A CLASSIFICATION OF NONSTANDARD MODELS OF PEANO ARITHMETIC BY GOODSTEIN'S THEOREM

DAN KAPLAN

Abstract. In this paper I intend to outline a method for finding nonstandard models of Peano Arithmetic (PA) that satisfy Goodstein's theorem. Goodstein's Theorem is an interesting result because, though it is expressible completely in the language of number theory, it is nonetheless independent of the axioms of PA. I begin by rehearsing a proof of Goodstein's theorem, followed by a proof of its independence, developing the necessary tools to do so along the way. Finally, using indicator theory, I show how that can classify the nonstandard models according to Goodstein's Theorem.

Contents

This paper was completed as the result of a two-semester senior independent study (MAT490) with Prof. Justin Brody in the Department of Mathematics at Franklin and Marshall College. The author's date of graduation is May 2012.

Introduction

This paper was heavily influenced by another undergraduate thesis by Justin Miller (see [13]). In the two semesters leading up to the writing of this paper I studied some very basic set theory, not so basic model theory, and went through Gödel's incompleteness theorems. One thing that I became quite interested in after working through the incompleteness theorems was what undecidable results might look like and the methods used to show that a specific theorem was undecidable or not. In the beginning of the Spring of 2012 semester, my adviser introduced me to Miller's paper which contains a detailed development of the set theory necessary to prove Goodstein's theorem and a brief introduction to model theory with an outline of how one would prove the independence of Goodstein's theorem. As a result of this, I decided an interesting project might be to expand on Miller's paper, especially as regards his section on model theory and the independence of Goodstein's theorem, and to classify the nonstandard models of Peano Arithmetic (PA)—the standard axioms given for arithmetic—according to whether Goodstein's theorem is true on them (an element that is not taken up in Miller's paper). Since I spend a lot more time in model theory than Miller does, I also thought it might be interesting and relevant to rehearse a proof of Gödel's theorems and some interesting and related results, including a sketch of what nonstandard models of PA must look like. This latter result seemed like a necessary step for my final goal in the paper, classification of such models. Unfortunately finding a indicator for models of Goodstein's theorem that was tractable and didn't involve introducing much more machinery is beyond the scope of this paper. As such, in the end I say, how given such an indicator (which I do show is guaranteed) we could classify the nonstandard models.

Thus, in this paper, I begin by rehearsing a proof of Goodstein's theorem, following those given in [10, 13]. In order to do this I will first have to take a quick detour through some basic set theory.[1] My overall intention is to make this paper as accessible as possible, meaning that I will start with the assumption that the reader only has an intuitive understanding of logic and some facts concerning the natural numbers and arithmetic. However, because I will try to start from the most basic level and work my way up I will unfortunately not have time to develop many interesting side results—though I do develop some, especially within model theory and I shall allude to them when possible. As always the reader should feel free to skip the introductory material with which she is already familiar and to only attend as much time to proofs as is needed for comprehension. In the later sections, any tedium will hopefully be welcomed as a good thing.

0.1. **Incompleteness.** It was a long standing problem in mathematics whether it was possible to axiomatize the rules for arithmetic. If it was possible to do so then in principle any true statement of arithmetic would be provable from the axioms. Further, because, as it turns out, we can codify statements of arithmetic in arithmetic, we could reduce the discovery of such proofs (i.e. verifying whether any statement is true or false) to the running of an algorithm. In 1931, Gödel's showed that Peano Arithmetic was incomplete for first-order logic. What this meant is that any consistent, recursive set of axioms containing the axioms of PA will be incomplete, i.e. there will be statements that cannot be proven or disproven from those axioms. As a result, we can find multiple models that realize those axioms that are not elementarily equivalent to each other: they will not realize all the same first-order statements. We call the natural numbers, $\mathbb{N}$, the intended model, and all other models of PA,

[1] See [12] for an axiomatic development. [11, 18] also contain good introductions to set theory.

nonstandard models. As it turns out all nonstandard models of PA have the form $\mathbb{N}+\mathbb{Z}\cdot\mathbb{Q}$, though we will have to develop quite a bit of machinery before we can prove that claim.

An initial question for logicians was what kinds of theorems would be independent—i.e. true for some models of PA, but false on some nonstandard models—and whether any of those results were of any interest. While some results were known, they were mostly combinatorial in nature, and thus did not seem very "natural". When Goodstein put forward his theory it was suspected that that theorem might be such a result, however, it was not until the development of indicator theory by Paris and Harrington that incompleteness was proven.[2]

0.2. **Goodstein's Theorem.** A Goodstein sequence is understood intuitively in the following manner. We start with a number m and a base n. The first term of a Goodstein sequence is m. In order to arrive at the next term, we first write m in *hereditary* base-n notation. In normal base-n notation, we write m as the sum of n to various powers multiplied by coefficients less than n, i.e. for some $j \in \mathbb{N}$ and some set of coefficients γ_0 to γ_j we can write m as

$$m = \sum_{i=0}^{j} \gamma_i n^i = \gamma_j n^j + \gamma_{j-1} n^{j-1} + \ldots + \gamma_1 n + \gamma_0.$$

In hereditary base-n notation however, we also write all the exponents in base n notation. So that in our above example we would write 0 to j in base-n as well. In the case where $j \leq n$ this makes no difference, but when $j > n$ it does. For example if 63 in base-2 notation is written as

$$2^5 + 2^4 + 2^3 + 2^2 + 2^1 + 1,$$

but in hereditary base-2 notation, this would be written as

$$2^{2^2+1} + 2^{2^2} + 2^{2^1+1} + 2^{2^1} + 2^1 + 1.$$

This is significant because once we have written m in hereditary base-n notation, we then change all the n's in this notation to $(n+1)$'s, and then subtract one from the result. So for example to get to the next term after 63 with hereditary base-2 notation would be

$$3^{3^3+1} + 3^{3^3} + 3^{3^1+1} + 3^{3^1} + 3^1 + 1 - 1 = 3^{3^3+1} + 3^{3^3} + 3^{3^1+1} + 3^{3^1} + 3^1 \approxeq 3.05023899 \cdot 10^{13}.$$

I define Goodstein sequences and Goodstein's theorem formally on page 18. This informal definition should suffice for now.

Goodstein's theorem is that all such sequences eventually terminate. It is easy to see from the above why such a result woud be counterintutive, for while the Goodstein sequence for 3 starting with 2 terminates in 6 steps:

$$\begin{aligned} m_0 &= 2^1 + 1 = 3 \\ m_1 &= 3^1 + 1 - 1 = 3^1 = 3 \\ m_2 &= 4^1 - 1 = 3 \\ m_3 &= 3 - 1 = 2 \\ m_4 &= 2 - 1 = 1 \\ m_5 &= 1 - 1 = 0, \end{aligned}$$

[2]See [10].

Goodstein sequences starting with larger numbers grow much, much quicker. For example the Goodstein sequence for 4 starting with 2 takes approximately $10^{121210700}$ steps to terminate.[3] We can, however, briefly give an intuition for why one might think the sequence does terminate, even when it gets quite large. For the following let $[a]_n$ stand for the base-n number a. Writing base-n with $n < 10$ is not difficult, and it's common practice to begin using the alphabet after that to obtain more numbers (as in hexidecimal we use $0, 1, \ldots, 8, 9, A, \ldots, F$), but if we are dealing with, for example, a base-500 number, it becomes difficult to form a representation we can easily read off.[4] Thus, we might represent a base-500 number as: $[1]_{500}[496]_{500}$, where such a number in base-10, would be $1 \cdot 500^1 + 496 \cdot 500^0 = 500 + 496 = 996$. Now, to motivate the intuition, consider the Goodstein sequence for $50,100,004$ starting at $n = 500$ (this number was chosen simply because it provides an easy example). We have the following:

$$\begin{aligned}
m_0 &= 200 \cdot 500^2 + 200 \cdot 500 + 4 = [200]_{500}[200]_{500}[4]_{500} \\
m_1 &= 200 \cdot 501^2 + 200 \cdot 501 + 3 = [200]_{501}[200]_{501}[3]_{501} \\
m_2 &= 200 \cdot 502^2 + 200 \cdot 502 + 2 = [200]_{502}[200]_{502}[2]_{502} \\
m_3 &= 200 \cdot 503^2 + 200 \cdot 503 + 1 = [200]_{503}[200]_{503}[1]_{503} \\
m_4 &= 200 \cdot 504^2 + 200 \cdot 504 = [200]_{504}[200]_{504}[0]_{504} \\
m_5 &= 200 \cdot 505^2 + 199 \cdot 505 + 504 = [200]_{505}[199]_{505}[504]_{505} \\
&\vdots \\
m_{508} &= 200 \cdot 1008^2 + 199 \cdot 1008 + 1 = [200]_{1008}[199]_{1008}[1]_{1008} \\
m_{509} &= 200 \cdot 1009^2 + 199 \cdot 1009 = [200]_{1009}[199]_{1009}[0]_{1009} \\
m_{510} &= 200 \cdot 1010^2 + 198 \cdot 1010 + 1009 = [200]_{1010}[198]_{1010}[1009]_{1010}.
\end{aligned}$$

While we might still be skeptical that it can terminate—because, for example, my example did not take advantage of hereditary base notation (I chose a large enough base)—this should at least give one the intuition that the sequence may eventually end, since looking at the base representations, we see that "in some sense" they aren't getting bigger (they are always

[3] See [13, 1] for more information on this.

[4] We might start using combination of letters to get higher, for example after Z, comes AA, AB, etc. but this is just using base-10 plus base-26, which is to say why not stick with base-10 and makes things easy.

getting smaller). The sequence:

$$\begin{gathered}
[200]_{500}[200]_{500}[4]_{500} \\
[200]_{501}[200]_{501}[3]_{501} \\
[200]_{502}[200]_{502}[2]_{502} \\
[200]_{503}[200]_{503}[1]_{503} \\
[200]_{504}[200]_{504}[0]_{504} \\
[200]_{505}[199]_{505}[504]_{505} \\
\vdots \\
[200]_{1008}[199]_{1008}[1]_{1008} \\
[200]_{1009}[199]_{1009}[0]_{1009} \\
[200]_{1010}[198]_{1010}[1009]_{1010},
\end{gathered}$$

appears as though eventually it will reach, for some $n, k \in \mathbb{N}$,

$$\begin{gathered}
[1]_n[1]_n[n-1]_n \\
\vdots \\
[0]_{n+k}[1]_{n+k}[1]_{n+k} \\
[0]_{n+k+1}[1]_{n+k+1}[0]_{n+k+1} \\
[0]_{n+k+2}[0]_{n+k+2}[(n+k+2)-1]_{n+k+2},
\end{gathered}$$

from which point we would be guaranteed to reach zero. As it turns out the proof of Goodstein's theorem essentially formalizes this intuition in set theory by creating a parallel sequence to Goodstein's sequence that is strictly decreasing.

0.3. **Classification of Nonstandard Models.** In closing I should like to outline that it is possible to classify the nonstandard models of PA by Goodstein's theorem. Accomplishing this will require going into a bit of detail concerning indicator theory.

0.4. **Notation.** Before I get started I should say a few words about the notation used in this paper. The only notational familiarity I assume on the part of the reader is basic sentential and propositional logic as well as perhaps the basic symbols that comprise arithmetic ($+$, $\cdot$, etc.). For the most part I define all other notation as it arises. There are several conventions for which symbols to use for which purpose, and I cannot claim to adhere any one definitely. However, for the most part I use x, y, z, w, u, v as variables (free or quantified), lowercase a, b, c, d, e to refer to members of a set or model, uppercase A, B, C, D, E, F to refer to sets, and lowercase Greek letters to refer to functions and sometimes sentences and ordinals. I use $\overline{x}$ if I mean to refer to some n-tuple $x_1, \ldots, x_n$ in contexts where referring to certain components of the n-tuple aren't important. I also use special fonts for any of the above if I am talking about a specific element, set, etc. that deserves to be named or to be referred to repeatedly. I define other notation as it arises and the context should hopefully make clear what I am referring to when I do deviate from these conventions.

