinals, so e.g. $\forall\alpha.\varphi(\alpha)$ stands for $\forall x \in \mathrm{Ord}.\varphi(x)$. The relation $\in$ on the class of ordinals is a linear ordering, and one usually writes $\alpha < \beta$ for $\alpha \in \beta$. $<$ is well-founded, i.e. for any formulae we have the principle of transfinite induction over ordinals: $(\forall\alpha.(\forall\beta < \alpha.\varphi(\beta)) \to \varphi(\alpha)) \to \forall\alpha.\varphi(\alpha)$. An ordering which is both linear and well-founded is called a well-ordering. The union of a set of ordinals A forms an ordinal which is the supremum of the ordinals, therefore written as sup A. There is a standard definition of addition, multiplication and exponentiation of ordinals. All natural numbers can be regarded as ordinals (using $0 = \emptyset$ and $\alpha + 1 = \{\alpha\} \cup \alpha$) and the set of natural numbers forms an ordinal ω, which is the least infinite ordinal. Every ordinal α is either 0, a successor ordinal, i.e. of the form $\beta + 1$, or a limit ordinal, which means that $\forall \beta < \alpha. \exists \gamma < \alpha. \beta < \gamma$. Furthermore, for every ordinal α there exists an $k \in \mathbb{N}$ and unique $\alpha \geq \alpha_1 > \cdots > \alpha_k$ and $n_i \in \mathbb{N}$ s.t. $n_i > 0$ and $\alpha = \omega^{\alpha_1} \cdot n_1 + \cdots + \omega^{\alpha_k} \cdot n_k$ (here we use the just mentioned operations of addition, multiplication and exponentiation on ordinals). If the expression on the right fulfils the previous conditions on α_i and n_i, it is called the Cantor Normal Form CNF of α. One can show that the ordering on ordinals in CNF is the lexicographic ordering: If $\alpha = \omega^{\alpha_1} \cdot n_1 + \cdots + \omega^{\alpha_k} \cdot n_k$, $\beta = \omega^{\beta_1} \cdot m_1 + \cdots + \omega^{\beta_l} \cdot m_l$ are the CNFs of α, β, then $\alpha < \beta$ iff $((\alpha_1, n_1), \ldots, (\alpha_k, n_k)) < ((\beta_1, m_1), \ldots, (\beta_l, m_l))$ with respect to the lexicographic ordering on pairs and descending sequences: one forms first the lexicographic ordering on pairs (α, n) s.t. $\alpha \in \mathrm{Ord}$ and $n \in \mathbb{N}$, and then the lexicographic ordering on descending sequences of such pairs. We will use this property in appendix A in a direct well-foundedness proof for the ordinal notation system up to ϵ_0. An ordinal α which has a CNF with ordinal coefficients $\alpha_i < \alpha$ can be considered to be constructed using CNF from smaller ordinals, and ϵ_0 is the least ordinal which does not have this property. It can be defined as $\epsilon_0 = \sup\{\underbrace{\omega^{\omega^{\cdot^{\cdot^{1}}}}}_{n \text{ times}} \mid n \in \omega\}$. One can easily show that $\epsilon_0 = \omega^{\epsilon_0}$.

An ordinal notation system is a pair $(\mathrm{OT}, <_{\mathrm{OT}})$ consisting of a set $\mathrm{OT} \subseteq \mathbb{N}$ and a primitive relation $<_{\mathrm{OT}} \subseteq \mathrm{OT}^2$, such that $<_{\mathrm{OT}}$ is linear and well-founded, i.e. the principle of transfinite induction over OT holds with respect to all sets: $\forall X \subseteq \mathbb{N}.(\forall x \in \mathrm{OT}.(\forall y \in \mathrm{OT}.y <_{\mathrm{OT}} x \to y \in X) \to x \in X) \to \mathrm{OT} \subseteq X$. We will write $<$ instead of $<_{\mathrm{OT}}$, if it will be clear from the context, whether $<$ or $<_{\mathrm{OT}}$ is meant by $<$. $(\mathrm{OT}, <)$ is primitive recursive, if both OT and $<$ are primitive recursive. In a context, in which only natural numbers and one ordinal notation system are mentioned, one writes Greek letters for elements of OT using the same convention as before – e.g. $\forall \beta. \varphi(\beta)$ stands for $\forall x \in \mathrm{OT}. \varphi(x)$.

Now it is easy to introduce an ordinal notation system based on CNF: One encodes sequences of natural numbers $(n_1, \ldots, n_k)$ as natural numbers $\langle n_1, \ldots, n_k \rangle$, and defines simultaneously inductively OT_{ϵ_0} and $<_{\epsilon_0} \subseteq \mathrm{OT}_{\epsilon_0} \times \mathrm{OT}_{\epsilon_0}$ as follows: If $k \in \mathbb{N}$, $a_1, \ldots, a_k \in \mathrm{OT}_{\epsilon_0}$, $a_k <_{\epsilon_0} \cdots <_{\epsilon_0} a_1$, $n_i > 0$ then $\langle \langle a_1, n_1 \rangle, \ldots, \langle a_k, n_k \rangle \rangle \in \mathrm{OT}_{\epsilon_0}$. (In the special case $k = 0$ we obtain $\langle \rangle$ representing 0). $\langle \langle a_1, n_1 \rangle, \ldots, \langle a_k, n_k \rangle \rangle <_{\epsilon_0} \langle \langle b_1, m_1 \rangle, \ldots, \langle b_k, m_k \rangle \rangle$, if the underlying sequences are in this order with respect to lexicographic ordering on pairs $\langle a, n \rangle \in \mathrm{OT}_{\epsilon_0} \times \mathbb{N}$ and lexicographic ordering on sequences formed from such pairs, which reduces to the underlying ordering. It is a standard exercise (assuming a standard properties of the encoding of sequences of natural numbers as natural numbers) to show that $(\mathrm{OT}_{\epsilon_0}, <_{\epsilon_0})$ forms a linear ordering which is primitive-recursive. Using set theory one can easily show that it is well-founded, by defining an embedding o of OT_{ϵ_0} into Ord by $\mathrm{o}(\langle \langle a_1, n_1 \rangle, \ldots, \langle a_k, n_k \rangle \rangle) = \omega^{\mathrm{o}(a_1)} \cdot n_1 + \cdots + \omega^{\mathrm{o}(a_k)} \cdot n_k$, and by then showing that $a <_{\epsilon_0} b \Leftrightarrow \mathrm{o}(a) < \mathrm{o}(b)$. We will, as usual in proof theory, identify ordinal notations with the ordinals they denote, and write $\omega^{a_1} \cdot n_1 + \cdots + \omega^{a_k} \cdot n_k$ instead of $\langle \langle a_1, n_1 \rangle, \ldots, \langle a_k, n_k \rangle \rangle$, if there is no confusion.

In Appendix A we will sketch a direct well-foundedness proof of the ordinal system of strength ϵ_0 – more precisely, we show that for every n the restriction of OT_{ϵ_0} to ordinals less than $\underbrace{\omega^{\omega^{\cdot^{\cdot^{1}}}}}_{n \text{ times}}$ is well-founded. By Meta-induction on n, this argument can be formulated directly in Peano Arithmetic, and therefore we can show for (Meta-)every $b < \epsilon_0$ transfinite induction over OT_{ϵ_0} restricted to ordinals less than b. The first proof carried out in PA we could find was in Hilbert/Bernays

([Hilbert and Bernays, 1939], §5, 3c). However we cannot show in PA that the union is well-founded, provided PA is consistent. This follows from Gödel's second incompleteness theorem and the fact that PRA plus the principle of transfinite induction over $(\mathrm{OT}_{\epsilon_0}, <_{\epsilon_0})$ proves the consistency of PA (therefore PA plus this principle proves the consistency as well), as shown by Gentzen using cut elimination:

Cut elimination. We will in the following give a modern version (in this compact form essentially due to Buchholz) of Gentzen's proof of the consistency of PA in PRA extended by quantifier free transfinite induction up to ϵ_0. First one can embed proofs of closed formulae of Peano arithmetic into a semi-formal system, i.e. a system of proof rules having rules with infinitely many premises, which we call PA*:

We take as set of formulae those constructed from prime formulae and negated prime formulae using $\wedge$, $\vee$, $\forall$ and $\exists$. The negation for non-prime formulae is *defined* by the deMorgan rules, e.g. $\neg(P(x) \wedge Q(x)) := (\neg P(x)) \vee (\neg Q(x))$. Similarly we *define* $A \to B := \neg A \vee B$.

PA* is a Tait-style sequent calculus. Here sequents Γ, Δ are sets of formulae $\{A_1, \ldots, A_n\}$, with the intended meaning being $A_1 \vee \cdots \vee A_n$. Especially the empty sequent, denoted by $\emptyset$, stands for falsity. One writes Γ, A for $\Gamma \cup \{A\}$ and Γ, Δ for $\Gamma \cup \Delta$. In PA* one derives closed sequents Γ, i.e. sequents such that the formulae don't contain any free variables.

The basic rules of the system are introduction rules for the logical connectives. Γ, A is an axiom, if A is a true prime formula. Furthermore, we have the following rules:

$$\frac{\Gamma, A \qquad \Gamma, B}{\Gamma, A \wedge B} \quad \frac{\Gamma, A}{\Gamma, A \vee B} \quad \frac{\Gamma, B}{\Gamma, A \vee B}$$

$$\frac{\Gamma, A(0) \quad \Gamma, A(1) \qquad \Gamma, A(2) \quad \cdots}{\Gamma, \forall x. A(x)}$$

$$\frac{\Gamma, A(t)}{\Gamma, \exists x. A(x)}$$

Note that $\forall$-introduction has infinitely many premises. The main formula in the conclusion of any rule (e.g. $A \wedge B$ in the $\wedge$-introduction rule) can be an element of Γ and therefore occur in the premise as well.

One can easily show that, if Γ is derivable, so is Γ, Δ. Therefore we can omit formulae in the premises of a rule which are not needed (e.g. if Γ, A and B are provable, then $\Gamma, A \wedge B$ is provable, since from a proof of B we can obtain a proof of the second premise Γ, B of the $\wedge$-introduction rule). We will therefore in derivations omit unnecessary formulae in sequents.

Additionally to the above rules, we add the cut rule and for technical purposes the repetition rule, which does not do anything (essentially due to Mints):

$$\frac{\Gamma, A \qquad \Gamma, \neg A}{\Gamma} \quad \frac{\Gamma}{\Gamma}$$

In the cut rule, A is called the cut-formula. If one allows non-well-founded proofs, i.e. proofs with infinite chains $\Gamma_1, \Gamma_2, \ldots$ s.t. Γ_1 is the conclusion and Γ_{i+1} is a premise of Γ_i, one can derive everything in this theory: derive Γ from Γ using the repetition rule, which is again derived from Γ etc. (Without the repetition rule one can instead apply any other rule applicable to the premise repetitively). Proofs without infinitely descending chains are called well-founded. Positively, one can define the set of well-founded proofs as the least set of proofs, such that if there are derivations d_i of Γ_i $(i \in I)$ in this set and a rule deriving Γ from $\Gamma_i (i \in I)$ then the derivation

$$\dfrac{\cdots \quad \overset{d_i}{\Gamma_i} \quad \cdots}{\Gamma} \; (i \in I)$$

is in this set.

Without using the cut- and repetition rule one can derive $A, \neg A$ by induction on the built-up of formulae. For prime formulae this is clear (since either A or $\neg A$ is true), and for instance in case $A = B \wedge C$ (note that $\neg(A \wedge B) = \neg A \vee \neg B$) this follows by:

$$\dfrac{\dfrac{A, \neg A}{A, \neg A \vee \neg B} \qquad \dfrac{B, \neg B}{B, \neg A \vee \neg B}}{A \wedge B, \neg A \vee \neg B}$$

Elimination rules for the logical connectives are provable using the cut rule. For instance, the proof that from $\Gamma, A \wedge B$ we can derive Γ, A is as follows (note that $\neg(A \wedge B) = \neg A \vee \neg B$):

$$\dfrac{\Gamma, A \wedge B \qquad \dfrac{\neg A, A}{\neg A \vee \neg B, A}}{\Gamma, A} \; \text{(Cut)}$$

Interpretation of PA **into** PA***.** We show, by induction over the derivation, that if a sequent Γ is derivable in PA depending on free variables $x_1, \ldots, x_k$, then for all $n_1, \ldots, n_k \in \mathbb{N}$ the sequent $\Gamma[x_1 := n_1, \ldots, x_k := n_k]$ is derivable in PA*. Here $\Gamma[x_1 := n_1, \ldots, x_k := n_k]$ denotes the result of substituting in all formulas in Γ the variables x_i by n_i. We consider some of the more important cases:

Case Γ is an axiom of PA: All instances of defining equations for primitive-recursive functions are true prime formulae, and therefore axioms in PA*. The equality axioms can be written as sequents of prime formulae (e.g. $x = y \to y = x$ can be rewritten as $x \neq y, y = x$), and then each instantiation is an axiom of PA* (if we substitute x by n and y by m in $x \neq y, y = x$, then we get $n \neq m, m = n$ and either $n \neq m$ or $m = n$ is true, hence an axiom). One minor problem is the transfer principle $x = y \to A(x) \to A(y)$ for arbitrary formulae A. One can easily see that, from the transfer principle restricted to prime formulae and negated prime formulae A one can prove using logical rules only the transfer principle for arbitrary formulae, so we can restrict in PA the principle to (negated) prime formulae. If we rewrite the transfer principle for (negated) prime formulae as $x \neq y, \neg A(x), A(y)$, we get a sequent, such that every instance is an axiom of PA*.

Case Γ is derived from logical rules: The introduction rules for $\wedge$ and $\vee$ follow from the corresponding rules of PA*, and the elimination rules are derivable, as we have seen above. We consider the case of $\forall$-introduction: Assume that $\Gamma = \Delta, \forall x.A(x)$, which is derived from $\Delta, A(x)$, where x is not free in Δ. Assume for simplicity that x is the only free variable in $\Delta, A(x)$. Then by induction hypothesis we know that we can show $\Delta, A(n)$ for all n, and obtain therefore the following proof of $\Delta, \forall x.A(x)$:

$$\dfrac{\Delta, A(0) \qquad \Delta, A(1) \qquad \cdots}{\Delta, \forall x.A(x)}$$

The case of an introduction rule for $\exists$ is similar.

Induction: We haven't specified precisely yet, how to formulate the induction principle in PA, and for our purposes the easiest way is to use the induction rule

$$\dfrac{\Delta, A(0) \qquad \Delta, \forall x(A(x) \to A(x+1))}{\Delta, \forall x.A(x)}$$

Assume that $\Gamma = \forall x.A(x)$ is derived by the above rule, and for simplicity assume that Δ is empty. By induction hypothesis, we have proofs in PA* of $A(0)$ and $\forall x(A(x) \to A(x{+}1))$. From $\forall x(A(x) \to A(x{+}1))$, which is the same as $\forall x(\neg A(x) \vee A(x+1))$, we obtain, using the fact that elimination for $\forall$ is derivable, a proof of $\neg A(n), A(n+1)$ as follows (note that $\neg(\neg A(n) \vee A(n+1)) = A(n) \wedge \neg A(n+1)$):

$$\dfrac{\dfrac{\forall x(\neg A(x) \vee A(x+1))}{\neg A(n) \vee A(n+1)} \qquad \dfrac{A(n), \neg A(n) \qquad \neg A(n+1), A(n+1)}{A(n) \wedge \neg A(n+1), \neg A(n), A(n+1)}}{\neg A(n), A(n+1)}\,(\text{Cut})$$

Now we obtain proofs in PA* of $A(n)$ for all n as follows:

$$\begin{array}{c}\dfrac{\dfrac{A(0) \qquad \neg A(0), A(1)}{A(1)} \qquad \neg A(1), A(2)}{A(2)} \\ \vdots \\ A(n)\end{array}$$

Using the introduction rule for $\forall$ we obtain therefore a proof of $\forall x.A(x)$ as follows (note that the proof of $A(n)$ has height at least n):

$$\dfrac{A(0) \qquad A(1) \qquad A(2) \quad \cdots}{\forall x.\phi(x)}$$

Since the proof of the nth premise has height at least n, the proof of $\forall x.A(x)$ has infinite height.

Every well-founded set can be linearised, and the resulting well-ordered set is order isomorphic to the set of ordinals, and therefore the height of infinite proofs can be measured by ordinals. One writes $\vdash^{\alpha}_{n} \Gamma$ for "Γ is provable in the system with ordinal height at most α and cut rank $< n$", where the rank n of a formula is a natural number which measures the size of the formula (for instance the number of connectives $\wedge, \vee, \forall, \exists$ in the formula), and the cut rank of a proof is the maximum rank of all cut formulas, if it exists. So, if we have $\vdash^{\alpha_i}_{n_i} \Gamma_i$ for the premises Γ_i of a rule with conclusion Γ, and if $\alpha_i < \alpha$, $n_i \leq n$, and, in case we have a cut, if the rank of the cut formula is $< n$, then $\vdash^{\alpha}_{n} \Gamma$ follows.

Every proof in Peano Arithmetic can now be interpreted in this system as a proof of height $< \omega + \omega$ and with finite cut rank. The main step in the consistency proof is to prove cut elimination, i.e. that from a derivation of a sequent in this calculus one obtains a cut-free derivation of the same sequent. As an example for how cut elimination is carried out, consider the following derivation:

$$\dfrac{\dfrac{\Gamma, A \wedge B, A \qquad \Gamma, A \wedge B, B}{\Gamma, A \wedge B} \qquad \dfrac{\Gamma, \neg A \vee \neg B, \neg A}{\Gamma, \neg A \vee \neg B}}{\Gamma}$$

This can be replaced by the following derivation, in which the original cut is reduced by ones with smaller cut rank or same cut rank and smaller natural sum of the heights of the subderivations:

$$\dfrac{\dfrac{\Gamma, A \wedge B, A \qquad \Gamma, \neg A \vee \neg B}{\Gamma, A} \qquad \dfrac{\Gamma, \neg A \vee \neg B, \neg A \qquad \Gamma, A \wedge B}{\Gamma, \neg A}}{\Gamma}$$

By systematically carrying out reduction steps like the above, one can eventually eliminate all cuts. More formally, one can show that from $\vdash^{\alpha}_{n+1} \Gamma$ it follows $\vdash^{2^{\alpha}}_{n} \Gamma$.

Therefore we have that if Peano arithmetic proves falsity, i.e. the empty sequent, then $\vdash^{\beta}_{n} \emptyset$ for $\beta = \underbrace{2^{2^{\cdot^{\cdot^{\rho}}}}}_{n \text{ times}}$ for some $\rho < \omega + \omega$. It follows that $\beta < \epsilon_0$. Since a cut-free proof of the empty sequent can only end by a repetition rule, it follows by transfinite induction up to ϵ_0 that there is no cut free proof of the empty sequent, and therefore no proof of an inconsistency in Peano Arithmetic.

One can formulate the above argument in PRA extended by the principle of transfinite induction up to ϵ_0 over quantifier-free formulae. The most elegant proof is due to Buchholz [Buchholz, 1991] (see as well [Michelbrink, 2000] for an extension to Kripke-Platek set theory plus Π_3-reflection). Buchholz introduces a primitive recursive notation system for infinitary derivations as follows: He starts with a notation system for proofs in Peano Arithmetic. Furthermore, for each lemma in the cut elimination proof he introduces a notation, which takes one or more infinitary derivations corresponding to the assumptions of that lemma, and has as result an infinitary derivation corresponding to the conclusion. Then he computes for every notation the last formula, the last rule, the cut rank, the height, and notations for the derivations of the premises. The notations for the subderivations might be longer than that of the derivation itself. The functions computing these results are primitive-recursive, and one can show in PRA that the computed derivations of the premises of a rule actually compute the premises of the last rule, have a smaller height and that the condition on the cut rank is fulfilled. We need now quantifier-free transfinite induction up to ϵ_0 (and this is the only place where this principle is needed) in order to show that there is no derivation of the empty sequent, i.e. that PA is consistent. Note that in this proof the use of the principle of quantifier-free transfinite induction up to ϵ_0 is concentrated in the last step of the proof.

