Arithmetics

A course by Lev Beklemishev
Personal notes of Joost J. Joosten

October 23, 2001

1 Lecture 1 (15-6-2001)

1.1 Fragments of Peano Arithmetic

In this lecture we fixed our subject of study which will be PA and fragments of PA. The language of PA will be $\{0, 1, +, \cdot, =\}$. Fragments T of arithmetic are subtheories of PA in the sense that all theorems of T are also theorems of PA. The language of T might be richer though than the language of PA. Some notions that are easily definable in PA (and their properties provable) might not be sufficiently definable in a weaker theory. So, typically, the richer language of the weaker theory contains a special symbol for a function and contains as axioms the defining properties of that function. We will encounter many such examples.

Roughly speaking fragments of arithmetics can be seen as divided into three categories of increasing strength.

- The strongest class is the class of the so-called strong fragments, like $\mathrm{I}\Sigma_n$ and $\mathrm{B}\Sigma_n$. Their classes of provably recursive total functions is less than the full class of recursive functions. These fragments are more or less characterized by their methods of proof employed like classical model theory and proof theory.

- The class of arithmetics below the strong arithmetics is referred to as weak, or bounded arithmetics. A typical example is Buss' S^1_2 where you only allow for induction over NP-predicates. (NP-predicates are represented by Σ^b_1-formula's which are built up from strictly bounded formulas by means of conjunction, disjunction, strictly bounded universal quantifiers and bounded existential quantifiers. Bounded quantifiers are quantifiers of the form $\exists x \leq t$ with x not occuring in t

and strictly bounded quantifiers are of the form $\exists x \leq |t|$, with $|t|$ denoting the length of the binairy representation of (the value of) t. ($|x| = \lceil \log_2(x+1) \rceil$, but normally $|\cdot|$ is a primitive symbol in the language.)) It is proved that the Σ_1^b definable sets are precisely the NP sets. So, the methods used in the field of bounded arithmetic are closely related to those of computational complexity. The provably total recursive functions of S_2^1 are precisely the P-time computable functions. Note that in order to get the P-time computable functions provably total one needs induction over NP predicates.

- The weakest possible fragments are referred to as systems of open induction like I(open). The proof methods here are purely algebraic. First considerable results in this field are those by Shepherdson.

Tennenbaum's theorem could be seen as a borderline between systems of open induction and of bounded arithmetic. So, the systems of open induction can have recursive nonstandard (countable) models whereas the stronger theories do not. A borderline between weak and strong fragments is more artificial and in our case will be laid at Elemantary Arithmetic, EA.

EA can be axiomatized in many ways. One can take Q (Robinson's arithmetic) plus function symbols for all the Kalmar elementary functions together with their defining axioms. The Kalmar elementary functions are those which are defined by bounded primitive recursion (just the regular recursion scheme where the function $f(x)$ should at any entry be mayorized by 2_n^x (see definition 5.4) for some fixed (standard) n). One could also define the Kalmar elementary functions as those obtained by adding to the language a bounded μ-operator and allow for definitions like $g(y) := \mu x \leq a.f(x,y) = 0$. (Here the $g(y)$ returns as value the smallest x not bigger than a such that $f(x,y) = 0$.) Alternatively one could say that the Kalmar elementary funcions are those which are computable by a Turing machine with a multiexponential time bound. We will often call upon EA by its equivalent formulation of $\mathrm{I}\Delta_0 + \mathsf{EXP}$. EXP is the one axiom expressing the totality of the exponential function and $\mathrm{I}\Delta_0$ can be taken to be Q plus the induction axioms for all bounded (Δ_0) formulas.

Work in the area of bounded arithmetic is very difficult. In especially the problem of separation of classes of sets is extremely hard. Whereas it is relatively easy to see that the classes of Σ_n, Δ_n and Π_n definable sets are all distinct, this analogous problem is far from being solved in the field of bounded arithmetic. Especially the question of $P \neq NP$ is still left unanswered.

Our basic theory will be EA. The language of EA is richer than that of PA and consists of $0, 1, +, \cdot, =, \leq$, and 2^x. Over PA, the latter two symbols are of course definable. $x \leq y \leftrightarrow \exists z\ z + x = y$ and 2^x is definable making use of coding techniques. (Actually the graph of 2^x can be defined by a Δ_0 formula, but this is not at all trivial.) Furthermore EA includes the defining axioms for these symbols:

1. $\leq$ is a discrete linear order, with $x+1$ the successor of x and 0 the least element and plus and times respect the order.

2. $2^0 = 1$,
   $2^{x+1} = 2^x + 2^x$.

Our notation will be as in the literture where it comes to the quantifier complexity of formulas (see for example Mendelsohn's book). One word of caution is in order here. Sometimes one finds in the literature that Σ_n formulas are closed under bounded quantification. We will not adhere to this convention because in the weaker systems one might not have enough collection principles to prove this closure properties. So, instead we will consider them as new axiom schemes, giving rise to the systems $\mathrm{B}\Sigma_n$ where the collection axioms B_ϕ: $\forall a \forall t (\forall x \leq t \quad \exists y \quad \phi(x,y,a) \rightarrow \exists z \forall x \leq t \exists y \leq z \quad \phi(x,y,a))$ are restricted to the case where ϕ is a Σ_n-formula. In our notation $\mathrm{I}\Sigma_n$ will stand for EA plus all the induction axioms for Σ_n formulas. Later we will see that EA is actually finitely axiomatizable. Note that we do allow extra parameters in all these axiom schemes. Over PA the extra parameters could be dispensed with but in weaker systems this is definitely not the case explaining why we distinguish between parameter free induction systems $\mathrm{I}\Sigma_n^-$, and normal induction systems $\mathrm{I}\Sigma_n$ with parameters. It will turn out that $\mathrm{I}\Sigma_n$ is finitely axiomatizable whereas $\mathrm{I}\Sigma_n^-$ is not. (The strength of induction rules is less dependent on the presence of parameters.)

1.2 Truth predicates

Church's thesis and the close relation between Σ_1-formulas and R.E. sets is used to make the arithmetical analogous of the universal Turing machine: a universal Σ_1-formula, a so-called Σ_1-truth predicate. We want to have a formula $\mathrm{T}r_{\Sigma_1}(x)$ which is true if and only if x is the code of a true Σ_1 sentence. We recall the following well known theorem relating Σ_1 formulas and R.E. sets.

Theorem 1.1 *A set of natural numbers X is R.E. iff there is $\phi(x) \in \Sigma_1$ such that* $[n \in X \leftrightarrow \mathbb{N} \models \phi(n)]$

So, we can consider R.E. sets as being indexed by the code of a Σ_1 formula in the sense that $X = W_{\ulcorner\sigma\urcorner} = \{n \mid \mathbb{N} \models \sigma(n)\}$. Now we know of the existence of a universal Turing machine and a coding of this protocol in arithmetic such that $\mathbb{N} \models W(e,x) \leftrightarrow$ '$x \in W_e$'. By Church's thesis and the above theorem we may assume that this $W(e,x)$ is actually a Σ_1 formula. A better inspection of the way $W(e,x)$ actually works reveals that it makes use of effective constructions only (and bounded (multi) exponential space and time), so that the validity of the equivalence is actually provable in EA. (This is merely a heuristics rather than a proof.) $W(e,x)$ is not yet our pursued truth-predicate. It is merely a satisfaction relation for Σ_1 formula's with one free variable. We want a truth predicate for sentences however. To get our predicate we employ the following elementary transformation

$$g: \ \ulcorner\sigma\urcorner \mapsto \ulcorner\sigma \wedge x = x\urcorner,$$

to define

$$\mathrm{T}r_{\Sigma_1}(y) := W(g(y), 0).$$

Indeed, now we have $\mathbb{N} \models \sigma \Leftrightarrow \mathbb{N} \models \sigma \wedge 0 = 0 \Leftrightarrow 0 \in W_{g(\ulcorner\sigma\urcorner)} \Leftrightarrow \mathbb{N} \models \mathrm{T}r_{\Sigma_1}(\ulcorner\sigma\urcorner)(= W(g(\ulcorner\sigma\urcorner), 0))$. (Working out all the details and not using Church's thesis is done in Kaye's book "Models of Peano Arithmetic" ([Kay91, Kaye 1991]) where actually a Δ_1 satisfaction predicate for Δ_0 formula's is designed. From this one also obtains a Σ_1 truth predicate for Σ_1 formula's.)

Note that we have $\mathrm{EA} \vdash \varphi(x) \leftrightarrow \mathsf{Tr}_{\Sigma_1}(\ulcorner\varphi(\dot{x})\urcorner)$ for all $\varphi \in \Sigma_1$. Once we have a Σ_1 truth predicate, truth predicates for higher complexities are readily concocted. For example if g is a function sending a code $x = \ulcorner\forall v \varphi(v)\urcorner$ to $g(x) = \ulcorner\varphi(v)\urcorner$, then $\mathrm{T}r_{\Pi_2}(x) := \forall w \mathrm{T}r_{\Sigma_1}(g(x)[\dot{w}/v])$.

Once we have the truth predicates it is not too hard to see that we can replace all of our induction axioms by one single instantiation. In other words we have the following.