Part 1. Goodstein's Theorem

1. Set Theory

To begin, I'll need to walk through some basic set theory that will prove necessary for proving Goodstein's theorem. I begin with an axiomatic development of Zermelo-Fraenkel set theory with the axiom of choice (ZFC).

Axiom 1: [5] *Existence.*

$$\exists x(x = x).$$

Existence says that there exists sets.

Axiom 2: *Extensionality.*

$$\forall x \forall y(\forall z(z \in x \leftrightarrow z \in y) \rightarrow x = y).$$

Extensionality says that if two sets include exactly the same elements, then those sets are equal.

Axiom 3: *Foundation*:

$$\forall x[\exists y(y \in x) \rightarrow \exists y(y \in x \wedge \neg\exists z(z \in x \wedge z \in y))].$$

Foundation says that every non-empty set has an element which is disjoint with that set.

Axiom 4: *Comprehension.* For each formula ϕ with free variables among $x, z, w_1, ..., w_n$,

$$\forall z \forall w_1, ..., w_n \exists y \forall x(x \in y \leftrightarrow (x \in z \wedge \phi)).$$

Comprehension says that if ϕ is a property that characterizes some of the elements x in the set z, then we can find a set y which contains *exactly* those elements.

Axiom 5: *Pairing.*

$$\forall x \forall y \exists z(x \in z \wedge y \in z).$$

Pairing says for any two sets there is a set which contains both sets as elements. Along with comprehension we can say that there is a set which contains exactly those sets as elements.

Axiom 6: *Union.*

$$\forall \mathscr{F} \exists A \forall Y \forall x((x \in Y \wedge Y \in \mathcal{F}) \rightarrow x \in A).$$

Given the previous two axioms, union allows us to understand set union. Union says that for any set $\mathscr{F}$ we can find a set which contains the elements of all sets in $\mathscr{F}$, i.e. the union of the elements of $\mathscr{F}$.

For the next axiom (as well as the axiom of choice), we introduce the following notation for uniqueness. We let $\exists! y \phi(y)$ be short for $\exists y \phi(y) \wedge (\forall y \forall x(\phi(y) \wedge \phi(x)) \rightarrow x = y)$, which is read as "there exists a unique y, such that $\phi(y)$".

Axiom 7: *Replacement.* For each formula ϕ with free variables among $x, z, w_1, ..., w_n$,

$$\forall A \forall w_1, ..., w_n[\forall x \in A \exists! y \phi \rightarrow \exists Y \forall x \in A \exists y \in Y \phi].$$

Replacement essentially says something about the range of a function. To see how this goes, let f be a function with arguments and values among $x, y, w_1, \ldots, w_n$, and take ϕ to a formula which holds whenever the arguments it takes satisfy f. Replacement

[5] Detailed explanations of what each axiom "intuitively accomplishes" can be found in [12, 10ff.].

says that if A is the domain of that function f, then we can find a set Y which contains the image of f.

Finally, for the last three axioms, we will need to once again introduce some more notation. We define '$\subset$' (subset) as $A \subset B$ if $\forall x(x \in A \to x \in B)$. We define '$\cup$' (union) as: $A \cup B = \{x | x \in A \vee x \in B\}$; and '$\cap$' (intersection) as: $A \cap B = \{x | x \in A \wedge x \in B\}$. We also introduce set difference here '$\setminus$' as $A \setminus B = \{x \in A | x \notin B\}$. Finally, we define the successor function for ordinals as: $S(x) := x \cup \{x\}$.

Axiom 8: *Infinity.*

$$\exists x(\emptyset \in x \wedge \forall y \in x(S(y) \in x)).$$

Infinity says that there is a set that is closed under the successor function, meaning a set which includes the successors of all its elements. Such a set would therefore not be finite.

Axiom 9: *Power Set.*

$$\forall x \exists \mathcal{Z} \forall y(y \subset x \to y \in \mathcal{Z}).$$

This establishes the existence of the powerset for all sets, which is the set of all subsets of a set. We normally write $\mathcal{P}(x)$ to mean the power set of x, i.e. $\mathcal{P}(x) := \{z | z \subset x\}$.

Axiom 10: *Choice.*[6]

$$\forall \mathscr{F}(\forall H \in \mathscr{F} \neg(H = \emptyset) \wedge \forall F \in \mathscr{F} \forall G \in \mathscr{F}(F = G \vee F \cap G = \emptyset) \to \exists S \forall F \in \mathscr{F} \exists! s(s \in S \wedge s \in F)).$$

The axiom of choice essentially says that if we can partition a set $\mathscr{F}$, then there will be some set S such that for all elements of $\mathscr{F}$, there will be a unique element shared by S and F. Informally this means we can pick or *choose* arbitrary elements out of set.

1.1. $\mathbb{N}$ **- The Natural Numbers.** Using the machinery developed so far, we can construct the natural numbers in such a way that '$\in$' serves as the more familar '$<$', which orders $\mathbb{N}$. We do so in the following manner. For $0 \in \mathbb{N}$, let $0 := \emptyset$ and let $n + 1 := S(n)$. The axiom of infinity guarantees the existence of a set closed under the successor function $S(n)$, and we call the smallest such set ω. The following theorem establishes that we are guaranteed its unique existence.

Theorem 1.1. $\exists! X \forall Y((\emptyset \in X \wedge \forall Z(Z \in X \to S(Z) \in X)) \wedge (\emptyset \in Y \wedge \forall Z(Z \in Y \to S(Z) \in Y))) \to (X \subset Y)$. *That is, there is a unique X, such that X contains 0 is closed under $S(n)$, and for any other set Y, such that Y contains 0 and is closed under $S(n)$, X is a subset of Y.*

Proof. As mentioned above, we are guaranteed the existence of at least one such X, and so let X be the set guaranteed to us by Axiom 8, and we define $\omega := \bigcap\{Y \in \mathcal{P}(X) | \emptyset \in Y \wedge \forall Z(Z \in Y \to S(Z) \in Y)\}$. In other words, we define ω to be the intersecton of all such Y's that are a subset of X and that satisfy Axiom 8. What this entails is that if W is a set that satisfies the infinity axiom, such that $W \in \{Y \in \mathcal{P}(X) | \emptyset \in Y \wedge \forall Z(Z \in Y \to S(Z) \in Y)\}$, then $\omega \subset W$. If W is closed under the successor and contains 0, but isn't a member of

[6]Another (less formal) way of defining the axiom of choice is as follows:

$$\forall A \exists R(R \text{ well-orders } A).$$

These two definitions are equivalent. I define what it means for a set to be well-ordered below.

$\{Y \in \mathcal{P}(X) | \emptyset \in Y \wedge \forall Z(Z \in Y \to S(Z) \in Y)\}$, then it must be larger than all members of that set. This means we still have that $\omega \subset W$. □

We can define the addition function $+ : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ recursively as follows. For $m, n \in \mathbb{N}$

- $n + 0 = n$
- $n + S(m) = S(n + m)$.

Actually, we can do the same for the multiplication function $\cdot : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ and exponentiation.

- $n \cdot 0 = 0$
- $n \cdot S(m) = (n \cdot m) + m$.

Because the natural numbers we are dealing with are sets, I will delay defining exponentiation until we have a better understanding of what is characteristic of these sets.

1.2. **Ordinals.** Consider the following properties a relation $R \subset X \times X$ may have:

(1) $\forall x \in X((x, x) \in R)$ (reflexivity)
(2) $\forall x \in X \neg((x, x) \in R)$ (antireflexivity)
(3) $\forall x, y \in X[((x, y) \in R \wedge (y, x) \in R) \to x = y]$ (antisymmetry)
(4) $\forall x, y, z \in X[((x, y) \in R \wedge (y, z) \in R) \to (x, z) \in R]$ (transitivity)
(5) $\forall x, y \in X(x = y \vee (x, y) \in R \vee (y, z) \in R)$ (trichotomy)
(6) $\forall x(\exists y((y, x) \in R) \to \exists y((y, x) \in R \wedge \forall w((w, x) \in R \to ((w, y) \in R \vee w = y))))$ (discreteness)
(7) $\forall x \exists z((x, z) \in R \wedge \forall w((x, w) \in R \to (z, w) \in R \vee z = w)$ (successor)
(8) $\exists x \forall y((x, y) \in R \vee x = y)$ (left bound)
(9) $\forall x, y((x, y) \in R \to \exists z((x, z) \in R \wedge (z, y) \in R)$ (denseness)
(10) $\forall x \exists y, z((y, x) \in R \wedge (x, z) \in R)$ (unboundedness) .

Note that 2 and 4 imply 3, though they are not equivalent. We can write 5 in this form because 5 together with 2 and 4 get us the stronger trichotomy property, where the $\vee$ is understood exclusively. As such, usually 5, 2, and 4 usually appear together. 7 gives us a way of talking about the successor of an element using only the "language" of the order. An order is understood to be a binary relation over the elements of a set. Which of the above properties obtain for that order, will determine what order type it has. 7 says that there is no greatest element and every element has an immediate successor (i.e. the order is not dense).

Definition 1.2. Now, we say that a relation $R \subset X \times X$ is a

- *total-order* on X if R satisfies 1, 3, 4, and 5.
- *linear-order* on X if R satisfies 2, 4, and 5.

I introduce a few more order types which will be important later on. We say that R is a

- *Discrete linear order with first element, but not last element (DIS)* if R on X satisfies 2, 4, 5, 6, 7, and 8.
- *Discrete linear order without endpoints* if R on X satisfies 2, 4, 5, 6, 7, and 10.
- *Dense linear order (DLO)* if R on X satisfies 2, 4, 5, 9, and 10.

$\mathbb{N}$ is a DIS; $\mathbb{Z}$ is a discrete linear order without endpoints; $\mathbb{Q}$ is a DLO. The above are linear orders as well since they all satisfy 2, 4, and 5.

Definition 1.3. A relation $\preceq$ on a set α is a *well-ordering*, and we say α is *well-ordered* by $\prec$, if $\prec$ is a linear ordering and $\forall Y \in \mathcal{P}(\alpha)\exists m \in Y \forall n \in Y(\exists x \in Y \rightarrow (m = n \vee m \prec n)$; meaning that all non-empty subsets of α have a smallest element.[7]

Definition 1.4. A set α is *transitive* if and only if $\forall x \in \alpha(x \subset \alpha)$.

Definition 1.5. A set α is an *ordinal* if and only if α is transitive and well ordered by $\in$.

Now we prove briefly that the ordinals are closed under the successor function.

Theorem 1.6. *If α is an ordinal, then $S(\alpha) = \alpha \cup \{\alpha\}$ is an ordinal.*

Proof. Suppose α is an ordinal and let $\beta = S(\alpha)$. β is transitive, since for all for all $x \in \beta$ either $x = \alpha$ or $x \in \alpha$. In the first case x is a subset of β, and in the second case, x is a subset of α since α is an ordinal. Thus, β is transitive. Further, β is well-ordered since the only subsets of β contain a combination of α and subsets of α. Regardless, because α is well ordered by supposition, β is well-ordered as well. □

Theorem 1.7. *Let $\mathbf{\Phi}$ be a collection of ordinals. There is no function $\alpha : \mathbb{N} \rightarrow \mathbf{\Phi}$, such that $\forall i, j \in \mathbb{N}(i < j \rightarrow \alpha(j) < \alpha(i))$.*

Proof. Assume there was, now construct $A = \{\alpha(i) | i \in \mathbb{N}\}$. By the foundation axiom and the fact that the ordinals are well-ordered by $\in$, we can find x, such that $x \in a$ for each $a \in A$, and for each $a \in A$, x and a are disjoint. This means we can find an $x \in A$ such that x is minimal. Now, let i be such that $\alpha(i) = x$. By supposition $\alpha(S(i)) \in A$ an $\alpha(S(i)) < \alpha(i) = x$ since $i < S(i)$, but this contradicts how x was chosen. □

Intuitively what this result tells us is that there cannot be an infinite sequence of decreasing ordinals, or that any strictly decreasing sequence of ordinals must terminate. This will prove important for proving Goodstein's Theorem. Next, a few important definitions.