Proof-theoretic strength. It follows therefore that PRA plus transfinite induction up to ϵ_0 proves the consistency of PA. Since PRA can be embedded into PA, it follows that PA does not prove transfinite induction up to ϵ_0. We have indicated above that PA proves transfinite induction up to every ordinal less than ϵ_0 (w.r.t. the ordinal notation system we use). Therefore it follows that, for the ordinal notation system chosen, ϵ_0 is the supremum of all ordinals, up to which transfinite induction can be shown.

One can obtain as well a sharper result, which refers to arbitrary ordinal notation systems. It can be shown that, if we add to PA one free predicate symbol (i.e. a symbol for a set X with no further axioms, but with all the logic and equality rules and the principle of induction extended to formulas containing X), then the supremum of the ordinals α, s.t. there exists an ordinal notation system of order type α and we can prove transfinite induction w.r.t. the free predicate in it, is ϵ_0.

For the above reasons, ϵ_0 is called the proof theoretic strength of Peano Arithmetic. Note that we were referring to two slightly different notions: one is referring to the limit w.r.t. a canonical notation system. The other one refers to the principle of transfinite induction for a new predicate X in an extended language, but w.r.t. arbitrary ordinal notation systems. For most theories considered, both notations coincide, and therefore the limit of transfinite induction provable in a theory T, understood in one of the two ways above, is called the proof theoretic strength $|T|$ of T.

In ordinal-theoretic proof theory, the techniques of Gentzen were extended further and the strength of increasingly strong theories was developed. One open question is what is meant by a canonical ordinal notation system, and one usually uses the notion "natural ordinal notation system". No conclusive answer has been found, and it might be in principal impossible to characterise all natural ordinal notation systems – if one has a mathematical precise notation, it is likely that one can diagonalise over it and then find a ordinal notation system, which is intuitively natural, but not covered by the definition. However, the standard ordinal notation systems used in ordinal theoretic analyses are regarded in general as natural ones, and it might be that this is the right approach to the notion of naturalness: to develop strong ordinal notation systems and investigate afterwards, whether they can be regarded as natural ones.

3 A Proof-Theoretic Programme

A foundational programme. What Gentzen has achieved, is to reduce the consistency of Peano Arithmetic to the principle of transfinite induction up to ϵ_0. The well-foundedness proof up to ϵ_0 is very perspicuous, and this gives strong evidence to the fact that PA is consistent. This proof together with other analyses of PA has contributed to the fact that not many people nowadays still have doubts about the consistency of PA, even so we cannot prove that fact.

In ordinal theoretic proof theory increasingly strong theories were analysed, with increasingly complicated ordinal notation systems. The corresponding well-foundedness proofs became less and less perspicuous. Although by Gödel's second incompleteness theorem, a real reduction of the principles needed in order to prove the consistency of a theory is not possible, one could at least hope for a reduction to principles which are more evident. By carrying out an ordinal analysis alone that has not necessarily been achieved – what one obtains is the concentration of the consistency strength to the well-ordering of an ordinal notation system. Therefore the author believes that the determination of the proof theoretic strength should only be a first step in a proof theoretic analysis. A second step is required, namely to develop theories, in which we can prove the well-foundedness of the ordinal notation system and therefore (assuming that PRA can be embedded into those theories) the consistency of the theories involved. And such theories should be formulated in such a way that in every proof step of such a theory there is direct evidence that the truth of the premises implies the truth of the conclusion. This insight needs to be, because of Gödel's theorem, a philosophical argument. If such an analysis has been carried out, we know that everything derived in such a theory is correct. Such theories will then be a substitute for Hilbert's finitary methods, we can call them *extended finitary methods*. The up to now most successful theories used for this purpose seem to be extensions of Martin-Löf type theory, and the philosophical argument are meaning explanations as given by Per Martin-Löf. An alternative approach taken have been Feferman's theories of explicit mathematics, but unfortunately up to now only partial philosophical analyses have been carried out for those theories. In the following, we will discuss how strong extensions of type theory are developed and how the corresponding reductions are carried out. However, in this article we will not investigate meaning explanations in detail.

Applications. Philosophical reasons are one motivation for following the programme described. Another major motivation for it comes from applications.

Martin-Löf type theory can be considered as a functional programming language, and there exists one fully developed functional language, Cayenne [1] based on dependent types. When considering Martin-Löf type theory as it stands, we already have data structures available which do not occur in other languages or are not often used there: we have the W-type, which represents infinitely branching trees (instances of this type can be represented in other languages, it is only when we want to introduce the W-type as a general concept that we need dependent types), and we have universes, which are types, the elements of which represent types. Both principles increase the strength of type theory substantially, and this is one of the reasons, why they were added. When investigating strong extensions of type theory, which allow to prove the well-foundedness of strong ordinal notation systems, we are searching for new data types, which we hope will be of use in general programming as well. One result of this programme is the development of the data type of inductive-recursive definitions (see Sect. 7), which uses principles developed first in the context of the Mahlo universe as part of this programme.

4 Type Theory with W-Type and Atom

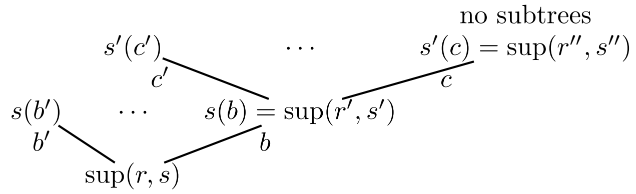
One step in this proof theoretic programme was the proof theoretic analysis of the theory $\mathrm{ML}_1\mathrm{W}$ of Martin-Löf type theory with one standard universe and one W-type by the author ([Setzer, 1993], [Setzer, 1998]). Independently, E. Griffor and M. Rathjen [Griffor and Rathjen, 1994] have analysed a slightly weaker theory, which has one standard universe closed under the W-type, and additionally Aczel's type V of iterated sets, but in which the formation of the W-type is restricted to elements of the universe only. Both theories do not include the logical framework, so there is no type Set and there are no judgements of the form $A : \mathrm{Set}$ or $A = B : \mathrm{Set}$. In our analysis, we obtained an upper bound for the proof theoretic strength of $\mathrm{ML}_1\mathrm{W}$ with extensional equality, and proved the lower bound for $\mathrm{ML}_1\mathrm{W}$ with intensional equality. Since both bounds coincide, this showed that both theories have the same strength.

In this section we make a first step towards this result, and consider the strength of type theory with W-type and a microscopic universe atom. In the Sect. 5 we will then look at the full theory.

The W-type. The assumptions for forming the W-type are $A : \mathrm{Type}$ and $x : A \Rightarrow B(x) : \mathrm{Type}$, from which one can form $(\mathrm{W}x : A.B(x)) : \mathrm{Type}$. (When writing $B(x)$ we mean that B might depend on x, and that later on $B(r)$ means the result of substituting x by r in B.) Its canonical elements are introduced by the following introduction rule:

$$\frac{\Gamma \Rightarrow r : A \qquad \Gamma \Rightarrow s : B(r) \rightarrow (\mathrm{W}x : A.B(x))}{\Gamma \Rightarrow \mathrm{sup}(r, s) : (\mathrm{W}x : A.B(x))}$$

An element $\mathrm{sup}(r, s)$ of $\mathrm{W}x : A.B(x)$ is a tree with label r and subtrees $s(b)$ for $b : B(r)$. The elements of $\mathrm{W}x : A.B(x)$ are therefore trees with branching degrees $(B(a))_{a:A}$. They can be visualised as follows (assume in this picture that $B(r'')$ is empty, therefore $\mathrm{sup}(r'', s'')$ has no subtrees $s''(y)$ for $y : B(r'')$):



The elimination rule for the W-type formalises induction over trees, which expresses that we have the "least type closed under the introduction rule", and that therefore elements of $\mathrm{W}x : A.B(x)$ are well-founded trees.

In standard models of the $\mathrm{W}x : A.B(x)$, one can define the height $\mathrm{height}(t)$ of a tree t, which is an ordinal, inductively as follows: $\mathrm{height}(\mathrm{sup}(r, s)) = \mathrm{sup}\{s(b) + 1 \mid b \in B'(b)\}$, where $B'(b)$ is the interpretation of $B(x)$ for $x = b$. The height of $\mathrm{W}x : A.B(x)$ (in a standard model) is defined as the supremum of the heights of its elements (in a standard model).

Finite types. In all type theories which we will consider in this article, we have types with finitely many elements, which we denote for the sake of simplicity by $\{0, 1, \ldots, n-1\}$ (or $\emptyset$ in case of $n = 0$). (The real type theoretic notation is N_n, having elements called i_k or A^k_i.) For the sake of readability, we sometimes write False for $\emptyset$ (the formula with no proof), True for $\{0\}$ (the formula with exactly on proof), tt for the element of True, Bool for $\{0, 1\}$, and true, false for the elements of Bool.

Atom and the type theory $\mathrm{MLW}_{\mathrm{atom}}$**.** We add a type constructor atom with typing $x : \mathrm{Bool} \Rightarrow \mathrm{atom}(x) : \mathrm{Type}$ to our type theory, together with equality rules $\mathrm{atom}(\mathrm{true}) = \mathrm{True}$ and $\mathrm{atom}(\mathrm{false}) = \mathrm{False}$. So atom takes a Boolean value and translates it into a formula corresponding to its value.

When looking later at universes, we will see that we have here the case of a microscopic universe Bool with two elements representing True and False. Without any universe at all one can show that one can model type theory in such a way that the interpretation of a type $A(x)$ does not depend on x. If we interpret in such a model $\mathrm{W}x : A.B(x)$ in a standard way, then we have that the interpretation of $B(x)$ is either empty for all x, or non-empty for all x. If $B(x)$ is empty for all x, then trees have no subtrees, hence have height 0. If the interpretation of $B(x)$ is non-empty for all x, the interpretation of $\mathrm{W}x : A.B(x)$ is empty, since in order to form an element $\sup(r, s)$, we need to have defined before $s(y) : (\mathrm{W}x : A.B(x))$ for one of the elements y of $B(x)$. In both cases, the height of $\mathrm{W}x : A.B(x)$ is 0. Type theory without a universe is known to be very weak – it does not even show Peano's fourth axiom, namely that in the type N of natural numbers, 0 is different from $\mathrm{S}(n)$.

We call the type theory having standard types, atom, and the W-type, $\mathrm{MLW}_{\mathrm{atom}}$.

Admissibles and $\aleph^{\mathrm{rec}}_\alpha$**.** In order to understand the strength of type theories involving the W-type, we need the notion of recursively regular or (an equivalent name) admissible ordinals, a notion originating from generalised computability theory. Classical computability is the theory of computable functions, where computable means computable by any mechanical device. There, one has developed a schema for defining so called partial recursive functions $f : \mathbb{N}^n \xrightarrow{\sim} \mathbb{N}$. Partial as expressed by the symbol $\xrightarrow{\sim}$ means that $f : \mathrm{dom}(f) \rightarrow \mathbb{N}$ for some set $\mathrm{dom}(f) \subseteq \mathbb{N}^n$. Recursive functions are total partial recursive functions, i.e. functions $f : \mathbb{N}^n \xrightarrow{\sim} \mathbb{N}$ s.t. $\mathrm{dom}(f) = \mathbb{N}^n$. One assumes that the set of partial recursive functions coincides with the set of computable functions. Note that computable is not a mathematical notion, therefore the fact that all computable functions are partial recursive is not a mathematical statement and can therefore not be proved – however, most researchers believe that all computable functions are partial recursive. One can encode partial recursive functions as natural numbers, and writes $f = \{e\}^n$, if $f : \mathbb{N}^n \xrightarrow{\sim} \mathbb{N}$ has code e.

In generalised computability theory, one extends the schema for defining partial recursive functions to ordinals, and obtains the notion of κ-partial recursive functions $f : \kappa^n \xrightarrow{\sim} \kappa$ (where κ is an ordinal). For a detailed description, see for instance Chapter VIII of [Hinman, 1978]. $f : \alpha \rightarrow \kappa$ is κ-recursive in parameters $< \kappa$, if $f = \lambda\gamma.g(\gamma, \beta_0, \ldots, \beta_{n-1})$ for some $\beta_i < \kappa$ and κ-partial recursive $g : \kappa^{n+1} \rightarrow \kappa$, and if $f(\gamma)$ is defined for all $\gamma < \alpha$. A limit ordinal κ is admissible, if it is closed under the formation of suprema of κ-recursive functions in parameters $< \kappa$: If $\alpha < \kappa$ and $f : \alpha \rightarrow \kappa$ is κ-recursive in parameters $< \kappa$, then $\sup_{\gamma<\alpha} f(\gamma) < \kappa$.

One defines now by recursion on the ordinal α ordinals $\aleph^{\mathrm{rec}}_\alpha$ as follows: $\aleph^{\mathrm{rec}}_0 = \omega$, the least infinite ordinal and least admissible ordinal; $\aleph^{\mathrm{rec}}_{\alpha+1}$ is the least admissible ordinal above $\aleph^{\mathrm{rec}}_\alpha$; and if λ is a limit ordinal, then $\aleph^{\mathrm{rec}}_\lambda = \sup_{\beta<\lambda} \aleph^{\mathrm{rec}}_\beta$. Usually, for limit ordinals λ, $\aleph_\lambda^{\mathrm{rec}}$ is not an admissible; but $\aleph_0^{\mathrm{rec}}$ and $\aleph_{\alpha+1}^{\mathrm{rec}}$ are admissible. Every admissible ordinal is of the form $\aleph_\alpha^{\mathrm{rec}}$ for some α.

Admissible ordinals are the recursive analogue of regular cardinals. An ordinal κ is a regular cardinal, if κ is not the supremum of $\beta < \kappa$ many ordinals $< \kappa$: There exists no $\beta < \kappa$ and $f : \beta \to \kappa$ s.t. $\kappa = \sup_{\alpha<\beta} f(\alpha)$. Note that here f can be arbitrary, whereas for admissibles it had to be recursive, having parameters $< \kappa$.

Admissibles and the W-type. The heights of W-types are closely related to admissible ordinals. Assume, that terms are encoded as natural numbers and assume a standard interpretation A' of A and of $B(x)$ for $x \in A'$ as $B'(x)$, where A' and $B'(x)$ are sets of natural numbers encoding the elements of this type. (A rough idea would be to interpret N as $\mathbb{N}$ and $A \to B$ as $\{e \mid \forall x \in A'.\{e\}(x) \in B'\}$. The detailed model is more complicated – in fact, more precisely one has to take equality of terms into account and interpret a type as a set of pairs of natural numbers, where a pair (n, m) being an element of the interpretation of A means that n and m are codes for equal elements of A). Then the standard interpretation of $\mathrm{W}x.A.B(x)$ is the least set C s.t., if $k \in A'$ and $\forall l \in B'(k).\{e\}^1(l) \in C$, then $\pi(k,e) \in C$. Here $\pi : \mathbb{N}^2 \to \mathbb{N}$ is the standard encoding of pairs of natural numbers as natural numbers. So C is the least set of *recursive* trees with branching degrees $B'(k)$ for $k \in A'$.

If we take $A = \{0,1\}$ and $B(0) = \emptyset$, $B(1) = \{0\}$ and let $\mathrm{O}_0 := (\mathrm{W}x : A.B(x))$, we can see that the elements of the interpretation of O_0 are of the form $\sup(1, \lambda x. \sup(1, \lambda x. \sup(1, \cdots \sup(0, \lambda x.e)) \cdots)))$ (more precisely, they are equal modulo the η-rule to such a tree). These trees have heights n for $n \in \omega$, so (the standard interpretation of) O_0 has height $\omega = \aleph_0^{\mathrm{rec}}$, the least admissible ordinal.

If we take $A = \{0,1,2\}$ and $B(2) = \mathrm{O}_0$, otherwise $B(x)$ as before, we obtain Kleene's O (in the usual definition, one has $B(2) = \mathbb{N}$ instead of O_0), which has height $\aleph_1^{\mathrm{rec}}$, the second admissible ordinal.

In general, the nth admissible ordinal $\aleph_n^{\mathrm{rec}}$ is the height of (the standard interpretation of) $\mathrm{O}_n := \mathrm{W}x : \{0,1,\ldots,n+1\}.B(x)$, where $B(0) = \emptyset$, $B(1) = \{0\}$, $B(k) = \mathrm{O}_{k-2}$ for $k > 1$.

We can obtain a more uniform version of this, by replacing $B(0)$ by $\mathrm{O}_{-2} := \mathrm{W}x : \emptyset.B(x)$, which is empty, and by replacing $B(1)$ by $\mathrm{O}_{-1} := \mathrm{W}x : \{0\}.B(x)$, where $B(0) = \mathrm{O}_{-2}$, which contains only trees of height 0. Then we obtain $\mathrm{O}_n = \mathrm{W}k : \{0,\ldots,n+1\}.\mathrm{O}_{k-2}$ for $n \geq -2$.

Lower bound for $|\mathrm{MLW}_{\mathrm{atom}}|$. Using the admissibles $(\aleph_n^{\mathrm{rec}})_{n\in\omega}$, one can form an ordinal notation system. The ordinal notations will be terms for expressions formed from 0, some standard operations on ordinals, and so called collapsing function ψ_κ for $\kappa > \omega$ admissible, mapping ordinals to ordinals $< \kappa$. We won't introduce ψ_κ in detail. ψ_κ collapses ordinals into the interval $[0, \kappa[$ and is weakly monotone ($\alpha < \beta \to \psi_\kappa(\alpha) \leq \psi_\kappa(\beta)$). Ordinal notation systems like this are usually constructed using regular cardinals instead of admissibles, but with some extra work, which has been carried out in some cases (e.g. [Schlüter]), one can see that one can replace those regular cardinals by their recursive analogues. One can simulate now these functions by replacing ordinals by elements of O_n and then show that the type theory in question proves transfinite induction up to $\psi_{\aleph_1^{\mathrm{rec}}}(\aleph_n^{\mathrm{rec}})$, which in the limit reaches $\psi_{\aleph_1^{\mathrm{rec}}}(\aleph_\omega^{\mathrm{rec}})$. This provides a lower bound for this theory. Details for the lower bound of a closely related type theory can be found in [Setzer, 1994].

Kripke-Platek set theory. For an upper bound, one interprets type theory into an extension of Kripke-Platek set theory, which we will introduce in the following.

This will show that the formation of O_n essentially exhausts the strength of type theory with W-type and a microscopic universe.

Kripke-Platek set theory KP is a weak version of set theory, which is closely connected to admissibles. It was developed by Kripke [Kripke, 1964] and Platek [Platek, 1966]. KP is obtained from standard ZF-set theory essentially as follows: one omits the existence of the power set of any set; one omits the infinity axiom, claiming the existence of an infinite set; one restricts the formation of $\{x \in a \mid \varphi(x)\}$ to Δ_0-formulae φ (i.e. formulae in which all quantifiers are bounded, i.e. of the form $\forall x \in a$ or $\exists x \in a$) – the corresponding principle is called Δ_0-separation; one restricts the so called collection principle expressing that if $\forall x \in a.\exists y.\varphi(x,y)$ then there exists a b s.t. $\forall x \in a.\exists y \in b.\varphi(x,y)$ to Δ_0-formulae φ – the new principle is correspondingly called Δ_0-collection. The precise axiomatisation can be found in the detailed monograph [2] on Kripke-Platek set theory.

The relationship to admissible ordinals is as follows: There exists an operation (called constructible hierarchy) which maps ordinals α to sets L_α, where $\mathrm{L}_0 = \emptyset$, $\mathrm{L}_\lambda = \bigcup_{\alpha<\lambda} \mathrm{L}_\alpha$ for λ limit ordinal, and $\mathrm{L}_{\alpha+1}$ is the result of applying certain operations for forming new sets (such as forming pairs, unions of sets or the domain and range of relations) to sets in $\mathrm{L}_\alpha \cup \{\mathrm{L}_\alpha\}$. Now one can show that L_α is a model of KP iff α is admissible, and introduce a notion of "admissible set" (as opposed to "admissible ordinal"), which holds for a set iff it is of the form L_α for an admissible α.