Theorem 1.2 *Most of our strong fragments are finitely axiomatizable.*

PROOF OF FINITE AXIOMATIZABILITY. While running through the argument we see why it is necessary to allow for parameters in the induction formula. We can restrict ourselves though to just one parameter as we know that for every n there exists a Δ_0 and 1-1 coding. We see that every induction axiom of a Σ_1-sentence can be obtained from the single instantiation of the formula $\mathrm{T}r_{\Sigma_1}(\dot{e}(x))$ which says:

$$\forall e(\mathrm{T}r_{\Sigma_1}(\dot{e}(0)) \wedge \forall x[\mathrm{T}r_{\Sigma_1}(\dot{e}(\dot{x})) \to \mathrm{T}r_{\Sigma_1}(\dot{e}(\dot{x}+1))] \to \forall x \mathrm{T}r_{\Sigma_1}(\dot{e}(\dot{x}))).$$

$\dot{e}(0)$ stands here for the code of 0 subtituted in the formula coded by e, et cetera. (Notice that this is highly ambiguish. Does the dot also applies to free variables occuring in the formula after the substitution has had taken place? We will not worry so much as for our purposes it is clear what we mean.) A priori EA could be infinite (we shall later on see that it is not) but all the induction axioms for Σ_1 sentences are derivable from that very specific instance by using only a fixed finite part of EA. So, indeed, $\mathrm{I}\Sigma_1$ is finitely axiomatizable. For the other systems the proof of their finite axiomatizability is very similar. QED

2 Lecture 2 (20-6-2001)

2.1 Arithmetization of Metamathematics

In the previous lecture we already saw some arithmetization in process. A truth-predicate was made in the language of arithmetic such that $\mathbb{N} \models \sigma \Leftrightarrow \mathbb{N} \models \mathrm{T}r_{\Sigma_1}(\ulcorner\sigma\urcorner)$ for Σ_1 sentences σ. We now want to arithmetize the notion of "provable" rather than that of "true". So, we would like to have a predicate $\mathsf{Pr}_T(x)$ or alternatively denoted by $\Box_T(x)$ which is to hold (in the standard model) precisely whenever the formula with code x is provable in T.

In the developement of our provability predicate we will closely follow the influential article of [Fef60, Feferman]. We will thus use a provability predicate which is uniformly associated to an arithmetical theory. We will assume that the languages of our theories are simple in the sense that they are Δ_0 definable. Oftern we will assume that we have some fixed Δ_0 predicates stating that " x is a logical axiom" and "x is obtained from y and z by means of an application of Modus Ponens". By the second we mean that y is the code of some formula φ, z of some formula of the form $\varphi \to \psi$ and x of the formula ψ. Of course we need certain strength of our theories to perform these codings and to prove certain properties of it like unique readability and the like but as we have induction over Δ_0 formulas in EA we have no worries. In the logical axioms we also include the equality axioms.

Now we want to be able to talk of a certain theory T inside another theory U. So certainly this theory T should be definable by some formula α in the sense that $\mathbb{N} \models \alpha(n)$ if and only if n is the code of an axiom of T. Definability however is not enough if the logical complexity of α is more than Σ_1 because U also has to "see" this that $\alpha(n)$. Therefore the following notion is introduced.

Definition 2.1 *A formula $\alpha(x)$ is said to be a* numeration *of a theory T in a theory U if we have*

- *φ is an axiom of T iff $U \vdash \alpha(\ulcorner\varphi\urcorner)$*

A formula $\alpha(x)$ is said to be a binumeration *of a theory T in a theory U if α is a numeration of T in U and*

- *If φ is not an axiom of T then $U \vdash \neg\alpha(\ulcorner\varphi\urcorner)$*

Here the theory U will be EA in most of the cases. Notice that if T is *R.E.* then the set of axioms is Σ_1 definable in the standard model. But EA proves all and only all true Σ_1 sentences so in this case numerabilty coincides with definability. If T is binumerable in U, and U is a sound *R.E.* theory, this implies that the axiom set of T is decidable. This does not necessarily imply that we can binumerate T in U with a $\Delta_1(U)$ predicate. The binumeration will externally be equivalent to a $\Delta_1(\mathbb{N})$ formula but the theory U need not necessarily be able to see this.[1]

Once we have fixed a numeration α of a theory T in U we can form an intensional proof predicate stating that there is some sequence of formula's where each formula is either an axiom (logical or non-logical, so this is where α comes into play) or is obtained from some rule using only formula's that occured earlier on in the sequence. We will denote this by $\exists y \mathsf{Prf}_\alpha(y,x)$ or simply Pr_α. Two observations are noteworthy. Firstly we see that the occurence of α is positive in Pr_α. You can see this either by inspecting the way Pr_α is constructed and conclude that Prf_α is $\Delta_0(\alpha)$. Alternetatively one can remark that taking a weaker α can only yield more theorems (or just as many). This can not happen if α occurs somewhere (essentially) negatively in Prf_α. Secondly we observe that different numerations yield different non-equivalent proof predicates. (cf. Feferman's proof predicate for which we don't have Gödel's second incompleteness theorem versus the regular proof predicate.) This non equivalence of different proof predicates

[1] Actually being a $\Delta_1(\mathbb{N})$-formula is a non-arithmetical notion, but the theory need also not necessarily see the $\Delta_1(U)$-ness. An example is given in the proof of Orey-Hajek's criterium for interpretability over a reflexive theory. Here β is some (say $\Delta_1(U)$) binumeration of a theory T in U. If you assume that $U \vdash \mathsf{Con}(T \upharpoonright n)$ for all $n \in \mathbb{N}$ you can make a Feferman-like binumeration α such that $U \vdash \mathsf{Con}_\alpha$. This is a binumeration of higher complexity, say Δ_2, of T in U. Using this binumeration in performing the Henkin construction in a formalized setting one obtains $U \rhd T$. (This α is defined as $\alpha(x) := \beta(x) \wedge \mathsf{Con}(T \upharpoonright n)$. To see that $U \vdash \mathsf{Con}_\alpha$ we reason in U and suppose $\neg\mathsf{Con}_\alpha$. So, we also have $\neg\mathsf{Con}_\beta$, so there is a least n_0 such that $\neg\mathsf{Con}_{T\upharpoonright n_0}$. By the minimality of n_0 we have that $\mathsf{Con}_{\beta(x)\wedge x<n_0}$ which is by the choice of n_0 just Con_α. This contradicts our initial assumption.)

already is manifestated in Σ_1 or even Δ_0 numerations. We already knew that strange things can happen when it comes to numerations. For example Craig's trick states that every *R.E.* theory is equivalent to one with a Δ_0 numeration.

In the paper of Feferman [Fef60, Feferman] one finds that if α is Σ_1, then Pr_α satisfies the derivability conditions. Actually this was already more or less known before. Let T be a theory containing EA that is numerated by α with $\alpha \in \Sigma_1$. (When we use the $\Box$ notation we automaically read that we have to take the code of the formula as the argument.)

1. $T \vdash \varphi \Leftrightarrow \mathrm{EA} \vdash \Box_\alpha \varphi$
2. $\mathrm{EA} \vdash \Box_\alpha(\varphi \to \psi) \to (\Box_\alpha(\varphi) \to \Box_\alpha(\psi))$
3. $\mathrm{EA} \vdash \Box_\alpha \varphi \to \Box_\alpha \Box_\alpha \varphi$
4. $T \vdash \Box_\alpha \varphi \to \varphi \Leftrightarrow T \vdash \varphi$

The first condition from left to right just expresses the numerability of theorems. The other direction follows from the Σ_1 soundness of EA plus the fact that $\Box_\alpha$ is an intensional coding of the concept of provability. The second one reflects that the provability predicate is closed under logical consequence. In the third clause it is essential that α is a Σ_1 numeration so that $\Box_\alpha$ is a Σ_1 predicate. Then the condition is a corrolary of a formalization of Σ_1 completeness which states that whenever $\mathbb{N} \models \sigma(n)$ for certain $n \in \mathbb{N}$ and $\sigma \in \Sigma_1$ that then $\mathrm{EA} \vdash \sigma(1 + 1 + \ldots + 1)$. ($n$ times addition.) The size of this proof can be elementary constructed from σ and n so that we actually have a formalization of this:

$$\mathrm{EA} \vdash \forall x \quad (\sigma(x) \to \Box_\alpha(\sigma(\dot{x}))).$$

This fact does actually hold for a larger class of formula's namely the smallest class containing Σ_1 and being closed under bounded quantifications conjunctions and disjunctions. This class is really bigger if you don't have Σ_1 collection. Notice that $\mathrm{EA} + \mathrm{B}\Sigma_1 \equiv \mathrm{EA} + \mathrm{B}\Delta_0$. The fourth condition can be viewed as a generalization of Gödel's second incompleteness theorem. Taking contraposition we obtain $T + \neg\varphi$ is consistent $\Rightarrow T \nvdash \neg\varphi \to \neg\Box_\alpha\varphi$ So, by the deduction theorem, $T + \neg\varphi \nvdash \neg\Box_\alpha(\neg\varphi \to \bot)$. If we set $U := T + \neg\varphi$, we can conceive the latter as $U \nvdash \mathsf{Con}_U$. As a natural numeration of U one could take $\alpha_U(x) := \alpha(x) \vee x = \ulcorner\neg\varphi\urcorner$. Now $\vdash \mathsf{Con}_U \leftrightarrow \neg\Box_\alpha\neg\neg\varphi$.

2.2 Reflection priciples

Reflection principles are a strong tool in studying systems of arithmetic. They quite naturally generalize consistency statements. Either way one intends to generalize the notion of so to say, the principle of "everything that is provable, is true", one is more or less bound to end up with one of the two following notions.

- **Local Reflection.** This consists of the scheme $\Box_\alpha \varphi \to \varphi$ for *sentences* φ and is denoted by Rfn_α or by Rfn_T whenever α numerates T.

