

CHAPTER 6

Okay it's undecidable, but it can't be incomplete too

In this chapter, we will prove a real big theorem, the kind of theorem that goes down in history books (especially books on the history of mathematics). We will need to use techniques from both first-order logic (things like definability, compactness, Löwenheim-Skolem) and from recursion theory (really, recursive functions are at the heart of things).

The main question we will want to answer is essentially Hilbert's problem that put us down this long and lonesome road. Namely:

- Is there a nice axiomatisation of arithmetic that is complete?
- Is there a decision procedure that, given an axiomatisation of arithmetic, allows us to decide if a sentence is a consequence of the axioms or not?

To answer these questions, we'll get back to our diagonal roots. To misquote something from a while back:

“As it applies to answering the second question, this argument reminds us of the famous paradox of Epimenides, the Cretan, who claimed that all Cretans are liars.”

Essentially, what we will do is construct a formula ϕ which literally says “The formula ϕ is not provable”. Assuming our set of axioms is sound, we must have that this formula is true (otherwise, if this formula is false, then ϕ is provable, and therefore the axioms prove something false). Of course, if ϕ is true, then ϕ cannot be provable, because that's what ϕ asserts, in the first place.

This result is extremely cool (and it's featured in many pop-sci expositions of mathematical logic), but we'll try to be rather careful about how we formulate everything.¹ The crucial fact here is that we wish for a set of axioms that is “nice” and what does

¹**Warning!** Don't believe everything you see on the internet. (If you're reading these notes on the internet, please refer to the liar paradox.) The *incompleteness* theorem will not contradict the *completeness* theorem! Recall that a theory T is called **complete** if for every formula ϕ we have that $T \vdash \phi$ or $T \vdash \neg\phi$. A proof system \vdash (like our Hilbert proof system from a while back) is called **complete** if whenever $\models \phi$ (i.e. ϕ is universally valid), we have that $\vdash \phi$. The incompleteness theorem refers to the FIRST notion of completeness not the second...

nice mean for a set of axioms? Weeeeell... it should mean that we can essentially write them all down (or at least have a computer write them all down), i.e. nice means recursively enumerable. Okay, enough blabbering, let's get to the maths.

1. Peano Lessons

1.1. The axioms. First and foremost, we need a language. I've mentioned the language of Peano arithmetic, \mathcal{L}_{Peano} , before, but it feels like it was a lifetime ago, so here we go:

- One constant symbol $\underline{0}$.
- One unary function symbol \underline{S} .
- Two binary function symbols $\underline{+}$ and $\underline{\times}$.

As I did before, I will write $x \underline{+} y$ and $x \underline{\times} y$ rather than the insane $\underline{+}(x, y)$ and $\underline{\times}(x, y)$.

We're all familiar with \mathbb{N} from the second chapter of our childhood, namely here \mathbb{N} will be the domain of an \mathcal{L}_{Peano} -structure \mathcal{N}_{st} where $\underline{0}$ is interpreted as 0, $\underline{+}$ and $\underline{\times}$ as the usual + and \times and \underline{S} is interpreted as the successor function:

$$S = \lambda x. x + 1.$$

And if you're worried that we did all this recursion theory only to learn that $\underline{0}$ is 0, fret not. Things will get more complicated.

The set of **Peano's axioms** consists of just seven (7) axioms and 1 (one) axiom scheme.

First, the axioms:

- $$\begin{aligned} A_1 \quad & (\forall x) \neg(\underline{S}x \doteq \underline{0}) \\ A_2 \quad & (\forall x)(\exists y)(\neg(x \doteq 0) \rightarrow (\underline{S}y \doteq x)) \\ A_3 \quad & (\forall x)(\forall y)((\underline{S}x \doteq \underline{S}y) \rightarrow (x \doteq y)) \\ A_4 \quad & (\forall x)(x \underline{+} \underline{0} \doteq x) \\ A_5 \quad & (\forall x)(\forall y)(\underline{S}(x \underline{+} y) \doteq x \underline{+} \underline{S}y) \\ A_6 \quad & (\forall x)(x \underline{\times} \underline{0} \doteq \underline{0}) \\ A_7 \quad & (\forall x)(\forall y)((x \underline{\times} y) \underline{+} x \doteq x \underline{\times} \underline{S}y) \end{aligned}$$

And now the axiom scheme:

$$\text{IS}_\phi \quad (\forall \bar{y}) (\phi(\underline{0}, \bar{y}) \wedge ((\forall x)(\phi(x, \bar{y}) \rightarrow \phi(\underline{S}x, \bar{y})) \rightarrow (\forall x)\phi(x, \bar{y}))),$$

for each \mathcal{L}_{Peano} -formula $\phi(x, y_1, \dots, y_n)$.