KP in Proof Theory. In proof theory, extensions of KP by axioms claiming the existence of many admissibles are often used as reference theories (see especially the monograph [Jäger, 1986]). More precisely, one adds a predicate $\mathrm{Ad}(x)$ standing for "x is an admissible set", axioms stating that $\mathrm{Ad}(x)$ implies that x is a model of KP, and axioms claiming the existence of elements fulfilling this predicate (for instance a fixed number of admissibles, arbitrarily finitely many admissibles, or axioms claiming the existence of admissibles closed under certain operations). Many of these extensions have been analysed proof theoretically, and are used as reference theories. The strength of other theories is usually obtained by comparing them with those reference theories.

General Technique for Developing Upper Bounds of Martin-Löf Type Theory. Our standard technique for determining upper bounds for the strength of variants of Martin-Löf type theories is to model them in extensions of KP having the same proof-theoretic strength. Assume that we have done this for a variant of type theory called $\mathrm{MLTT}^{\mathrm{var}}$ and a variant of KP called $\mathrm{KP}^{\mathrm{var}}$. Then the above provides us with a model of $\mathrm{MLTT}^{\mathrm{var}}$ in a set theory of minimal strength. The main purpose for developing this model is to obtain an upper bound for the proof theoretic strength of $\mathrm{MLTT}^{\mathrm{var}}$. We will be able to show using this model that, if $\mathrm{MLTT}^{\mathrm{var}}$ proves transfinite induction up to a certain ordinal (more precisely up to an ordinal notation), then the same holds for $\mathrm{KP}^{\mathrm{var}}$. Therefore $|\mathrm{MLTT}^{\mathrm{var}}| \leq |\mathrm{KP}^{\mathrm{var}}|$. If the variants of KP chosen have been analysed proof theoretically, one obtains a concrete ordinal $\alpha = |\mathrm{KP}^{\mathrm{var}}|$, and we have $|\mathrm{MLTT}^{\mathrm{var}}| \leq \alpha$.

Note that KP and its extensions are classical theories and therefore not constructive. However, apart from the proof theoretic result, which measures the strength of the theory, we believe that modelling a theory in a set theory of minimal strength provides additional insight into what can be achieved in type theory. Furthermore, once we have shown the other direction, i.e. that $|\mathrm{MLTT}^{\mathrm{var}}| \geq \alpha = |\mathrm{KP}^{\mathrm{var}}|$, one can easily show that $\mathrm{MLTT}^{\mathrm{var}}$ shows the consistency of approximations of $\mathrm{KP}^{\mathrm{var}}$, such that each proof in $\mathrm{KP}^{\mathrm{var}}$ can be formalised in one of these approximations. This provides us with a constructive understanding of $\mathrm{KP}^{\mathrm{var}}$.

A more refined analysis often shows as well that $\mathrm{MLTT}^{\mathrm{var}}$ and $\mathrm{KP}^{\mathrm{var}}$ show the same arithmetic Π_2-sentences, i.e. the same formulae $\forall x \in \mathbb{N}.\exists y \in \mathbb{N}.\varphi(x,y)$, where $\varphi(x,y)$ is a quantifier-free arithmetic formula. Such formulae can be considered as the specifications of programs, and from each proof of such a formula in $\mathrm{MLTT}^{\mathrm{var}}$ one obtains a program computing a function $f : \mathbb{N} \to \mathbb{N}$ s.t. $\forall x \in \mathbb{N}.\varphi(x, f(x))$ holds. So one can say that the provably total programs in $\mathrm{KP}^{\mathrm{var}}$ and $\mathrm{MLTT}^{\mathrm{var}}$ coincide.

Upper bound for $|\mathrm{MLW}_{\mathrm{atom}}|$. In $\mathrm{MLW}_{\mathrm{atom}}$, we can define O_n, which corresponds to the possibility of forming finitely many admissible, and we claimed that this essentially exhausts the strength of this theory. A variant of Kripke-Platek set theory, which allows to form finitely many admissibles, is the theory KPl. Its standard model is $\mathrm{L}_{\aleph_\omega^{\mathrm{rec}}}$. $\aleph_\omega^{\mathrm{rec}}$ is not an admissible, so the standard model of KPl doesn't fulfil the axioms KP, and we can't include all axioms of KP into those of KPl. What is omitted is that the set theoretic universe fulfils Δ_0-collection – however, restricted to any a s.t. $\mathrm{Ad}(a)$, Δ_0-collection will hold. Further one demands that every set is contained in an admissible set. So we can form a sequence $\mathrm{Ad}_0 = \emptyset$, and Ad_{n+1} as being one admissible above Ad_n for Meta-$n \in \omega$, but without Δ_0-collection it is not possible to form this sequence inside the theory. If one could, then one could form $\mathrm{Ad}_\omega := \bigcup_{n\in\omega} \mathrm{Ad}_n$ and hence an admissible above Ad_ω.

One can now form a model of $\mathrm{MLW}_{\mathrm{atom}}$ in KPl. Essentially, we interpret each type $A(x)$ as an element of Ad_n for some n uniformly for all x. The main step is to interpret $\mathrm{W}x : A.B(x)$. If A and $B'(x)$ are interpreted as A' and $B'(x)$, which are elements of Ad_n, and if κ is the supremum of the ordinals in Ad_n, then $\mathrm{W}x : A.B(x)$ is interpreted by iterating a certain operator (which forms trees with subtrees having been formed before) κ many times. The result is an element of Ad_{n+1}.

In general, in order to obtain a fixed point of such kind of operators (more precisely Σ_1-operators), we need to iterate the operator up to an admissible ordinal corresponding to the least admissible set containing all set parameters used by this operator.

5 Type Theory with W-Type and one Universe

In this section we consider the theory $\mathrm{ML}_1\mathrm{W}$, which is Martin-Löf type theory with W-type and a universe closed under the W-type.

Universes and the type theory $\mathrm{ML}_1\mathrm{W}$. A universe is the type theoretic formalisation of a type the elements of which are types. This suggests that a universe should be a type U, s.t. for $x : \mathrm{U}$ we have $x : \mathrm{Type}$, as in so called "universes à la Russell". However, this causes conceptual problems, since in this step a term x becomes a type, and changes therefore its category. (Note that in a term model, a term is interpreted by itself, whereas a type is interpreted for instance as a set of terms; something similar happens in meaning explanations.) In order to avoid such conceptual problems (note that Martin-Löf type theory is considered as well as a foundation of mathematics), it is better to keep terms and types separted and work with "universes à la Tarski": a universe is a type U, the elements of which are codes for types. We need therefore an additional decoding function T, the typing of which is giving by the judgement $x : \mathrm{U} \Rightarrow \mathrm{T}(x) : \mathrm{Type}$. If x is an element of U, i.e. a code for type, $\mathrm{T}(x)$ is the type, x denotes.

The microscopic universe atom introduced in Sect. 4 is a universe in this sense: the underlying set is Bool and the decoding function is atom: we have $x : \mathrm{Bool} \Rightarrow \mathrm{atom}(x) : \mathrm{Set}$.

A standard universe U is a universe closed under all standard type constructions. This means for instance that there is a code $\widehat{\mathrm{N}} : \mathrm{U}$ (i.e. a constructor $\widehat{\mathrm{N}}$ of U) for the type N of natural numbers, so $\mathrm{T}(\widehat{\mathrm{N}}) = \mathrm{N}$. Similarly, we have codes for the types $\{0, \ldots, l-1\}$. Standard universes are closed under $+$, Π, Σ, so for instance we have that, if $r : \mathrm{U}$, and s is s.t. $x : \mathrm{T}(r) \Rightarrow s(x) : \mathrm{U}$, then $(\widehat{\Sigma} x : r.s(x)) : \mathrm{U}$ and $\mathrm{T}(\widehat{\Sigma} x : r.s(x)) = \Sigma x : \mathrm{T}(r).\mathrm{T}(s(x))$ for a constructor $\widehat{\Sigma}$ of U.

$\mathrm{ML}_1\mathrm{W}$ has now the standard types, the W-type and a standard universe U, which is closed additionally under the W-type: if r and $s(x)$ are as for the premises of the $\widehat{\Sigma}$-introduction rule above, then $(\widehat{\mathrm{W}} x : r.s(x)) : \mathrm{U}$ and $\mathrm{T}(\widehat{\mathrm{W}} x : r.s(x)) = \mathrm{W} x : \mathrm{T}(r).\mathrm{T}(s(x))$ for a constructor $\widehat{\mathrm{W}}$ of U.

Strength of $\mathrm{ML}_1\mathrm{W}$. We will in this section provide some intuition concerning the strength of $\mathrm{ML}_1\mathrm{W}$, more details can be found in Appendix B. In the presence of a universe closed under the W-type, we can define by induction on $n : \mathrm{N}$ codes $\widehat{\mathrm{O}}_n : \mathrm{U}$ for the finitely iterated trees O_n, and therefore form $\mathrm{O}_{\omega+1} := \mathrm{W} x : \mathrm{N}.\mathrm{T}(\widehat{\mathrm{O}}_n)$, which is a W-type of height $\aleph^{\mathrm{rec}}_{\omega+1}$, the first admissible bigger than $\aleph^{\mathrm{rec}}_n$ for $n \in \omega$. $\mathrm{O}_{\omega+1}$ can be represented by an element of U. We can iterate the above process as well over any W-type which has a code in U. By iterating it over $\mathrm{W}_1 := \mathrm{O}_1$, we can form trees of height $\aleph^{\mathrm{rec}}_\alpha$ for $\alpha < \aleph^{\mathrm{rec}}_1$, and then form a W-type W_1 of height $\aleph^{\mathrm{rec}}_{\aleph^{\mathrm{rec}}_1+1}$. Doing the same with W_2 instead of O_1, we reach $\aleph^{\mathrm{rec}}_{\aleph^{\mathrm{rec}}_{\aleph^{\mathrm{rec}}_1+1}+1}$ by a tree W_3. Continuing this process, we obtain W-types W_n for $n \in \omega$, s.t. the supremum of their heights is $\Lambda^{\mathrm{rec}} = \sup_{n\in\omega} \underbrace{\aleph^{\mathrm{rec}}_{\aleph^{\mathrm{rec}}_{\aleph^{\mathrm{rec}}_{\cdot_{\cdot_{\cdot_1}}}}}}_{n \text{ times}}$. Λ^{rec} is the first (non-admissible) fixed point of $\lambda\alpha.\aleph^{\mathrm{rec}}_\alpha$. W_n can be formed as elements of U, and therefore we can form a W-type inside U, which has as height the least admissible $\aleph^{\mathrm{rec}}_{\Lambda^{\mathrm{rec}}+1}$ above Λ^{rec}. In general, if we have formed a W-type of height α, then we can iterate the formation of O_γ for $\gamma < \alpha$ as elements of U and form therefore a W-type of height $\aleph^{\mathrm{rec}}_{\alpha+1}$. So the height of the W-types, we can form as elements of U, must be an admissible κ, which is closed under $\lambda\beta.\aleph^{\mathrm{rec}}_\beta$. In order to obtain this is property, it suffices to demand that for every $\alpha < \kappa$ there exists an admissible π s.t. $\alpha < \pi < \kappa$, and admissibles with this property are called *recursively inaccessible ordinal.* Admissible sets corresponding to such ordinals are called *recursively inaccessible sets.* Recursively inaccessible ordinals are the recursive analogue of strongly inaccessible cardinals. One can form a model of ZFC (Zermelo-Fraenkel set theory with axiom of choice), s.t. the supremum of ordinals in this model is the first strongly inaccessible cardinal, therefore the existence of strongly inaccessible cardinals cannot be shown in ZFC.

So we have now obtained some intuition that we can form inside the universe W-types, s.t. the supremum of their heights reaches (in the standard model) $\mathrm{I}^{\mathrm{rec}}$, the least recursively inaccessible ordinal. We can form further W-types on top of U, which are no longer elements of U, and which have heights $\kappa_n := \aleph^{\mathrm{rec}}_{\mathrm{I}^{\mathrm{rec}}+n}$, the nth admissible above $\mathrm{I}^{\mathrm{rec}}$. We can do this only by Meta-induction over n, and the supremum of the heights of W-types, we can form, is $\mathrm{I}^+ := \aleph^{\mathrm{rec}}_{\mathrm{I}^{\mathrm{rec}}+\omega} = \sup_{n\in\omega} \kappa_n$.

Upper bound for $|\mathrm{ML}_1\mathrm{W}|$. An upper bound for the strength of $\mathrm{ML}_1\mathrm{W}$ can be obtained by modelling it in a theory KPI^+, which has standard model $\mathrm{L}_{\mathrm{I}^+}$. In KPI^+ we have one recursively inaccessible set Ad_I, and constants for finitely many admissibles above it. (Alternatively, one can define it as the theory KPl plus the existence of one recursively inaccessible.) We can model U by iterating an operator (which essentially forms new sets representing $\widehat{\mathrm{W}} x : a.b(x)$, $\widehat{\Sigma} x : a.b(x)$ etc. from sets previously defined). In order to reach a fixed point, we have to iterate this operator up to $\mathrm{I}^{\mathrm{rec}}$, which is the union of all ordinals in the recursively inaccessible set Ad_I. If we form the representation of $\mathrm{W} x : A.B(x)$ from sets A and $B(x)$ s.t. A

and $B(x)$ are elements of L_α, one can see that the representation of $\mathrm{W}x : A.B(x)$ is an element of $\mathrm{L}_{\alpha^{++}}$, where α^{++} is the second admissible above α. A fixed point of this operator is obtained if we iterate it up to an admissible, which is closed, in order to accommodate for the formation of $\mathrm{W}x : A.B(x)$, under the step from α to α^{++}. That's why we need a recursively inaccessible ordinal, in order to reach the fixed point. The interpretation of U is an element of L_{κ_1}. We can form W-types making use of U and sets constructed from it, and each W-type construction corresponds to iterating an operator up to the next admissible. An n-times nested W can be interpreted by iterating an operator up to κ_{n+1}, and the interpretation is an element of κ_{n+2}. This way we can model it in KPI^+.

A lower bound for $|\mathrm{ML_1W}|$ can be obtained by carrying out a well-ordering proof for an ordinal notation system which has the strength of KPI^+. Details for this proof can be found in Appendix B.

6 The Mahlo Universe

In proof theory, the next major step taken after treating theories of strength KPI was the analysis of KPM, Kripke Platek set theory plus the recursive Mahloness of the set theoretic universe, by Michael Rathjen [Rathjen, 1991]. An ordinal M is a *recursively Mahlo ordinal*, if it is admissible, and if for all $f : M \to M$, which are M-recursive with parameters in M, there exists an admissible $\kappa < M$ s.t. $\forall\alpha < \kappa.f(\alpha) < \kappa$. If one replaces in this definition "admissible" by "recursively inaccessible", one obtains an equivalent definition. *Recursively Mahlo sets* are sets of the form L_M for recursively Mahlo ordinals M. In order to extend dependent type theory by a principle which reaches the strength of KPM, the author introduced in [Setzer, 2000] a type theory with one Mahlo universe. This type theory is substantially stronger than Martin-Löf type theory extended by standard types (including W-type and standard universes). We will in the following develop the rules for this type theory from the definition of recursively Mahloness.

In order to translate the Mahlo principle into type theory, we replace M by a family of types (V, T), i.e. we have $\mathrm{V} : \mathrm{Type}$ and $x : \mathrm{V} \Rightarrow \mathrm{T}(x) : \mathrm{Type}$.

A recursively inaccessible ordinal corresponds in type theory to a standard universe, so we add rules expressing that (V, T) is closed under the universe constructions and under the W-type.

The function $f : M \to M$ in the definition of recursively Mahloness can be translated as having a function $f : \mathrm{Fam}(\mathrm{V}, \mathrm{T}) \to \mathrm{Fam}(\mathrm{V}, \mathrm{T})$, where $\mathrm{Fam}(\mathrm{V}, \mathrm{T}) := \Sigma a : \mathrm{V}.\mathrm{T}(a) \to \mathrm{V}$ is the set of families of sets in V.

The existence of a recursively inaccessible κ can be translated into the existence of a subuniverse $(\mathrm{U}_f, \mathrm{s}_f)$ of (V, T). This means that we have $\mathrm{U}_f : \mathrm{Type}$ and $\mathrm{s}_f : \mathrm{U}_f \to \mathrm{V}$, which interprets each code in U_f as a code in V. For $a : \mathrm{U}_f$ we define $\mathrm{S}_f(a) := \mathrm{T}(\mathrm{s}_f(a))$, which is the type corresponding to the code a.

We demand that $(\mathrm{U}_f, \mathrm{S}_f)$ is a standard universe closed under the W-type, and that codes for the standard universe constructions in U_f correspond to codes in V. For instance, $\widetilde{\mathrm{N}} : \mathrm{U}_f$ and $\mathrm{s}_f(\widetilde{\mathrm{N}}) = \widehat{\mathrm{N}}$ for the code $\widehat{\mathrm{N}}$ of N in V. $\mathrm{s}_f : \mathrm{U}_f \to \mathrm{V}$ can now be lifted to a function $\mathrm{s}_f^{\mathrm{Fam}} : \mathrm{Fam}(\mathrm{U}_f, \mathrm{S}_f) \to \mathrm{Fam}(\mathrm{V}, \mathrm{T})$, where $\mathrm{s}_f^{\mathrm{Fam}}(\langle x, y\rangle) = \langle \mathrm{s}_f(x), \lambda y.\mathrm{s}_f(y(x))\rangle$.

That $\forall\alpha < \kappa.f(\alpha) < \kappa.$ will be interpreted as rules expressing that $f : \mathrm{Fam}(\mathrm{V}, \mathrm{T}) \to \mathrm{Fam}(\mathrm{V}, \mathrm{T})$ is reflected by a function $\mathrm{Res} : \mathrm{Fam}(\mathrm{U}_f, \mathrm{S}_f) \to \mathrm{Fam}(\mathrm{U}_f, \mathrm{S}_f)$, i.e. $\mathrm{s}_f \circ \mathrm{Res} = f \circ \mathrm{s}_f$. Res constructs new elements of U_f, and we obtain corresponding constructors by splitting Res into two parts, namely $\mathrm{Res0} : \mathrm{Fam}(\mathrm{U}_f, \mathrm{S}_f) \to \mathrm{U}_f$ and $\mathrm{Res1} : (x : \mathrm{Fam}(\mathrm{U}_f, \mathrm{S}_f)) \to \mathrm{S}_f(\mathrm{Res0}(x)) \to \mathrm{U}_f$. So Res as before is $\lambda x.\langle \mathrm{Res0}(x), \lambda y.\mathrm{Res1}(x, y)\rangle$.

In type theory we split $\mathrm{s}_f \circ \mathrm{Res} = f \circ \mathrm{s}_f$ into two equality rules, one for Res0, namely $\mathrm{s}_f(\mathrm{Res0}(\langle x, y\rangle)) = \pi_0(f(\mathrm{s}_f^{\mathrm{Fam}}(\langle x, y\rangle)))$, and one for Res1, namely $\mathrm{s}_f(\mathrm{Res1}(\langle x, y\rangle, z)) = \pi_1(f(\mathrm{s}_f^{\mathrm{Fam}}(\langle x, y\rangle)))(z)$. Here π_0, π_1 stand for the first and second projection.

Up to now the rules do not reach more strength than $\mathrm{ML}_1\mathrm{W}$, since we could easily model $\mathrm{U}_f := \mathrm{V}$ and $\mathrm{s}_f := \lambda x.x$. Strength is reached by modelling the condition that $\kappa \in M$. This can be modelled as the existence of a constructor $\widehat{\mathrm{U}}$ with argument f of V and the condition $\mathrm{T}(\widehat{\mathrm{U}}_f) = \mathrm{U}_f$. Note that this means that V has now a constructor which depends negatively on V, namely $\widehat{\mathrm{U}} : (f : \mathrm{Fam}(\mathrm{V}, \mathrm{T}) \to \mathrm{Fam}(\mathrm{V}, \mathrm{T})) \to \mathrm{V}$.