- **Uniform reflection.** This is the same scheme as local reflection except that we now allow for formulas. It is denoted by $\mathrm{RFN}(T)$ or RFN_α. So, a reflection principle is one of the form $\forall x(\Box_\alpha \varphi(\dot{x}) \to \varphi(x))$. Here we assume that $FV(\varphi) = \{x\}$. Possible other variables can be dispensed with by means of encoding.

Often one considers restrictions on the formulas or sentences that occur in the reflection principles giving rise to the so-called restricted reflection schemes (RFN_{Σ_n}, et cetera). For Π_1 reflection we have the following easy but interesting result over EA.

Lemma 2.2 (EA $\vdash$) $\mathrm{RFN}_{\Pi_1}(T) \Leftrightarrow \mathsf{Rfn}_{\Pi_1}(T) \Leftrightarrow \mathsf{Con}(T)$

Proof of Lemma. We always have the arrows from left to right, that is to say for any nonempty formula class. Thus it suffices to show $\mathsf{Con}(T) \Rightarrow \mathrm{RFN}_{\Pi_1}$. So, assume $\Box_\alpha \varphi(\dot{x}) \wedge \neg\varphi(x)$. By provable Σ_1 completeness we have, as $\neg\varphi$ is Σ_1, that $\Box_\alpha \neg\varphi(\dot{x})$ and hence $\Box_\alpha(\varphi(\dot{x}) \wedge \neg\varphi(\dot{x}))$, i.e. $\Box_\alpha \bot$, which contradicts the assumption of consistency. QED

Another easy fact is that $\mathrm{RFN}_{\Pi_{n+1}} \equiv \mathrm{RFN}_{\Sigma_n}$ which follows directly from the fact that

$$\Box_\alpha \forall u \varphi(u, \dot{x}) \to \forall u \Box_\alpha \varphi(\dot{u}, \dot{x}).$$

We also have the following.

Fact 2.3 *$T + \mathsf{Rfn}_{\Pi_n}$ is not contained in any finite consistent Σ_n extension of T, and dually we have that $T + \mathsf{Rfn}_{\Sigma_n}$ is not contained in any finite consistent Π_n extension of T.*

Proof of fact 2.3. Take any finite extension of T by Σ_n sentences. (The other case goes completely the same.) By taking conjunctions we may assume we have only one sentence φ. We see that $T + \varphi$ does not prove the reflection principle for $\neg\varphi$ $(\in \Pi_n)$. For suppose it did, i.e., $T + \varphi \vdash \Box_\alpha \neg\varphi \to \neg\varphi$. Thus, $T \vdash \varphi \to (\Box_\alpha\neg\varphi \to \neg\varphi)$ which amounts to the same as $T \vdash \Box_\alpha\neg\varphi \to \neg\varphi$. By the fourth provability condition (Löb's rule) we obtain $T \vdash \neg\varphi$. So, indeed $T+\varphi$ can not prove all local Π_n reflection axioms. QED

This fact can even be extended to one where the condition on finiteness is replaced by a condition of Σ_n-numerability to obtain the following.

Proposition 2.4 *Let T be an R.E. theory containing* EA. *No Σ_n numerable set of Π_n sentences U such that $T+U$ is consistent, contains $T+\mathsf{Rfn}_{\Sigma_n}$*
.

PROOF OF PROPOSITION 2.4. The trick is done by the fact that one can include every Σ_n numerable Π_n extension of T in some $T+\varphi$ for some Π_n-formula φ. For this you consider a fixed point φ with a built in Rosser trick

$$\varphi \leftrightarrow \forall\pi[\mathsf{Ax}_U(\pi) \wedge \forall y \leq \pi \neg\mathsf{Prf}_{T+\varphi}(y, \ulcorner\bot\urcorner) \to \mathsf{Tr}_{\Pi_n}(\pi)].$$

There are two important observations:

(1.) $T+\varphi$ is consistent,

(2.) $U \subseteq T+\varphi$.

We have that (1.) $\Rightarrow$ (2.). If $T+\varphi$ is consistent and π is an element of U then (π is standard) $\forall y \leq \pi \ \ \neg\mathsf{Prf}_{T+\varphi}(y, \ulcorner\bot\urcorner)$ is true and this is also provable, so we have $T+\varphi \vdash \mathsf{Tr}_{\Pi_n}(\pi)$ hence $T+\varphi \vdash \pi$. So, it remains to show that $T+\varphi$ is consistent. Suppose it were not. Let m be the smallest (standard!) proof of $\bot$ in $T+\varphi$. Then $\mathsf{Prf}_{T+\varphi}(m, \ulcorner\bot\urcorner)$ is true and also provable (and also provably the smallest). So, φ is provably equivalent to $\forall\pi[\mathsf{Ax}_U(\pi) \wedge \pi < m \to Tr_{\Pi_n}(\pi)]$, i.e., $\bigwedge_{\pi \in \mathsf{Ax}_U \restriction m} \mathsf{Tr}_{\Pi_n}(\pi)$. But certainly this is provable in $T+U$ but this contradicts the consistency of $T+U$. Now we can proceed as before to arrive a contradiction assuming that $T+\varphi$ proves all Σ_n instances of the reflection principle. QED

The dual of proposition 2.4 is somewhat harder and one needs another fix point.

We also have the following uniform version of the above fact.

Fact 2.5 *$T+\mathrm{RFN}_{\Pi_n}$ is not contained in any consistent Σ_n extension of T, and dually we have that $T+\mathrm{RFN}_{\Sigma_n}$ is not contained in any consistent Π_n extension of T.*

PROOF OF FACT. Let S be some collection of Σ_n sentences such that $T+S$ extends $T + \mathrm{RFN}_{\Pi_n}$. We also have $T + S \vdash \forall x(\Box_\alpha(\mathsf{Tr}_{\Pi_n}(\dot{x})) \to \mathsf{Tr}_{\Pi_n}(x))$. By compactness we have for some single Σ_n-sentence σ that $T+\sigma \vdash \forall x(\Box_\alpha(\mathsf{Tr}_{\Pi_n}(\dot{x})) \to \mathsf{Tr}_{\Pi_n}(x))$. We conclude that $T+S$ can not prove the reflection axiom for $\mathsf{Tr}_{\Pi_n}(\neg\sigma)$, for if it would then by an application of Löb's rule and the fact that $\mathsf{Tr}_{\Pi_n}(\neg\sigma) \leftrightarrow \neg\sigma$ we would be able to prove $\neg\sigma$ in $T+S$ which can not be so. QED

3 Lecture 3 (27-06-2001)

3.1 Provably total recursive functions

An important proof theoretic question about arithmetical theories is about the recursive functions that they prove to be total. This is also closely related to proof theoretic ordinal analysis of theories. We know that recursive functions are represented on the standard model by Σ_1 formulas so we can say what it means for a function $f(\vec{x})$ to be provably recursive and total in some theory T namely

1. For some Σ_1 formula $\varphi(\vec{x}, y)$ $: f(\vec{x}) = y \leftrightarrow \mathbb{N} \models \phi(\vec{x}, y)$,
2. $T \vdash \forall\vec{x}\exists y\varphi(\vec{x}, y)$.

This φ need not be unique. It might be the case that for some φ you have 1 and 2 and for some other φ' you only have 1. Or even if you have both 1 and 2 for two different formulas representing the same function, they need not be provably equivalent. If we have 2 for some formula φ, we can switch to some formula φ' so that we actually have

2'. $T \vdash \forall\vec{x}\exists!y\varphi'(\vec{x}, y)$.

This is done as follows. What we would like to do is to take as our unique y the minimal y such that $\varphi(\vec{x}, y)$. But we do not want to use $L\Sigma_1$. To avoid that we use coding. As φ is Σ_1 we have that $\varphi = \exists u\varphi_0(\vec{x}, y, u)$ for some φ_0 in Δ_0. So, we have that $T \vdash \forall\vec{x}\exists y\exists u\varphi_0(\vec{x}, y, u)$. By means of coding we now take the minimum pair y, u. This takes $L\Delta_0$ which we have as we assumed that $\mathrm{EA} \subset T$.

We denote the set of p.t.c.f.'s (provably total computable (=recursive) functions) of some theory T by $\mathcal{F}(T)$. If T is some RE theory $\mathcal{F}(T)$ is properly contained in the set of all total recursive functions. Intuitively this is clear as checking totality is a Π_2-complete task and the set $\mathcal{F}(T)$ is RE if T

is. A more formal demonstration of this fact is by means of a diagonalization argument. Constrain the attention to unary functions and let $f_i(x)$ be the recursive function for which we have found a proof in T of its totality after having found precisely i such proofs before in some cannonical enumeration of all possible proofs in T. Clearly the function $g(x) := f_x(x) + 42$ is total and recursive. However this is impossible to prove in T. (Behold! A proof of Gödel's first incompleteness theorem without making use of the fixed point theorem.) Sometimes one can hear erroneous remarks like ZFC proves all the recursive functions to be total. (Recall that per definition all recursive functions are total!)

The set $\mathcal{F}(T)$ is always closed under composition but need not be closed under primitive recursion. $\mathcal{F}(\mathrm{EA})$ is not closed under primitive recursion. Clearly one also has $\mathcal{F}(T) = \mathcal{F}(T \restriction \Pi_2)$. So, if one would like to prove the apartness of two theories which are Π_2 equivalent their respective classes of p.t.c.f.'s will not help. In bounded arithmetics the $\mathcal{F}$ operator yields classes of functions of different complexity, so, $\mathcal{F}(S^1_2)$ is the class of P-time computable functions et cetera.