Definition 1.8. For α an ordinal, we say that α is a successor ordinal if and only if there is some ordinal β such that $S(\beta) = \alpha$. Otherwise we say that α is a limit ordinal.

An example of a limit ordinal discussed so far is ω. To see that ω is a limit ordinal, assume it is not. Then $\omega = S(\delta)$ for some δ. Because ω is an ordinal, $\delta \in \omega$. Further, because ω is defined so that it is closed under the successor function, $\S(\delta) \in \omega$. This means that $\omega \in \omega$, a contradiction. That sets cannot contain themselves is a result of the axiom of foundation.

We can find other limit ordinals as well. For example, because ω is an ordinal, its successor must be an ordinal as well. Thus by the infinity axiom, we can find a smallest set which contains $\omega, \omega + 1, \omega + 2, \ldots$, which we call $\omega + \omega$ or $\omega \cdot 2$. Likewise, if we take the sequence $\omega, \omega \cdot 2, \omega \cdot 3, \ldots$, we can find the limit ordinal $\omega \cdot \omega$ or ω^2. Similarly, we can find ω^ω and so on.

Definition 1.9. Cardinal numbers are used to measure the size of sets. The natural numbers $1, 2, 3, \ldots$ are all cardinal numbers, and we say that any set has cardinality n for some $n \in \mathbb{N}$ just in case there is a bijection from the ordinal n to that set.

We also define transfinite cardinal numbers, which correspond to the ordinals $\omega, \omega_1, \ldots$. It is convention when speaking of cardinality to use $\aleph_0, \aleph_1, \aleph_2, \ldots$ to refer to the cardinality of $\omega_0, \omega_1, \omega_2, \ldots$ respectively. Now, we denote ω_0 to be ω and ω_{n+1} to be the smallest ordinal with cardinality greater than $\aleph_n$.

[7] An alternate formulation is to say that $\preceq$ *well-orders* α if and only if $\preceq$ is total and that every non-empty subset of α has a $\preceq$-least element.

Note: ω_1 is the smallest ordinal which contains all countable ordinals. I do not prove the existence of these ordinals. Further because of space limits, the only transfinite cardinal that we will concern ourselves with is $\aleph_0$. While we will at later points talk of $2^{\aleph_0}$, but we do not take any stand in this paper on whether $2^{\aleph_0} = \aleph_1$ (i.e. the continuum hypothesis). In fact, we only introduce cardinals here because of their relevance generally and their use later on. Now, we move onto transfinite recursion.

1.3. **Transfinite induction and recursion.** First we introduce some notation and definitions.

Definition 1.10. Given a nonempty set X which is well-ordered by $<$, and $Y \in X$, we call the set $I(Y) = \{x \in X | x < Y\}$ an *initial segment* of X. In the case when $I(Y) \neq X$,then we say it is a *proper initial segment.*

Theorem 1.11 (Transfinite induction)**.** *Let A be well-ordered and $B \subset A$ such that for all $x \in A$, $I(x) \subset B \to x \in B$. It follows that $A = B$.*

Transfinite induction allows us a way to do proof by induction. Since if what is characteristic of B is that $\forall x \in B \phi(x)$, for some formula ϕ, then if we can show $A = B$, it will follow that $\forall x \in A \phi(x)$. This allows us to prove something about all the elements of a set. Transfinite recursion below allows and establishes a way for us to make recursive definitions of sets.

Note that we do not need to include a requirement for a base case in our theorem. This is because if $B = \emptyset$, then B will fail since $I(S(\emptyset)) = \emptyset \subset B$, but $S(\emptyset) \notin B$.

Proof. Suppose not. Then let $x \in A \setminus B$ be least element. We are guaranteed such an x since A is well-ordered. Now, $I(x) \subset B$ since all $y \in A$, such that $y < x$ are also in B. Thus $x \in B$, which is a contradiction. □

Definition 1.12. [8] Let f be a function such that $f : A \to B$, we say that $f \cap (C \times B)$ is the *restriction* of f to C, written $f \lceil C$.

A proof by transfinite induction is similar to a normal proof by induction as encountered in number theory except that we must also add a case for limit ordinals. For example, suppose we wish to show that $\forall x \phi(x)$ for some property ϕ. A proof by transfinite induction proceeds as follows:

Base case: In the base case we must prove that $\phi(0)$ holds.
Successor Case: In the successor case we assume for some ordinal α that $\phi(\beta)$ for all $\beta \leq \alpha$. Then we must show that $\phi(\alpha + 1)$.
Limit case: For the limit case we assume that $\phi(\beta)$ holds for all $\beta < \kappa$, where κ is a limit ordinal. Then we must show that $\phi(\kappa)$.

I demonstrate a proof using this method in the proof of transfinite recursion. First a brief bit of notation to make that proof more accessible. Define $\mathbf{V}$ and $\mathbf{O}$ informally in the following way:

Definition 1.13.

$$\mathbf{V} = \{x | x = x\}$$

and

$$\mathbf{O} = \{x | x \text{ is an ordinal}\}.$$

[8] It is convention to use $\restriction$ for restriction, but we reserve the use of this symbol for a model theoretic notion.

Technically these are not well defined in set theory as they are too large. But we lose no rigor if we should instead think of expressions involving them as being abbreviations for longer expressions. So for example $x \in \mathbf{O}$ should just be thought of as an abbreviation for the formula with one free variable that says x is an ordinal (see definition above). Likewise, we should think of $\mathbf{O} \cap y$ as $\{x \in y | x \text{ is an ordinal } \}$. The reason we introduce this notation is that it allows us to prove transfinite recursion in a much simpler way. That there is a function $g : \mathbf{O} \to \mathbf{V}$ just means that there is a set of ordered pairs $\{(x, y) | x \text{ is an ordinal } \}$. It is of course much simpler if we use this notation. A more rigorous formulation of transfinite recursion is presented in the note below.[9]

Theorem 1.14 (Transfinite recursion)**.** *If $f : \mathbf{V} \to \mathbf{V}$, then there is a unique function $g : \mathbf{O} \to \mathbf{V}$ such that $\forall \alpha (g(\alpha) = f(g \lceil \alpha))$.*

Proof. I will prove uniqueness first. Suppose the theorem holds for functions g_1 and g_2. We prove $\forall x (g_1(x) = g_2(x))$ (i.e. uniqueness) by transfinite induction (this also gives us an excellent opportunity to demonstrate transfinite induction). We get the base case as follows

$$\begin{aligned} g_1(0) &= f(g_1 \lceil 0) \\ &= f(g_1 \cap \emptyset \times \mathbf{V}) \\ &= f(\emptyset \times \mathbf{V}) \\ &= f(g_2 cap(\emptyset \times \mathbf{V}) \\ &= g_2(0). \end{aligned}$$

[9] A more rigorous version, which I shall not prove here, would go as follows:

Theorem (Transfinite recursion (robust version))**.** *Given a formula $f(x, y, \overline{w})$, we explicitly define a formula $g(v, y)$ such that*

$$\forall x \exists ! y f(x, y, \overline{w}) \to (\forall x \exists ! y g(x, y) \wedge \forall x \exists y \exists z (g(x, z) \wedge f(y, z) \wedge y = g \lceil x)). \tag{1.1}$$

Further for any formula $g'(v, y)$ which also satisfies 1.1, we have the following, which essentially says that g is unique (we could combine these two statements, but it would be quite unwieldy. No rigor is lost by doing it this way). For brevity, let $\mathbf{G}$ be the sentence which says that $g'(v, y)$ satisfied 1.1 if we replace all the instances of g in 1.1 with g'.

$$(\forall x \exists ! y f(x, y, \overline{w}) \wedge \mathbf{G}) \to \forall x \forall y (g(x, y) \leftrightarrow g'(x, y)). \tag{1.2}$$

This says exactly what we said in the first theorem in a more rigorous form. I do not prove this version here, however a proof (and more details concerning this formulation) can be found in [12, pp.25-27]. The informal version presented here is also taken from [12].

Now, for the successor case, assume that $g_1(\beta) = g_2(\beta)$ for all $\beta < \alpha$, and let $S(\beta) = \alpha$. Again, we have that

$$\begin{aligned} g_1(\alpha) &= f(g_1\lceil\alpha) \\ &= f(g_1 \cap (\alpha \times \mathbf{V})) \\ &= f(g_1 \cap (S(\beta) \times \mathbf{V})) \\ &= f(g_1 \cap ((\beta \cup \{\beta\}) \times \mathbf{V})) \\ &= f(g_1 \cap [(\beta \times \mathbf{V}) \cup (\{\beta\} \times \mathbf{V})]) \\ &= f([g_1 \cap (\beta \times \mathbf{V})] \cup [g_1 \cap (\{\beta\} \times \mathbf{V}])) \\ &= f([g_2 \cap (\beta \times \mathbf{V})] \cup [g_2 \cap (\{\beta\} \times \mathbf{V}])) \\ &= f(g_2 \cap S(\beta) \times \mathbf{V}) \\ &= f(g_2 \cap (\alpha \times \mathbf{V})) \\ &= g_2(\alpha). \end{aligned}$$

The limit case follows even faster, since if we assume it holds for all $\alpha < \kappa$, where κ is a limit ordinal, then since $\kappa = \bigcup_{\alpha<\kappa} \alpha$, we have our result. The details go as follows

$$\begin{aligned} g_1(\kappa) &= f(g_1\lceil\kappa) \\ &= f(g_1 \cap ((\cup_{\alpha<\kappa}\alpha) \times \mathbf{V})) \\ &= f(g_1 \cap (\cup_{\alpha<\kappa}\alpha \times \mathbf{V})) \\ &= f(\cup_{\alpha<\kappa} g_1 \cap (\alpha \times \mathbf{V})) \\ &= f(\cup_{\alpha<\kappa} g_2 \cap (\alpha \times \mathbf{V})) \\ &= f(g_2 \cap ((\cup_{\alpha<\kappa}\alpha) \times \mathbf{V})) \\ &= f(g_2\lceil\kappa) \\ &= g_2(\kappa). \end{aligned}$$

Thus, we have uniqueness.

Now, we must establish the existence of g. Let us call g a γ-approximation if and only if g holds for all $\alpha < \gamma$, i.e.

$$\forall\alpha < \gamma(g(\alpha) = f(g\lceil\alpha).$$

Now we show by transfinite induction[10] that there is γ-approximation for each γ. The base case follows vacuously since we can find g_0 such that $g_0(\emptyset) = f(g\lceil\emptyset)$.

For the successor case, let $\gamma = \alpha + 1$ be a successor ordinal and suppose that there is a β-approximation for all $\beta < \gamma$. Now, define g_γ such that for each $\beta < \alpha$, $g_\gamma(\beta) = f(g\lceil\beta)$, where g is a β-approximation, and finally let $g_\gamma(\alpha) = f(g_\gamma\lceil\alpha)$. The limit case follow similarly.

Therefore, let g be defined such that $g(\gamma)$ takes the value of $g'(\gamma)$ where g' is a $\gamma + 1$-approximation. We are guaranteed that g is unique by the above. □

Now, transfinite recursion allows us to define a function recursively for all ordinals. We do so by again specifying a base, a successor case, and finally a limit case. We define it in this way in order to guarantee that such a function exists uniquely. Next, we introduce ordinal addition, multiplication, and exponentiation, which we define recursively. The definitions

[10]This particular proof using transfinite induction is rather simple. I have included it in full detail just to provide another illustration of transfinite induction.

are similar to those we introduced in section 1.1. The major difference is that we define them for all ordinals and not just the natural numbers.