In type theory, it is more natural to replace $f : \mathrm{Fam}(\mathrm{V}, \mathrm{T}) \to \mathrm{Fam}(\mathrm{V}, \mathrm{T})$ by two functions $f_0 : (a : \mathrm{V}, b : \mathrm{T}(a) \to \mathrm{V}) \to \mathrm{V}$ and $f_1 : (a : \mathrm{V}, b : \mathrm{T}(a) \to \mathrm{V}, \mathrm{T}(f_0(a, b))) \to \mathrm{V}$. In the same way, one replaces the type of Res0 by $(a : \mathrm{U}_{f_0,f_1}, b : \mathrm{S}_{f_0,f_1}(a) \to \mathrm{U}_{f_0,f_1}) \to \mathrm{U}_{f_0,f_1}$, similarly for Res1. The type of $\langle f_0, f_1\rangle$ is $(\mathrm{Fam}(\mathrm{V}, \mathrm{T}) \to \mathrm{Fam}(\mathrm{V}, \mathrm{T}))' := ((a : \mathrm{V}, b : \mathrm{T}(a) \to \mathrm{V}) \to \mathrm{V}) \times ((a : \mathrm{V}, b : \mathrm{T}(a) \to \mathrm{V}, \mathrm{T}(f_0(a, b))) \to \mathrm{V})$. By $\vec{f} : (\mathrm{Fam}(\mathrm{V}, \mathrm{T}) \to \mathrm{Fam}(\mathrm{V}, \mathrm{T}))'$ we mean in the following that $\langle f_0, f_1\rangle : (\mathrm{Fam}(\mathrm{V}, \mathrm{T}) \to \mathrm{Fam}(\mathrm{V}, \mathrm{T}))'$, similarly for $\vec{g}$.

MLM denotes the type theory with W-type and one Mahlo universe.

Constructive understanding of the Mahlo universe. There are two approaches in order to get a constructive understanding of the Mahlo universe.

The first approach uses partial functions. In order to give meaning explanations, we have to say what it means to be an element of the Mahlo universe, when two elements of the Mahlo universe are equal, and we have to understand for every element a of the Mahlo universe $\mathrm{T}(a)$ as a set. The understanding of the standard constructors for the universe is as usual. For instance, if a is an element of the Mahlo universe and for x in $\mathrm{T}(a)$, b is as well an element of it, then $\widehat{\Sigma} x : a.b$ is an element of the Mahlo universe, and $\mathrm{T}(\widehat{\Sigma} x : a.b)$ is defined as $\Sigma x : \mathrm{T}(a).\mathrm{T}(b)$. Here we refer to the fact that we have understood already how to form from a set A and a set $B(x)$ depending on $x : A$ the set $\Sigma x : A.B(x)$. $\widehat{\Sigma} x : a.b$ and $\widehat{\Sigma} x : a'.b'$ are equal, if a and a' are equal, and if for $x : \mathrm{T}(a)$, b and b' are equal.

In order to define, when $\widehat{\mathrm{U}}_{\vec{f}}$ is an element of the Mahlo universe, we introduce sets $\mathrm{U}_{\vec{f}}$ together with functions $\mathrm{s}_{\vec{f}} : \mathrm{U}_{\vec{f}} \to \mathrm{V}$ for arbitrary terms $\vec{f}$. Note the reference to arbitrary terms. As an abbreviation, let in the following $\mathrm{s}_{\vec{f}}^{\mathrm{Fam}} : \mathrm{Fam}(\mathrm{U}_{\vec{f}}, \mathrm{S}_{\vec{f}}) \to \mathrm{Fam}(\mathrm{V}, \mathrm{T})$ be defined as $\mathrm{s}_{\vec{f}}^{\mathrm{Fam}}(\langle x, y\rangle) = \langle \mathrm{s}_f(x), \lambda y.\mathrm{s}_f(y(x))\rangle$. We write loosely $f_0(\mathrm{s}_{\vec{f}}^{\mathrm{Fam}}(\langle a, b\rangle))$ for $f_0(\pi_0(\mathrm{s}_{\vec{f}}^{\mathrm{Fam}}(\langle a, b\rangle)), \pi_1(\mathrm{s}_{\vec{f}}^{\mathrm{Fam}}(\langle a, b\rangle)))$, similarly for $f_1(\mathrm{s}_{\vec{f}}^{\mathrm{Fam}}(\langle a, b\rangle), c)$.

By $\langle a, b\rangle$ being an element of $\mathrm{Fam}(\mathrm{U}_{\vec{f}}, \mathrm{S}_{\vec{f}})$ we mean that a is an element of $\mathrm{U}_{\vec{f}}$ and that for x in $\mathrm{S}_{\vec{f}}(a)$ it follows that $b\,x$ is an element of $\mathrm{U}_{\vec{f}}$.

First of all, we demand that $\mathrm{U}_{\vec{f}}$ is closed under the usual universe constructions. For instance, if a is an element of $\mathrm{U}_{\vec{f}}$, and for x in $\mathrm{T}(\mathrm{s}_{\vec{f}}(a))$, b is an element of $\mathrm{U}_{\vec{f}}$, then $\widetilde{\Sigma} x : a.b$ is an element of $\mathrm{U}_{\vec{f}}$. Furthermore, $\mathrm{s}_{\vec{f}}(\widetilde{\Sigma} x : a.b) = \widehat{\Sigma} x : \mathrm{s}_{\vec{f}}(a), \mathrm{s}_{\vec{f}}(b)$, of which we know already that it is an element of the Mahlo universe.

We demand as well that $\mathrm{U}_{\vec{f}}$ is closed under $\vec{f}$, provided $\vec{f}$ applied to the corresponding elements in V has a result in V: Assume $\langle a, b\rangle$ is an element of $\mathrm{Fam}(\mathrm{U}_{\vec{f}}, \mathrm{S}_{\vec{f}})$. Assume that $f_0(\mathrm{s}_{\vec{f}}^{\mathrm{Fam}}(\langle a, b\rangle))$ is an element of the Mahlo universe. Then we reflect this in $\mathrm{U}_{\vec{f}}$ as an element $\mathrm{Res0}(a, b)$. So we demand that $\mathrm{Res0}(a, b)$ is an element of $\mathrm{U}_{\vec{f}}$, and decode $\mathrm{s}_{\vec{f}}(\mathrm{Res0}(a, b)) = f_0(\mathrm{s}_{\vec{f}}^{\mathrm{Fam}}(\langle a, b\rangle))$.

Assume additionally c is an element of $\mathrm{T}(f_0(\mathrm{s}_{\vec{f}}^{\mathrm{Fam}}(\langle a, b\rangle)))$, and $f_1(\mathrm{s}_{\vec{f}}^{\mathrm{Fam}}(\langle a, b\rangle), c)$ is an element of the Mahlo universe. Then we reflect this in U_f and demand that

$\mathrm{Res1}(a, b, c)$ is an element of $\mathrm{U}_{\vec{f}}$ and $\mathrm{s}_{\vec{f}}(\mathrm{Res1}(a, b, c)) = f_1(\mathrm{s}^{\mathrm{Fam}}_{\vec{f}}(\langle a, b\rangle), c)$. Note that $\mathrm{U}_{\vec{f}}$ depends on V.

Assume now $\mathrm{U}_{\vec{f}}$ is closed under f, i.e. for every $\langle a, b\rangle$, which is an element of $\mathrm{Fam}(\mathrm{U}_{\vec{f}}, \mathrm{S}_{\vec{f}})$ it is the case that $f_0(\mathrm{s}^{\mathrm{Fam}}_{\vec{f}}(\langle a, b\rangle))$ is an element of the Mahlo universe, and that for every c as above, $f_1(\mathrm{s}^{\mathrm{Fam}}_{\vec{f}}(\langle a, b\rangle), c)$ is an element of V. Then we have a complete understanding of $\mathrm{U}_{\vec{f}}$ independently of any further elements added to V, and we demand that $\widehat{\mathrm{U}}_{\vec{f}}$ is an element of V and $\mathrm{T}(\widehat{\mathrm{U}}_{\vec{f}}) = \mathrm{U}_{\vec{f}}$.

$\widehat{\mathrm{U}}_{\vec{f}}$ and $\widehat{\mathrm{U}}_{\vec{g}}$ are equal iff for every element $\langle a, b\rangle$ of $\mathrm{Fam}(\mathrm{U}_{\vec{f}}, \mathrm{S}_{\vec{f}})$ we have that $f_0(\mathrm{s}^{\mathrm{Fam}}_{\vec{f}}(\langle a, b\rangle))$ and $g_0(\mathrm{s}^{\mathrm{Fam}}_{\vec{f}}(\langle a, b\rangle))$ are equal elements of V, and, if for every c in $\mathrm{T}(f_1(\mathrm{s}^{\mathrm{Fam}}_{\vec{f}}(\langle a, b\rangle)))$ we have that $f_1(\mathrm{s}^{\mathrm{Fam}}_{\vec{f}}(\langle a, b\rangle), c)$ and $g_1(\mathrm{s}^{\mathrm{Fam}}_{\vec{f}}(\langle a, b\rangle), c)$ are equal elements of V. We observe it is correct to identify the two sets since $\mathrm{U}_{\vec{f}}$ and $\mathrm{U}_{\vec{g}}$ are the same set, and therefore $\mathrm{T}(\widehat{\mathrm{U}}_{\vec{f}})$ and $\mathrm{T}(\widehat{\mathrm{U}}_{\vec{g}})$ are equal.

We have to show now that V fulfils the rules for the Mahlo universe. Assume $\vec{f} : (\mathrm{Fam}(\mathrm{V}, \mathrm{T}) \to \mathrm{Fam}(\mathrm{V}, \mathrm{T}))'$. Then it follows that $\mathrm{U}_{\vec{f}}$ is closed under $\vec{f}$ in the sense as stated above, and therefore $\widehat{\mathrm{U}}_{\vec{f}} : \mathrm{V}$. If $\vec{f}$ and $\vec{g}$ are equal elements of $(\mathrm{Fam}(\mathrm{V}, \mathrm{T}) \to \mathrm{Fam}(\mathrm{V}, \mathrm{T}))'$, then it follows that for a, b as above $f_0(\mathrm{s}^{\mathrm{Fam}}_{\vec{f}}(\langle a, b\rangle))$ $g_0(\mathrm{s}^{\mathrm{Fam}}_{\vec{f}}(\langle a, b\rangle))$ are equal, similarly for f_1 and g_1, and therefore $\widehat{\mathrm{U}}_{\vec{f}}$ and $\widehat{\mathrm{U}}_{\vec{g}}$ are equal.

The problem with this construction is that we need to refer to the collection of all terms, and that we construct elements of the Mahlo universe, which we cannot derive in the type theory, namely those elements $\widehat{\mathrm{U}}_{\vec{f}}$, for which $\vec{f}$ is only total on the restriction to the elements $\mathrm{s}_{\vec{f}}(a)$ of V, but not on the whole of $\mathrm{Fam}(\mathrm{V}, \mathrm{T})$. This differs from ordinary meaning explanations, in which the rules are in one to one correspondence with the explanations.

A second approach to a constructive understanding of the Mahlo universe is directly in accordance with the rules for the Mahlo universe. However, up to now no mathematical precise model corresponding to it has been developed. Here we understand the Mahlo universe as an open concept. The Mahlo universe is again closed under the usual universe constructions. Assume now that we know from our knowledge about the Mahlo universe up to now that it is closed under some function $\vec{f}$ in the sense of being a function from $\mathrm{Fam}(\mathrm{V}, \mathrm{T})$ into itself (more precisely in the sense of the primed version of the function space), independently of any further elements to be added later, and therefore not assuming a complete knowledge about V. Then we define $\mathrm{U}_{\vec{f}}$ and $\mathrm{s}_{\vec{f}}$ as before, and then $\widehat{\mathrm{U}}_{\vec{f}}$ is an element of the Mahlo universe and $\mathrm{T}(\widehat{\mathrm{U}}_{\vec{f}}) = \mathrm{U}_{\vec{f}}$. Assume now that, independent of any elements to be added later to the Mahlo universe, $\vec{f}$ and $\vec{g}$ coincide as functions from $\mathrm{Fam}(\mathrm{V}, \mathrm{T})$ into itself. Then $\widehat{\mathrm{U}}_{\vec{f}}$ and $\widehat{\mathrm{U}}_{\vec{g}}$ are equal elements of V.

Inconsistency of Mahlo universe with elimination rules. Erik Palmgren discovered in [Palmgren, 1998] that the Mahlo universe, extended by elimination rules, is inconsistent. This was shown in the following way. Let C be any fixed element of V. Let for $\langle f_0, f_1\rangle : (\mathrm{Fam}(\mathrm{V}) \to \mathrm{Fam}(\mathrm{V}))'$, $F_\downarrow(\langle f_0, f_1\rangle) := \lambda x.f_0(x, \lambda y.C) : \mathrm{V} \to \mathrm{V}$, and for $g : \mathrm{V} \to \mathrm{V}$, $F_\uparrow(g) := \langle \lambda x, y.g(x), \lambda x, y, z.C\rangle : (\mathrm{Fam}(\mathrm{V}) \to \mathrm{Fam}(\mathrm{V}))'$. Then we have $F_\downarrow(F_\uparrow(g)) = g : \mathrm{V} \to \mathrm{V}$, so we have (in a trivial way) embedded $\mathrm{V} \to \mathrm{V}$ into $(\mathrm{Fam}(\mathrm{V}) \to \mathrm{Fam}(\mathrm{V}))'$. Let for $\langle f_0, f_1\rangle : (\mathrm{Fam}(\mathrm{V}) \to \mathrm{Fam}(\mathrm{V}))'$ $G_\downarrow(\langle f_0, f_1\rangle) := \widehat{\mathrm{U}}_{f_0,f_1} : \mathrm{V}$, and define, using the elimination rules of the Mahlo universe, $G_\uparrow : \mathrm{V} \to (\mathrm{Fam}(\mathrm{V}) \to \mathrm{Fam}(\mathrm{V}))'$ s.t. $G_\uparrow(\widehat{\mathrm{U}}_{f_0,f_1}) = \langle f_0, f_1\rangle$. Then we have $G_\uparrow(G_\downarrow(\langle f_0, f_1\rangle)) = \langle f_0, f_1\rangle$, so we have embedded $(\mathrm{Fam}(\mathrm{V}) \to \mathrm{Fam}(\mathrm{V}))'$ into V. Define for $f : \mathrm{V} \to \mathrm{V}$ $f^- := G_\downarrow(F_\uparrow(f)) : \mathrm{V}$, and for $v : \mathrm{V}$, $v^+ := F_\downarrow(G_\uparrow(v)) : \mathrm{V} \to \mathrm{V}$.

Then we have $(f^-)^+ = f : \mathrm{V} \to \mathrm{V}$, so we have embedded $\mathrm{V} \to \mathrm{V}$ into V. Now we can interpret the untyped lambda-calculus into V by having as code for λ-abstraction the expression $\widehat{\lambda} x.t := (\lambda x.t)^-$, and as code for the application of s to t the expression $\widehat{\mathrm{Ap}}(s,t) := s^+(t)$. It is easy to verify that for $s : \mathrm{V}$, $x : \mathrm{V} \Rightarrow t : \mathrm{V}$ β-equality holds: $\widehat{\mathrm{Ap}}(\widehat{\lambda} x.t, s) = t[x := s] : \mathrm{V}$. Now one defines as usual in the untyped lambda-calculus the Y-combinator as $\mathrm{Y} := \widehat{\lambda} x.\widehat{\mathrm{Ap}}(V,V)$ where $V := \widehat{\lambda} y.\widehat{\mathrm{Ap}}(x, \widehat{\mathrm{Ap}}(y,y))$ and one gets for $a : \mathrm{V}$, $\widehat{\mathrm{Ap}}(\mathrm{Y}, a) = \widehat{\mathrm{Ap}}(a, \widehat{\mathrm{Ap}}(\mathrm{Y}, a)) : \mathrm{V}$. Let $\widehat{\mathrm{False}}$ be a code for False in U. Let $u := \widehat{\mathrm{Ap}}(\mathrm{Y}, \widehat{\lambda} x.x \widehat{\to} \widehat{\mathrm{False}}) : \mathrm{V}$. Then $u = u \widehat{\to} \widehat{\mathrm{False}} : \mathrm{V}$ and therefore with $A := \mathrm{T}(u)$ $A = A \to \mathrm{False} : \mathrm{Type}$. Now assuming $a : A$ we obtain $a : A \to \mathrm{False}$ and therefore $a(a) : \mathrm{False}$. This shows $f := \lambda a.a(a) : A \to \mathrm{False}$ and therefore $f(f) : \mathrm{False}$, so False is inhabited.

A model for the Mahlo universe and the upper bound for the proof-theoretic strength. A Model for the Mahlo universe was introduced in [Setzer, 1996] and used in order to determine an upper bound for the proof theoretic strength of the Mahlo universe, and to show the consistency relative to the corresponding set theory. The model was constructed using KPM^+ as Meta theory.

KPM^+ is similar to KPI^+ Kripke-Platek set theory plus the existence of one admissible set Ad_{M} , for which the Mahlo axiom holds, i.e. if $\forall x \in \mathrm{Ad}_{\mathrm{M}}.\exists y \in \mathrm{Ad}_{\mathrm{M}}.\varphi(x,y)$ for some Δ_0-formula φ, then there exists a $b \in \mathrm{Ad}_{\mathrm{M}}$, which is admissible, and such that $\forall x \in b.\exists y \in b.\varphi(x,y)$. Furthermore, similarly as for KPI^+, we demand that there exist finitely many admissibles above Ad_{M}. (Again one could alternatively define KPM^+ as KPl plus the existence of one recursively Mahlo ordinal.) One can now model the Mahlo universe by iterating an operator up to M, which is the first recursively Mahlo ordinal and the union of ordinals in Ad_{M}. That the universe is closed under the Mahlo operation follows by the fact that we have iterated the operator up to the first recursively Mahlo ordinal. More details about this model can be found in Appendix C.

Lower Bound. The lower bound [Setzer, 2000] was carried out similarly to the lower bound for $\mathrm{ML}_1\mathrm{W}$ by carrying out a well-ordering proof. That proof is rather technical, and to go into details is beyond the scope of this article.

7 Application: Inductive-Recursive Definitions

Induction-recursion. The concept of induction-recursion is due to Dybjer [Dybjer, 2000]. It is an abstract formalisation of the general principles used for introducing new sets in Martin-Löf type theory (excluding the Mahlo principle).

The principle of strictly positive inductive definitions in simple type theory has been studied since long time ago. An algebraic data type A (introduced by constructors C_i) is defined strictly positive inductively, if the constructors are of the form $C_i : B_1 \to \cdots \to B_n \to A$, where B_i either do not depend on A, or are of the form $D_1 \to \cdots \to D_k \to A$, where D_i do not depend on A. Both n and k can be 0. An argument of a constructor of type B_i not referring to A is called a *non-inductive argument*, an argument of type $D_1 \to \cdots \to D_k \to A$ is called an *inductive argument.* Examples are the finite sets N_l with constructors $\mathrm{A}_l^i : \mathrm{N}_l$ for $i = 0, \ldots, l-1$ (the constructors have no arguments); the set of natural numbers N with constructors $0 : \mathrm{N}$ and $\mathrm{S} : \mathrm{N} \to \mathrm{N}$ (the constructor S has one inductive argument where k as above is 0); the set Nlist of lists of natural numbers with constructors $\mathrm{nil} : \mathrm{Nlist}$ (no arguments) and $\mathrm{cons} : \mathrm{N} \to \mathrm{Nlist} \to \mathrm{Nlist}$ (one non-inductive and one inductive argument).