3.2 The p.t.c.f's of $\mathrm{I}\Sigma_1$

In this part it will be shown that $\mathcal{F}(\mathrm{I}\Sigma_1)$ consists of precisely the primitive recursive functions. In order to do so we will actually switch to another arithmetic namely $\mathrm{EA} + \Sigma_1$-IR. This is the logic consisting of EA plus the induction rule for Σ_1 sentences, that is, from $\varphi(0)$ and $\forall x(\varphi(x) \to \varphi(x + 1))$ conclude $\forall x\ \varphi(x)$ whenever $\varphi(x)$ is a Σ_1 formula. Note that this is really different from the induction axiom. For example we do not have the deduction theorem for $\mathrm{EA} + \Sigma_n$-IR. $\mathrm{EA} + \Sigma_1$-IR has in a certain way, as we shall see, nicer properties than $\mathrm{I}\Sigma_1$. One thing is that from Π_2 premises one obtains a Π_2 consequence using this rule. When using a Σ_1 induction axiom you actually use a Π_3 formula.

The arithmetic $\mathrm{EA} + \Sigma_1$-IR is Π_2 equivalent to $\mathrm{I}\Sigma_1$ so it has the same class of p.t.c.f's. To see this equivalence we need to do some work. It is easy to see that we have $\mathrm{EA} + \Sigma_1\text{-IR} \subseteq \mathrm{I}\Sigma_1$. But we also have the following.

Theorem 3.1 $\mathrm{I}\Sigma_1$ *is* Π_2 *conservative over* $\mathrm{EA} + \Sigma_1$-IR.

PROOF OF Π_2 CONSERVATIVITY. In this proof we will use a Tait sequent caculus of first order logic which is exposed in Swichtenberg's contribution to the Handbook of Mathematical Logic. (See [Sch77, Schwichtenberg]). It works with sequents which are sets and should be read disjunctively in the sense that $\Gamma = \{\varphi_1, \ldots, \varphi_n\}$ stands for $\varphi_1 \vee \ldots \vee \varphi_n$. All formulas are written in a form that uses only $\wedge, \vee, \forall, \exists$ and literals, that is, atoms or negations of atoms. Negation of composed formulas is defined by the de Morgan laws. The axioms are:

$$\Gamma, \varphi, \neg\varphi \quad \text{for atomic } \varphi \ ,$$

$$\frac{\Gamma, \varphi \quad \Gamma, \psi}{\Gamma, \varphi \wedge \psi}, \qquad \frac{\Gamma, \varphi}{\Gamma, \varphi \vee \psi}, \qquad \frac{\Gamma, \psi}{\Gamma, \varphi \vee \psi},$$

$$\frac{\Gamma, \varphi(a)}{\Gamma, \forall x \varphi(x)}, \qquad \frac{\Gamma, \varphi(t)}{\Gamma, \exists x \varphi(x)},$$

plus the cut rule

$$\frac{\Gamma, \varphi \quad \Gamma, \neg\varphi}{\Gamma}.$$

In the axiom of the universal quantifier introduction it is necessary that the a does not occur free somewhere else in Γ. And in the axiom for the introduction of the existential quantifier one requires t to be substitutable for x in φ. This Tait calculus is sound and complete and serves us well in providing a proof of our theorem.

So, we need to prove that if $\mathrm{I}\Sigma_1 \vdash \pi$ then $\mathrm{EA} + \Sigma_1\text{-IR} \vdash \pi$ whenever π is a Π_2 sentence. We reason as follows. If $\mathrm{I}\Sigma_1 \vdash \pi$, we have by the compactness and deduction theorem that $\vdash \sigma \to \pi$ where σ is the conjunction of a finite number of axioms of $\mathrm{I}\Sigma_1$. Or equivalently $\vdash \neg\sigma \vee \pi$. As the Tait calculus is complete this amounts to the same as saying that the sequent $\neg\sigma, \pi$ is derivable within the calculus. By the cut elimination theorem for this Tait calculus we know that there exists a cut free derivation of the sequent. Cut free proofs have all sorts of pleasant properties like the subformula property (modulo substitution of terms). The proof is concluded by showing by induction on the length of cut free derivations that if a sequent of the form Σ, Π is derivable then $\mathrm{EA} + \Sigma_1\text{-IR} \vdash \bigvee \Pi$. Where Σ is a finite set of negations of induction axioms of Σ_1 formulas and Π is a finite set of Π_2 formulas.

So, suppose we have a cut free proof of Σ, Π. What can be above this sequent? Either the last rule yielded something in the Π part of the sequent or in the Σ part of it. In the first case nothing interesting happens and we almost automatically obtain the desired result by the induction hypothesis. So, suppose something had happened in the Σ part. We can assume that the Σ part only contains formulas of the form $\exists a[\varphi(0) \wedge \forall x(\varphi(x) \to \varphi(x+1)) \wedge \neg\varphi(a)]$, whith $\varphi \in \Sigma_1$. The last deduction step thus must have been the introduction of the existential quantifier and we can by a one step shorter proof derive the following sequent

$$\Sigma', \varphi(0) \wedge \forall x(\varphi(x) \rightarrow \varphi(x+1)) \wedge \neg\varphi(t), \Pi$$

By the inversion property of the Tait calulus (for a proof and precise formulation of the statement consult [Sch77, Schwichtenber] page 873) we obtain proofs of the same length of the following sequents

$$\Sigma', \varphi(0), \Pi \ , \quad \Sigma', \forall x(\varphi(x) \rightarrow \varphi(x+1)), \Pi \quad \text{and} \ \ \Sigma', \neg\varphi(t), \Pi.$$

As all of $\varphi(0), \forall x(\varphi(x) \rightarrow \varphi(x+1))$ and $\neg\varphi(t)$ are Π_2 formulas, we can apply the induction hypothesis to conclude that we have

$$\begin{array}{lcl} \mathrm{EA} + \Sigma_1\text{-IR} & \vdash & \varphi(0) \vee \bigvee \Pi, \\ \mathrm{EA} + \Sigma_1\text{-IR} & \vdash & \forall x(\varphi(x) \rightarrow \varphi(x+1)) \vee \bigvee \Pi, \\ \mathrm{EA} + \Sigma_1\text{-IR} & \vdash & \neg\varphi(t) \vee \bigvee \Pi. \end{array}$$

Recall that Π consists of Π_2 statements. So, Π is of the form $\forall u \exists v \Pi_0(u,v)$. In our context we can omit the outer quantifier. If we now define $\varphi'(x,u) := \varphi(x) \vee \bigvee \exists v \Pi_0(u,v)$, we obtain a Σ_1 formula to which we can apply the induction rule to obtain $\forall x \varphi'(x)$ and thus also $\varphi'(t)$. Combining this with $\mathrm{EA} + \Sigma_1\text{-IR} \vdash \neg\varphi(t) \vee \bigvee \exists v \Pi_0(u,v)$ yields $\mathrm{EA} + \Sigma_1\text{-IR} \vdash \bigvee \exists v \Pi_0(u,v)$ by one application of the cut rule and thus the desired $\mathrm{EA} + \Sigma_1\text{-IR} \vdash \bigvee \Pi$. QED

The above proof is actually proving a somewhat stronger statement.

Corollary 3.2 *Let T be some collection of Π_3 sentences in the language of arithmetic and let Π be some Π_2 sentence.*
We have that $T + \mathrm{I}\Sigma_1 \vdash \Pi \Rightarrow T + \mathrm{EA} + \Sigma_1\text{-IR} \vdash \Pi$.

Corollary 3.3 $\mathrm{I}\Sigma_1$ *is Σ_3-conservative over* $\mathrm{I}\Sigma_1^-$.

PROOF OF COROLLARY 3.3. So, suppose $\mathrm{I}\Sigma_1 \vdash \sigma$ for some Σ_3 sentence σ. Then we also have $\mathrm{I}\Sigma_1 + \neg\sigma \vdash \perp$. $\perp$ is a Π_2 sentence, so we may apply the above corollary to conclude $\neg\sigma + \mathrm{EA} + \Sigma_1\text{-IR} \vdash \perp$ As shall be proved later we have that $\Sigma_1\text{-IR} \equiv \Sigma_1\text{-IR}^-$ whence $\neg\sigma + \mathrm{EA} + \Sigma_1\text{-IR}^- \vdash \perp$. We don't have a deduction theorem but we do have $\neg\sigma + \mathrm{I}\Sigma_1^- \vdash \perp$. Now we may apply the deduction theorem to obtain $\mathrm{I}\Sigma_1^- \vdash \sigma$. QED

We will now determine the provably total recursive functions of EA. The strategy in doing so is as follows. First we enrich our language in such a way that we can axiomatize an equivalent variant of EA by open formulas. Proving a recursive function to be total amounts to proving some Π_2 statement. But as our theory is open we can apply Herbrand's theorem to obtain terms expressing the possible function values. Our language will turn out to be rich enough as to reduce this to a single term. Thus the provably total recursive functions will be precisely all the terms of the enriched language that was employed to "openize" EA. A central ingredient thus is Herbrand's theorem.

Theorem 3.4 Herbrand's Theorem for Predicate Logic
Let T be an open set of axioms, that is, no quantifiers occuring in it. If $T \vdash \exists x\varphi(x,a)$ with $\varphi(x,a)$ some open formula, then there are terms $t_1(a),\ldots,t_k(a)$ of the language of T such that $T \vdash \bigvee_{i=1}^{k} \varphi(t_i(x),a)$.