Let T_{PA} be the \mathcal{L}_{Peano} -theory consisting of $A_1 - A_7$ and IS_ϕ , for all \mathcal{L} -formulas $\phi(x, y_1, \dots, y_n)$.

The point of IS_ϕ is that it allows us to argue (in models of T_{PA}) using our usual intuition of induction. Indeed, using IS_ϕ , to show that a property $\phi(x, t_1, \dots, t_n)$ is provably true from T_{PA} (for some \mathcal{L}_{Peano} -terms t_1, \dots, t_n) for all x , we can do induction (who'd have thunk?), more precisely, we just need to prove:

- The **base case**: $T_{PA} \vdash \phi(\underline{0}, \bar{t})$.
- The **inductive step**: $(\forall x)(\phi(x, \bar{t}) \rightarrow \phi(\underline{Sx}, \bar{t}))$.

Remark 1.1.1. What does $T_{PA} \vdash \phi$ (for some \mathcal{L}_{Peano} -sentence ϕ) mean? It means that there is some derivation in our good-old Hilbert proof system which uses the axioms of T_{PA} , the quantifier axioms, the equality axioms, and propositional tautologies that ends with ϕ . BUT we have proved that \vdash is sound and complete, i.e. that $T_{PA} \vdash \phi$ if and only if $T_{PA} \vDash \phi$. Thus, to show that $T_{PA} \vdash \phi$ we may well show that $T_{PA} \vDash \phi$, i.e. that **for every** model $\mathcal{N} \vDash T_{PA}$ we have that $\mathcal{N} \vDash \phi$. As we will see (and actually should have already seen, if we were paying attention), T_{PA} has many models. The only

Induction, as you may be able to tell already, is a very strong axiom scheme. We will also consider a weakening (i.e. a sub-theory) of T_{PA} , which I will denote by T_{PA_0} (referred to sometimes as **weak Peano arithmetic**) and which consists only of axioms $A_1 - A_7$. There is a crucial fact about T_{PA_0} , which is not hard to prove:

FACT. T_{PA_0} is a finite theory.

Exercise 1.1.2. Show that \mathcal{N}_{st} , as described before is a model of T_{PA} and of T_{PA_0}

Of course, \mathcal{L}_{Peano} only “sees” $\underline{0}$, but really, it sees more elements, it “sees” $\underline{S} \underline{0}$ and also $\underline{S} \underline{S} \underline{0}$ and also $\underline{S} \underline{S} \underline{S} \underline{0}$ and also... okay you see where I’m going for this. For each integer $n \in \mathbb{N}$ (where \mathbb{N} is the usual² mathematical object not a model of T_{PA} !) we will let \underline{n} denote the \mathcal{L}_{Peano} -term

$$\underbrace{\underline{S} \underline{S} \cdots \underline{S}}_{n \text{ times}} \underline{0}.$$

²I tried pretty hard not to use the word standard here. Footnote ² exists only because I failed. Göd(el), why can’t I help myself with self-reference?

So the terms I wrote before were $\underline{1}$, $\underline{2}$, $\underline{3}$. Anyway, given a model $\mathcal{N} \models T_{PA}$, we will say that an element of N is **standard** if it is the interpretation of a term of the form \underline{n} , for some $n \in \mathbb{N}$ (again, as a mathematical object, I'll stop saying it though). The little subscript in \mathcal{N}_{st} was meant to indicate that this is the **standard model** of T_{PA} (and T_{PA_0}), the one where every element is standard. We call models of T_{PA} (or T_{PA_0}) in which there are elements that are not standard **non-standard**.