In dependent type theory, the argument types B_i can depend on previous non-inductive arguments. So here we have $C_i : (x_1 : B_1) \to (x_2 : B_2) \to \cdots \to (x_n : B_n) \to A$ where B_i are either independent of A or of the form $(y_1 : D_1) \to \cdots \to (y_k : D_k) \to A$. Examples are the set $\Sigma x : A.B(x)$ with constructor $\mathrm{p} : (x : A) \to B \to \Sigma x : A.B(x)$ (two non-inductive arguments, the second depends on the first); the set $\Pi x : A.B(x)$ with constructor $\lambda : ((x : A) \to B(x)) \to (\Pi x : A.B(x))$ (no inductive argument); the set $\mathrm{W} x : A.B(x)$ with constructor $\mathrm{sup} : (x : A) \to (B(x) \to (\mathrm{W} x : A.B(x))) \to (\mathrm{W} x : A.B(x))$ (the first argument is non-inductive, the second is inductive and depends on the first).

The above can be generalised to *indexed* inductive definitions, where we define several sets A_i simultaneously inductively. This can take the form of finitely many sets, each of which has different constructors. An example for this is the set of finitely branching trees FinTree together with the set of lists of such trees FinTreeList (the example is due to U. Berger): lists of finite trees are introduced in a standard way by having constructors $\mathrm{nil} : \mathrm{FinTreeList}$, $\mathrm{cons} : \mathrm{FinTree} \to \mathrm{FinTreeList} \to \mathrm{FinTreeList}$. Furthermore, if we have a list of trees, we can form a tree, having the elements of the list as subtrees, so we have a constructor $\mathrm{maketree} : \mathrm{FinTreeList} \to \mathrm{FinTree}$. A more general form of indexed inductive definition is, when we introduce simultaneously possibly infinitely many sets $A(i)$ $(i : I)$ indexed over a set I. One degenerate example for this is the equality set $\mathrm{I}(A, a, b)$, which can be considered as a set $C(\langle a, b\rangle)$ indexed over $\langle a, b\rangle : A \times A$. It has constructor $\mathrm{refl} : (a : A) \to C(\langle a, a\rangle)$. An example, which is really inductive, is the predicate $\mathrm{Even}(n)$ for $n : \mathrm{N}$, meaning "n is even". It has constructors $\mathrm{zeroproof} : \mathrm{Even}(0)$ and $\mathrm{succproof} : (n : \mathrm{N}, \mathrm{Even}(n)) \to \mathrm{Even}(\mathrm{S}(\mathrm{S}(n)))$. The general form of indexed inductive definition is that we have constructors of type $(x_1 : B_1) \to \cdots \to (x_n : B_n) \to A(j)$, and where B_l either does not depend on $A(i)$, or is of the form $(y_1 : D_1) \to \cdots \to (y_k : D_k) \to A(j')$, where j' might depend on $y_1, \ldots, y_k$. j might depend on the noninductive arguments x_i of type B_i, where B_i does not depend on $A(i)$.

In inductive definitions, the argument types cannot depend on previous inductive arguments: Before introducing a new set (or sets) inductively, we have to introduce the argument types of the constructors, which cannot refer to the set(s) to be introduced. Inductive-recursive definitions go beyond inductive definition, and will allow an indirect dependency of argument types on previous inductive arguments. An example of a truly inductive-recursive definition is a standard universe. The constructor introducing the code for the Σ-set as an element of the universe has the form $\widehat{\Sigma} : (a : \mathrm{U}) \to (b : \mathrm{T}(a) \to \mathrm{U}) \to \mathrm{U}$. When looking at a process for constructing the elements of such a universe, we see that, whenever one constructs an element $a : \mathrm{U}$, one has to define immediately $\mathrm{T}(a)$. Otherwise one cannot use this element for forming further elements of U. This means that we define elements of U *inductively*, while defining simultaneously *recursively* $\mathrm{T}(a)$ for every element $a : \mathrm{U}$ introduced – therefore the terminology *inductive-recursive definition*. In the above example for instance, one defines $\mathrm{T}(\widehat{\Sigma}(a, b)) = \Sigma x : \mathrm{T}(a).\mathrm{T}(b\, x)$.

In inductive-recursive definitions, later arguments can depend on arbitrary previous non-inductive arguments and on the recursively defined function applied to previous inductive arguments. The result of T can depend in the same way on the arguments, as can later arguments depend on previous ones.

A closed formalisation of inductive-recursive definitions. In P. Dybjer's original formalisation, the dependency of arguments on previous arguments was syntactic, and meant essentially the occurrence of a variable for one argument in the term for a later argument. Therefore, his type theory is a schema, which allows to introduce for each inductive-recursively defined set new rules. However, we can see that, before we can introduce a new inductive-recursive definition, we often need first to carry out a proof using the rules defined before.

We give an example: Assume we want to define inductive-recursively a non-standard universe U, T (i.e. a universe which is not closed under standard type theoretic constructions). Instead it should contain a code for N and, if $a_0, a_1, a_2 : \mathrm{U}$, and $\mathrm{T}(a_i) = A_i$, then U should contain a code for $A_0 + (A_1 + A_2)$. For this to be a good definition, we need to know, before adding the rules for such a universe to type theory, that we have that $A + (B + C) : \mathrm{Type}$, provided $A, B, C : \mathrm{Type}$. In this toy example, this is of course obvious, but one can easily construct more complicated examples, which require a long derivation. So in general, before introducing an inductive-recursive definition, one has to derive certain type theoretic judgements. Therefore we see that in P. Dybjer's original framework one has to work as follows: one starts with basic inductive-recursive definitions. Then one derives type theoretic formulae, using which one can introduce additional inductive-recursive definitions. Then one can introduce further inductive-recursive definitions, and so on.

It is difficult to analyse a framework like this proof theoretically and to construct models of it. Therefore P. Dybjer and myself developed in a series of articles [Dybjer and Setzer, 1999, Dybjer and Setzer, 2001, Dybjer and Setzer, 2003b, Dybjer and Setzer, 2003a] a closed formalisation of inductive-recursive definitions. "Closed" means that we have a fixed set of rules, which we can introduce from the beginning, which don't depend on previous proofs, and which allow to introduce all inductive-recursive definitions. The formalisation taken makes use of ideas used in the definition of the Mahlo universe. In fact, the resulting theory reaches the strength of a slightly weakened version of Mahlo type theory.

We will in the following only consider the non-indexed case. We will make use of the logical framework. We used as logical framework operations the dependent function-type with η-rule (written as $(x : A) \to B$ – we reserve the notation $\Pi x : A.B(x)$ for the definition of a set having essentially the same rules, but no η-rule) the dependent product with η-rule (written as $(x : A) \times B$ – again, $\Sigma x : A.B(x)$ is reserved for a corresponding set, which has no η-rule), and the types having zero, one and two elements, written as $\mathbf{0}$, $\mathbf{1}$, $\mathbf{2}$. The canonical element of $\mathbf{1}$ is $*$ and the canonical elements of $\mathbf{2}$ are $*_0$ and $*_1$. We add the η-rule for $\mathbf{1}$, which expresses that for $x : \mathbf{1}$, $x = * : \mathbf{1}$ We have as well case distinction $\mathrm{case}_2 : \mathbf{2} \to A \to A \to A$ for any type A, with the equalities $\mathrm{case}_2(*_0, a, b) = a$, $\mathrm{case}_2(*_1, a, b) = b$. Furthermore, we have Set : Type, containing inductive-recursively defined sets. All sets are types. Both Set and Type are closed under the operations of the logical framework.[1]

In the following, we will uncurry the arguments of the constructors, so we have $C_i : ((x_1 : B_1) \times \cdots \times (x_n : B_n)) \to A$ instead of $C_i : (x_1 : B_1) \to \cdots \to (x_n : B_n) \to A$. We can code several constructors into one by having one additional argument, which is an element of a finite set and indicates, which constructor was chosen. Depending on this argument, the types for the other arguments of the constructors are taken. Therefore, our inductive-recursively defined sets will have only one constructor.

Rules for inductive-recursive definitions. In the new type theory, we will replace the notion of dependency of a type B on $x : A$, as it occurred in P. Dybjer's original schema by a judgement $\Gamma, x : A, \Delta \Rightarrow B : \mathrm{Type}$. In order to able to derive that something is an inductive recursive definition, we need a corresponding judgement, and therefore we will introduce a type OP_D of codes for inductive-recursive definitions, depending on a type D (the meaning of D will be explained

[1] In our original articles besides Set an additional type stype was used. Set contained only sets introduced inductive-recursively, whereas stype contained all elements of Set, but not Set itself, and was closed under the operation of the logical framework. This distinction is not really necessary, and we omit it in this article.

later on). To derive an inductive recursive definition means to derive an element $\gamma : \mathrm{OP}_D$. For every inductive recursive definition $\gamma : \mathrm{OP}_D$ we introduce the inductive recursively defined set U_γ introduced by it together with its decoding function T_γ, having typing rule $a : \mathrm{U}_\gamma \Rightarrow \mathrm{T}_\gamma(a) : D$. This explains now the meaning of the parameter D in OP_D: D is the codomain of the recursively defined functions T_γ for $\gamma : \mathrm{OP}_D$. For instance, if we take $D = \mathrm{Set}$, then $(\mathrm{U}_\gamma, \mathrm{T}_\gamma)$ will be a universe, but not necessarily closed under type theoretic operations. If we take $D = \mathbf{1}$, then we have $\mathrm{T}_\gamma(a) = *$, so T_γ doesn't carry any information – this is nothing but an ordinary inductive definition (as opposed to an inductive-recursive definition). If we take $D = ((X : \mathrm{Set}) \times (X \to \mathrm{Set})) \to ((X : \mathrm{Set}) \times (X \to \mathrm{Set}))$, U_γ will be a universe of operations, where an operation maps families of sets to families of sets. In general D can be any type. In the following, D will be kept fixed, and we assume globally $D : \mathrm{Type}$. When fully spelled out, all rules will have an additional premise $D : \mathrm{Type}$.

Note that OP_D is therefore some kind of big universe, having two decoding functions, namely $\lambda\gamma.\mathrm{U}_\gamma : \mathrm{OP}_D \to \mathrm{Set}$ and $\lambda\gamma.\lambda x.\mathrm{T}_\gamma(x) : (\gamma : \mathrm{OP}_D) \to (\mathrm{U}_\gamma \to \mathrm{Set})$. OP_D itself cannot be defined inductive-recursively.

The argument type of the constructor is given by $\mathbb{F}^{\mathrm{U}}_\gamma$, having the formation rule

$$\frac{\gamma : \mathrm{OP}_D \qquad U : \mathrm{Set} \qquad T : U \to \mathrm{Set}}{\mathbb{F}^{\mathrm{U}}_\gamma(U, T) : \mathrm{Set}}$$

The result of T_γ applied to a constructor element, is given by $\mathbb{F}^{\mathrm{T}}_\gamma$, having the formation rule

$$\frac{\gamma : \mathrm{OP}_D \qquad U : \mathrm{Set} \qquad T : U \to \mathrm{Set} \qquad a : \mathbb{F}^{\mathrm{U}}_\gamma(U, T)}{\mathbb{F}^{\mathrm{T}}_\gamma(U, T, a) : D}$$

So when introducing $\gamma : \mathrm{OP}_D$ we have to define $\mathbb{F}^{\mathrm{U}}_\gamma$ and $\mathbb{F}^{\mathrm{T}}_\gamma$.

Once this is defined we have the following formation and equality rules for inductive-recursively defined sets:

$$\mathrm{U}_\gamma : \mathrm{Set} \qquad\qquad \mathrm{T}_\gamma : \mathrm{U}_\gamma \to D$$

The introduction rules for U_γ and equality rules for T_γ are

$$\mathrm{intro}_\gamma : \mathbb{F}^{\mathrm{U}}_\gamma(\mathrm{U}_\gamma, \mathrm{T}_\gamma) \to \mathrm{U}_\gamma \qquad\qquad \mathrm{T}_\gamma(\mathrm{intro}_\gamma(a)) = \mathbb{F}^{\mathrm{T}}_\gamma(\mathrm{U}_\gamma, \mathrm{T}_\gamma, a)$$

We have the following rules for generating elements of OP_D:

- **Addition of a non-inductive argument:** Assume $A : \mathrm{Set}$ and $\gamma : A \to \mathrm{OP}_D$. Then we can form a new code $\sigma(A, \gamma) : \mathrm{OP}_D$ for the inductive-recursive definition, having a first non-inductive argument $a : A$, and depending on it, the other arguments taken from $\gamma(a)$. So we have

  $$\mathbb{F}^{\mathrm{U}}_{\sigma(A,\gamma)}(U, T) = (a : A) \times \mathbb{F}^{\mathrm{U}}_{\gamma(a)}(U, T)$$

  The result of T_γ for an element $\mathrm{intro}_\gamma(a)$ is the result obtained for the remaining arguments with respect to $\gamma(a)$. Therefore we have:

  $$\mathbb{F}^{\mathrm{T}}_{\sigma(A,\gamma)}(U, T, \langle a, b\rangle) = \mathbb{F}^{\mathrm{T}}_{\gamma(a)}(U, T, b)$$

- **Addition of an inductive argument:** Assume $A : \mathrm{Set}$ and $\gamma : (A \to D) \to \mathrm{OP}_D$. Then we can form a new code $\delta(A, \gamma)$ for the inductive-recursive definition, having a first inductive argument indexed over A, i.e. $f : A \to \mathrm{U}_\gamma$. The further arguments depend on T_γ applied to the elements of U_γ, to which

f is referring, i.e. on $\mathrm{T}_\gamma \circ f$, and are therefore taken from $\gamma(\mathrm{T}_\gamma \circ f)$. So we have

$$\mathbb{F}^{\mathrm{U}}_{\delta(A,\gamma)}(U,T) = (f : A \to U) \times \mathbb{F}^{\mathrm{U}}_{\gamma(T\circ f)}(U,T)$$

The result of T_γ for an element $\mathrm{intro}_\gamma(a)$ is the result obtained for the remaining arguments with respect to $\gamma(\mathrm{T}_\gamma \circ f)$. Therefore we have:

$$\mathbb{F}^{\mathrm{T}}_{\delta(A,\gamma)}(U,T,\langle f,b\rangle) = \mathbb{F}^{\mathrm{T}}_{\gamma(T\circ f)}(U,T,b)$$

- **Base case:** This corresponds to the inductive-recursive definition with no arguments. We only have to determine the result of T, which is an element of type D. Assuming $\psi : D$, we have therefore $\iota(\psi) : \mathrm{OP}_D$ and the rules

$$\begin{array}{rcl} \mathbb{F}^{\mathrm{U}}_{\iota(\psi)}(U,T) & = & \mathbf{1} \ , \\ \mathbb{F}^{\mathrm{T}}_{\iota(\psi)}(U,T,*) & = & \psi \ . \end{array}$$

Elimination and equality rules. In order to define the elimination and equality rules, one has to define first for every $\gamma : \mathrm{OP}_D$ two more types:

$$\dfrac{\begin{array}{c}\gamma : \mathrm{OP}_D \qquad U : \mathrm{Set} \qquad T : U \to D \\ u : \mathbb{F}^{\mathrm{U}}_\gamma(U,T) \qquad x : U \Rightarrow E[x] : \mathrm{Type}\end{array}}{\mathbb{F}^{\mathrm{IH}}_\gamma(U,T,E,u) : \mathrm{Type}}$$

$$\dfrac{\begin{array}{c}\gamma : \mathrm{OP}_D \qquad U : \mathrm{Set} \qquad T : U \to \mathrm{Set} \\ x : U \Rightarrow E[x] : \mathrm{Type} \qquad h : (x : U) \to E[x]\end{array}}{\mathbb{F}^{\mathrm{map}}_\gamma(U,T,E,h) : (u : \mathbb{F}^{\mathrm{U}}_\gamma(U,T)) \to \mathbb{F}^{\mathrm{IH}}_\gamma(U,T,E,u)}$$

Then the elimination rule for U_γ is as follows:

$$\dfrac{\begin{array}{c}x : \mathrm{U}_\gamma \ \Rightarrow E[x] : \mathrm{Type} \\ g : (u : \mathbb{F}^{\mathrm{U}}_\gamma(\mathrm{U}_\gamma, \mathrm{T}_\gamma), \mathbb{F}^{\mathrm{IH}}_\gamma(\mathrm{U}_\gamma, \mathrm{T}_\gamma, E, u)) \to E[\mathrm{intro}_\gamma(u)]\end{array}}{\mathrm{R}_{\gamma,E}(g) : (u : \mathrm{U}_\gamma) \to E[u]}$$

The equality rule is as follows:

$$\dfrac{\begin{array}{c}x : \mathrm{U}_\gamma \Rightarrow E[x] : \mathrm{Type} \\ g : (u : \mathbb{F}^{\mathrm{U}}_\gamma(\mathrm{U}_\gamma, \mathrm{T}_\gamma), \mathbb{F}^{\mathrm{IH}}_\gamma(\mathrm{U}_\gamma, \mathrm{T}_\gamma, E, u)) \to E[\mathrm{intro}_\gamma(u)] \\ u : \mathbb{F}^{\mathrm{U}}_\gamma(\mathrm{U}_\gamma, \mathrm{T}_\gamma)\end{array}}{\mathrm{R}_{\gamma,E}(g, \mathrm{intro}_\gamma(u)) = g(u, \mathbb{F}^{\mathrm{map}}_\gamma(\mathrm{U}_\gamma, \mathrm{T}_\gamma, E, \mathrm{R}_{\gamma,E}(g), u)) : E[\mathrm{intro}_\gamma(u)]}$$

We won't give the equality rules for $\mathbb{F}^{\mathrm{IH}}_\gamma$ and $\mathbb{F}^{\mathrm{map}}_\gamma$ here, the straighforward and boring details can be found in [Dybjer and Setzer, 2003b].

Examples. The first examples will be inductive definitions, so in this case $D := \mathbf{1}$. Let $\iota_* := \iota(*) : \mathrm{OP}_{\mathbf{1}}$. The finite sets are defined by

$$\begin{array}{rcl} \gamma_{\mathrm{N}_0} & := & \sigma(\mathbf{0}, \lambda x.\iota_*) : \mathrm{OP}_{\mathbf{1}} \ . \\ \gamma_{\mathrm{N}_1} & := & \iota_* : \mathrm{OP}_{\mathbf{1}} \ , \\ \gamma_{\mathrm{N}_{n+2}} & := & \sigma(\mathbf{2}, \lambda x.\mathrm{case}_2(x, \gamma_{\mathrm{N}_{n+1}}, \iota_*)) : \mathrm{OP}_{\mathbf{1}} \ . \end{array}$$

$A + B$ and $\Sigma x : A.B(x)$ have codes

$$\begin{array}{rcl} \gamma_{A+B} & := & \sigma(\mathbf{2}, \lambda x.\mathrm{case}_2(x, \sigma(A, \lambda y.\iota_*), \sigma(B, \lambda y.\iota_*))) \ , \\ \gamma_{\Sigma x:A.B(x)} & := & \sigma(A, \lambda x.\sigma(B(x), \lambda y.\iota_*)) \ . \end{array}$$

N has code

$$\gamma_{\mathrm{N}} \quad := \quad \sigma(\mathbf{2}, \lambda x.\mathrm{case}_2(x, \iota_*, \delta(\mathbf{1}, \lambda y.\iota_*))) \ \ .$$

Zero is here $\mathrm{intro}_{\gamma_{\mathrm{N}}}(\langle *_0, *\rangle)$, and the successor of n is $\mathrm{intro}_{\gamma_{\mathrm{N}}}(\langle *_1, \langle n, *\rangle\rangle)$.