Definition 1.15. We define ordinal addition for the ordinal α as follows (we must define a new function for each ordinal)

- $\alpha + 0 = \alpha$
- $\alpha + S(\beta) = S(\alpha = \beta)$
- For γ a limit ordinal, $\alpha + \gamma = \bigcup_{\mu<\gamma} \alpha + \mu$.

Definition 1.16. We define ordinal multiplication for the ordinal α as follows (we must define a new function for each ordinal)

- $\alpha \cdot 0 = 0$
- $\alpha \cdot 1 = \alpha$
- $\alpha \cdot S(\beta) = (\alpha \cdot \beta) + \beta$
- For γ a limit ordinal, $\alpha \cdot \gamma = \bigcup_{\mu<\gamma} \alpha \cdot \mu$.

Definition 1.17. We define ordinal exponentiation for the ordinal α as the function $exp(\alpha, \beta)$. For convenience we write this as α^β.

- $\alpha^0 = 1$
- $\alpha^{S(\beta)} = (\alpha^\beta) \cdot \beta$
- For γ a limit ordinal, $\alpha^\gamma = \bigcup_{\mu<\gamma} \alpha^\mu$.

I have chosen to write these definitions more informally because our main goal in working through set theory is to prove Goodstein's theorem. As such, a rigorous working out of the definitions of these functions is not necessary.Similarly we do not prove the uniqueness of these functions. Because of how they are defined uniqueness follows almost immediately by transfinite recursion.[11]

Theorem 1.18 (Ordinal Addition)**.** *There is a unique function $+$ as specified in definition 1.15.*

Theorem 1.19 (Ordinal Multiplication)**.** *There is a unique function $\cdot$ as specified in definition 1.16.*

Theorem 1.20 (Ordinal Exponentiation)**.** *There is a unique function $exp(a, b) = a^b$ as specified in definition 1.17.*

1.3.1. *Normal Form.* The following is extremely important for Goodstein's theorem. Essentially it lets us say that there is a unique base-n representation for every number.

Theorem 1.21. *For all ordinals $\alpha, \beta > 1$, we can find unique ordinals $\gamma_0 > \gamma_1 > \cdots > \gamma_n$ and $0 < \mu_0, \mu_1, \ldots, \mu_n < \alpha$ such that*

$$\beta = \alpha^{\gamma_0}\mu_0 + \alpha^{\gamma_1}\mu_1 + \cdots + \alpha^{\gamma_n}\mu_n.$$

Proof. Fix α. If $\beta < \alpha$ then $\beta = \alpha^0\beta$, and if $\beta = \alpha$ then $\beta = \alpha^1 \cdot 1$, thus assume that $\alpha < \beta$.

By induction, assume the theorem holds for all ordinals less than β. Let γ be minimal such that $\alpha^\gamma > \beta$. I claim γ must be a successor ordinal. To see this, assume not. It follows then that $\alpha^\gamma = \bigcup_{\delta<\gamma} \alpha^\delta$. Note that $\alpha^\delta \leq \beta$ whenever $\delta < \gamma$, thus $\bigcup_{\delta<\gamma} \alpha^\delta < \beta$, a contradiction.

[11] See [12, 13] for worked out proofs of the uniqueness of these functions.

Thus let γ be the successor of γ_0. Note that $\alpha^{\gamma_0} \leq \beta$, and because the successor is unique, γ_0 will be unique here.

Now, let μ be minimal such that $\alpha^{\gamma_0}\mu > \beta$. By the exact same argument used above, μ must be a successor ordinal and thus let μ_0 be its unique predecessor. Note again that $\alpha^{\gamma_0}\mu_0 < \beta$. Finally, I claim we can find a unique β_1 such that $\alpha^{\gamma_0}\mu_0 + \beta_1 = \beta$. This follows from the fact that additional is well-defined, meaning that for fixed ordinals a, b, there is exactly one solution to the equation $a + x = b$. Thus, we can find a unique β_1. Since $\beta_1 < \beta$ it must have a unique representation by our hypothesis, and so we are done. □

The above notation for β we call the normal form of β with respect to base α. Another form, called the Cantor Normal Form is similar to the above, except that we also express $\gamma_0, \ldots, \gamma_n$ in normal form with respect to base α and carry on this process indefinitely (i.e. since each γ_n will have $\mu_0, \ldots, \mu_n$, we also write each μ_n in normal form with respect to base α, and so on). The following theorem formalizes this.

Theorem 1.22 (Cantor Normal Form)**.** *Given an ordinal Γ with $\alpha, \beta \in \Gamma$, we can find by Theorem unique ordinals $\gamma_0 > \gamma_1 > \cdots > \gamma_n$ and $0 < \mu_0, \mu_1, \ldots, \mu_n < \alpha$ such that $\beta = \alpha^{\gamma_0}\mu_0 + \alpha^{\gamma_1}\mu_1 + \cdots + \alpha^{\gamma_n}\mu_n$.. Next, we define the Cantor normal form function $CNF : \Gamma \times \Gamma \to \Psi$ for some ordinal Ψ such that*

$$CNF(\beta, \alpha) = \sum_{i=0}^{n} \alpha^{CNF(\gamma_i, \alpha)}\mu_i.$$

Note that if $\beta < \alpha$ then it follows that $CNF(\beta, \alpha) = \beta$.

Every ordinal has a unique cantor normal form representation.

Proof. Follows from theorem 1.22. □

The Cantor normal form is what allows us to prove Goodstein's theorem, since the cantor normal form allows us to uniquely express ordinals in hereditary base notation.

1.4. **Predecessor Sequences.** Before we turn to a proof of Goodstein's sequence, we must introduce one more thing: predecessor sequences. Though while we are at it, there are several important related idea that we can introduce here and which will be necessary for the independence proof below.

To begin, we let ϵ_0 be the smallest ordinal such that $\epsilon_0 = \omega^{\epsilon_0}$, i.e.

$$\epsilon_0 = \omega^{\omega^{\omega^{\cdot^{\cdot^{\cdot}}}}}$$

where ω has ω iterated exponents. It is the limit ordinal of the sequence $\omega, \omega^\omega, \omega^{\omega^\omega}, \omega^{\omega^{\omega^\omega}}, \ldots$.

It is not difficult to prove the existence of ϵ_0 given the above definitions, but we do not do so here.

Definition 1.23. We define the *predecessor operator*, for $\alpha < \epsilon_0$, $\{\alpha\} : \omega \to \epsilon_0$ recursively such that

$$\{0\}(n) = 0$$
$$\{\alpha + 1\}(n) = \alpha$$
$$\{\omega^{\delta+1}(\alpha + 1)\}(n) = \omega^{\delta+1}\alpha + \omega^\delta n, \text{ and}$$
(δ limit) $$\{\omega^\delta(\alpha + 1)\}(n) = \omega^\delta\alpha + \omega^{\{\delta\}(n)}.$$

Definition 1.24. We define the *Goodstein predecessor operator*, for $\alpha < \epsilon_0$, $\langle\alpha\rangle : \omega \to \epsilon_0$ recursively such that

$$\begin{aligned}\langle 0\rangle(n) &= 0\\ \langle\alpha+1\rangle(n) &= \alpha, \text{ and}\\ \langle\omega^\delta(\alpha+1)\rangle(n) &= \omega^\delta\alpha + \omega^{\langle\delta\rangle(n)}n + \langle\omega^{\langle\delta\rangle(n)}\rangle(n).\end{aligned}$$

For convenience we understand $\{\alpha\}(n_1, n_2, \ldots, n_k)$ to mean $\{\cdots\{\{\alpha\}(n_1)\}(n_2)\cdots\}(n_k)$, and similarly for $\langle\alpha\rangle(n_1, n_2, \cdots, n_k)$, which is how we define predecessor and Goodstein predecessor sequences. What is important about predecessor operators is that they give us a way of creating decreasing sequences that are defined even for limit ordinals. In both cases the predecessor operators are simply the inverse of the successor except in the limit cases, in which case we must "jump" down below the limit ordinal. We claim that both predecessor operators are well-defined. Since it is obvious for the base case and successor ordinals, we shall show that the predecessor operator is well-defined for limit ordinals (the proof is similar for the Goodstein operator).

Suppose κ is a limit ordinal. By theorem 1.22, κ has a unique normal form representation with respect to the base ω. Write κ in this form, i.e.

$$\kappa = \omega^{\gamma_n}\mu_n + \omega^{\gamma_{n-1}}\mu_{n-1} + \cdots + \omega^1\mu_1 + \omega^0\mu_0.$$

With $\gamma_n > \gamma_{n-1} > \cdots > \gamma_0$. Now, let m be minimal such that $\mu_m \neq 0$. Note that if $\mu_0 \neq 0$, then $\omega^0\mu_0$ is non-zero, but less than ω, which would mean that κ is a successor ordinal, thus assume $m > 0$. We can therefore drop all the terms of the above representation that have 0 as a co-efficient. Thus, we rewrite κ as

$$\kappa = \omega^{\gamma_n}\mu_n + \omega^{\gamma_{n-1}}\mu_{n-1} + \cdots + \omega^{\gamma_m} + \mu_m.$$

Now, we factor out ω^{γ_m} and we have

$$\begin{aligned}\kappa &= \omega^{\gamma_m}\left(\omega^{\gamma_n-\gamma_m}\mu_n + \cdots\omega^{\gamma_m-\gamma_m}\mu_m\right)\\ &= \omega^{\gamma_m}\left(\omega^{\gamma_n-\gamma_m}\mu_n + \cdots\omega^0\mu_m\right)\\ &= \omega^{\gamma_m}\left(\omega^{\gamma_n-\gamma_m}\mu_n + \cdots + \mu_m\right)\\ &= \omega^{\gamma_m}\left((\omega^{\gamma_n-\gamma_m}\mu_n + \cdots + \mu_m - 1) + 1\right).\end{aligned}$$

We can rewrite it in this form because μ_m must be a successor ordinal (otherwise $\mu_m \not< \omega$). It therefore follows that we can rewrite all limit ordinals as $\omega^\delta(\alpha+1)$, and so the predecessor operators are well-defined. As an example, suppose that $\kappa = \omega^\omega + \omega^2$, then we rewrite as follows:

$$\begin{aligned}\kappa &= \omega^2(\omega^\omega + 1)\\ &= \omega^2\cdot\omega^\omega + \omega^2\\ &= \omega^{2\cdot\omega} + \omega^2\\ &= \omega^\omega + \omega^2.\end{aligned}$$

We introduce two more pieces of notation before going on to prove several theorems which will prove imporant later on.

Definition 1.25. Given two ordinals $\alpha, \beta \in \epsilon_0$, we say α *n-supercedes*, written as $\alpha \underset{n}{\to} \beta$, if we can find $n_1, n_2, \ldots, n_k \leq n$, such that $\{\alpha\}(n_1, n_2, \ldots, n_k) = \beta$. As a special case, if we have that $n = n_1 = n_2 = \cdots = n_k$, then we say that α *n-succeeds* β, written $\alpha \underset{n}{\Rightarrow} \beta$.

Definition 1.26. We say that a set $A \subset \mathbb{N}$ is α-large iff

- if $\alpha = 1$, then $|A| \geq 2$, otherwise
- $\{a_i \in A | A \setminus \{a_i\}$ is $\{\alpha\}(a_i)$-large $\}$ is 1-large.

This definition gives us a notion of set-size based on the predecessor operator. In order to see if A is α-large, we must see if there are at least two elements $a_i \in A$, such that $A \setminus \{a_i\}$ is $\{\alpha\}(a_i)$-large. For example, if we let $A = (a_0, a_1, a_2, \ldots, a_k)$, then $A \setminus \{a_0\}$ is $\{\alpha\}(a_0)$-large if and only if $A \setminus \{a_0, a_1\}$ is $\{\alpha\}(a_0, a_1)$-large, which holds if and only if $\ldots$ $A \setminus \{a_0, \ldots, a_{k-2}\}$ is $\{\alpha\}(a_0, \ldots, a_{k-2})$-large, which holds if and only if $\{\alpha\}(a_0, \ldots, a_{k-2}) = 2$, since $A \setminus \{a_0, \ldots, a_{k-2}\}$ has only two elements: a_{k-1} and a_k. The following theorem establishes an equivalent definition based on this.