THEOREM 1.1.3. *Non-standard models of T_{PA} (and T_{PA_0}) exist.*

PROOF. This follows immediately by the upwards Löwenheim-Skolem theorem. Anyway, it's been a minute since we discussed this, and we didn't cover the proof in class, so let's prove it. Let Γ be the following set of sentences:

$$\Gamma := \{\neg(\underline{c} \doteq \underline{n}) : n \in \mathbb{N}\} \cup T_{PA},$$

in $\mathcal{L}_{Peano} \cup \{\underline{c}\}$. This is finitely satisfiable in \mathcal{N}_{st} and hence it has a model. The interpretation of \underline{c} in that model cannot be standard. \square

It is a thesis (i.e. a non-formalisable conjecture) that all theorems of arithmetic follow from T_{PA} . In the next exercise, you will be asked to do some formal proving, because I'm worried you may have missed it after all the recursion theory we did:

Exercise 1.1.4 (Long and hard!). Let $\mathcal{N} \models T_{PA}$. Prove that in \mathcal{N} :

- (1) Addition and multiplication are associative and commutative.
- (2) The cancellation laws for addition and multiplication hold, namely:

$$\mathcal{N} \models (\forall x)(\forall y)(\forall z)((x \pm y \doteq x \pm z) \rightarrow (y \doteq z))$$

and

$$\mathcal{N} \models (\forall x)(\forall y)(\forall z)((x \neq 0 \wedge (x \times y \doteq x \times z) \rightarrow (y \doteq z)).$$

It turns out that in weak Peano arithmetic, we cannot prove many of the properties above (in some HW you *may* be asked to build a model of T_{PA_0} in which addition is not even commutative). Nonetheless, we still have the following:

THEOREM 1.1.5. *Let $\mathcal{N} \models T_{PA}$. The formula $\phi(x, y)$ given by $(\exists z)(z \pm x \doteq y)$ defines total order on \mathcal{N} , and this order is compatible with addition and multiplication.*

PROOF (SKETCH). That this formula defines a partial order compatible with $+$ and \times is true in the usual \mathbb{N} , and the way one proves this in \mathbb{N} is essentially using the Peano axioms.³ \square

Hmph... Fine, let's write a more proper proof.

PROOF. Let's gather some facts that T_{PA} can prove:

- (1) $T_{PA} \vdash (\forall x)(\underline{0} \pm x \doteq x)$ [Use A_4 , A_5 and IS with the formula $\underline{0} \pm x \doteq x$]
- (2) $T_{PA} \vdash (\forall x)(\forall y)(\underline{S}(y \pm x) \doteq \underline{S}y \pm x)$ [Use A_4 a couple of times, A_5 and IS]
- (3) $T_{PA} \vdash (\forall x)(\underline{1} \pm x \doteq \underline{S}x)$ [Use (1) and (2)]
- (4) $T_{PA} \vdash (\forall x)(\forall y)(x \pm y \doteq y \pm x)$ [Use (1), A_5 , (2) and IS]
- (5) $T_{PA} \vdash (\forall x)(\forall y)(\forall z)(x \pm (y \pm z) \doteq (x \pm y) \pm z)$ [Use IS (on each of the variables). The base case is by A_4 and the inductive step, essentially, by A_5]
- (6) $T_{PA} \vdash (\forall x)(\underline{0} \times x \doteq \underline{0})$ [Use A_6 , A_7 and IS]
- (7) $T_{PA} \vdash (\forall x)(x \times \underline{1} \doteq x)$ [A_7 , A_6 and (1)]
- (8) $T_{PA} \vdash (\forall x)(\underline{1} \times x \doteq x)$ [Use IS and up to four of the axioms of T_{PA_0}]
- (9) $T_{PA} \vdash (\forall x)(\forall y)(\forall z)(x \times (y \pm z) \doteq (x \times y) \pm (x \times z))$ [We use IS again. This one is more of an exercise than the others]
- (10) $T_{PA} \vdash (\forall x)(\forall y)(\forall z)(x \times (y \times z) \doteq (x \times y) \times z))$ [You guessed it, this is by IS and (9)]
- (11) $T_{PA} \vdash (\forall x)(\forall y)(x \times y \doteq y \times x)$ [Okay, I won't solve the whole "Long and hard" exercise here.]
- (12) $T_{PA} \vdash (\forall x)(\forall y)(\forall z)(x \pm z \doteq y \pm z \rightarrow x \doteq y)$
- (13) $T_{PA} \vdash (\forall x)(\forall y)(\neg(x \doteq 0) \rightarrow \neg(x \pm y \doteq 0))$. Okay, let's prove this one:

- From A_2 and A_5 we have that:

$$T_{PA} \vdash \neg(x \doteq \underline{0}) \rightarrow (\exists z)((x \doteq \underline{S}z) \wedge (x \pm y \doteq \underline{S}(x \pm z)))$$

- From A_1 we then have that:

$$T_{PA} \vdash \neg(\underline{S}(x \pm z \doteq \underline{0}))$$

In particular, this is a consequence of T_{PA_0} .