$\mathrm{W}x : A.B(x)$ has code

$$\gamma_{\mathrm{W}x:A.B(x)} := \sigma(A, \lambda x.\delta(B(x), \lambda y.\iota_*)) \ \ .$$

Finally, the first universe (consisting of U_0 : Set and $\mathrm{T}_0 : \mathrm{U}_0 \to$ Set and for simplicity closed under N and Σ only) has code

$$\gamma_{\mathrm{U}_0,\mathrm{T}_0} := \sigma(\mathbf{2}, \lambda x.\mathrm{case}_2(x, \iota(\mathrm{N}), \delta(\mathbf{1}, \lambda A.\delta(A(*), \lambda B.\iota(\Sigma x : A(*), B\,x))) : \mathrm{OP}_{\mathrm{Set}} \ \ .$$

Application in generic programming. The theory developed has a data type for inductive-recursive definitions. If one considers this type theory as a functional programming language, it is possible to write programs, which have a higher degree of polymorphism, and take as input a data type (an element of OP_D), analyse it and generate a new data type (a new element of OP_D). Such kind of programming is called generic programming. Examples for its use are: a function, which takes an inductive recursive definition and adds one constructor to it, together with an embedding of the original one into the new one; and the definition of a defined equality relation on a data type. This is an area of ongoing research; see [Dybjer and Setzer, 1998, Benke et al., 2003, 3] for details.

Inductive-recursive definitions and the Mahlo universe. The constructor δ of OP_D has type $\delta : (A : \mathrm{Set}, \gamma : (A \to D) \to \mathrm{OP}_D) \to \mathrm{OP}_D$ and refers therefore negatively to D. Note that D can be Set, and that from elements of OP_D we introduce new elements of Set. Therefore elements of Set can be introduced by referring negatively to Set.

In fact, Set is essentially a weak variant of the Mahlo universe (the strength of a type theory with a weak Mahlo universe is only slightly below that of the type theory with a full Mahlo universe): Assume $f_0 : (A : \mathrm{Set}, B : A \to \mathrm{Set}) \to \mathrm{Set}$ and $f_1 : (A : \mathrm{Set}, B : A \to \mathrm{Set}, f_0(A, B)) \to \mathrm{Set}$. Let

$$\begin{aligned}
\gamma_0(\vec{f}) &:= \delta(\mathbf{1}, \lambda A'.\delta(A'(*), \lambda B'.\iota(f_0(A'(*), B')))) : \mathrm{OP}_{\mathrm{Set}} \ \ , \\
\gamma_1(\vec{f}) &:= \delta(\mathbf{1}, \lambda A'.\delta(A'(*), \lambda B'.\sigma(f_0(A'(*), B'), \lambda C.\iota(f_1(A'(*), B', C))))) : \mathrm{OP}_{\mathrm{Set}} \ \ , \\
\gamma(\vec{f}) &:= \sigma(\mathbf{2}, \lambda x.\mathrm{case}_2(x, \gamma_0(\vec{f}), \gamma_1(\vec{f}))
\end{aligned}$$

Then $\mathrm{U}_{\gamma(\vec{f})}$ will be, similar to $\mathrm{U}_{\vec{f}}$ in case of the Mahlo universe, a universe closed under f_0, f_1. Here $\mathrm{intro}_{\gamma(\vec{f})}(\langle *_0, \cdots\rangle)$ and $\mathrm{intro}_{\gamma(\vec{f})}(\langle *_1, \cdots\rangle)$ will play the rôle of Res_0, Res_1, respectively, in the Mahlo universe. In this form, the universe $\mathrm{U}_{\gamma(\vec{f})}$ will be empty, but one can easily expand $\gamma(\vec{f})$ and guarantee that $\mathrm{U}_{\gamma(\vec{f})}$ is closed under the standard universe constructions as well. Therefore, for every pair of functions $\vec{f}$ from families of elements of Set into families of elements of Set there exists a universe in Set closed under $\vec{f}$. Note that, in the presence of the logical framework, it is possible to have $\vec{f}$ as elements of the context.

In appendix D we will show, using this observation, that the theory of inductive-recursive definitions reaches the strength of KPM, Kripke-Platek set theory with Mahloness of the universe.

Model. In [Dybjer and Setzer, 1999], a model of the theory of inductive-recursive definitions was developed in set theory plus the existence of one strongly Mahlo cardinal. There we interpreted Set as V_{M}, where $(\mathrm{V}_\alpha)_{\alpha\in\mathrm{Ord}}$ is the commulative hierarchy of sets and M is one strongly Mahlo cardinal. The usual set constructions were interpreted by their naïve interpretation, e.g. $[\![\, A \to B \,]\!]$ was interpreted as the set theoretic function space $[\![\, A \,]\!] \to [\![\, B \,]\!]$. OP_D was interpreted as an appropriate inductive definition. We defined approximations of U_γ^α, T_γ^α of the interpretation of U_γ, T_γ, and interpreted U_γ as $\mathrm{U}_\gamma^{\mathrm{M}}$, T_γ as $\mathrm{T}_\gamma^{\mathrm{M}}$. The definition of U_γ^α was based on the interpretation of $\lambda U, T.\langle \mathbb{F}_\gamma^{\mathrm{U}}(U,T), \mathbb{F}_\gamma^{\mathrm{T}}(U,T)\rangle$. For every $\alpha < \mathrm{M}$ there exists a $\beta < \mathrm{M}$ s.t. if $U \in \mathrm{V}_\alpha$, $T \in U \to [\![\, D \,]\!]$, then $[\![\, \mathbb{F}_\gamma^{\mathrm{U}}(U,T) \,]\!]_{U \mapsto U, T \mapsto T} \in \mathrm{V}_\beta$. By the Mahlo property one obtained a $\kappa < \mathrm{M}$ recursively inaccessible, s.t. if $\alpha < \kappa$, then the β as above is $< \kappa$. Using the recursive inaccessibility of κ it followed then that $\mathrm{U}_\gamma^\kappa = \mathrm{U}_\gamma^{\mathrm{M}}$ and therefore $[\![\, \mathrm{U}_\gamma \,]\!] \in [\![\, \mathrm{Set} \,]\!]$.

Upper Bound. We introduced the above mentioned model, because it is rather natural. Using this model one obtains an upper bound, which is far too big, namely that of ZF plus the existence of one strongly Mahlo cardinal. We have not yet spelled out a model, which uses only the strength of KPM^+. In order to define such a model, one would have to interpret Set as the iteration of an operator up to the first recursively Mahlo ordinal M in a similar way as the interpretation of the Mahlo universe. The interpretation of OP_D would require in such a model finitely many admissibles above the recursively Mahlo ordinal. We do not expect any major difficulties in carrying this out in detail.

Precise strength. Assuming the model in KPM^+ has been developed, we obtain an interval for the strength of the theory, namely $[|\mathrm{KPM}|, |\mathrm{KPM}^+|]$. We do not know yet currently, what the exact strength is. This depends on, whether the types OP_D actually contribute to the strength of the theory of inductive-recursive definitions.

Appendix A: Direct Well-Foundedness Proof of the Ordinal Notation System of Strength ϵ_0

We sketch here a direct well-ordering proof for the ordinal notation system of strength ϵ_0, developed in Sect. 2. This argument doesn't refer to ordinals, and can be formalised, restricted to ordinals $< \epsilon_0$, in Peano Arithmetic.

The argument proceeds as follows: First one shows that, if $(A, <_A), (B, <_B)$ are well-founded linear orderings, so are $(A \times B, <_{A\times B})$ and $(A_{\mathrm{dec}}, <_{\mathrm{lex}})$. Here $<_{A\times B}$ is the lexicographic ordering on pairs $\langle a, b\rangle$, where $a \in A$ and $b \in B$, A_{dec} is the set of w.r.t. $<$ descending sequences of A (i.e. the set of sequences $\langle a_0, \ldots, a_{n-1}\rangle$ s.t. $a_0 > \cdots > a_{n-1}$), and $<_{\mathrm{lex}}$ is the lexicographic ordering on these sequences (we suppress the dependencies of A_{dec} on $<$). Define for orderings $(X, <)$ on $\mathbb{N}$ the operation $\Gamma(X, <) := ((X \times \mathbb{N})_{\mathrm{dec}}, (<_{X\times\mathbb{N}})_{\mathrm{lex}})$ as defined before. Then $(\mathrm{OT}_{\epsilon_0}, <_{\epsilon_0}) = (\bigcup_{n\in\omega} A_n, \bigcup_{n\in\omega} <_n)$ with $(A_n, <_n) := \Gamma^n((\emptyset, \emptyset))$. Observe that $(A_n, <_n)$ is the set of ordinal notations $< \underbrace{\omega^{\omega^{\cdot^{\cdot^{\cdot^{1}}}}}}_{n \text{ times}}$. $(\emptyset, \emptyset)$ is trivially well-founded, therefore each $(A_n, <_n)$ are well-founded. Furthermore, one can easily see that $(A_n, <_n) \sqsubseteq (A_{n+1}, <_{n+1})$ where $\sqsubseteq$ means "initial segment" defined as $(A, <) \sqsubseteq (B, <') :\Leftrightarrow A \subseteq B \wedge <' \restriction A \times A = < \wedge \forall a \in A. \forall b \in B. b <' a \to b \in A$. It is easy to see that if $(B_n, <_n)$ are well-founded and transitive, $(B_n, <_n) \sqsubseteq (B_{n+1}, <_{n+1})$ for all n, then $(\bigcup_{n\in\omega} B_n, \bigcup_{n\in\omega} <_n)$ is well-founded. Therefore it follows that $(\mathrm{OT}_{\epsilon_0}, <_{\epsilon_0})$ is well-founded.

From the above argument one can develop in Peano Arithmetic a proof of the principle of transfinite induction for formulae of PA over $(A_n, <_n)$ for Meta-each n, and therefore for OT_{ϵ_0} restricted to any ordinal $b < \epsilon_0$.

Appendix B: Determination of the Strength of Type Theory with one Universe and the W-Type

Upper bound for the proof theoretic strength of $\mathrm{ML}_1\mathrm{W}$**.** We will in the following construct a model of $\mathrm{ML}_1\mathrm{W}$ with extensional equality in a theory of the same strength, namely KPI^+, and use therefore KPI^+ as Meta-theory.

We form a simple PER (= partial equivalence relation) model: every type expression A in dependent type theory is modelled as a set $[\![A]\!]_\rho$ of pairs of terms, namely those terms which are considered to be equal, and, if $\mathrm{ML}_1\mathrm{W}$ proves $A : \mathrm{Type}$, then $[\![A]\!]_\rho$ will be a partial equivalence relation, i.e. transitive and symmetric. In the model, we identify terms with their Gödel-numbers. ρ is an environment, i.e. a finite map from variables to closed terms, s.t. all free variables of A are in the domain of ρ. Let for sets of pairs of terms A, B $A \to B := \{\langle s, s'\rangle \mid \forall \langle r, r'\rangle \in A. \langle s(r), s'(r')\rangle \in B\}$. We form the model for the restriction of type theory, where the W-rank of type expressions is $\leq n$ for some $n \in \mathbb{N}$. Here the W-rank of A is 0, if A does not contain U. Otherwise, the rank of U is 1, the rank of $\mathrm{W}x : A.B(x)$ is the maximum of the W-rank of A, $B(x)$ incremented by 1, and the rank of all other type expression is the maximum of the W-rank of its immediate subterms which are subtypes (e.g. the W-rank of $\Pi x : A.B(x)$ is the maximum of the W-rank of A and $B(x)$).

We introduce some notations: $x_1 \mapsto r_1, \ldots, x_n \mapsto r_n$ denotes the environment, mapping x_i to r_i. If ρ is an environment, $\rho(x \mapsto r)$ is the environment, mapping x to ρ and $y \neq x$ to $\rho(y)$, provided y is in the domain of ρ. We write $[x_1 := r_1, \ldots, x_n := r_n]$ for the simultaneous substitution of x_i by r_i. If $\Gamma = x_1 : A_1, \ldots, x_n : A_n$, we write $\forall \langle \vec{r}, \vec{r}'\rangle \in [\![\Gamma]\!]$ for
$\forall \langle r_1, r_1'\rangle \in [\![A_1]\!], \forall \langle r_2, r_2'\rangle \in [\![A_2]\!]_{x_1 \mapsto r_1}, \ldots, \forall \langle r_n, r_n'\rangle \in [\![A_n]\!]_{x_1 \mapsto r_1, \ldots, x_{n-1} \mapsto r_{n-1}}$.
Furthermore, we write $\vec{x} \mapsto \vec{r}$ for the environment $x_1 \mapsto r_1, \ldots, x_n \mapsto r_n$ and $[\vec{x} := \vec{r}]$ for $[x_1 := r_1, \ldots, x_n := r_n]$, assuming that the choice of x_i, r_i is obvious from the context.

We first state what it means for a derived judgement to be correct in this model. We define the set of immediate presuppositions (ips) of a judgement as follows:

- $\emptyset \Rightarrow \mathrm{Context}$ has no ips.
- The ips of $\Gamma, x : A \Rightarrow \mathrm{Context}$ is $\Gamma \Rightarrow A : \mathrm{Type}$.
- The ips of $\Gamma \Rightarrow A : \mathrm{Type}$ is $\Gamma \Rightarrow \mathrm{Context}$.
- The ips of $\Gamma \Rightarrow A = B$ are $\Gamma \Rightarrow A : \mathrm{Type}$ and $\Gamma \Rightarrow B : \mathrm{Type}$.
- The ips of $\Gamma \Rightarrow s : A$ is $\Gamma \Rightarrow A : \mathrm{Type}$.
- The ips of $\Gamma \Rightarrow r = s : A$ are $\Gamma \Rightarrow r : A$ and $\Gamma \Rightarrow s : A$.

Then one defines the presuppositions of a judgement as follows:

- The ips of a judgement are presuppositions of it.
- The ips of a presupposition of a judgement are as well presuppositions of that judgement.

The correctness condition for $\Gamma \Rightarrow \theta$ is defined as the conjunction of the immediate correctness conditions (icc) of all its presuppositions and of the judgement itself, where the icc of a judgement $\Gamma \Rightarrow \theta$ is defined as follows:

- If $\theta = \mathrm{Context}$, then the icc is the true formula.

- If θ is $A = B$: Type, then the icc is
  $\forall\langle\vec{r},\vec{r}'\rangle \in [\![\Gamma]\!].[\![A]\!]_{\vec{x}\mapsto\vec{r}} = [\![B]\!]_{\vec{x}\mapsto\vec{r}'} \wedge \mathrm{Equiv}([\![A]\!]_{\vec{x}\mapsto\vec{r}})$,
  where $\mathrm{Equiv}(X)$ means that A is a partial equivalence relation on terms.

- If $\theta = A$: Type, then the icc is the same as that of $\Gamma \Rightarrow A = A$: Type.

- If θ is $r = s : A$, the icc is
  $\forall\langle\vec{r},\vec{r}'\rangle \in [\![\Gamma]\!].\langle r[\vec{x} := \vec{r}], s[\vec{x} := \vec{r}']\rangle \in [\![A]\!]_{\vec{x}\mapsto\vec{r}}$.

- If θ is $r : A$, the icc is the same as that of $\Gamma \Rightarrow r = r : A$.

Standard types are modelled in a straightforward way, e.g.

$$\begin{aligned}[\![\Pi x : A.B]\!]_\rho := \{\langle r, r'\rangle \,|\forall\langle s, s'\rangle \in [\![A]\!]_\rho([\![B]\!]_{\rho[x\mapsto s]} = [\![B]\!]_{\rho[x\mapsto s']}\\ \wedge\langle r(s), r'(s')\rangle \in [\![B]\!]_{\rho[x\mapsto s]})\}\end{aligned}$$

Interpretation of the W-type. The W-type corresponds in Kripke-Platek set theory to the step to the next admissible. In order to interpret this type, we will introduce for every type A, when defining $[\![A]\!]_\rho$, additionally an $\alpha \in \mathrm{Ord}$ s.t. $[\![A]\!]_\rho \in \mathrm{L}_\alpha$ for any environment ρ. This will be done in such a way that $\alpha < \kappa_n (= \aleph^{\mathrm{rec}}_{\mathrm{I}^\delta+n})$, if the W-rank of A is n. The definition of α is straightforward, except in case of $\mathrm{W}x : A.B(x)$:

Assume $[\![A]\!]_{\rho'}$ and $[\![B]\!]_{\rho'}$ have already been defined for environments ρ'. Then we define $[\![\mathrm{W}x : A.B(x)]\!]_\rho$ as follows: First, we define for environments ρ an operator Γ_ρ on sets of pairs of terms by

$$\begin{aligned}\Gamma_\rho(X) := \mathrm{Cl}(\{\langle \sup(r, s), \sup(r', s')\rangle \mid \\ \langle r, r'\rangle \in [\![A]\!]_\rho \wedge [\![B]\!]_{\rho(x\mapsto r)} = [\![B]\!]_{\rho(x\mapsto r')} \wedge \langle s, s'\rangle \in [\![B]\!]_{\rho(x\mapsto r)} \to X\})\end{aligned}$$

Here $\mathrm{Cl}(X)$ is the closure of X under reductions, i.e.

$$\mathrm{Cl}(X) := \{\langle r, r'\rangle \mid \exists\langle s, s'\rangle \in X.(r \longrightarrow s \wedge r' \longrightarrow s')\}$$

If α is s.t. for all ρ $[\![A]\!]_\rho, [\![B]\!]_\rho \in \mathrm{L}_\alpha$ and α^+ is the least admissible above α, we can define the least fixed point of Γ_ρ as the iteration $\Gamma_\rho^{\alpha^+}$ of Γ_ρ α^+-times (starting with the empty set and taking at limit points the union), and therefore define

$$[\![\mathrm{W}x : A.B(x)]\!]_\rho := \Gamma_\rho^{\alpha^+} \quad .$$

It is easy to see that if $\langle r, r'\rangle \in [\![A]\!]_\rho$, and $\langle s, s'\rangle \in [\![B]\!]_{\rho(x\mapsto r)} \to [\![\mathrm{W}x : A.B(x)]\!]_\rho$, then $\langle \sup(r, s), \sup(r', s')\rangle \in [\![\mathrm{W}x : A.B(x)]\!]_\rho$: We have $[\![B]\!]_{\rho(x\mapsto r)} \in \mathrm{L}_{\alpha^+}$, and for all t, t' s.t. $\langle t, t'\rangle \in [\![B]\!]_{\rho(x\mapsto r)}$ it follows $\langle s(t), s'(t')\rangle \in \Gamma_\rho^{\alpha^+}$, therefore there exists by the admissibility of κ a $\gamma < \alpha^+$ s.t. for all such t, t' we have $\langle s(t), s'(t')\rangle \in \Gamma_\rho^\gamma$, and therefore $\langle \sup(r, s), \sup(r', s')\rangle \in \Gamma_\rho^{\gamma+1} \subseteq \Gamma_\rho^{\alpha^+}$.

It is easy to show that $[\![\mathrm{W}x : A.B(x)]\!]_\rho$ is a partial equivalence relation. In order to show that the correctness condition w.r.t. the induction rule is fulfilled, one shows first, assuming the correctness of the premises of the induction rule, by induction over γ the correctness of the conclusion for elements of Γ_ρ^γ instead of $[\![\mathrm{W}x : A.B(x)]\!]_\rho$. Then the correctness of the conclusion without this replacement follows.

Interpretation of the universe. In order to interpret the universe, we define simultaneously a set of pairs of terms U^α, and for r, r' s.t. $\langle r, r'\rangle \in \mathrm{U}^\alpha$ a set of pairs of terms $\mathrm{T}^\alpha(r)$, s.t. U^α and $\mathrm{T}^\alpha(r)$ are partial equivalence relations, $\mathrm{U}^\alpha \subseteq \mathrm{U}^\beta$ and $\mathrm{T}^\beta(r) = \mathrm{T}^\alpha(r)$ for $\alpha < \beta$ and r s.t. $\langle r, r\rangle \in \mathrm{U}^\alpha$, and s.t. U^α and $\mathrm{T}^\alpha(r)$ are in $\mathrm{L}_{\aleph^{\mathrm{rec}}_{\alpha+2}}$. Then we interpret $\llbracket \mathrm{U} \rrbracket_\rho := \mathrm{U}^{\mathrm{I}}$ and $\llbracket \mathrm{T}(r) \rrbracket_\rho := \mathrm{T}^{\mathrm{I}}(r_\rho)$ (which is empty for $\langle r_\rho, r_\rho \rangle \not\in \mathrm{U}^{\mathrm{I}}$), where I is the first recursively inaccessible ordinal, and r_ρ is the result of substituting in r free variables according to ρ.