Theorem 1.27. *$A = \{a_1, a_2, \ldots, a_k\}$ is α-large if and only if $\{\alpha\}(a_1, a_2, \ldots, a_k) = 0$.*

Proof. Follows from induction on α. For the base case let $\alpha = 1$. A is α-large if and only if $|A| \geq 2$. Let $A = a_0, a_1, \ldots$. Then by definition $\{\alpha\}(a_0, a_1, \ldots) = 0$

Now, assume it holds for all $\beta < \alpha$. By definition A is α-large if and only if the following term is 1-large:

$$B = \{a_i \in A | A \setminus \{a_i\} \text{ is } \{alpha\}(a_i)\text{-large}\}$$

This is equivalent to saying that if and only if $|B| \geq 2$. Let $B = a_0, a_1, \ldots$. By definition this means for each $a_0 \in B$, that $A \setminus \{a_0\}$ is $\{\alpha\}(a_0)$-large. By supposition then $\{\{\alpha\}(a_0)\}(a_1, \ldots, a_k) = 0$, thus

$$\{\alpha\}(a_0, \ldots a_k) = 0.$$

□

Theorem 1.28. *Suppose $\beta \underset{n}{\to} \alpha$ and $0 < n \leq n_1 < n_2 < \cdots < n_k$, then $\{\beta\}(n_1, \ldots, n_k) \geq \{\alpha\}(n_1, \ldots, n_k)$.*

Proof. Proof is by induction on β. For the base case $\beta = 0$, we note that $\beta \underset{n}{\to} \alpha$ implies that $\alpha = 0$, from which the result is trivial.

Therefore, assume the theorem holds for all $\gamma < \beta$. Now suppose $\beta \underset{n}{\to} \alpha$. It follows that $\beta \underset{n_1}{\to} \alpha$ since $n_1 \geq n$. Likewise, we have $\alpha \underset{n_1}{\to} \{\alpha\}(n_1)$ and $\beta \underset{n_1}{\to} \{\beta\}(n_1)$ by definition. We therefore have that $\{\beta\}(n_1) \underset{n_1}{\to} \{\alpha\}(n_1)$. Recall that this just says that $\{\{\beta\}(n_1)\}(n_2, \ldots, n_k) = \{\alpha\}(n_1)$. Thus, by assumption, we get

$$\{\{\beta\}(n_1)\}(n_2, \ldots, n_k) \geq \{\{\alpha\}(n_1)\}(n_2, \ldots, n_k),$$

which is equivalent to

$$\{\beta\}(n_1, \ldots, n_k) \geq \{\alpha\}(n_1, \ldots, n_k).$$

□

Theorem 1.29. *For all $\alpha < \epsilon_0$ and $j, k, n \in \mathbb{N}$, $\langle\alpha\rangle(n, n+1, \ldots, n+k) \underset{j}{\to} \{\alpha\}(n, n+1, \ldots, n+k)$.*

Proof. If $\alpha = 0$ or $\alpha = \gamma + 1$, then we have our result immediately by definition. If $\alpha = \omega^{\gamma+1}(\beta+1)$, then $\{\alpha\}(j) = \omega^{\gamma+1} + \omega^{\gamma} j$ and $\langle\alpha\rangle(j) = \{\alpha\}(j) + \langle\omega^{\gamma}\rangle(j)$, from which the result follows. Finally, we consider the limit case where $\alpha = \omega^{\delta}(\beta+1)$, with δ a limit ordinal. Assume that the theorem holds for all $\mu < \alpha$. Follows that $\langle\delta\rangle(j) \underset{j}{\to} \{\delta\}(j)$, and it follows from this that $\omega^{\langle\delta\rangle(j)} \underset{j}{\to} \omega^{\{\delta\}(j)}$. Now, we calculate

$$\begin{aligned}\langle\alpha\rangle(j) &= \omega^{\delta}\beta + \omega^{\langle\delta\rangle(j)} j + \langle\omega^{\langle\delta\rangle(j)}\rangle(j) \\ &\underset{j}{\to} \omega^{\delta}\beta + \omega^{\langle\delta\rangle(j)} \\ &\underset{j}{\to} \omega^{\delta}\beta + \omega^{\{\delta\}(j)} \\ &= \{\alpha\}(j).\end{aligned}$$

Thus, we have our result:

$$\langle\alpha\rangle(j) \underset{j}{\to} \{\alpha\}(j).$$

□

2. Goodstein's Theorem

Now, we turn to the proof of Goodstein's theorem. Formally, we can define Goodstein sequences in the following manner. First, we note that we can write each $m \in \mathbb{N}$, m has a unique normal representation in base n. That is to say that we can find unique $\mu_0, ..., \mu_k < n$ and $\gamma_0 < \gamma_1 < ... < \gamma_k < \omega$ such that

$$m = \sum_{i=0}^{k} n^{\gamma_i}\mu_i. \tag{2.1}$$

Next we define $f_{m,n}(x) : \omega + 1 \to \epsilon_0$ by

$$f_{m,n}(x) = \sum_{i=0}^{k} \mu_i x^{f_{\gamma_i,n}(x)}. \tag{2.2}$$

$f_{m,n}$ replaces each n in the cantor normal form representation of n with x, which is a necessary component of building Goodstein sequences. For example, say that we have written in 728 (which is $3^6 - 1$) in hereditary base 3 notation as follows:

$$728 = 2 \cdot 3^{3^1+2} + 2 \cdot 3^{3^1+1} + 2 \cdot 3^{3^1} + 2 \cdot 3^{2 \cdot 3^0} + 2 \cdot 3^{1 \cdot 3^0} + 2 \cdot 3^0.$$

Now, say we want to replace all the 3's in that representation with 4's. We do so as follows:

$$f_{728,3}(4) = 2 \cdot 4^{4^1+2} + 2 \cdot 4^{4^1+1} + 2 \cdot 4^{4^1} + 2 \cdot 4^{2 \cdot 4^0} + 2 \cdot 4^{1 \cdot 4^0} + 2 \cdot 4^0.$$

This is the first step in defining a Goodstein sequence, which we now define formally.

Definition 2.1. First, let $G_n : \omega \to \omega$, be such that $G_n(m) = f_{m,n}(n+1) - 1$. We define the Goodstein sequence for m beginning with n recursively, letting $m_k(n)$ represent the kth term of the sequence, with

$$\begin{aligned} m_0(n) &= m, \\ m_{i+1} &= G_{n+i}(m_i(n)). \end{aligned}$$

Theorem 2.2 (Goodstein's Theorem)**.**

$$\forall m, n \in \omega \exists k \in \omega (m_k(n) = 0).$$

As I described in the introduction Goodstein's theorem is essentially proven by formalizing the intuition that a certain representation of Goodstein sequences don't *really* decrease. The way we formalize this is to create a parallel sequence of ordinals to any Goodstein sequence which is actually strictly decreasing, and by Theorem 1.7, therefore terminate. In order to prove Goodstein's theorem, we will need to prove two brief lemmas. The following notation will prove helpful in that respect. Define $o_n : \omega \to \epsilon_0$ such that $o_n(m) = f_{m,n}(\omega)$; in other words $o_n(m)$ replaces n in the base-n representation of m with ω.

Lemma 2.3. *For $m, n \in \omega$ with $n > 1$, if $\alpha = o_{n+1}(m)$, then $o_{n+1}(m-1) = \langle \alpha \rangle(n)$.*

Proof. We proceed by induction. First we note that since we are dealing with ordinals, we let $(m-1) := 0$, when $m = 0$. Now, for the base case let $m = 0$. We have our result since $\alpha = o_{n+1}(0) = 0$, and thus $\langle \alpha \rangle(n) = 0 = o_{n+1}(0)$.

Next, for the inductive step assume that our lemma holds for all $m' < m$ and $m > 0$. We may express m as $m = \sum_{i=0}^{k} a_i (n+1)^{f_{i,n+1}(n+1)}$. First we consider the case when $a_0 > 0$. Applying definitions yields $o_{n+1}(m-1) \sum_{i=1}^{k} a_i \omega^{f_{i,n+1}}(\omega) + (a_0 - 1) = \langle o_{n+1}(m) \rangle(n) = \langle \alpha \rangle(n)$., which is the result we want.

Now, we consider the case where $a_0 = 0$. Let be j be minimal, such that $a_j \neq 0$ and $\forall x < j (a_x = 0)$. In this case, we have

(2.3)

$$m - 1 = \left(\sum_{i=j+1}^{k} a_i (n+1)^{f_{i,n+1}(n+1)} \right) + \left((n+1)^{f_{j,n+1}(n+1)} (a_j - 1) + (n+1)^{f_{j,n+1}(n+1)-1} n + ((n+1)^{f_{j,n+1}(n+1)-1} - 1) \right)$$

Applying o_{n+1} to the above yields

(2.4)

$$o_{n+1}(m-1) = \left(\sum_{i=j+1}^{k} a_i \omega^{f_{i,n+1}(n+1)} \right) + \left(\omega^{f_{j,n+1}(n+1)} (a_j - 1) + o_{n+1}((n+1)^{f_{j,n+1}(n+1)-1} n) + o_{n+1}((n+1)^{f_{j,n+1}(n+1)-1} - 1) \right)$$

By our inductive hypothesis, we have that $o_n + 1(f_{j,n+1}(n+1) - 1) = \langle f_{j,n+1}(\omega) \rangle(n)$ and that $o_{n+1}((n+1)^{f_{j,n+1}(n-1)-1} - 1) = \langle \omega^{f_{j,n+1}(\omega)-1} \rangle(n) = \langle \omega^{\langle f_{j,n+1}(\omega) \rangle(n)} \rangle(n)$. We therefore calculate

$$\begin{aligned}\langle\alpha\rangle(n) &= \langle o_{n+1}(m)\rangle(n)\\ &= \langle f_{m,n+1}(\omega)\rangle(n)\\ &= \left\langle \sum_{i=j}^{k} \omega^{f_{i,n+1}(\omega)} a_i \right\rangle(n)\end{aligned}$$

Because taking the predecessor only affects the smallest term, i.e. the term whose ω component is the smallest, taking the predecessor of the above is equivalent to taking the predecessor of the smallest term. I detail this above when defining predecessor sequences. Now, we calculate further, therefore, that

$$\begin{aligned}\left\langle \sum_{i=j}^{k} \omega^{f_{i,n+1}(\omega)} a_i \right\rangle(n) &= \sum_{i=j+1}^{k} \omega^{f_{i,n+1}(\omega)} a_i + \langle \omega^{f_{j,n+1}(\omega)} a_j \rangle(n)\\ &= \sum_{i=j+1}^{k} \omega^{f_{i,n+1}(\omega)} a_i + \omega^{f_{j,n+1}(\omega)}(a_j - 1) + \omega^{\langle f_{j,n+1}(\omega)\rangle(n)} n + \langle \omega^{\langle f_{j,n+1}(\omega)\rangle(n)} \rangle(n)\\ &= o_{n+1}(m-1).\end{aligned}$$

Thus, $o_{n+1}(m-1) = \langle\alpha\rangle(n)$. □

Lemma 2.4. *If $n > 1$ then $\langle o_n(m)\rangle(n) = o_{n+1}(G_n(m))$.*

Proof. By straight calculation using the previous result, we find

$$\begin{aligned}o_{n+1}(G_n(m)) &= o_{n+1}(f_{m,n}(n+1) - 1)\\ &= \langle o_{n+1}(f_{m,n}(n+1))\rangle(n)\\ &= \langle o_n(m)\rangle(n).\end{aligned}$$

□

Now, we turn to Goodstein's theorem.