³You are not allowed to use this terrible reasoning when solving Exercise 1.1.4.

- (14) $T_{PA} \vdash (\forall x)(\forall y)(x \pm y \dot{=} 0 \rightarrow (x \dot{=} 0 \wedge y \dot{=} 0))$ [This is immediate from (13) and A_4] – thus this is also a consequence of T_{PA_0} .
- (15) $T_{PA} \vdash (\forall x)(\forall y)(x \pm y \dot{=} x \rightarrow y \dot{=} 0)$ [This is a consequence of (12), A_4 and (4).]

So far, I've actually pretty much given you a blueprint of the solution to the “Long and hard” exercise. Let's use a bunch of the facts we've proved above, together with more facts (to be proved below), to show that:

$$T_{PA} \vdash “(\exists z)(z \pm x \dot{=} y) \text{ is a total order compatible with } \pm \text{ and } \leq”$$

In the remainder of the proof, I will use $T_{PA} \vdash x \leq y$ to mean that $T_{PA} \vdash (\exists z)(z \pm x \dot{=} y)$. The usual abbreviations $<, \geq, >$ are all defined as usual.

Ah, also, by (4), I'm allowed to use the following:

$$T_{PA} \vdash (\forall x)(\forall y)(x \leq y \leftrightarrow (\exists z)(x \pm z \dot{=} y)).$$

Okay? Alright.

- (16) $T_{PA} \vdash (\forall x)(x \leq x)$ [This is just because $T_{PA} \vdash 0 \pm x \dot{=} x$]
- (17) $T_{PA} \vdash (\forall x)(\forall y)(\forall z)((x \leq y \wedge y \leq z) \rightarrow x \leq z)$ [By (5)]
- (18) $T_{PA} \vdash (\forall x)(\forall y)((x \leq y \wedge y \leq x) \rightarrow x \dot{=} y)$ [Use (5),(15),(4), and (14), I think]
- (19) $T_{PA} \vdash (\forall x)(\forall y)(\forall z)((x \pm z \leq y \pm z) \leftrightarrow x \leq y)$ [Probably (5) and (12)]
- (20) $T_{PA} \vdash (\forall x)(\forall y)((x \leq y) \vee (y \leq x))$ [Hard exercise. You should use IS, A_1 , (1), A_5 , A_4 , (17), A_4 , A_2 , A_5 , and (2).]
- (21) $T_{PA} \vdash (\forall x)(\forall y)(\forall z)((x \leq y) \rightarrow (x \leq z \leq y \leq z))$ [(9) and (11) – bad choice of numbering?]
- (22) $T_{PA} \vdash (\forall x)(\forall y)(\neg(y \dot{=} 0) \rightarrow (x \leq y \geq x))$ [A_2 and A_7 should do the trick]
- (23) $T_{PA} \vdash (\forall x)(\forall y)(\neg((x \dot{=} 0) \vee (y \dot{=} 0)) \rightarrow \neg(x \leq y \dot{=} 0))$ [If you've trusted me thus far, you should trust that this follows from A_7 , A_5 , A_2 and A_1 , if not you should try it yourselves.]
- (24) $T_{PA} \vdash (\forall x)(\forall y)(\forall z)((x \leq z \dot{=} y \leq z) \rightarrow ((x \dot{=} y) \vee (z \dot{=} 0)))$ – here's a different kind of argument. Let $\mathcal{N} \models T_{PA}$ and $a, b, c \in N$ arbitrary. Suppose that $ac = bc$ (i.e. $\mathcal{N} \models (x \leq z \dot{=} y \leq z)[a/x, b/y, c/z]$). Then, by (20) we have that $a \leq b$ or $b \leq a$. In the first case, there is some d such that $d + a = b$, hence by (11) and (9) we have $bc = dc + ac$. By (4) and (15) we have that

$dc = 0$ and then we're done by (23). [Try out the argument in the other case.]

□

It turns out that in the proof I didn't give and then gave above, to prove that T_{PA} proves that $\phi(x, y)$ is a total order, we need that addition is commutative, which is not necessarily true in models of T_{PA_0} . Crucially, to prove that:

$$T_{PA} \vdash "(\exists z)(z \pm x \doteq y) \text{ is a total order}"$$

we *need* the induction scheme, in particular:

Remark 1.1.6. The theorem above is not true for models of T_{PA_0} .