The inductive definition is straightforward. E.g. if $\langle r, r' \rangle \in \mathrm{U}^{<\alpha}$ (where $\mathrm{U}^{<\alpha} = \bigcup_{\beta<\alpha} \mathrm{U}^\beta$) and for $\langle t, t' \rangle \in \mathrm{T}^{<\alpha}(r)$, $\langle s[x := t], s'[x' := t'] \rangle \in \mathrm{U}^{<\alpha}$, then $\langle \widehat{\mathrm{W}}x : r.s, \widehat{\mathrm{W}}x' : r'.s' \rangle \in \mathrm{U}^\alpha$ and $\mathrm{T}^\alpha(\widehat{\mathrm{W}}x : r.s) = \llbracket \mathrm{W}x : \mathrm{T}(r).\mathrm{T}(s) \rrbracket'$, where $\llbracket \mathrm{W}x : \mathrm{T}(r).\mathrm{T}(s) \rrbracket'$ is defined as above, but interpreting $\llbracket \mathrm{T}(r) \rrbracket$ as $\mathrm{T}^{<\alpha}(r)$, similarly for $\llbracket \mathrm{T}(s) \rrbracket_{x \mapsto t}$. Furthermore, U is closed under reductions.

We have to show $\mathrm{U}^\alpha, \mathrm{T}^\alpha(r) \in \aleph^{\mathrm{rec}}_{\alpha+2}$. The crucial part of the proof is when we add $\widehat{\mathrm{W}}x : a.b$ to U^α. By IH $\mathrm{T}^{<\alpha}(r)$ and $\mathrm{T}^{<\alpha}(s[x := t])$ are in L_β for $\beta := \sup_{\alpha'<\alpha} \aleph^{\mathrm{rec}}_{\alpha'+2}$. We have $\beta \leq \kappa := \aleph^{\mathrm{rec}}_{\alpha+1} < \aleph^{\mathrm{rec}}_{\alpha+2}$ and κ is admissible. By the admissibility of κ and $\mathrm{L}_\kappa = \bigcup_{\gamma<\kappa} \mathrm{L}_\gamma$, there exists a $\gamma < \kappa$ s.t. $\mathrm{T}^{<\alpha}(r) \in \mathrm{L}_\gamma$, and for $\langle t, t \rangle \in \mathrm{T}^{<\alpha}(r)$, $\mathrm{T}^{<\alpha}(s[x := t]) \in \mathrm{L}_\gamma$. Therefore the fixed point of the operator defining $\llbracket \mathrm{W}x : \mathrm{T}(r).\mathrm{T}(s) \rrbracket$ can be obtained by iterating the operator up to κ, and therefore $\llbracket \mathrm{W}x : \mathrm{T}(r).\mathrm{T}(s) \rrbracket \in \mathrm{L}_{\aleph^{\mathrm{rec}}_{\alpha+2}}$.

We show now that $\mathrm{U}^{<\mathrm{I}}$ is closed under the introduction rules for the universe. In case of the W-type, this is done as follows: Assume $\langle r, r' \rangle \in \mathrm{U}^{<\mathrm{I}}$ and for $\langle t, t' \rangle \in \mathrm{T}^{<\mathrm{I}}(r)$, $\langle s[x := t], s'[x' := t'] \rangle \in \mathrm{U}^{<\mathrm{I}}$. Then $\mathrm{T}^{<\mathrm{I}}(r) \in L_{\mathrm{I}}$, and therefore there exists an $\alpha <$ I s.t. for all t, t' as above $\langle s[x := t], s'[x' := t'] \rangle \in \mathrm{U}^\alpha$. Here we used that I is an admissible closed under $\lambda\alpha.\aleph^{\mathrm{rec}}_\alpha$. Now it follows that $\langle \widehat{\mathrm{W}}x : r.s, \widehat{\mathrm{W}}x' : r'.s' \rangle \in \mathrm{U}^{\alpha+1}$.

Completion of the proof of the upper bound. Now one shows that every arithmetic statement provable in $\mathrm{ML}_1\mathrm{W}$ can be shown in KPI^+. Let ψ be any arithmetic formula. We extend the set of terms by additional terms $\mathrm{C}_{\varphi,\vec{x}}$ for all subformulae φ of ψ and variables $\vec{x} = x_1, \ldots, x_k$ containing the free variables of φ, together with reduction rules for $\mathrm{C}_{\varphi,\vec{x}}$. This will be done in such a way that in the model we have provable in KPI^+ $\forall n_1, \ldots, n_k \in \omega.\langle \mathrm{C}_{\varphi,\vec{x}}(n_1, \ldots, n_k), \mathrm{C}_{\varphi,\vec{x}}(n_1, \ldots, n_k)\rangle \in \llbracket \varphi \rrbracket_{\vec{x} \mapsto \vec{n}} \Leftrightarrow \llbracket \varphi \rrbracket_{\vec{x} \mapsto \vec{n}} \neq \emptyset \Leftrightarrow \varphi[x_1 := n_1, \ldots, x_k := n_k]$. Especially, if $\mathrm{ML}_1\mathrm{W}$ proves φ, then for all $\vec{n}$ $\llbracket \varphi \rrbracket_{\vec{x} \mapsto \vec{n}}$ is inhabited and therefore $\varphi[\vec{x} := \vec{n}]$ holds in KPI^+. If $\varphi \equiv r = s$ then $\mathrm{C}_{\varphi,\vec{x}}(\vec{n}) \longrightarrow \mathrm{refl}$, if $r[\vec{x} := \vec{n}] = s[\vec{x} := \vec{n}]$ is true, otherwise the term does not reduce. Here refl is the canonical element of the identity type $r =_{\mathrm{N}} s$ between r and s. If $\varphi \equiv \varphi_0 \wedge \varphi_1$, $\mathrm{C}_{\varphi,\vec{x}}(\vec{n}) \longrightarrow \langle \mathrm{C}_{\varphi_0,\vec{x}}(\vec{n}), \mathrm{C}_{\varphi_1,\vec{x}}(\vec{n})\rangle$, if $\varphi \equiv \varphi_0 \vee \varphi_1$ then $\mathrm{C}_{\varphi,\vec{x}}(\vec{n}) \longrightarrow \mathrm{inl}(\mathrm{C}_{\varphi_0,\vec{x}}(\vec{n}))$ if $\varphi_0[\vec{x} := \vec{n}]$ holds, and $\mathrm{C}_{\varphi,\vec{x}}(\vec{n}) \longrightarrow \mathrm{inr}(\mathrm{C}_{\varphi_1,\vec{x}}(\vec{n}))$ otherwise. If $\varphi \equiv \varphi_0 \rightarrow \varphi_1$ then $\mathrm{C}_{\varphi,\vec{x}}(\vec{n}) \longrightarrow \lambda x.\mathrm{C}_{\varphi_1,\vec{x}}(\vec{n})$. if $\varphi \equiv \forall x.\psi$, $\mathrm{C}_{\varphi,\vec{x}}(\vec{n}) \longrightarrow \lambda y.\mathrm{C}_{\psi,\vec{x},x}(\vec{n}, y)$, and if $\varphi \equiv \exists x.\psi$, then $\mathrm{C}_{\varphi,\vec{x}}(\vec{n}) \longrightarrow \langle n, \mathrm{C}_{\psi,\vec{x},x}(\vec{n}, n)\rangle$, if n is minimal s.t. $\psi[\vec{x} := \vec{n}, x := n]$ holds; if there is no such n, there is no reduction for $\mathrm{C}_{\varphi,\vec{x}}(\vec{n})$.

It follows that, if an arithmetic sentence ψ is provable in $\mathrm{ML}_1\mathrm{W}$, then by forming the above model w.r.t. ψ it follows that ψ holds in KPI^+. Therefore the limit of transfinite induction provable for all arithmetic formulae in $\mathrm{ML}_1\mathrm{W}$ is less than or equal that for KPI^+, therefore $|\mathrm{ML}_1\mathrm{W}| \leq |\mathrm{KPI}^+|$.

Ordinal notation system. Whereas for the upper bound one is relying on the proof theoretic analysis of KPI^+, the lower bound will be carried out explicitly, and we need first to set up an ordinal notation system of appropriate strength.

The ordinal notation system will have as basic constants 0 and I, where I is the first strongly inaccessible cardinal.[2] One takes as basic functions addition of

[2] One could replace all cardinals in the following by their recursive analogues (admissibles), but setting up an ordinal notation system like this is much more complicated. What eventually matters is only the resulting ordinal notation system, which is primitive recursive and w.r.t. which we prove upper and lower bounds for the proof theoretic strength of corresponding theories. The set theoretic development can be considered as mere heuristic. It is a however a very valuable heuristic, since it has contributed to a much better understanding of the ordinal notation systems

ordinals, the Veblen function φ (where $\varphi 0\beta = \omega^\beta$ and for $\alpha > 0$ $\varphi\alpha\beta$ is the βth common fixed point of $\lambda\beta'.\varphi\alpha'\beta'$ for $\alpha' < \alpha$) and $\lambda\alpha.\Omega_\alpha$ (where $\Omega_0 = 0$ and $\Omega_\alpha = \aleph^{\mathrm{rec}}_\alpha$ otherwise). Furthermore, one adds the collapsing function ψ. Here one defines simultaneously for $\kappa \in \mathrm{R}$ (where R is the set of regular cardinals) by recursion on α

$$\begin{aligned} \psi_\kappa\alpha &:= \min\{\beta \mid \kappa \in C(\alpha,\beta) \wedge C(\alpha,\beta) \cap \kappa \subseteq \beta\} \ , \\ C(\alpha,\beta) &:= \text{Closure of } \beta \text{ under } 0, \mathrm{I}, +, \varphi, \lambda\alpha.\Omega_\alpha, \lambda\pi \in \mathrm{R}.\lambda\xi < \alpha.\psi_\pi\xi \ . \end{aligned}$$

It is outside the scope of this article to give a detailed explanation of the ψ-function, here we give only a few remarks. If $\kappa = \Omega_{\beta+1}$, then $\psi_\kappa\alpha$ is the least ordinal $\geq \Omega_\alpha$, s.t., if we form the closure under basic constants, basic functions and all collapsing functions we have defined before, i.e. with arguments $< \alpha$, we do not get any new ordinals $< \kappa$. In case of $\kappa = \mathrm{I}$, I is automatically contained in the closure and we need only that the same closure as before does not contain any new ordinals $<$ I. Let Γ_α the αth fixed point of $\lambda\alpha.\varphi\alpha 0$, i.e. the αth ordinal, which cannot be defined from smaller ones using 0, + and the Veblen function. Let β be the first fixed point of $\lambda\alpha.\Gamma_\alpha$. For $\alpha < \beta$, $\psi_{\Omega_1}\alpha = \Gamma_\alpha$, and $\beta = \psi_{\Omega_1}(\Omega_1)$. Let I_α be the αth (not necessary regular) fixed point of $\lambda\alpha.\Omega_\alpha$ and γ be the first (non-regular) fixed point of $\lambda\alpha.\mathrm{I}_\alpha$. For $\alpha < \gamma$, $\psi_\mathrm{I}(\alpha) = \mathrm{I}_\alpha$ and $\psi_\mathrm{I}(\mathrm{I}) = \gamma$.

If one adapts the analysis of KPI in [Buchholz, 1992] to an analysis of KPI$^+$, one can see that $|\mathrm{KPI}^+| = \psi_{\Omega_1}(\Omega_{\mathrm{I}+\omega})$.

Lower bound for the strength of $\mathrm{ML_1W}$. The lower bound is obtained by proving directly in $\mathrm{ML_1W}$ transfinite induction up to $\psi_{\Omega_1}(\Omega_{\mathrm{I}+n})$ for (Meta-)every $n < \omega$. Then it follows $|\mathrm{ML_1W}| \geq \sup_{n\in\omega}\psi_{\Omega_1}(\Omega_{\mathrm{I}+n}) = \psi_{\Omega_1}\Omega_{\mathrm{I}+\omega} = |\mathrm{KPI}^+|$. We will use the technique of distinguished sets, which is due to Buchholz. Before introducing it, we start with some basic definitions.

An expression $C[x]$ is a type expression possibly depending on a free variable x, and we write $C[t]$ for $C[x := t]$ (possibly applying some α-conversion to C first in order to avoid clashes of bound and free variables). A subset B of a set A is a function $B : A \to \mathrm{U}$, and a subclass is an expression $B[x]$ s.t. $x : A \Rightarrow B[x] : \mathrm{Type}$. In case of subsets we write $x \in B$ for $\mathrm{T}(B(x))$ and in case of subclasses $x \in B$ for $B[x]$. $\forall x \in B.C[x] := \forall x : A.x \in B \to C[x]$, similarly for $\exists x \in B.C[x]$. The following definitions can be carried out both for sets and for classes, although we we will explicitly only define them for sets. If B, C are subsets of A, then $B \subseteq C :\Leftrightarrow \forall x \in B.x \in C$, $B \cong C :\Leftrightarrow B \subseteq C \wedge C \subseteq B$. A partially ordered set (class) $(A, <)$ is a type A together with a binary relation $<$ on A, i.e. a type expression $x < y$ s.t. $x : A, y : A \Rightarrow x < y : \mathrm{Type}$. For partially ordered sets A and $a, b \in A$, $a \leq b :\Leftrightarrow a = b \vee a < b$. For $a \in A$, $B, C \subseteq A$, $a \leq B :\Leftrightarrow \exists b \in B.a \leq b$, $C \leq B :\Leftrightarrow \forall c \in C.c \leq B$. For partially ordered sets $(A, <)$ we identify $a \in A$ with $\{b \in A \mid b < a\}$ and write $a + 1$ for $\{b \in A \mid b \leq a\}$; this explains for instance notions like $A \cap (a + 1)$. If $B, C \subseteq A$, $B \sqsubseteq C$ ("B is an initial segment of C") iff $\forall b \in B.B \cap (b + 1) \cong C \cap (b + 1)$. If $(A, <)$ is a partially ordered set (i.e. $<$ is a binary relation on a set A), we can form the accessible part $\mathrm{Acc}(A)$ as the largest well-founded segment of A.

Let $(\mathrm{OT}, <)$ be the ordinal notation system constructed from the above mentioned functions. We define for $A \subseteq \mathrm{OT}$ the set $\mathrm{C}^a(A)$ as the closure of $A \cap a$ under 0, I, +, φ, $\lambda\gamma.\Omega_\gamma$ and $\lambda\kappa > a.\lambda\gamma.\psi_\kappa\gamma$. Let $\mathrm{M}(A) := \{\alpha \mid \alpha \in \mathrm{C}^\alpha(A)\}$. So the elements of $\mathrm{M}(A)$ are those, which can be formed from $A \cap \alpha$ from basic functions and collapsing functions ψ_κ with $\kappa > \alpha$.

developed.

Distinguished sets. In order to get an idea of what a distinguished set is, we will give first some examples of distinguished sets, as they were introduced originally. Later we will slightly change this definition.

The first one A_0 is the accessible part of the ordinals below Ω_1. The next one, A_1 is the union of A_0 with the accessible part of the ordinals $\alpha \in [\Omega_1, \Omega_2[$ s.t. their components below Ω_1 are in A_0, i.e. s.t. $\alpha \in \mathrm{C}^{\Omega_1}(A_0)$. The next one, A_2 is the union of A_1 with the accessible part of the ordinals $\alpha \in [\Omega_2, \Omega_3[$ having components below Ω_2 in A_1, i.e. s.t. $\alpha \in \mathrm{C}^{\Omega_2}(A_1)$. This series can be iterated transfinitely in an obvious way. Sets A introduced in this way can be characterised as having the following property: for all α s.t. $\Omega_\alpha \leq A$ we have that $A \cap [\Omega_\alpha, \Omega_{\alpha+1}[$ is the accessible part of $\mathrm{C}^{\Omega_\alpha}(A) \cap [\Omega_\alpha, \Omega_{\alpha+1}[$. This was essentially the original definition of a distinguished set. In order to avoid the jumps at Ω_α, we introduce the following variations of this definition: first one replaces $\beta \in \mathrm{C}^{\Omega_\alpha}(A)$ by $\beta \in \mathrm{C}^{\beta}(A)$, i.e. $\beta \in \mathrm{M}(A)$ – this has only a minor effect on the definition. Furthermore, one consideres the definition of the accessible part: This is an inductive definition of the form: "if $\beta \in \mathrm{M}(A) \cap [\Omega_\alpha, \Omega_{\alpha+1}[$ and $\beta \cap \mathrm{M}(A) \cap [\Omega_\alpha, \beta[$ is a subset of the accessible part, then β is in the accessible part". This definition, in which one examines A in slices of the form $[\Omega_\alpha, \Omega_{\alpha+1}[$, will now replaced by the following inductive definition of an unsliced set $\mathrm{W}(A)$: $\mathrm{W}(A)$ is the least set Y s.t., if $\alpha \in \mathrm{C}^\alpha(A)$ and $\mathrm{C}^\alpha(A) \cap \alpha \subseteq Y$, then $\alpha \in Y$. Our final definition of distinquished is now as follows:

$$A \text{ is distinguished } \Leftrightarrow A \sqsubseteq \mathrm{W}(A)$$

Assuming that $A \cap \Omega_\alpha \cong \mathrm{W}(A) \cap \Omega_\alpha$ (therefore $A \cap \Omega_\alpha$ is distinguished) and $A \cap \Omega_{\alpha+1} \cong \mathrm{M}(A) \cap \Omega_{\alpha+1}$ (i.e. $A \cap \Omega_{\alpha+1}$ is sufficiently closed), it follows that $\mathrm{W}(A) \cap [\Omega_\alpha, \Omega_{\alpha+1}[$ is the accessible part of the set $[\Omega_\alpha, \Omega_{\alpha+1}[\cap \mathrm{M}(A)$, therefore the new definition is essentially the same as the original definition of distinguished sets. We will below show how to introduce $\mathrm{W}(A)$ and therefore as well the notion of "A is distinguished" in $\mathrm{ML}_1\mathrm{W}$.

Distinguished sets are well-ordered (since, if A is distinguished, $\alpha, \beta \in A$, $\alpha < \beta$, then $\alpha \in \mathrm{C}^\beta(A) \cap \beta$). Now one can show using transfinite induction over distinguished sets that distinguished sets are essentially approximations of the same class: If A, B are distinguished sets, $\alpha \in A$ and $\alpha \leq B$, then $A \cap (\alpha + 1) \cong B \cap (\alpha + 1)$. In type theory, we take now $\mathcal{P}(\mathrm{N}) := \mathrm{N} \to \mathrm{U}$ as the notion of the powerset of N and form the union of all distinguished sets. This union will be a class $\mathcal{W}$. $\mathcal{W}$ is a distinguished class, i.e. if we define W for classes, we obtain $\mathcal{W} \sqsubseteq \mathrm{W}(\mathcal{W})$.

If one forms by Meta-induction on n the classes $\mathcal{W}_0 := (\mathcal{W} \cap \mathrm{I}) \cup \{\mathrm{I}\}$, $\mathcal{W}_{n+1} := (\mathrm{W}(\mathcal{W}_n) \cap \Omega_{\mathrm{I}+n+1}) \cup \{\Omega_{\mathrm{I}+n+1}\}$, one obtains as well distinguished classes. It can be shown, using induction over distinguished sets, that distinguished sets and classes A are closed under 0, I, +, the step to the next cardinal, the Veblen function, the collapsing functions and $\lambda\alpha.\Omega_\alpha$ bounded by A, i.e. if the result of applying of these operations is $\leq A$, then it is in A. It follows that $\Omega_1, \Omega_{\mathrm{I}+n} \in \mathcal{W}_n$, therefore $\psi_{\Omega_1}(\Omega_{\mathrm{I}+n}) \in \mathcal{W}_n \cap \Omega_1$. Furthermore, $\mathcal{W}_n \cap \Omega_1$ is an initial segment of OT which is well-ordered, and therefore we obtain transfinite induction up to $\psi_{\Omega_1}(\Omega_{\mathrm{I}+n})$, provable in $\mathrm{ML}_1\mathrm{W}$.