Proof. Using our previous result, we can construct a parallel sequence to any given Goodstein sequence, which we claim has the property of being strictly decreasing. Call this sequence γ_n and define it as follows

$$\begin{aligned}\gamma_0 &= o_n(m_0(n)), \text{ and}\\ \gamma_{i+1} &= o_{n+i+1}(m_{i+1}(n)).\end{aligned}$$

In order to demonstrate how this works, I shall create the parallel sequence for a Goodstein sequence we know to terminate fairly quickly: 3 starting with two. In the left column I have rewritten the Goodstein sequence and the right I write the parallel sequence:

$$\begin{aligned}m_0 &= 3 = 2^1 + 1\\ m_1 &= 3 = 3^1 + 1 - 1 = 3^1\\ m_2 &= 3 = 4^1 - 1\\ m_3 &= 2 = 3 - 1\\ m_4 &= 1 = 2 - 1\\ m_5 &= 0 = 1 - 1\end{aligned}$$

$$\begin{aligned}
\gamma_0 &= o_2(m_0) = \omega^1 + 1 \\
\gamma_1 &= o_3(m_1) = \omega^1 \\
\gamma_2 &= o_4(m_2) = \omega^0 \cdot 3 = 3 \\
\gamma_3 &= o_5(m_3) = 2 \\
\gamma_4 &= o_6(m_4) = 1 \\
\gamma_5 &= o_7(m_5) = 0.
\end{aligned}$$

As we can see, the ω-representation, like the base representation isn't really getting bigger, and there is a similarity in form between how the ω sequence decreases and the base sequence in the introduction decreased. The follow calculations show γ_n to be strictly decreasing:

$$\begin{aligned}
\gamma_{i+1} &= o_{n+i+1}(m_{i+1}(n)) \\
&= o_{n+i+1}(g_{n+1}(m_i(n))) \\
&= \langle o_{n+i}(m_i(N))\rangle(n+i) \\
&= \langle\langle o_{n+i-1}(m_{i-1}(n))\rangle(n+i-1)\rangle(n+i) \\
&\vdots \\
&= \langle\langle o_n(m_0(n))\rangle(n)\rangle(n+1,\ldots,n+i) \\
&= \langle o_n(m_0(n))\rangle(n)(n,n+1,\ldots,n+i) = \langle\gamma_0\rangle(n,n+1,\ldots,n+i),
\end{aligned}$$

which is strictly decreasing. Because γ_n is strictly decreasing, we have from Theorem 1.7, that it must eventually terminate. Let k be minimal, therefore, such that $\gamma_k = 0$. Thus, have $\gamma_k = o_{n+k}(m_k(n)) = 0$. By definition then, $m_k = 0$, and so we have our result. □

Part 2. Independence of Goodstein's Theorem

Now that we have shown Goodstein's Theorem is a true theorem of the natural numbers, we shall show that it is independent of PA, meaning that it is not provable from the axioms of Peano Arithmetic (which we spell out in detail below). To begin we shall have to develop some background in model theory, recursion theory, and finally indicator theory. Along the way I intend to give a thorough proof of Gödel's incompleteness theorem, because, after all, Goodstein's theorem is the topic of this paper precisely because it is an interesting example of an independent result.

3. Background Model Theory

Before we can jump into incompleteness we shall need to introduce some basic results from model theory. In particular I will need to motivate enough machinery to make plausible completeness, compactness, Löwenheim-Skolem theorem, and the Łoś-Vaught Test.[12] Model theory is commonly characterized as "universal algebra plus logic", though Wilfrid Hodges[13] has suggested that "algebraic geometry minus fields" might be a more accurate characterization. At any rate, the point of these metaphors is that model theory is concerned with studying classes of mathematical structures (or models), meaning that it is not so much concerned with particular truths of a mathematical structure, so much as what makes that structure the structure it is in relation to other structures. What is distinctive about fields, rings, graphs, etc. rather than waht is distinctive about *this* particular field. A better way of putting all this might be to say that model theory studies any structure describable by first-order logic. What Gödel's incompleteness theorem says is that the first-order theory of arithmetic does not pick out a unique structure up to elementary equivalence (a notion we define below).

In Model theory, there are three fundamental objects of study: languages, theories, and models. A language is literally the symbols we use to give meaning to sentences. All languages, therefore, include all the symbols of first order logic, plus some nonlogical symbols. Using a language we may define a theory (using those symbols naturally). Finally we say that a model, which is an interpretation of a language and a universe, is a model of a theory if that theory is true on that model. For the most part, we can be somewhat ambiguous about these three objects, as group theorests often do when talking of say $\mathbb{Z}_n$ (it refers to groups that are isomorphic even if these groups don't involve arithemetic modulo n); however, in laying out the foundations we should be quite formal. Usually it is fine to be ambiguous about a language in model theory since we can say more or less the same things in different languages.

Now for some formal definitions. Since we will be discussing Peano Arithmetic, I will motivate the following definitions with the corresponding language, theory, and so on.

Definition 3.1. A first order language $\mathscr{L}$ has the following set of symbols:

(1) A set of m-ary function symbols for each $m > 0$
(2) A set of constant symbols
(3) A set of m-ary relation symbols for each $m > 0$
(4) A binary relation symbol for equality: $\equiv$

[12]Most of the results here are taken from [11], though [7, 6, 18] also contain excellent in-depth explorations of model theory.

[13]See [7, 6].

(5) A countably infinite list of variable symbols: $x_0, x_1, \ldots, x_n, \ldots$ for all $n \in \omega$
(6) The logical connectives: $\neg, \wedge, \vee, \rightarrow, \leftrightarrow$ and the quantifiers: $\forall, \exists$.
(7) Parentheses: (and).

We call 1-3 the non-logical symbols of a language, $\mathscr{L}^{nl}$, since they will vary from language to language and the rest of the symbols the logical symbols of a language which do not vary.

So the language of arithmetic, which we specify with $\mathscr{L}_A$ has two binary function symbols: $+$ and $\cdot$, one binary relation symbol $<$, and two constant symbols $0, 1$.

Definition 3.2. We call $\mathrm{Tm}_{\mathscr{L}}$ the set of terms of a first order language $\mathscr{L}$, which we define as follows:

(1) $x_n \in \mathrm{Tm}_{\mathscr{L}}$ for each $n \in \omega$
(2) $c \in \mathrm{Tm}_{\mathscr{L}}$ for each constant symbol c of $\mathscr{L}$
(3) For each m-ary function symbol F of $\mathscr{L}$ and $t_1, \ldots, t_m \in \mathrm{Tm}_{\mathscr{L}}$, $F(t_1, \ldots, t_m) \in \mathrm{Tm}_{\mathscr{L}}$.

Keeping as our example $\mathscr{L}_A$, the terms of this language, $\mathrm{Tm}_{\mathscr{L}_A}$, include all the variables and constants in our language, and are furthermore closed under the functions of the language. So $0, 1 \in \mathrm{Tm}_{\mathscr{L}_A}$, and also $1+1, (1+1)+1, ((1+1)+1)+1, \ldots \in \mathrm{Tm}_{\mathscr{L}_A}$, and so on.

Definition 3.3. We call $\mathrm{Fm}_{\mathscr{L}}$ the set of formulas of a first order language $\mathscr{L}$, which we define as follows:

(1) $S \in \mathrm{Fm}_{\mathscr{L}}$ for each atomic sentence S, where we understand the atomic sentences of $\mathscr{L}$ to be expressions of the form: $Rt_1 \ldots t_m$ for all m-ary relation symbols R in $\mathscr{L}$, $m \in \omega$, and $t_1, \ldots, t_m \in \mathrm{Tm}_{\mathscr{L}}$
(2) for each $\phi, \psi \in \mathrm{Fm}_{\mathscr{L}}$,
    - $\neg\phi \in \mathrm{Fm}_{\mathscr{L}}$
    - $\phi \wedge \psi \in \mathrm{Fm}_{\mathscr{L}}$
    - $\phi \vee \psi \in \mathrm{Fm}_{\mathscr{L}}$
    - $\phi \rightarrow \psi \in \mathrm{Fm}_{\mathscr{L}}$
    - $\phi \leftrightarrow \psi \in \mathrm{Fm}_{\mathscr{L}}$
(3) for each $\phi \in \mathrm{Fm}_{\mathscr{L}}$ and $n \in \omega$: $\forall x_n\phi, \exists x_n\phi \in \mathrm{Fm}_{\mathscr{L}}$.

Definition 3.4. The set of sentences of $\mathscr{L}$, $\mathrm{Sn}_{\mathscr{L}}$ is the set: $\{\phi \in \mathrm{Fm}_{\mathscr{L}} | \phi \text{ has no free variables}\}$.

Now, we turn to one of the most important definitions of model theory, the definition of a model.

Definition 3.5. A *model* for a structure of a first order language $\mathscr{L}$ is a pair $\mathfrak{A} = (A, \mathcal{I})$ such that A is the *universe* of the structure and I is a function mapping the non-logical symbols of the language, $\mathscr{L}^{nl}$, to their interpretation in the universe. Building up recursively, we set the atomic for formulas $\phi \in \mathrm{Fm}_{\mathscr{L}}$, $I(\phi) = \phi^{\mathfrak{A}}$, and so on for relations and constant symbols. It is common practice to use Fraktur to refer to structures and to leave out I explicitly. Thus we might specify the model of a discrete linear order with one endpoint as $\mathfrak{S} = (\omega, <, s, 0)$, or the model of arithmetic as $\mathfrak{A} = (\mathbb{N}, +, \cdot, <, 0, 1)$. It is unnecessary to specify exactly what $\mathcal{I}$ is in these cases because we assume that we interpret them in the normal way. The main point of a model is therefore to connect the syntactic notions describable in a language with the semantic notions (i.e. notions of truth) that we actually encounter in mathematical structures.

As such, in the model of a discrete linear order with one endpoint $\mathfrak{S}$, the language in question would have the following nonlogical symbols: $\mathscr{L}_S^{nl} = \{<_S, s_S, 0_S\}$, i.e. it would have one binary relation symbol $<_S$, one unary function symbol s_S and one constant 0_S. We understand its terms then as repeated applications of the successor function to 0_S and the variables. We interpret then 0_S as 0, and then likewise 1 would be the interpretation $s_S(0_S)$ and so on.

We adopt the following notation for making clear interpretations, leaving it out where the interpretation of a model is obvious. Given an $\mathscr{L}$ structure $\mathfrak{A}$, we define the interpretation of a closed term $t^{\mathfrak{A}_A}$ as follows:

(1) If t is a constant symbol of $\mathscr{L}$ then $t^{\mathfrak{A}_A} = c^{\mathfrak{A}_A}$.
(2) if $t = F(t_1, \ldots, t_m)$ for some m, then $t^{\mathfrak{A}_A} = F^{\mathfrak{A}}((t_1^{\mathfrak{A}_A}, \ldots, t_m^{\mathfrak{A}_A})$

It will rarely be necessary to be this formal when talking about interpretations of a model, but we do so in defining some of these basic ideas.

Of course this is not the only model that could be interpreted on that language. We might specify the following model: $\mathfrak{P} = \{A, <, s, 0\}$, where $A = 0, 1, 2, 3, 4$, and then we interpret 0_S as 0, $s_S(0_S)$ as 1, and so on, except we also interpret 0 as $s(4)$, i.e. $s_S(s_S(s_S(s_S(s_S(0_S)))))$. On this model we would not have the same statements true. As such, we introduce the following notation for talking about truth.

For any model $\mathfrak{A}$ and a sentence S, we say that S is true on $\mathfrak{A}$ or $\mathfrak{A}$ satisfies S as $\mathfrak{A} \vDash S$. For atomic sentences $\mathfrak{A} \vDash S$ just in case S holds, and we understand truth for nonatomic sentences ($S \wedge T$, $S \vee T$, ...) in the normal way. Further, for a formula F, we write $\mathfrak{A} \vDash F$ just in case $\mathfrak{A} \vDash \phi(a_1, \ldots, a_n)$ for all $a_1, \ldots, a_n \in A$ the universe of $\mathfrak{A}$. The following definition establishes what I have just said informally.