1.2. Models of varying standards. Having established that the formula $\phi(x, y)$ given by $(\exists z)(z \pm x \doteq y)$ is a (\emptyset -definable) total order on models of T_{PA} , for $\mathcal{N} \models T_{PA}$, I am now justified when I write $x \leq y$ to mean that $\mathcal{N} \models \phi(x, y)$, as I did in the last proof. The usual abbreviations $<$, \geq , $>$ are all still defined as usual.

Definition 1.2.1. Let $\mathcal{N}, \mathcal{N}' \models T_{PA_0}$, and assume that \mathcal{N} is a substructure of \mathcal{N}' . We say that \mathcal{N} is an *initial segment* of \mathcal{N}' if for all $a \in N$ and all $b \in N'$ we have that:

- (1) If $\mathcal{N}' \models b \leq a$ then $b \in N$.
- (2) If $b \notin N$ then $\mathcal{N}' \models a \leq b$.

In this case, we also say that \mathcal{N}' is an *end extension* of \mathcal{N} .

Remark 1.2.2. As I discussed earlier, \leq is not necessarily an order in models of T_{PA_0} . Nonetheless, in the definition above, when we write $x \leq y$ we simply mean $\phi(x, y)$ were $\phi(x, y)$ was the formula $(\exists z)(z \pm x \doteq y)$.

The remark above is not all that important, since:

Proposition 1.2.3. *Let $\mathcal{N} \models T_{PA_0}$. The set:*

$$M = \{x \in N : x \text{ is standard}\}$$

is the universe of a substructure \mathcal{M} of \mathcal{N} that is an initial segment of \mathcal{N} and is isomorphic to \mathcal{N}_{st} .

Before getting into the details of the proof, let me remark that in this proposition we are using the usual integers, so we are allowed to use INDUCTION ON \mathbb{N} – this is not the same thing as IS.

PROOF. Recall that $a \in N$ is standard if there is some $n \in \mathbb{N}$ such that $a = \underline{n}^{\mathcal{N}}$. Consider the (obvious map):

$$\begin{aligned} f : \mathbb{N} &\rightarrow M \\ n &\mapsto \underline{n}^{\mathcal{N}} \end{aligned}$$

We should note straight away that f is surjective, since for every $n \in \mathbb{N}$, the term \underline{n} has an interpretation in \mathcal{N} , and thus belongs to M .

We need to show that this map is injective, and in fact, a homomorphism of \mathcal{L}_{PA} -structures. We will do this by slowly writing down a bunch of things that T_{PA_0} proves.

- $T_{PA_0} \vdash \underline{n} \pm \underline{1} \doteq \underline{S}n$ and $T_{PA_0} \vdash \underline{n} + \underline{1} \doteq \underline{S}n$. This follows once we observe that all expressions here represent the same term, namely the term consisting of $n + 1$ occurrences of \underline{S} followed by a single occurrence of the symbol $\underline{0}$.
- For all $m, n \in \mathbb{N}$ we have that:

$$T_{PA_0} \vdash \underline{m} + \underline{n} \doteq \underline{m} + \underline{n}$$

We prove this by (actual factual) induction on n (keeping m fixed). For $n = 0$ we are done by A_4 , since we have that:

$$T_{PA_0} \vdash \underline{m} + \underline{0} \doteq \underline{m}.$$

For $n + 1$, we have to show that:

$$T_{PA_0} \vdash \underline{m} + \underline{n} + \underline{1} \doteq \underline{m} + (\underline{n} + \underline{1})$$

assuming, by induction, that:

$$T_{PA_0} \vdash \underline{m} + \underline{n} \doteq \underline{m} + \underline{n}.$$

Since addition is associative in the actual factual natural numbers, we can drop the brackets in the RHS.

On the one hand, by the first bullet, we have that:

$$T_{PA_0} \vdash \underline{n} + \underline{1} \doteq \underline{S}n \text{ and } T_{PA_0} \vdash \underline{m} + \underline{n} + \underline{1} \doteq \underline{S}(\underline{m} + \underline{n})$$

But also, by (A_5) we have that:

$$T_{PA_0} \vdash \underline{m} + \underline{S}n \doteq \underline{S}(\underline{m} + \underline{n}).$$

Putting everything together, this bullet follows.

- For all $m, n \in \mathbb{N}$ we have:

$$T_{PA_0} \vdash \underline{m} \times \underline{n} \doteq mn.$$

This follows again by induction.