Theorem 3.5 *The provably total recursive functions of* EA *are precisely the Kalmar elementary functions.*

PROOF OF THEOREM 3.5. We will start defining a new theory EA$'$. The language of EA$'$ is the language of EA plus two new binary funtion symbols for the characteristic functions of the order and equality.
So, $\mathcal{L}(\mathrm{EA}') = \{0,1,+,\cdot,=,\leq,2^{(\cdot)},\chi_=(\cdot,\cdot),\chi_\leq(\cdot,\cdot),\mu(\cdot,\cdot,\cdot)\}$. Let Λ be a countable infinite set of individual variables. The terms of EA$'$ are defined inductively.

$$T := \quad \Lambda \mid 0 \mid 1 \mid 2^T \mid T+T \mid T\cdot T \mid \chi_=(T,T) \mid \chi_\leq(T,T) \mid \mu(\Lambda,\Lambda,T) \ .$$

To the logical symbols we add one new symbol called a bounded μ-operator. So, formulas in EA$'$ are formed just as usual exept that we now also can have things like $(\mu x \leq y.t(x)=0)=z$. The defining axioms will tell us that $\mu x \leq y.t(x)=0$ gives us the least x such that $t'(x)=0$ if such an x exists, if not it will return us the value of $y+1$. The axioms of EA$'$ comprise the following; All the open versions of the axioms of Q leaving out the axiom $\neg(x=0)\rightarrow \exists y\ y+1=x$ which is actually a theorem having induction over bounded formulas; The open axioms defining 2^x and the open axioms stating that $\leq$ is a discrete linear order respected by $+$ and $\cdot$; axioms defining the characteristic functions like $\chi_=(x,y)=0 \leftrightarrow x=y$; open axioms explaining the behavior of the bounded μ operator to the effect of adding for all terms t of EA$'$ an axiom

$$(\mu x \leq y.t(x)=0)=z \rightarrow$$
$$z\leq y \wedge t(z)=0 \wedge (u<z \rightarrow t(u)\neq 0)) \vee (z=y+1 \wedge (v\leq y\ t(v)\neq 0)) \ .$$

To each bounded formula φ in the language of EA we will assign a characteristic function χ_φ inductively.

$$\begin{array}{rcl} t_1 = t_2 & \longmapsto & \chi_=(t_1, t_2)\ , \\ t_1 \leq t_2 & \longmapsto & \chi_\leq(t_1, t_2)\ , \\ \varphi \wedge \psi & \longmapsto & \chi_\varphi \dot{+} \chi_\psi\ , \\ \neg\varphi & \longmapsto & 1 - \chi_\varphi\ , \\ \forall x \leq y \varphi(x) & \longmapsto & \chi_=(\mu x \leq y . \chi_{\neg\varphi}(x) = 0, y+1)\ . \end{array}$$

Having done all this it is easy to see that EA$'$ is at least as strong as EA for EA$'$ proves the least number principle for bounded formulas. For suppose $\varphi(a)$ for a bounded formula φ. Thus $\chi_\varphi(a) = 0$. But then we get the existence of a least element satisfying $\varphi(x)$ almost for free as $\mu x \leq a . \chi_\varphi(x) = 0$.

If EA proves the totality of some recursive function we have that for some bounded $\varphi(x,y)$ we have EA $\vdash \exists y \varphi(x,y)$. So, EA$' \vdash \exists y \chi_\varphi(x,y) = 0$. Herbrand's theorem now tells us that there are $t_1(x), \ldots, t_k(x)$ such that EA$' \vdash \chi_\varphi(x, t_1(x)) = 0 \vee \ldots \vee \chi_\varphi(x, t_k(x)) = 0$. It is not hard to convince oneself that the language is actually strong enough to give one single term $t(x)$ incorperating this disjunction such that EA$' \vdash \chi_\varphi(x, t(x)) = 0$. Thus the function is a term in our extended language. The collection of all these possible terms is called the set of Kalmar Elementary functions. One can prove that this formulation is equivalent to any of the previous descriptions of the Kalmar Elementary functions.

QED

Theorem 3.6 $\mathcal{F}$(EA+Σ_1-IR) $=$ *p.r., where p.r. is the class of primitive recursive functions.*

Proof of theorem 3.6. It is not so hard to see that any primitive recursive function is indeed provably total in EA + Σ_1-IR. (See for example [HP93, Hájek and Pudlák].) The other inclusion is proved by induction on the number of nested applications of the Σ_1-IR. If this number is zero we have that the function was already provably total in EA and thus by theorem 3.5 was Kalmar Elementery and thus primitive recursive. So, suppose that we have some occurence of the Σ_1-IR. By the induction hypothesis we get primitive recursive functions g and h such that $\mathbb{N} \models \varphi(g(a), 0, a)$ and $\mathbb{N} \models \varphi(y, x, a) \rightarrow \varphi(h(x, y, a), x+1, a)$. Thus we define

$$\begin{array}{rcl} f(0,a) & = & g(a), \\ f(x+1, a) & = & h(x, f(x,a), a). \end{array}$$

By induction on x in the standard model we now see that $\mathbb{N} \models \varphi(x, f(x, a), a)$.

QED

In this proof we see a very close relation between an application of the induction rule and an application of primitive recursion. Now that we have classified the provably total recursive functions of EA + Σ_1-IR we have also classified the provably total recursive functions $\mathrm{I}\Sigma_1$ because these two systems are Π_2 equivalent. In a certain way we also have that $\mathsf{PRA} \equiv \mathrm{EA} + \Sigma_1$-IR. We actually mean here that there exist interpretations in both directions. (Recall that PRA has for every primitive recursive function a symbol plus the defining axioms. These symbols will be interpreted in the canonical way.) So, we also know the provably total recursive functions of PRA. One could consider $\mathcal{F}$ as an arrow going from fragments to classes of functions. It has some interesting behavior though. For example $\mathcal{F}$ does not remain constant under taking unions e.g. $\mathcal{F}(\mathrm{I}\Pi_2^-) = \mathcal{F}(\mathrm{I}\Sigma_1) =$ *p.r.* but $\mathrm{I}\Pi_2^- + \mathrm{I}\Sigma_1$ proves the totality of the Ackerman function which is known to be not primitive recursive.

4 Lecture 4

4.1 Reflection principles relating fragments

Reflection principles provide us with a powerful tool in comparing fragments of arithmetic. It turns out that many fragments are characterizable by the amount of reflection they poses. The bulk of table 1 will be proved in this lecture.

Arithmetic	Axiomatized by EA plus the indicated reflection
PA	RFN(EA)
$\mathrm{I}\Sigma_1, \mathrm{I}\Pi_1$	$\mathrm{RFN}_{\Pi_3}(\mathrm{EA})$
$\mathrm{I}\Sigma_1(=\mathrm{I}\Pi_n)$	$\mathrm{RFN}_{\Pi_{n+2}}$
[2]$\mathrm{I}\Pi_n^-$	$\varphi \rightarrow \mathrm{RFN}_{\Pi_n}(\varphi)$ with $\varphi \in \Pi_{n+1}$
$\mathrm{I}\Sigma_n^-$	$\varphi \rightarrow \mathrm{RFN}_{\Pi_{n+1}}(\varphi)$ with $\varphi \in \Pi_{n+1}$
$\mathrm{EA} + \Sigma_n$-IR	$\mathrm{RFN}^{\omega}_{\Pi_{n+1}}$

Table 1: Characterizing fragments using reflection

The parameter-free induction scheme systems are characterized in so-called relativized reflection. $\mathrm{RFN}(\varphi)$ thus stands for $\Box(\varphi \rightarrow \psi) \rightarrow \psi$.

[2]For $n = 1$ one needs to add an axiom stating the totality of the superexponentiation function.

$\mathrm{RFN}^{\omega}_{\Pi_{n+1}}$ stands for omega times iteration of the Π_{n+1} reflection principle. This is defined by recursion over the ordinals.[3]

$$\begin{array}{rcl} \mathrm{RFN}^0(T) & = & T \\ \mathrm{RFN}^{\alpha+1}(T) & = & \mathrm{RFN}(\mathrm{RFN}^{\alpha}(T)) \\ \mathrm{RFN}^{\gamma}(T) & = & \bigcup_{\alpha<\gamma} \mathrm{RFN}^{\alpha}(T) \end{array}$$

We will not prove the axiomatization of the systems based on rules but all the other systems will be treated here. A good treatment of the axiomatization of the induction rules is found in [Bek97] Beklemishev. A good indication for what reflection is needed to axiomatize a certain system is the logical complexity of the axioms. Let us consider for example the system $\mathrm{I}\Sigma_n$. The induction axioms have the form $\varphi(0) \wedge \forall x(\varphi(x) \rightarrow \varphi(x+1)) \rightarrow \forall x \varphi(x)$ where φ is a Σ_n formula. So, the axiom has the complexity (slightly abusing notation) $\Sigma_1 \wedge \Pi_{n+1} \rightarrow \Pi_{n+1}$, that is, a Boolean combination of Π_{n+1} sentences. In the system $\mathrm{I}\Sigma_n$ we allow free variables so the complexity is $\forall\ [\Pi_{n+1} \rightarrow \Pi_{n+1}]$ that is Π_{n+2}. In the table indeed we see that $\mathrm{I}\Sigma_n$ corresponds to Π_{n+2} reflection. (Recall that our base system EA is Π_2 axiomatizable.)

The table will be used to obtain an overview of the arithmetics and their interaction as depicted in picture 1.