Definition 3.6. Given an $\mathscr{L}$-structure $\mathfrak{A}$, we define the truth-value of the interpretation of a sentence $\theta \in \mathrm{Sn}_{\mathscr{L}}$, i.e. $\theta^{\mathfrak{A}_A}$, so that $\theta^{\mathfrak{A}_A} \in \{T, F\}$ (i.e. the set of truth values, true and false) recursively as follows

(1) if θ is $Rt_1, \ldots, t_m$ for closed terms $t_1, \ldots, t_m$ and relation R, then $\theta^{\mathfrak{A}_A} = T$ if and only if $R^{\mathfrak{A}}(t_1^{\mathfrak{A}_A}, \ldots, t_m^{\mathfrak{A}_A})$ holds.
(2) if θ is $t_1 \equiv t_2$, then $\theta^{\mathfrak{A}_A} = T$ if and only if $t_1^{\mathfrak{A}_A} = t_2^{\mathfrak{A}_A}$.
(3) if $\theta = \neg\phi$ then $\theta^{\mathfrak{A}_A} = T$ if and only if $\phi^{\mathfrak{A}_A} = F$.
(4) if $\theta = (\phi \wedge \psi)$ then $\theta^{\mathfrak{A}_A} = T$ if and only if $\phi^{\mathfrak{A}_A} = \psi^{\mathfrak{A}_A} = T$, and likewise for the other sentential connections: $\vee, \rightarrow$, etc.
(5) if $\theta = \forall x\phi$ then $\theta^{\mathfrak{A}_A} = T$ if and only if $\phi^{\mathfrak{A}_A}(x) = T$ for all $a \in A$; and if $\theta = \exists x\phi$ then $\theta^{\mathfrak{A}_A} = T$ if and only if $\phi^{\mathfrak{A}_A}(x) = T$ ofr some $a \in A$.

Truth value is defined then exactly as we would expect. A sentence is true just in case it holds on the model. Now with this in mind, we can define the double turn style $\vDash$ formally. We write $\mathfrak{A} \vDash \theta$ if and only if $\theta^{\mathfrak{A}_A} = T$ and $\mathfrak{A} \nvDash \theta$ if and only if $\theta^{\mathfrak{A}_A} = F$.

Note: this still leaves room for independence because while on any particular structure all first-order sentences will either be true or false, it does not follow that any set of recursive axioms can specify all of those sentences. More on this in the following sections.

We also introduce the following notation. If Σ is a set of sentences and θ a sentence, then we write $\Sigma \vDash \theta$ to mean that on any model in which Σ holds, θ holds as well.

Now, we will introduce two relations between structures.

Definition 3.7. Let $\mathscr{L}_1, \mathscr{L}_2$ be first order languages such that $\mathscr{L}_1 \subset \mathscr{L}_2$. Let $\mathfrak{A}$ be a $\mathscr{L}_1$-structure and $\mathfrak{B}$ be a $\mathscr{L}_2$-structure. If $\mathfrak{A}, \mathfrak{B}$ have the same universe and interpret $\mathscr{L}^{nl}$ in the same way then we say that $\mathfrak{A}$ is the *reduct* of $\mathfrak{B}$ to $\mathscr{L}_1$ or that $\mathfrak{B}$ is the expansion of $\mathfrak{A}$ to $\mathscr{L}_2$, written $\mathfrak{A} = \mathfrak{B} \restriction \mathscr{L}_1$.

Once again let $\mathscr{L}_A$ be the language of arithmetic, and now let $\mathscr{L}_O$ be the language of the order, which has as its only nonlogical symbol the relation $<$. Now, $\mathscr{L}_A \subset \mathscr{L}_O$, and so $(\mathbb{N}, <) = (\mathbb{N}, 0, 1, +, \cdot, <) \restriction \mathscr{L}_O$.

Finally, we introduce the single-turn style $\vdash$ which we use to indicate that a set of formulas proves a formula. For the following definition let Λ be the set of tautologies and logical truths of first order logic. In general, we will appeal to more rules, but for formal purposes we say that our only inference rule is modus-ponens (MP), i.e. ψ if ϕ and $\phi \to \psi$.

Definition 3.8. Let $\Gamma \subset \mathrm{Fm}_{\mathscr{L}}$. A *deduction* from Γ is a finite set sequences $\phi_1, \ldots, \phi_n$ of formulas such that for each $i \leq n$, $\phi_i \in \Lambda \cup \Gamma$ or there are $j, k < i$, such that $\phi_k = \phi_j \to \phi_i$ (i.e. MP). We say the formula ϕ is *deducible* from Γ if and only if there is a deduction $\phi_1, \ldots, \phi_n, \phi$. We write this as $\Gamma \vdash \phi$.

3.1. **Theory and Model Equivalence.** In this section we will need to define two important ideas. Equivalence between theories (defined below) and equivalence between models. First some preliminary definitions.

Definition 3.9. Let $\Sigma \subset \mathrm{Sn}_{\mathscr{L}}$. We define the set of consequences of Σ as

$$\mathrm{Cn}_{\mathscr{L}}(\Sigma) = \{S \in \mathrm{Sn}_{\mathscr{L}} | \Sigma \vDash S\}.$$

For example, on the theory of a discrete linear order with lower bound $\mathfrak{S}$, if $\Sigma = \{0 < 1, 1 < 2, 2 < 3, \forall x, y, z((x < y \wedge y < z) \to x < z)\}$, then $0 < 2, 0 < 3, 1 < 3 \in \mathrm{Cn}_{\mathscr{L}}$, because these sentences are all consequences of the sentences in Σ. The following definition gives us a way of talking about all the models that satisfy some set of sentences. For example, $0 < 1$ is satisfied by both the theory of a discrete linear order with lower bound and the theory of arithmetic.

Definition 3.10. Let $\Sigma \subset \mathrm{Sn}_{\mathscr{L}}$ again. We define the class of models of Σ as

$$\mathrm{Mod}_{\mathscr{L}}(\Sigma) = \{\mathscr{L}\text{-structures } \mathfrak{A} | \mathfrak{A} \vDash \Sigma\}.$$

Definition 3.11. A theory of $\mathscr{L}$ is a set of sentence T such that $T = \mathrm{Cn}_{\mathscr{L}}(\Sigma)$ for some $\Sigma \subset \mathrm{Sn}_{\mathscr{L}}$. In this case we call Σ the set of axioms of T.

Definition 3.12. A theory T is said to be complete if and only if T has a model and for all $S \in \mathrm{Sn}_{\mathscr{L}}$ either $S \in T$ or $\neg S \in T$.

Naturally, we will want a way of reading a theory off from a model, which the following definition captures.

Definition 3.13. Let $\mathfrak{A}$ be a $\mathscr{L}$-structure. The theory of $\mathfrak{A}$ is

$$Th(\mathfrak{A}) = \{S \in \mathrm{Sn}_{\mathscr{L}} | \mathfrak{A} \vDash S.\}$$

The following establishes what it means for two models to have the same theories.

Definition 3.14. Let $\mathfrak{A}, \mathfrak{B}$ be $\mathscr{L}$-structures. We say $\mathfrak{A}$ and $\mathfrak{B}$ are elementarily equivalent, written $\mathfrak{A} \equiv \mathfrak{B}$ if and only if $Th(\mathfrak{A}) = Th(\mathfrak{B})$.

While elementary equivalence is an important idea in model theory, elementarily equivalent models are not necessarily isomorphic. For example, the models of a theory may have different cardinality. The following established what it means for two models to be isomorphic.

Definition 3.15. Let $\mathfrak{A}, \mathfrak{B}$ be $\mathscr{L}$-structures. We say $\mathfrak{A}$ and $\mathfrak{B}$ are isomorphic, written $\mathfrak{A} \cong \mathfrak{B}$, if and only if there is a bijection $h : A \to B$, where A is the universe of $\mathfrak{A}$ and B is the universe of $\mathfrak{B}$, and the following are true of h:

(1) $h(c^{\mathfrak{A}}) = c^{\mathfrak{B}}$, for all constant symbols $c \in \mathscr{L}$. We understand $c^{\mathfrak{A}}$ to be the interpretation of c in $\mathfrak{A}$; we follow a similar convention below for functions and relations
(2) $h(F^{\mathfrak{A}}(a_1, \ldots, a_m)) = F^{\mathfrak{B}}(h(a_1), \ldots, h(a_m))$ for all m-ary function symbols $F \in \mathscr{L}$ and all $a_1, \ldots, a_m \in A$
(3) $\mathfrak{A} \vDash R^{\mathfrak{A}}(a_1, \ldots, a_m)$ if and only if $\mathfrak{B} \vDash R^{\mathfrak{B}}(h(a_1), \ldots, h(a_m))$ for all m-ary relation symbols $F \in \mathscr{L}$ and all $a_1, \ldots, a_m \in A$.

Elementary equivalence is meant to express the fact that two models say the same things in first order logic; isomorphism is meant to express the fact that two models have the exact same structure. One famous and counter-intuitive (to those unfamiliar with model theory) example is that while the theory of the order on $\mathbb{Q}$: $(\mathbb{Q}, <) \ncong (\mathbb{R}, <)$, $Th((\mathbb{Q}, <)) \equiv Th((\mathbb{R}, <$)). The reason for this is that what distinguishes $\mathbb{R}$ and $\mathbb{Q}$ (besides their cardinality) is that in the reals, all non-empty subsets have a supremum that is also in the reals; however, this is not a fact we can express using first order logic and the relation symbol '$<$'.

3.2. **Important theorems in model theory.** The following is a short list of important theorems in model theory. I do not supply a proof since it is not essential for our purposes.

Theorem 3.16 (Completeness)**.** *Let $\Gamma \subset Fm_{\mathscr{L}}$ and $\phi \in Fm_{\mathscr{L}}$. Then $\Gamma \vDash \phi$ if and only if $\Gamma \vdash \phi$, and Γ is consistent if and only if it is satisfiable.*

The gist of completeness is that any consistent set of formulas will have a model. That is to say that the syntactic notion of $\vdash$ corresponds with the semantic notion of $\vDash$. The following are two interesting results of this theorem. Compactness says that any consistent set of formulas has a model if and only if every finite subset of that that set has a model. Löwenheim-Skolem Theorem says that if a theory has a model of infinite cardinality, it has a model of every infinite cardinality.

Theorem 3.17 (Compactness)**.** *Let $\Gamma \subset Sn_{\mathscr{L}}$*

(1) *For any $\Gamma \in Fm_{\mathscr{L}}$, $\Gamma \vDash \phi$ if and only if $\Gamma_0 \vDash \phi$ for some finite $\Gamma_0 \subset \Gamma$.*
(2) *Γ has a model if and only if every finite $\Gamma_0 \subset \Gamma$ has one.*

Theorem 3.18 (Löwenheim-Skolem Theorem)**.** *Let $\kappa = |\mathscr{L}|$ and assume that $\Gamma \subset Sn_{\mathscr{L}}$ has a model. Then Γ has a model $\mathfrak{A}$ with $|A| \leq \kappa$.*

Definition 3.19. We say a theory T is *κ-categorical*, for some cardinal κ, if and only if T has exactly one model of cardinality κ up to isomorphism.

Theorem 3.20 (Łoś-Vaught Test)**.** *Let T be a theory of $\mathscr{L}$ such that T has no finite models and κ a cardinal, such that $|\mathscr{L}| \leq \kappa$ and let T be κ-categorical. It follows then that T is complete.*

4. Background recursion theory

I will have to be even briefer in my treatment of recursion theory.[14] We include recursion theory in order to make formal what we mean when we say that a function or relation is recursive. The original hope in defining axioms of arithmetic was that we could recursively define axioms that would deductively entail all truths of arithmetic. Gödel's theorems, which I get to later, show that such a recursive set of axioms cannot exist.