- For all $n \in \mathbb{N}_{>0}$ we have:

$$T_{PA_0} \vdash \neg(\underline{n} \doteq \underline{0}).$$

This follows essentially from A_1 , since if $n \in \mathbb{N}_{>0}$, then we can write $n = m + 1$, for some m , and since $T_{PA_0} \vdash \underline{n} \doteq \underline{Sm}$, we're good.

- For all distinct $n, m \in \mathbb{N}$ we have that:

$$T_{PA_0} \vdash \neg(\underline{m} \doteq \underline{n}).$$

The proof is by induction on $\min\{m, n\}$. If either m or n is zero, then we are done by the previous bullet, otherwise it follows easily by inductive hypothesis and the T_{PA_0} -fact that successor is injective, i.e. A_3 (since m and n are both non-zero we think of them as the successors of their predecessors).

Putting all the bullets from above together, we see that \mathcal{M} is indeed closed under the interpretations of the functions and moreover that f is a homomorphism, and moreover moreover that f is injective.

Finally, we ought to show that \mathcal{M} is an initial segment of \mathcal{N} . For this, it will suffice to prove the following two bullets:

- For each $n \in \mathbb{N}$:

$$T_{PA_0} \vdash (\forall x) \left(x \leq \underline{n} \rightarrow \bigwedge_{i=1}^n x \doteq \underline{i} \right),$$

We, of course, argue by induction on n . Suppose first that $n = 0$. Then we need to show:

$$T_{PA_0} \vdash (\forall x)(\forall y)(x \pm \underline{y} = \underline{0} \rightarrow (x \doteq \underline{0} \wedge y \doteq \underline{0})).$$

This is a consequence of A_2, A_5, A_1 and A_4 (in that order).

For the inductive step, assume that the result holds for n . We must show it for $n + 1$. Let us now invoke soundness and completeness to make life a little easier. We wish to show that $T_{PA_0} \vdash \phi$, for some formula ϕ (of course, as we know, T_{PA_0} is consistent), so it suffices to show that for any model of $\mathcal{I} \models T_{PA_0}$ we have $\mathcal{I} \models \phi$ (of course, on that model we will only be allowed to use the properties guaranteed to us by the axioms). Let $\mathcal{I} \models T_{PA_0}$, and $a \in \mathcal{I}$ such that $\mathcal{I} \models a \leq \underline{n+1}$. It suffices to show that there is some $p \in \mathbb{N}$ such that $p \leq n + 1$ and $\mathcal{I} \models a \doteq \underline{p}$.

We know that there is a point $b \in I$ such that $\mathcal{M} \models b \dot{+} a \dot{=} \underline{S}n$ (that's just what it means to have $\mathcal{I} \models a \leq \underline{n+1}$). If $a = \underline{0}$ then we are done, and if not, by A_2 there is some $c \in M$ such that $\mathcal{I} \models a \dot{=} \underline{Sc}$. By A_5 and A_3 it follows that $\mathcal{I} \models b \dot{+} c \dot{=} \underline{n}$. But this just says that $\mathcal{I} \models c \leq \underline{n}$, so by inductive hypothesis, there is some $m \leq n$ such that $\mathcal{I} \models c \dot{=} \underline{m}$, and hence $\mathcal{I} \models \underline{Sc} \dot{=} \underline{Sm}$ and after a moment's thought, that's enough to show the result.

- For each $n \in \mathbb{N}$ we have:

$$T_{PA_0} \vdash (\forall x)(x \leq \underline{n} \vee x \geq \underline{n}).$$

This is again by induction on n . Of course, if $n = 0$ then we are done. Otherwise, we may argue as in the previous bullet. Let $\mathcal{I} \models T_{PA_0}$ and let $a \in I$. We need to show that $\mathcal{I} \models a \leq \underline{n+1}$ or $\mathcal{I} \models \underline{n+1} \leq a$. This is of course obvious if $a = \underline{0}$, so we may assume otherwise, and like before find some $b \in I$ such that $\mathcal{I} \models a \dot{=} \underline{S}b$. By our inductive hypothesis, we know that $\mathcal{I} \models b \leq \underline{n}$ or $\mathcal{I} \models \underline{n} \leq b$. In the first case, there is some $c \in I$ such that $\mathcal{I} \models c \dot{+} b \dot{=} \underline{n}$ and by A_5 and the very first bullet we can deduce that $\mathcal{I} \models c \dot{+} a \dot{=} \underline{n+1}$, which gives the result. In the second case, similarly, there is some d such that $\mathcal{I} \models d \dot{+} \underline{n} \dot{=} b$ and we can similarly conclude.