It is known that $\mathrm{I}\Sigma_n = \mathrm{I}\Pi_n$. Later it will also be proved that $\mathrm{EA} + \Sigma_n\text{-IR} = \mathrm{EA} + \Sigma_n\text{-IR}^- = \Pi_{n+1}\text{-IR} = \Pi_{n+1}\text{-IR}^-$. One remark is in order here on how to compare rules. We consider a rule R to be a set of instances, i.e. a set of expressions of the form

$$\frac{A_1, \ldots, A_n}{B}.$$

The effect of a rule is of course dependent on the base theory. A rule R_1 is said to be *derivable* from another rule R_2 in some theory T if any derivation in T containing occurences of R_1 can be replaced by a derivation in T possibly containing occurences of R_2 rather than R_1. We call two rules, R_1 and R_2, *equivalent* with respect to some theory T if R_1 is derivable from R_2 in T and R_2 derivable from R_1 in T. We write $T + R_1 \equiv T + R_2$. A refinement of this notion of equivalence arises when attention is being paid to the number of nested occurences. A rule R_1 is said to *reduce* to a rule R_2 over some theory T if any instance of the rule R_1 in a derivation in T

[3] A natural question seems to be what is going on with this progression of theories if it is propagated through all the ordinals. It turns out that.... stay tuned!

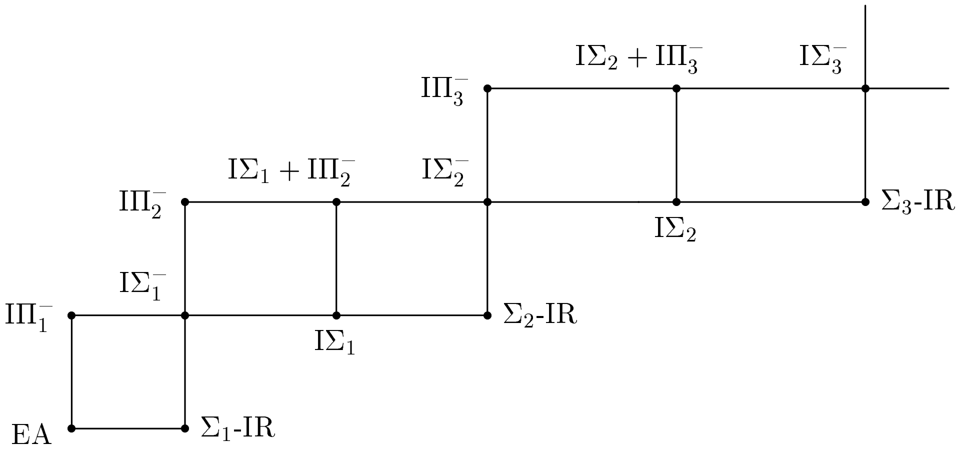

Figure 1: The fragments related.

can be replaced by a subderivation in T without nested applications of R_2. We write $R_1 \le R_2$ or $R_1 \le_T R_2$ if necessary. If both R_1 is reducible to R_2 over T and R_2 reducible to R_1, we call them *congruent* with respect to T and write $R_1 \cong R_2$ or $R_1 \cong_T R_2$ if necessary. The notion of congruence is more informative than that of equivalence. For example it is known that Π_n-IR $\cong_{\mathrm{EA}} \frac{\varphi\ (\in \Pi_{n+1})}{\mathrm{RFN}_{\Pi_n}(\varphi)}$. Also it is known that Σ_n-IR $\le_{\mathrm{EA}} \frac{\varphi\ (\in \Pi_{n+1})}{\mathrm{RFN}_{\Pi_{n+1}}(\varphi)}$, but it is unknown if $\frac{\varphi\ (\in \Pi_{n+1})}{\mathrm{RFN}_{\Pi_{n+1}}(\varphi)} \le_{\mathrm{EA}} \Sigma_n$-IR. It is only known that n nestings of the rule $\frac{\varphi\ (\in \Pi_{n+1})}{\mathrm{RFN}_{\Pi_{n+1}}(\varphi)}$ can be simulated by $n+1$ nested applications of Σ_n-IR in EA. Hence the two rules are equivalent with respect to EA.

4.2 Re-axiomatizing PA and some subsystems

Most of the entries of table 1 will be proved in this section.

Theorem 4.1 $\mathsf{PA} \equiv \mathrm{EA} + \mathrm{RFN}$

PROOF OF THEOREM 4.1. The easiest direction is to show that $\mathsf{PA} \subset \mathrm{RFN}$.

It is a formalisation of how children in Russian primary (nursery) school are persuaded to accept the principle of induction. Informally we see the correctness of induction as follows. Let n be an arbitrary number and suppose we know that

1. $\varphi(m) \rightarrow \varphi(m+1)$

for any m. If we moreover know that $\varphi(0)$ holds, we can conclude $\varphi(n)$ after n applications of different instantiations of 1. So, we deduce $\varphi(0), \varphi(1), \varphi(2), \ldots$ till we reach the conclusion $\varphi(n)$. Let P be the formula $\varphi(0) \wedge \forall x\ [\varphi(x) \rightarrow \varphi(x+1)]$. It is easy to see that

$$\mathrm{EA} \vdash \forall x \Box_{\mathrm{EA}}(P \rightarrow \varphi(\dot{x}))$$

as the proof of $P \rightarrow \varphi(x)$ grows linearly in the size of x. Applying reflection now yields the required $\forall x(P \rightarrow \varphi(x))$. By looking a bit closer we see that actually the same proof gives us

Corollary 4.2 $\mathrm{I}\Pi_n \subseteq \mathrm{RFN}_{\Pi_{n+2}}$.

Corollary 4.3 $\mathrm{I}\Sigma_n^- \subseteq \ \mathrm{EA} + \varphi \rightarrow \mathrm{RFN}_{\Pi_{n+1}}(\varphi)\ \ (\varphi \in \Pi_{n+1})$ *and* $\mathrm{I}\Pi_n^- \subseteq \ \mathrm{EA} + \varphi \rightarrow \mathrm{RFN}_{\Pi_n}(\varphi)\ \ (\varphi \in \Pi_{n+1})$

Now we prove $\mathrm{EA} + \mathrm{RFN} \subseteq \mathsf{PA}$. We thus have to prove that $\mathsf{PA} \vdash \forall \mathsf{x}[\Box\varphi(\dot{\mathsf{x}}) \rightarrow \varphi(\mathsf{x})]$. The sketch of the proof goes as follows. Consider a proof object for $\varphi(x)$ in EA. It is easy to obtain (we have superexponentiation!) from this a cut free proof object in the Tait calculus where the final sequent is $\neg\mathrm{EA}, \varphi(x)$. This proofobject has the subformula. As EA is finitely axiomatized, there is a standard k (here we see that this proof could not be used to proof reflection of PA) which majorizes the quantor rank of all subformulas. By induction on the length of the proof one can show that for all subsequents Γ we have that $\mathsf{PA} \vdash \mathrm{T}\mathsf{r}_{\Pi_k}(\bigvee \Gamma)$. We also thus have $\mathrm{T}r_{\Pi_k}(\neg\mathrm{EA} \vee \varphi(\dot{x}))$. As $\mathsf{PA} \vdash \mathrm{EA}$ and $\mathsf{PA} \vdash \mathrm{T}\mathsf{r}_{\Pi_k}(\mathrm{EA})$ we conclude that $\mathrm{T}r_{\Pi_k}(\varphi(\dot{x}))$ and thus $\varphi(x)$. QED

This proof also serves to obtain that $\mathrm{RFN}_{n+2} \subseteq \mathrm{I}\Pi_{n+1}$ but that is not sufficient for us. It is possible though to better analize the structure of the cut free proof in the Tait calculus to obtain the following.

Corollary 4.4 $\mathrm{RFN}_{n+2} \subseteq \mathrm{I}\Pi_n$ *and thus* $\mathrm{RFN}_{n+2} \equiv \mathrm{I}\Pi_n$

The axiomatization of the parameter free fragments by means of reflection principles goes through the results on the systems based on the induction rules. As said before these results are more difficult and will not be proved here. The following is not so hard.

Proposition 4.5 $\mathrm{EA} + \Pi_n\text{-IR} \equiv \mathrm{EA} + \Pi_n\text{-IR}^-$ *and* $\mathrm{EA} + \Sigma_n\text{-IR} \equiv \mathrm{EA} + \Sigma_n\text{-IR}^-$

PROOF OF PROPOSITION 4.5. The parameter free rule is just a special case of the normal rule so, we are to prove two inclusions. The Π_n-IR case is the simplest and is obtained by "pushing a universal quatifier inside". So, suppose we have that $\mathrm{EA} + \Pi_n\text{-IR}^- \vdash \varphi(0, y)$ and $\mathrm{EA} + \Pi_n\text{-IR}^- \vdash \forall x[\varphi(x, y) \to \varphi(x + 1, y)]$. We also thus have $\mathrm{EA} + \Pi_n\text{-IR}^- \vdash \forall y\varphi(0, y)$ and $\mathrm{EA} + \Pi_n\text{-IR}^- \vdash \forall x, y[\varphi(x, y) \to \varphi(x + 1, y)]$. Hence we also have $\mathrm{EA} + \Pi_n\text{-IR}^- \vdash \forall x, y\varphi(x, y) \to \forall x, y\varphi(x + 1, y)]$. We thus have all the premisses for the parameterfree induction rule, so we get that $\mathrm{EA} \vdash \forall x, y\varphi(x, y)$, i.e., $\mathrm{EA} \vdash \forall x\varphi(x, y)$.