Definition of $\mathrm{W}(A)$**.** First we define $\mathrm{C}^a(A)$ in type theory. We have that $b \in \mathrm{C}^a(A)$, if $b \in A \cap a$ or b can be formed by one of the operations from other elements of $\mathrm{C}^a(A)$, where the latter terms are smaller. When we unfold this, we get a finite formula of the shape $b \in \mathrm{C}^a(A) \Leftrightarrow b \in A \cap a \vee ((c_0 \in A \cap a \vee \cdots) \wedge (c_1 \in A \cap a \vee \cdots))$. This formula can be transformed into disjunctive normal form with atomic formulae of the form $c \in A \cap a$ and therefore one can define for $b \in \mathrm{OT}$ a finite set $\mathrm{K}_a(b)$ of finite sets of elements of OT s.t. $b \in \mathrm{C}^a(A) \Leftrightarrow \exists C \in \mathrm{K}_a(b).C \subseteq A \cap a$. $\mathrm{K}_a(b)$ can be introduced in type theory directly by induction over OT. This way we obtain a formalisation of $\mathrm{C}^a(A)$ in type theory.

$\mathrm{W}(A)$ itself can be defined using the W-type as follows: First a system of rules for deriving statements of the form $a \in \mathrm{W}(A)$ for $a : \mathrm{N}$ can be given by having for each $a \in \mathrm{C}^a(A)$ one rule

$$\frac{\cdots \qquad a' \in \mathrm{W}(A) \qquad \cdots \ (a' \in \mathrm{C}^a(A) \cap a)}{a \in \mathrm{W}(A)}$$

In type theory we can represent such derivations as elements of $\mathrm{W}b : B.C(b)$, where $B := \Sigma a : \mathrm{N}.a \in \mathrm{C}^a(A)$ and $C(\langle a, p\rangle) := \Sigma a' : \mathrm{N}.a' \in \mathrm{C}^a(A) \cap a$. An element $w := \sup(\langle a, p\rangle, q)$ derives $a \in \mathrm{W}(A)$ from subderivations $q(c)$ $(c : C(\langle a, p\rangle))$, provided at each subtree of w (including w) the labels of the trees are respected: Define $\mathrm{label} : (\mathrm{W}b : B.C(b)) \to \mathrm{N}$, $\mathrm{label}(\sup(\langle a, p\rangle, q)) = a$. Then the local correctness needed at each subtree w' is: If $w' = \sup(\langle a', p'\rangle, q')$, then the $\langle a'', q''\rangle$th subtree of w' has label a'', i.e. $\mathrm{label}(q'(\langle a'', q''\rangle)) =_{\mathrm{N}} a''$. This local property for subtree w' will be called $\mathrm{LocCor}(w')$. In order to define the notion of a subtree, we first define the notion $w \prec^1_{\mathrm{Tree}} w'$, "$w$ is an immediate subtree of w'" recursively as $w \prec^1_{\mathrm{Tree}} \sup(b, p) := \exists c : C(b).w =_{\mathrm{W}b:B.C(b)} p(c)$. Here $a =_D b$ denotes the intensional equality type for $a, b : D$. Now we define $w \preceq_{\mathrm{Tree}} w'$, "$w$ is a subtree of w' or equal to w'", as "there exists a sequence of trees $(w_0, \ldots, w_n)$ s.t. $n \geq 0$, $w_0 =_{\mathrm{W}b:B.C(b)} w'$, $w_n =_{\mathrm{W}b:B.C(b)} w$ and for $k = 0, \ldots, n-1$ $w_{k+1} \prec^1_{\mathrm{Tree}} w_k$. We define $\mathrm{Cor}(w) := \forall w' : (\mathrm{W}b : B.C(b)).w' \preceq_{\mathrm{Tree}} w \to \mathrm{LocCor}(w')$ and

$$a \in \mathrm{W}(A) :\Leftrightarrow \exists w : (\mathrm{W}b : B.C(b)).\mathrm{Cor}(w) \wedge \mathrm{label}(w) =_A a \ .$$

It is an easy exercise to verify that $a \in \mathrm{C}^a(A) \to (\forall a' \in \mathrm{C}^a(A) \cap a.a' \in \mathrm{W}(A)) \to a \in \mathrm{W}(A)$ holds and that we obtain the least such set. The latter can be expressed by the following induction principle (assuming $a : \mathrm{N} \Rightarrow D(a) : \mathrm{Type}$):
$(\forall a \in \mathrm{W}(A).(\forall a' \in \mathrm{C}^a(A) \cap a.D(a')) \to D(a)) \to \forall a \in \mathrm{W}(A).D(a)$. This completes the well-ordering proof.

Appendix C: A Model for Type Theory with W-Type and one Mahlo Universe

We will in the following give the details of the model of MLM in KPM^+. The main construction is as for KPI^+, the only difference is of course how to interpret the Mahlo universe itself. This is done as follows:

Similarly as when interpreting the universe of $\mathrm{ML_1W}$, one defines simultaneously sets of pairs of terms V^α together with a set of pairs of terms $\mathrm{T}^\alpha(r)$ for r s.t. $\langle r, r\rangle \in \mathrm{V}_\alpha$, with similar conditions as before. We identify in the following the names of the constructors for sets in V and in $\mathrm{U}_{\vec{f}}$, i.e. we identify $\widetilde{\mathrm{N}}$ with $\widehat{\mathrm{N}}$, $\widetilde{\Sigma}$ with $\widehat{\Sigma}$ etc.

The inductive definition is for standard universe constructors as for an ordinary universe. Assume now terms $\vec{f}$, $\vec{g}$, and an α is s.t. V^α is closed under the standard universe constructions, and s.t. if $a \in \mathrm{V}^\alpha$, $b \in \mathrm{T}^\alpha(a) \to \mathrm{V}^\alpha$, then $\langle f_0(a, b), g_0(a, b)\rangle \in \mathrm{V}^\alpha$ and for $c : \mathrm{T}^\alpha(f_0(a, b))$, $\langle f_1(a, b, c), g_1(a, b, c)\rangle \in \mathrm{V}^\alpha$. Then $\langle \widehat{\mathrm{U}}_{\vec{f}}, \widehat{\mathrm{U}}_{\vec{g}}\rangle \in \mathrm{V}^{\alpha+1}$ and $\mathrm{T}(\widehat{\mathrm{U}}_{\vec{f}}) = \mathrm{V}^\beta$ for the minimal $\beta \leq \alpha$, s.t. the above conditions hold with α replaced by β.

Now let $\mathrm{M} := \bigcup\{\alpha \mid \alpha \in \mathrm{Ad_M}\}$ be the recursively Mahlo ordinal corresponding to $\mathrm{Ad_M}$ and introduce a model with $[\![\,\mathrm{V}\,]\!] := \mathrm{V}^{\mathrm{M}}$. Since M is recursively inaccessible, it follows as before that V^{M} is closed under the usual universe constructions. We verify now that V^{M} is closed under the introduction rule for $\widehat{\mathrm{U}}$. For simplicity, we work as if the interpretation of sets were a set of terms rather than a set of pairs of terms considered to be equal, and we assume that we do not have any context. Assume $f_0 \in [\![(a : \mathrm{V}, b : \mathrm{T}(a) \to \mathrm{V}) \to \mathrm{V}]\!]$ and $f_1 \in [\![(a : \mathrm{V}, b : \mathrm{T}(a) \to \mathrm{V}, \mathrm{T}(f_0'(a,b))) \to \mathrm{V}]\!]_{f_0' \mapsto f_0}$. This means that for $a \in \mathrm{V}^{\mathrm{M}}$, $b \in \mathrm{T}^{\mathrm{M}}(a) \to \mathrm{V}^{\mathrm{M}}$, $f_0(a,b) \in \mathrm{V}^{\mathrm{M}}$ and for $c \in \mathrm{T}^{\mathrm{M}}(f_0(a,b))$, $f_1(a,b,c) \in \mathrm{V}^{\mathrm{M}}$. Assume now $\alpha < \mathrm{M}$. For every $a \in \mathrm{V}^{\alpha}$ and $b \in \mathrm{T}^{\alpha}(a) \to \mathrm{V}^{\alpha}$ there exists a $\beta < \mathrm{M}$ s.t. $f_0(a,b) \in \mathrm{V}^{\beta}$. For every $c \in \mathrm{T}^{\beta}(f_0(a,b))$ there exists a $\gamma < \mathrm{M}$ s.t. $f_1(a,b,c) \in \mathrm{V}^{\gamma}$. Using admissibility of M it follows that there exists a $\beta < \mathrm{M}$ s.t. $f_0(a,b) \in \mathrm{V}^{\beta}$ and for $c \in \mathrm{T}^{\beta}(a,b)$, $f_1(a,b,c) \in \mathrm{V}^{\beta}$. Let $\varphi(\alpha,\beta) := \forall a \in \mathrm{V}^{\alpha}.\forall b \in \mathrm{T}^{\alpha}(a) \to \mathrm{V}^{\alpha}.f_0(a,b) \in \mathrm{V}^{\beta} \wedge \forall c \in \mathrm{T}^{\alpha}(f_0(a,b)).f_1(a,b,c) \in \mathrm{V}^{\beta}$. Then $\forall \alpha < \mathrm{M}.\exists \beta < \mathrm{M}.\varphi(\alpha,\beta)$, where $\varphi(\alpha,\beta)$ is Δ_0. Furthermore, we have $\forall \alpha < \mathrm{M}.\exists \beta < \mathrm{M}.\beta = \aleph^{\mathrm{rec}}_{\alpha}$ which is as well a Δ_0-formula. By the Mahlo axiom there exists a $\kappa < \mathrm{M}$, which is admissible, s.t. $\forall \alpha < \kappa.\exists \beta < \kappa.\beta = \aleph^{\mathrm{rec}}_{\alpha}$, i.e. s.t. κ is recursively inaccessible, and s.t. $\forall \alpha < \kappa.\exists \beta < \kappa.\varphi(\alpha,\beta)$. Since we have demanded $\mathrm{T}^{\alpha}(a) \in \mathrm{L}_{\aleph^{\mathrm{rec}}_{\alpha}}$ it follows $\mathrm{T}^{\kappa}(a) \in \mathrm{L}_{\kappa}$ for $a \in \mathrm{V}^{\kappa}$. Using admissibility of κ it follows from the above that if $a \in \mathrm{V}^{\kappa}$, $b \in \mathrm{T}^{\kappa}(a) \to \mathrm{V}^{\kappa}$ then $f_0(a,b) \in \mathrm{V}^{\kappa}$ and for $c \in \mathrm{T}^{\kappa}(f_0(a,b))$, $f_1(a,b,c) \in \mathrm{V}^{\kappa}$, and therefore $\widehat{\mathrm{U}}_{\vec{f}} \in \mathrm{V}^{\kappa+1} \subseteq \mathrm{V}^{\mathrm{M}}$.

Appendix D: Proof that Inductive-Recursive Definitions Reach the Strength of KPM

We show here how to adapt the well-ordering proof for the Mahlo universe in order to show that the theory of inductive-recursive definitions reaches the strength of KPM, Kripke-Platek set theory plus recursive Mahloness of the set-theoretic universe. We make use of the fact that Set is a weak variant of the Mahlo universe, as described in Sect. 7. In the well-ordering proof for type theory with one Mahlo universe, similarly as in that for $\mathrm{ML}_1\mathrm{W}$, one introduces the union of distinguished sets $\mathcal{W}$, where sets were elements of $\mathrm{N} \to \mathrm{V}$ and V is the Mahlo universe. Then one introduces finitely many distinguished classes on top of $\mathcal{W}$, which corresponds to the step to finitely many admissibles above the recursively Mahlo ordinal. The formation of those sets makes use of the W-type, formed using sets depending on V. In the theory of inductive-recursive definitions, we can from $\mathcal{W}$ as union of all distinguished sets as elements of $\mathrm{N} \to \mathrm{Set}$, but we cannot form the nth admissible set above $\mathcal{W}$. We cannot express that $\mathcal{W}$ is distinguished, however $\mathcal{W}$ will inherit the closure properties of distinguished sets. One can show transfinite induction over $\mathcal{W}' := (\mathcal{W} \cap \mathrm{M}) \cup \{\mathrm{M}\}$. Using induction into types one can then show transfinite induction up to the $n$times nested Cantor Normal Form over elements in $\mathcal{W}'$, which is essentially transfinite induction up to $\underbrace{\omega^{\omega^{\cdot^{\cdot^{\omega^{\mathrm{M}+1}}}}}}_{n \text{ times}}$ This allows then show that $\alpha_n := \psi_{\Omega_1}(\underbrace{\omega^{\omega^{\cdot^{\cdot^{\omega^{\mathrm{M}+1}}}}}}_{n \text{ times}})$ is in $\mathcal{W}$, and, since $\mathcal{W} \cap (\alpha_n + 1)$ is a segment of OT, transfinite induction up to α_n. Since $\psi_{\Omega_1}(\epsilon_{\mathrm{M}+1}) = \sup_{n \in \omega} \alpha_n$, it follows that the strength of the type theory of inductive-recursive definitions is at least $\psi_{\Omega_1}(\epsilon_{\mathrm{M}+1})$, which is the proof theoretic ordinal of KPM. The strength of KPM is only slightly below that of KPM^{+}, the strength of type theory with one Mahlo universe.

References

AUGUSTSSON, L., “Cayenne - a language with dependent types”. In “International Conference on Functional Programming”, 1998, p. 239–250.

BARWISE, J., “Admissible Sets and Structures. An Approach to Definability Theory”, Omega-series, Springer, 1975.

BENKE, M., "Towards generic programming in type theory". Presentation at Annual ESPRIT BRA TYPES Meeting, Berg en Dal, submitted for publication, 2002. Available via `http://www.cs.chalmers.se/~marcin/Papers/Notes/nijmegen.ps.gz`.

BENKE, M., DYBJER, P., and JANSSON, P., "Universes for generic programs and proofs in dependent type theory", submitted, 2003.

BUCHHOLZ, W., "Notation systems for infinitary derivations", *Arch. Math. Logic* 30, 1991, p. 277 – 296.

BUCHHOLZ, W., "A simplified version of local predicativity". In Aczel, P., Simmons, H., and Wainer, S. S., editors, "Proof Theory. A selection of papers from the Leeds Proof Theory Programme 1990", Cambridge University Press, 1992, p. 115 – 147.

DYBJER, P., "A general formulation of simultaneous inductive-recursive definitions in type theory", *J.Sym.Log.* 65(2), 2000, p. 525 – 549.

DYBJER, P. and SETZER, A., "Finite axiomatizations of inductive and inductive-recursive definitions". In "Workshop on Generic Programming, Marstrand, Sweden, 18 June 1998", http://www.cs.ruu.nl/people/johanj/programme_wgp98.html, 1998.

DYBJER, P. and SETZER, A., "A finite axiomatization of inductive-recursive definitions". In Girard, J.-Y., editor, "Typed Lambda Calculi and Applications", volume 1581 of Springer Lecture Notes in Computer Science, p. 129–146, 1999.

DYBJER, P. and SETZER, A., "Indexed induction-recursion". In Kahle, R., Schroeder-Heister, P., and Stärk, R., editors, "Proof Theory in Computer Science", LNCS 2183, 2001, p. 93 – 113.

DYBJER, P. and SETZER, A. (2003a), "Indexed induction-recursion", long version, submitted, 2003.

DYBJER, P. and SETZER, A. (2003b), "Induction-recursion and initial algebras", *Annals of Pure and Applied Logic* 124, 2003, p. 1 – 47.

GENTZEN, G., "Die Widerspruchsfreiheit der reinen Zahlentheorie", *Mathematische Annalen* 112, 1936, p. 493 – 565.

GÖDEL, K., "Über formal unentscheidbare Sätze der Principia mathematica und verwandter Systeme I", *Monatshefte für Mathematik und Physik* 38, 1931, p. 173 – 198.

GOODSTEIN, R. L., "Recursive Number Theory", North-Holland, 1964.

GRIFFOR, E. and RATHJEN, M., "The strength of some Martin-Löf type theories", *Arch. math. Log.* 33, 1994, p. 347 – 385.

HEIJENOORT, J. v., "From Frege to Gödel", Harward University Press, 1967.

HILBERT, D., "Mathematische probleme", *Gött. Nachr.*, 1900, p. 253–297.

HILBERT, D. and BERNAYS, P., "Grundlagen der Mathematik. Zweiter Band", Julius Springer, Berlin, 1939.

HINMAN, P. G., "Recursion-Theoretic Hierarchies", Springer, 1978.

JÄGER, G., "Theories for Admissible Sets: A Unifying Approach to Proof Theory", Bibliopolis, Naples, 1986.

KRIPKE, S., "Transfinite recursion on admissible ordinals, I, II (abstracts)", *J. Symbolic Logic* 29, 1964, p. 161 – 162.

MICHELBRINK, M., "Zur endlichen Behandlung der Beweistheorie schwacher Fragmente der Mengenlehre: KP + Π_3-Reflexion", PhD thesis, Fachbereich Mathematik und Informatik, Hannover, 2000.

MINTS, G. and TUPAILO, S., "Epsilon-substitution method for the ramified language and Δ^1_1-comprehension rule". In Cantini, A., editor, "Logic and foundations of mathematics. Proceedings of the congress of logic, methodology and philosophy of science, Florence, 1995", Synthese Library 280, Kluwer, 1999, p. 107 – 130.

MINTS, G., TUPAILO, S., and BUCHHOLZ, W., "Epsilon substitution method for elementary analysis", *Arch. Math. Logic* 35(2), 1996, p. 103 – 130.

PALMGREN, E., "On universes in type theory", In Sambin, G. and Smith, J., editors, "Twenty five years of constructive type theory", Oxford University Press, 1998, p. 191 – 204.

PLATEK, R., "Foundations of recursion theory", Doctorial Dissertation and Supplement, Stanford, CA: Stanford University, 1966.

RATHJEN, M., "Proof-theoretical analysis of KPM", *Arch. math. Logic* 30, 1991, p. 377 – 403.

SCHLÜTER, A., "On provability in set theories with reflection," preprint.

SETZER, A., "Proof theoretical strength of Martin-Löf Type Theory with W-type and one universe", PhD thesis, Universität München, 1993. Available via http://www.cs.swan.ac.uk/∼csetzer.

SETZER, A., "A type theory for iterated inductive definitions", draft, 14 pp., 1994. Available via http://www.cs.swan.ac.uk/∼csetzer.

SETZER, A., "A model for a type theory with one Mahlo Universe," preprint, 10pp, 1996. Available via
http://www.cs.swan.ac.uk/∼setzer/articles/uppermahlo.pdf.

SETZER, A., "Well-ordering proofs for Martin-Löf type theory", *Annals of Pure and Applied Logic* 92, 1998, p. 113 – 159.

SETZER, A., "Extending Martin-Löf type theory by one Mahlo-universe", *Arch. Math. Log.* 39, 2000, p. 155 – 181.

SKOLEM, T., "Begründung der elementaren Arithmetik durch die rekurrierende Denkweise ohne Anwendung scheinbarer Veränderlichen mit unendlichem Ausdehnungsbereich", *Videnskapsselskapets skripfter, I. Matematisk-naturvidenskabelig klasse* 6, 1923. See as well [Heijenoort, 1967], p. 302 – 333.

TAIT, W., "Normal derivability in classical logic". In Barwise, J., editor, "The syntax and semantics of infinitary languages", Springer Lecture Notes in Mathematics 72, 1968, p. 204 – 236.