Putting everything together, the result follows. \square

In particular, since every model of T_{PA} is a model of T_{PA_0} we have the following corollary:

Corollary 1.2.4. *Let $\mathcal{N} \models T_{PA}$, then the set of all standard elements of \mathcal{N} is an initial segment of \mathcal{N} isomorphic to \mathcal{N}_{st} .*

2. Some dirty work

2.1. Representation... Recall that \mathcal{F}_p is the set of all (total) functions from \mathbb{N}^p to \mathbb{N} . We will start by defining a weak form of definability, for such functions:

Definition 2.1.1. Let $f \in \mathcal{F}_p$ and $\phi(x, y_1, \dots, y_p)$ be an \mathcal{L}_{Peano} formula. We say that ϕ represents f if for all $n_1, \dots, n_p \in \mathbb{N}$ we have that:

$$T_{PA_0} \vdash (\forall x)(\phi(x, \underline{n}_1, \dots, \underline{n}_p) \leftrightarrow x \dot{=} f(\underline{n}_1, \dots, \underline{n}_p)).$$

In this case, we say that f is *representable*.

The point is that ϕ represents f if and only if for every model \mathcal{N} of weak Peano arithmetic and every p -tuple \bar{n} (of actual factual) naturals we plug into f (well f

can only take actual factual naturals as inputs), there is a unique element in N satisfying $\phi(x, \bar{n})$ and that element is STANDARD, but not only that, it also is the correct element, in that it interprets the term $\underline{f}(n_1, \dots, n_p)$.

Of course, a subset $A \subseteq \mathbb{N}^p$ is called **representable** if its characteristic function is representable. Equivalently, $A \subseteq \mathbb{N}^p$ is representable if and only if there is an \mathcal{L}_{Peano} -formula $\phi(x_1, \dots, x_p)$ such that for all $\bar{n} \in \mathbb{N}^p$ we have:

- If $\bar{n} \in A$ then $T_{PA_0} \vdash \phi(\bar{n})$.
- If $\bar{n} \notin A$ then $T_{PA_0} \vdash \neg\phi(\bar{n})$.

It should be clear why these two notions are equivalent, but just to be safe:

Let $A \subseteq \mathbb{N}^p$.

- If ϕ represents A , then the formula:

$$(\phi(x_1, \dots, x_p) \wedge x \doteq \underline{1}) \vee (\neg\phi(y_1, \dots, y_p) \wedge x \doteq \underline{0})$$

represents $\mathbb{1}_A$.

- If $\mathbb{1}_A$ is representable by the formula $\psi(x, y_1, \dots, y_p)$, then, the formula $\psi(1, y_1, \dots, y_p)$ represents A .

You may be wondering, at this point, why didn't I write:

$$\bar{n} \in A \text{ if and only if } T_{PA_0} \vdash \phi(\bar{n}).$$

Is it because I'm trying my hardest to up the page count of these notes? Well, no! The point is that we don't know if T_{PA_0} is complete or not at this point (but if you're reading between the lines we actually know that it's not!) so we can't quite say that:

$$T_{PA_0} \not\vdash \phi(\bar{n}) \text{ if and only if } T_{PA_0} \vdash \neg\phi(\bar{n}).$$

We are asking, thus, for something strong. Not only that T_{PA_0} proves membership in A (when it has to) but also that it *proves* non-membership in A (when it has to).

Here are some examples of representable functions:

- (1) The successor function.
- (2) The constant function 0.
- (3) The projection functions.

We essentially proved all of these in the proof that the map:

$$\begin{aligned} f : \mathbb{N} &\rightarrow M \\ n &\mapsto \underline{n}^{\mathcal{N}} \end{aligned}$$

is an injective homomorphism (where M is the set of standard elements of some model $\mathcal{N} \models T_{PA_0}$).

Do you see where I'm going with this?

THEOREM 2.1.2 (The Representation Theorem). *Every total recursive function is representable.*

Going back to our proof that all register machine computable functions are recursive, we actually wrote down an explicit formula:

$$g(x_1, \dots, x_p) = \text{return}(\text{Run}_R\text{-at}(\text{term-time}_R(x_1, \dots, x_p), x_1, \dots, x_p)).$$

and this formula is clearly built using the μ operator, constant functions, projections and primitive recursive functions. Of course, as we already know, all recursive functions are (register machine) computable, and thus, given a recursive function, we can find the register machine that computes it and then reconstruct the function we started with using the formula above.