The proof of $\mathrm{EA} + \Sigma_n\text{-IR} \subseteq \mathrm{EA} + \Sigma_n\text{-IR}^-$ is a bit more subtle. Again suppose that $\mathrm{EA} + \Sigma_n\text{-IR}^- \vdash \varphi(0, y)$ and $\mathrm{EA} + \Sigma_n\text{-IR}^- \vdash \forall x[\varphi(x, y) \to \varphi(x + 1, y)]$. $\varphi(x, y)$ is a Σ_n formula and can thus be written as $\varphi(x, y) = \exists z\psi(x, z, y)$, with $\psi \in \Pi_{n-1}$. We now consider the Σ_n *sentence* $\exists u\forall v \leq w\psi(\{v\}_0, (u)_v, \{v\}_1)$. (We need Π_{n-1} collection here!) Here $\{v\}_i$ is the primitive recursive function giving the i-th coordinate when considering v as the code of a pair of numbers. $(u)_v$ is just the v-th coordinate of the sequence of numbers coded by u. Recall that the pairing function enumerates the pairs of numbers by "walking over consecutive diagonals" as depicted in

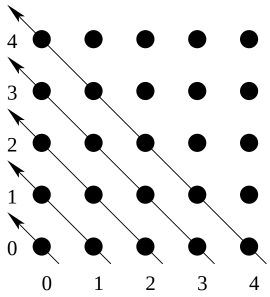

Figure 2: The pairing function

figure 4.2. The properties that we will need from the pairing function are understood well by looking at the picture but they are all provable in EA. As $\{0\}_0 = 0$ and $\mathrm{EA} + \Sigma_n\text{-IR}^- \vdash \varphi(0, y)$ we have that $\mathrm{EA} + \Sigma_n\text{-IR}^- \vdash \exists u \forall v \leq 0 \psi(\{v\}_0, (u)_v, \{v\}_1)$. Now reason in $\mathrm{EA} + \Sigma_n\text{-IR}^-$ and assume that $\exists u \forall v \leq w \psi(\{v\}_0, (u)_v, \{v\}_1)$. Let b be such that $\forall v \leq w \psi(\{v\}_0, (b)_v, \{v\}_1)$. We would like to have $\exists u \forall v \leq w + 1 \psi(\{v\}_0, (u)_v, \{v\}_1)$. For some $a \leq w$ we have that $\{a\}_0 + 1 = w + 1$. Now we can use $\varphi(\{a\}_0, y) \to \varphi(\{a\}_0 + 1, y)$ to get some c such that $\psi(\{w + 1\}_0, c, \{w + 1\}_1)$. Let $b * (c)$ denote the code of c concatenated to the string coded by b. We thus have $\forall v \leq w + 1 \psi(\{v\}_0, (b * (c))_v, \{v\}_1)$. Now we can apply the induction rule to obtain $\forall w \exists u \forall v \leq w \psi(\{v\}_0, (u)_v, \{v\}_1)$. From the latter we obtain $\forall x, y \exists z \psi(x, z, y)$ and thus also $\forall x \varphi(x, y)$. QED

We now want to see that $\mathrm{I}\Pi_n^- \vdash \varphi \to \mathrm{RFN}_{\Pi_n}(\varphi)$ or equivalently that $\mathrm{I}\Pi_n^- + \varphi \vdash \mathrm{RFN}_{\Pi_n}(\varphi)$. By theorem 4.5 and some elementary logic we know that $\mathrm{EA} + \varphi \Pi_n\text{-IR} = \mathrm{EA} + \varphi \Pi_n\text{-IR}^- \subseteq \mathrm{I}\Pi_n^- + \varphi$. From the work of [Bek97] we know that $\mathrm{EA} + \varphi \Pi_n\text{-IR}^- \vdash \mathrm{RFN}_{\Pi_n}(\varphi)$. [4] Combining this with corrolories 4.2 and 4.3 results to the following.

Theorem 4.6 $\mathrm{I}\Sigma_n^- \equiv \quad \mathrm{EA} + \varphi \to \mathrm{RFN}_{\Pi_{n+1}}(\varphi) \ (\varphi \in \Pi_{n+1})$ *and*
$\mathrm{I}\Pi_n^- \equiv \quad \mathrm{EA} + \varphi \to \mathrm{RFN}_{\Pi_n}(\varphi) \ (\varphi \in \Pi_{n+1})$

4.3 Comparing subsystems

We can now apply the results from table 1 to compare different arithmetics. For example we know that $\mathrm{EA} + \Sigma_{n+1}\text{-IR} equiv \mathrm{RFN}^{\omega}_{\Pi n+2}$. Also we know that $\mathrm{RFN}_{\Pi_{n+2}} \vdash \mathit{Cons}(\mathrm{I}\Sigma_n)$. But $\mathrm{I}\Sigma_n \nvdash \mathit{Cons}(\mathrm{I}\Sigma_n)$ so we have the following.

Fact 4.7 $\mathrm{I}\Sigma_n \nvdash \Sigma_{n+1}\text{-IR}$

We know that $\mathrm{I}\Sigma_n^-$ is axiomatizable by Σ_{n+2} sentences. We have just seen that $\mathrm{I}\Sigma_n \equiv \mathrm{RFN}_{n+2}$. combining this with 2.5 we get

Fact 4.8 $\mathrm{I}\Sigma_n^- \nvdash \mathrm{I}\Sigma_n$.

It is clear that reformulating two theories in terms of reflection gives a good means to compare. Another application yields the following.

[4] We use a result here that says that if T is a finite Π_{n+1} axiomatized theorem then $T + \Pi_n\text{-IR} \equiv \mathrm{RFN}^{\omega}_{\Pi_n}$. Combining this with $\mathrm{EA} + \Pi_{n+1}\text{-IR} \equiv \mathrm{EA} + \Sigma_n\text{-IR}$ also gives us an axiomatisation for $\mathrm{EA} + \Sigma_n\text{-IR}$.

Theorem 4.9 $\mathrm{I\Pi}_2^-$ *is* Π_2 *conservative over* $\mathrm{EA} + \Sigma_1$-IR

Proof of theorem 4.9. It can be shown that $\mathsf{GL} \vdash \Diamond^{n+1}\top \to \Diamond(\bigwedge_{i=1}^{n}(P_i \to \Diamond P_i))$. This will be proved semantically here. (For a good treatment of GL and its semantics see for example [Boo93].) So, suppose that at some world x in a GL model $\Diamond^{n+1}\top$ holds. We can thus find a path of length $n+1$ going up from this world. We easily see that along this path at most at one world $P_i \to \Diamond P_i$ can fail to be true. For that would mean that $P_i \wedge \Box\neg P_i$ would hold at such a world, excluding the possibility that $P_i \wedge \Box\neg P_i$ is true at some world below or above it. As we have only n sentences $P_i \to \Diamond P_i$ we have by the pigeon hole principle that for some world along the path we have that $(\bigwedge_{i=1}^{n} P_i \to \Diamond P_i)$ holds. We thus know that at x we have $\Diamond(\bigwedge_{i=1}^{n} P_i \to \Diamond P_i)$ and we are done.

We will need to consider generalized provability predicate now. Let therefore $[n]$ denote the provability predicate that uses as axioms apart from the axioms of some theory T also all true Π_n sentences. So, if α is a numeration of T the new predicate $[n]$ would just be $\Box_{\alpha(x)\vee\mathsf{Tr}_{\Pi_1}(x)}$. As a generalization of provable Σ_1 completeness we have in the new formalism provable Σ_2 completeness, that is, $\mathrm{EA} \vdash \sigma \to [1]\sigma$ whenever $\sigma \in \Sigma_2$. We also have an analogon of 2.2 stating $\mathrm{RFN}_{\Pi_2}(T) \leftrightarrow \langle 1\rangle T$. Where as always $\langle 1\rangle\psi$ is to be read as $\neg[1]\neg\psi$.

The theory $\mathrm{I\Pi}_2^-$ is axiomatized by making use of Π_2 reflection. In the new language we can thus express this as follows

$$\mathrm{I\Pi}_2^- \equiv \mathrm{EA} + \{\varphi \to \langle 1\rangle\varphi \mid \varphi \in \Pi_3\}.$$

It is not hard to see that GL is sound if we interpred the modal $\Box$ operator as $[1]$. By the above we thus know that

$$\mathrm{EA} \vdash \langle 1\rangle^{n+1}\top \to \langle 1\rangle(\bigwedge_{i=1}^{n}(\varphi_i \to \langle 1\rangle\varphi_i))$$

for any choice of the φ_i.

Now suppose $\mathrm{I\Pi}_2^- \vdash \pi$ for some $\pi \in \Pi_2$. By our result on the axiomatisation of $\mathrm{I\Pi}_2^-$, compactness and the deduction theorem we get that for some Π_3 formulas φ_i we have that

$$\mathrm{EA} \vdash (\bigwedge_{i=1}^{n}(\varphi_i \to \langle 1\rangle\varphi_i)) \to \pi$$

and thus also

$$\mathrm{EA} \vdash \langle 1\rangle(\bigwedge_{i=1}^{n}(\varphi_i \to \langle 1\rangle\varphi_i)) \to \langle 1\rangle\pi.$$

By the above we may conclude that $\mathrm{EA} + \langle 1\rangle^{n+1}\top \vdash \langle 1\rangle\pi$. By provable Σ_2 completeness we may thus also conclude $\mathrm{EA} + \langle 1\rangle^{n+1}\top \vdash \pi$. As we know that $\mathrm{EA} + \Sigma_1\text{-IR} \vdash \mathrm{RFN}^{n+1}_{\Pi_2} \vdash \langle 1\rangle^{n+1}\top$ we have $\mathrm{EA} + \Sigma_1\text{-IR} \vdash \pi$ as we wanted. QED

Theorem 4.10 *$I\Pi_2^-$ is not finitely axiomatizable.*

Proof of theorem 4.10. So, suppose it were. Then for finitely many Π_3 formulas φ_i we have that $\mathrm{I}\Pi_2^- \equiv \bigwedge_{i=1}^{n}(\varphi_i \to \langle 1\rangle\varphi_i)$. Thus, $\mathrm{EA} + \langle 1\rangle^{n+1}\top \vdash \langle 1\rangle \mathrm{I}\Pi_2^-$. But $\mathrm{I}\Pi_2^- \vdash \langle 1\rangle^{n+1}\top$, so this would imply that $\mathrm{I}\Pi_2^-$ would prove its own consistnecy, which it does not. QED

5 Lecture 5, 15-10-2001

5.1 Bounded Induction

Weak systems are weak in the sense that they do not prove very much. When it comes to interpreting (this will be defined later) weak theories are already quite strong in the sense that strong theories can be interpreted in them. $\mathrm{I}\Delta_0$ has a special place among the weaker fragments. This special status was obtained from model theoretic considerations. In model theory it is very natural to look at initial segments of some model of an arithmetical theory. If now $I \subset_e M$ and $M \models \mathsf{PA}$ ($I \subset_e M$ is read as "M is an *end extension* of I" and implies that if $y \in I$ and $M \models x < y$ then $x \in I$) we can not draw the conclusion that $I \models \mathsf{PA}$. We do know however that $I \models \mathrm{I}\Delta_0$ whenever $M \models \mathrm{I}\Delta_0$. This is due to the fact that $\mathrm{I}\Delta_0$ is Π_1 axiomatizable. Alternative induction axioms are

$$\forall x[\varphi(0) \wedge \forall y < x(\varphi(y) \to \varphi(y+1)) \to \varphi(x)]$$

which are all Π_1 formulas if φ is Δ_0. The language of $\mathrm{I}\Delta_0$ is basically that of EA leaving out the special symbol for exponentiation. The terms in this language are precisely the polynomials. A somewhat weaker theorem than 3.5 for $\mathrm{I}\Delta_0$ goes by the name of Parikh's theorem.

Theorem 5.1 *(Parikh) Let $\varphi \in \Delta_0$ and $I\Delta_0 \vdash \forall x \exists y \varphi(x,y)$. Under these conditions there exists a term $t(x)$ in the language of $I\Delta_0$ such that $I\Delta_0 \vdash \forall x \exists y \leq t(x) \varphi(x,y)$.*

Proof of theorem 5.1. The proof is a model theoretic one and goes by compactness. So, suppose that for any term $t(x)$ there is some model M_t of $\mathrm{I}\Delta_0$ such that $M_t \models \exists x \forall y \leq t(x) \neg\varphi(x, y)$. Now consider the theory $T := \mathrm{I}\Delta_0 \cup \{\forall y \leq t(c) \neg\varphi(x, c)\}$. Every finite subset of T is satisfiable. For let $t_1(x), \ldots, t_n(x)$ be terms of $\mathrm{I}\Delta_0$. $t_0(x) := t_1(x) + \ldots + t_n(x)$ is also a term and thus by our assumption there is a model $M_{t_0,c}$ of $\mathrm{I}\Delta_0$ such that $M_{t_0,c} \models \forall y \leq t(c) \neg\varphi(c, y)$. As T is finitely satisfiable it is satisfiable. Let M_c therefore be some model of T. Let I be the initial segment in M_c generated by the interpretation of this non-standard element c (which we will call c too). So, basically I consists of all the elements of M_c that are smaller than $t(c)$ for some term t. Actually it has to be verified that I is indeed an initial segment. Let therefore $a < t_1(c)$ and $b < t_2(c)$. We would like to also have $a + b < t_1 + t_2$. But fortunately we can prove monotonicity of the order, so, that for example $\mathrm{I}\Delta_0 \vdash x \leq a \wedge y \leq b \rightarrow x + y < a + b$. So, indeed I is an initial segment of M_c and thus a model of $\mathrm{I}\Delta_0$ too. But $\forall x \exists y \varphi(x, y)$ is not valid in I because every element in I is below some $t(c)$ and $M_, c \models \forall y \leq t(c) \neg\varphi(c, y)$. QED

As in the proof of 3.5 on could proceed the same here by first openizing $\mathrm{I}\Delta_0$ by enriching the language. The result thus obtained does not yield such a nicely defined class of functions ....

Parikh's theorem says quite a lot on the strength of $\mathrm{I}\Delta_0$. Or better, it says a lot about it's weakness. We know that every polynomial is eventually majorized by the exponential function. By Parikh's theorem we get

Corollary 5.2 *There does not exist a Δ_0 formula $\chi(x, y)$ that defines the exponential function such that $\mathrm{I}\Delta_0 \vdash \forall x \exists y \chi(x, y)$.*

On the other hand we know that there does exist a Δ_0 formula that defines the exponential function. It might be the case that some other formula of some higher complexity also defines the exponential function and that for this very formula $\mathrm{I}\Delta_0$ does prove the totality. However this seems rather unlikely. Another consequence of 5.1 is for example that $\mathrm{I}\Delta_0$ can never prove a Π_2 formulation of Ramsey's theorem. That is because the bounds in Ramsey's theorem are exponentially. (Again it might be formulated in a more complex but provable way but this seems rather unlikely.) Another implication concerns the prime numbers. There many proofs of the infinitude of primes (at least 5). All of these proofs make use of some exponential bounds though. For example Euclid's proof employs the factorial function.

These bounds are of course not (in it's easiest formulation) available in $I\Delta_0$. But a-priori this does not exclude that the infinitude of primes is provable in $I\Delta_0$. It is an open question whether the infinitude of primes is provable in $I\Delta_0$ or not. In some sense the question boils down to the essentialness of the occurence of exponential bounds in a proof of the infinitude of primes. This is one of the three big open questions on $I\Delta_0$. The others are if the MRDP theorem is provable in $I\Delta_0$ and if $I\Delta_0$ is finitely axiomatizable. A result by Paris, Wilkie and Woods tells us that the infinitude of primes is provable in $I\Delta_0 + \Omega_1$.

5.2 Definable cuts and their shortenings

In model theory initial segments play an important role. The counterpart of this concept in proof theory is the notion of a *definable cut.*

Definition 5.3 *A formula $I(x)$ is a T-cut if*

1. $T \vdash I(0) \wedge \forall x(I(x) \rightarrow I(x+1))$,
2. $\forall x, y(I(x) \wedge y \leq x \rightarrow I(y))$.

Notice that we do not demand here that I is closed under plus and times. As it will turn out, for our purposes this is not necessary as we can always find a smaller cut that is closed under plus and times. (And many more functions!) If we have full induction at our disposal clause 1 tells us that the only cut is the trivial one, that is, $I(x) \leftrightarrow (x = x)$. [5] Cuts are closely related to the notion of relative interpretability as we shall see later on. In some sense they will compensate for a lack of induction. If some argument can't be executed in the theory due to a lack of sufficient induction we can switch to a cut where the argument can be proceded.

Definition 5.4

1. $2_0^x = x$,
2. $2_{n+1}^x = 2^{2_n^x}$.

By $|x|$ we mean the length of the binary representation of x, that is, $|x| = \lceil {}^2\log(x+1) \rceil$.

[5] A question from Jaap at this moment was if there exixts a non-standard model of $I\Delta_0$ such that the natural numbers are definable. The general feel is that this is very unlikely.

Definition 5.5

1. $\omega_0(x) = 2x$,
2. $\omega_{n+1}(x) = 2^{\omega_n(|x|-1)}$.

We see that $\omega_1(x) \sim x^2$, $\omega_2(x) \sim x^{|x|}$, $\omega_3(x) \sim x^{|x|^{||x||}}$, et cetera. The limit of these functions is in a certain way 2^x. In IΔ_0 we can verify a great deal of the properties of the ω_n's. We have to bare in mind here that we don't actually have 2^x at present in IΔ_0 but of course you have to read all the properties with a bit more care using the Δ_0 relation that we mentioned before. We have that

Fact 5.6

1. $\omega_n(x) \leq 2^x$,
2. $\omega_n(2^x_{n+1}) = 2^{x+1}_{n+1}$,
3. $\omega_n(x) \leq \omega_{n+1}(x)$
4. $\omega_n(x) \leq \omega_n(x+1)$

The proofs are easy and go by induction. We now come to a famous theorem concerning the shortening of cuts.

Theorem 5.7 *Let I be a T-cut and k a natural number, then there is a T-cut J_k such that*

- $T \vdash \forall x[J_k(x) \rightarrow I(2^x_k)]$,
- $T \vdash \forall x[J_k(x) \rightarrow J_k(\omega_k(x))]$.

The proof can be found, modulo the usual missprints, in [HP93, Hájek and Pudlák] page 173. The axiom that states the totality of the function ω_{n+1} is abbreviated by Ω_n.

Corollary 5.8 $I\Delta_0 \rhd I\Delta_0 + \Omega_n$

This will be treated later in more detail when also a formal definition of interpretability is given but the idea is simple. If we just restrict our domain to the definable cut of sufficient size we get the required interpretation.

References

[Bek97] L.D. Beklemishev. Induction rules, reflection principles, and provably recursive functions. *Annals of Pure and Applied Logic*, 85:193–242, 1997.

[Boo93] G. Boolos. *The Logic of Provability.* Cambridge University Press, Cambridge, 1993.

[Fef60] S. Feferman. Arithmetization of metamathematics in a general setting. *Fundamenta Mathematicae*, 49:35–92, 1960.

[HP93] P. Hájek and P. Pudlák. *Metamathematics of First Order Arithmetic.* Springer-Verlag, Berlin, Heidelberg, New York, 1993.

[Kay91] R. Kaye. *Models of Peano Arithmetic.* Oxford University Press, Oxford, 1991.

[Sch77] H. Schwichtenberg. Some applications of cut-elimination. In J. Barwise, editor, *Handbook of Mathematical Logic*, pages 867–896. North Holland, Amsterdam, 1977.