We are now visibly interested in *total* functions. Let's define closure under the *total* unbounded μ operator. Let $\mathcal{R}_{\text{total}}$ be the smallest subset of \mathcal{F} which contains the basic functions, is closed under primitive recursion, composition, and moreover is closed under the following (which we dub the total unbounded μ operator):

Let $A \subseteq \mathbb{N}^{n+1}$ be such that $\mathbb{1}_A \in \mathcal{R}$ and such that the function:

$$f(\bar{x}) := \mu y. [(y, \bar{x}) \in A],$$

is total. Then $f \in \mathcal{R}_{\text{total}}$.

It is not hard to see that $\mathcal{R}_{\text{total}} \subseteq \mathcal{R}$ and, in fact, $\mathcal{R}_{\text{total}} = \mathcal{R} \cap \mathcal{F}$.

What's the point of all of this? Well, in showing that every total recursive function is representable, it will suffice to show that:

- (1) The set of representable functions contains all the basic functions (Ha! We did this already.)
- (2) The set of representable functions is closed under composition.
- (3) The set of representable functions is closed under the total μ operator.
- (4) The set of representable functions is closed under primitive recursion.

We'll discuss points (2) and (3) here and point (4) in the next section, which will be rather technical and subtle.

Proposition 2.1.3. *The set of representable functions is closed under composition.*

PROOF. Let $f_1, \dots, f_n \in \mathcal{F}_p$ and $g \in \mathcal{F}_n$ be representable functions, and suppose that for $1 \leq i \leq n$, the formula $\phi_i(x, y_1, \dots, y_p)$ represents f_i and that the formula $\psi(x, y_1, \dots, y_n)$ represents g . Then, $g(f_1, \dots, f_n)$ is represented by:

$$(\exists z_1) \cdots (\exists z_n) \left(\psi(y, z_1, \dots, z_n) \wedge \bigwedge_{i=1}^n \phi_i(z_i, x_1, \dots, x_p) \right).$$

□

The μ business is harder, but we can do it!

Proposition 2.1.4. *The set of representable functions is closed under the total μ operator.*

PROOF. We have to show that if $A \subseteq \mathbb{N}^{p+1}$ is representable and the function $f(x_1, \dots, x_p)$ given by $\mu y. [(y, x_1, \dots, x_p) \in A]$ is total, then f is representable.

Suppose that $\phi(y, x_1, \dots, x_p)$ represents A . I claim that the formula $\psi(x, y_1, \dots, y_p)$ given by

$$\phi(y, x_1, \dots, x_p) \wedge ((\forall z)(z < y \rightarrow \neg\phi(z, x_1, \dots, x_p)))$$

represents f . What does this mean? Well... It means that given any model $\mathcal{M} \models T_{PA_0}$, the interpretation b of $f(n_1, \dots, n_p)$ in \mathcal{M} is the only element of \mathcal{M} that satisfies the formula $\psi(x, \underline{n}_1, \dots, \underline{n}_p)$. Obviously, since $\mathcal{M} \models T_{PA_0}$ and ϕ represents f we have that

$$\mathcal{M} \models \phi(b, \underline{n}_1, \dots, \underline{n}_p).$$

Also, if $c \in M$ is smaller than b , then (since b was standard) c is standard, so by definition of f we have that:

$$\mathcal{M} \models \neg\phi(c, \underline{n}_1, \dots, \underline{n}_p).$$

In particular, we have that:

$$\mathcal{M} \models \psi(b, \underline{n}_1, \dots, \underline{n}_p).$$

Finally, if $d \in M$ also satisfies $\psi(x, \underline{n}_1, \dots, \underline{n}_p)$, then neither $d < b$ nor $b < d$ can hold in \mathcal{M} , but since b is standard, we must have that $d = b$. □

Now, for the crux of the argument, we have to prove the following:

Proposition 2.1.5. *Let $f \in \mathcal{F}_p$ and $g \in \mathcal{F}_{p+2}$ be representable functions. Then, the function h defined by recursion from f and g , by:*

$$h(x_1, \dots, x_k, x_{k+1}) = \begin{cases} f(x_1, \dots, x_k) & \text{if } x_{k+1} = 0 \\ g(x_1, \dots, x_k, x_{k+1} - 1, h(x_1, \dots, x_k, x_{k+1})) & \text{otherwise} \end{cases}$$

is also representable. In particular, the set of representable functions contains all primitive recursive functions.