Definition 4.1. The class of *partial recursive functions*, $\mathscr{C}$, is the smallest class of functions $f : \mathbb{N}^k \to \mathbb{N}$, $k \geq 1$, such that

(1) $\mathscr{C}$ contains the zero functon, i.e. $0(x) = 0$.
(2) $\mathscr{C}$ contains the successor function, i.e. $S(x) = x + 1$ for all $x \in \mathbb{N}$.
(3) $\mathscr{C}$ contains the projection functions, i.e. for each $1 \leq i \leq n \in \mathbb{N}$, $P_i^n(x_1, \ldots, x_n) = x_i$.
(4) $\mathscr{C}$ is closed under function composition, i.e. for each $f(x), g(x) \in \mathscr{C}$, $f(g(x)) \in \mathscr{C}$.
(5) $\mathscr{C}$ is closed under primitive recursion, meaning if $f(x), g(x, y, z) \in \mathscr{C}$ then $h(x, y) \in \mathscr{C}$ such that h is defined by:
   - $h(x, 0) = f(x)$
   - $h(x, y + 1) = \begin{cases} g(x, y, h(x, y)) \\ \text{undefined, if } h(x, y) \text{ undefined} \end{cases}$
(6) $\mathscr{C}$ is closed under minimalization, i.e. if $g(x, y) \in \mathscr{C}$, then so is

$$h(x) = \begin{cases} \min\{y | g(x, y) = 0\} \\ \text{undefined if there is no such } y. \end{cases}$$

We define also the class of *primitive recursive functions*, $\mathscr{PR}$, such that $\mathscr{PR}$ satisfies 1, 2, 3, 4, 5. I therefore follows that $\mathscr{PR} \subset \mathscr{C}$.

The following functions are in $\mathscr{PR}$ (and thus $\mathscr{C}$):

- $x + y$
- $x \cdot y$
- x^y
- $x - 1$, with $x - 1 = 0$ if $x \leq 1$
- $x - y$, with $x - y = 0$ if $y \geq x$
- $\min(x, y)$ and $\max(x, y)$.

The following definition simplifies checking if a set or functon is recursive.

Definition 4.2. For a set $S \subset \mathbb{N}^k$ is recusrive if and only if its characteristic function is recursive, where its characteristic function is defined as follows:

$$\mathcal{CF}(x) = \begin{cases} 1 \text{ if } x \in S \\ 0 \text{ if } x \notin S. \end{cases}$$

Theorem 4.3 (Church-Turing Thesis)**.** *A function is f is in $\mathscr{C}$ if and only if there is an algorithm for computing f in finite time.*

There is no proof the Church-Turing thesis as the rather informal statement of it would suggest. It was put forward initially as a conjecture saying that any function is computable

[14]The information on recursion theory was taken primarily from [11, 3]. [8] contains a very brief introduction to recursion theory, which was helpful in deciding which material was relevant for inclusion here.

only if it can be computed by a Turing machine. Nonetheless it is largely accepted as true since it succinctly captures what we have in mind when we think of computability/recursion, is equivalent with some other notions of computability, and has no counterexamples.. While it is less formal than we might wish, we can use it to justify our intuitons that a function is recursive, and of course we can delve deeper into what such an algorithm might look like, as the proof demands.

5. Peano Arithmetic

With that basic model theory out of the way, I now turn to the theory of Peano Arithmetic. Since we have the foundations laid out, we can become a bit more relaxed concerning notation. The follow axiomatization is the most commonly given axiomatization of PA (and also the most intuitive).[15] Nonetheless, alternative axiomatizations exist, which I address briefly at the end of this section.

PA1: *Associativity* $(+)$: $\forall x, y, z((x + y) + z = x + (y + z))$
PA2: *Commutivity* $(+)$: $\forall x, y(x + y = y + x)$
PA3: *Associativity* $(\cdot)$: $\forall x, y, z((x \cdot y) \cdot z = x \cdot (y \cdot z))$
PA4: *Commutivity* $(\cdot)$: $\forall x, y(x \cdot y = y \cdot x)$
PA5: *Distribution*: $\forall x, y, z(x \cdot (y + z) = x \cdot y + x \cdot z)$
PA6: *Identity* $(+)$: $\forall x((x + 0 = x) \wedge (x \cdot 0 = 0))$
PA7: *Identity* $(\cdot)$: $\forall x(x \cdot 1 = x)$
PA8: *Transitivity* $(<)$: $\forall x, y, z((x < y \wedge y < z) \rightarrow x < z)$
PA9: *Antireflexivity* $(<)$: $\forall x \neg(x < x)$
PA10: *Trichotomy* $(<)$: $\forall x, y(x < y \vee x = y \vee y < x)$
PA11: *Respect of the order* $(+)$: $\forall x, y, z(x < y \rightarrow x + y < y + z)$
PA12: *Respect of the order* $(\cdot)$: $\forall x, y, z((0 < z \wedge x < y) \rightarrow x \cdot y < y \cdot z)$
PA13: *Subtraction*: $\forall x, y(x < y \rightarrow \exists z(x + z = y))$
PA14: *Discreteness* $(<)$: $0 < 1 \wedge \forall x(0 < x \rightarrow 1 \leq x)$
PA15: *Foundation*: $\forall x(0 \leq x)$

We call these first 15 axioms PA^-. In the next section I will show that the PA1-PA15 (plus the induction schema; see below) are insufficient for picking out the intended model of arithemetic, $\mathbb{N}$. If, however, we add the following second-order axiom we can show that any model of that theory can be reduced to PA. We shall not make much use of this stronger theory, though it is worth mentioning. Let us call the previous 15 axioms plus the following *second-order induction axiom* PA^+:

PA16$^+$: *Induction*: $\forall X(0 \in X \wedge \forall x(x \in X \rightarrow x + 1 \in X) \rightarrow \forall y(y \in X))$

Intutively, what this axiom says is that the only elements in our model are 0 and those elements which can be "reached" by repeated application of the successor function to 0, i.e. all and only $\mathbb{N}$. Another way of putting this point is to say that all the elements of this second-order theory are exactly those elements contained in the set X which only contains 0 and is closed under the successor. We can exclude nonstandard models because any nonstandard model will still have $\{0\}$ as a subset. This axioms says that all elements of our model must be reachable via a finite number of applications of the successor on $\{0\}$, and this will not hold in nonstandard models.

[15] For a more in-depth development of these axioms see [18, 8].

We call this axiom second-order because we have to quantify over sets in order to express it. The disadvantage of second-order logic, unfortunately is that many basic results in model theory and proof theory fail to hold. Theorem 3.17 fails, for example. Hence we continue to work in the first-order theory of PA. On this note, one might think that the above axioms can be simplified. While alternative axiomatizaton schemes do exist that are simpler, it is necessary to include multiplication as well as addition because we cannot define multiplication from addition or the successor. While we define, for example $x + y$ as $S(\cdots(S(x))\cdots)$, we would need to include such a definition for every such y, which is to say we would need to quantify over formulas—something that cannot be done in first-order logic.

Instead, we include an induction scheme to get us PA. Essentially for each formula, we incude the following axiom, and PA is taken to mean PA1-PA15, all the induction schema for formulas of language. Because all the formulas are built out of symbols in our language, we can include the following class of formulas in our axioms:

PA16$_\phi$: $\forall\overline{x}((\phi(\overline{x},0) \wedge \forall y(\phi(\overline{x},y) \to \phi(\overline{x},y+1))) \to \forall z(\phi(\overline{x},z)))$

is the axiom of induction for the formula $\phi(\overline{x},y)$.

5.1. **Alternative Schema.** Another way one might axiomatize PA is as follows:

PA'1: $\forall x\neg(x+1=0)$
PA'2: $\forall x,y(x+1=y+1 \to x=y)$
PA'3: $\forall x(x+0=0)$
PA'4: $\forall x,y(x+(y+1)=(x+y)+1)$
PA'5: $\forall x(x\cdot 0=0)$
PA'6: $\forall x,y(x\cdot(y+1)=x\cdot y+x)$
PA'7$_\phi$: $\forall\overline{x}((\phi(\overline{x},0) \wedge \forall y(\phi(\overline{x},y) \to \phi(\overline{x},y+1))) \to \forall z(\phi(\overline{x},z)))$.

To see that they are equivalent, we would need only to show that the above 16 axioms follow from these 6. It is not proven here, since the first axiomatization schema is more convenient. The reader may consult [18, 115f.] for a proof of several of them.

6. Gödel's Theorems

6.1. **Δ_0, Σ_1, and Π_1 Formulas.** The following allow us to define classes of formulas based upon the amount of quantifiers they have

Definition 6.1. If a formula, in which the free variable is x has either of the following forms

- $\exists y(y \leq x \wedge \phi(x,y))$, or
- $\forall y(y \leq x \to \phi(x,y))$

then we say that the quantifier is *bounded* in this formula. Essentially, to check if a formula if this form is true, we need only check values of y up to n, and so we can determine the truth of such a formula in a finite number of steps. For clarity, we write formulas with bounded variables as follows and treat them as equivalent to the above:

- $\exists y \leq x\phi(x,y))$, or
- $\forall y \leq x\phi(x,y))$.

Definition 6.2. A formula is called Δ_0, or *with bounded quantifiers*, if all of its quantifiers are bounded. We call a formula Σ_1 if it has the form $\exists x\phi$, where ϕ is Δ_0. Likewise, we call a formula Π_1 if it has the form $\forall x\phi$ where ϕ is Δ_0. Further, for $n > 1$, we say a formula is

Σ_{n+1} if it has the form $\exists x\phi$, where ϕ is Π_n; and we say it is Π_{n+1} if it has the form $\forall x\phi$, where ϕ is Σ_n.

Therefore, n tells us that there are n unbounded quantifiers in front of the formula, which are alternatively $\exists$ and $\forall$. Σ formulas start with $\exists$; Π formulas start with $\forall$. As an example, if we have $\exists x\exists y f(x)$, where $f(x)$ is Δ_0, that formula would be equivalent to a Σ_1 since we may rewrite it as $\exists z\exists x \leq z\exists y \leq z f(x)$.

From here we can define classes of formulas. We can speak of the Δ_0, Σ_n, Π_n classes of formulas to refer to the set of all formulas of that form. Further, we say that a set of formulas is Δ_n if it is equivalent to both a Σ_n and Π_n set of formulas.

The following definition will prove useful in the next section.

Definition 6.3. Let $f : \mathbb{N}^k \to \mathbb{N}$ be a total function and T an $\mathscr{L}_A$ theory extending PA. We say that a function f or set $S \subset \mathbb{N}^k$ is represented in T if and only if there is a $\phi(\overline{x}, y) \in \mathrm{Fm}_{\mathscr{L}_A}$ such that for each $\overline{n} \in \mathbb{N}$

- $T \vdash \exists! y\phi(\overline{n}, y)$
- if $f(\overline{n}) = d$, then $T \vdash \phi(\overline{n}, d)$

Similarly, for S:

- if $\overline{s} \in S$ then $T \vdash \phi(\overline{s})$
- if $\overline{s} \notin S$, then $T \vdash \neg\phi(\overline{s})$.

We say that f or S is Σ_n-represented in T just in case ϕ is equivalent to a Σ_n formula. Likewise, we say that s or f is Π_n-represented in T just in case ϕ is equivalent to a Π_n formula.

What representation does is gives us a way of either reducing a function to a formula. For example, say that f is the function for addition by 2, i.e. $f(x) = x + 2$. Then ϕ represents f just in case for all x, $T \vdash \phi(x, x + 2)$, meaning ϕ says $f(x) = x + 2$.

Likewise we say a set is represented just in case there is a formula that picks out that set, i.e. is only true for members of that set. A rather mundane example might go as follows. $S = \{0, 1, 2, 3\}$ and $\phi(x) = x = 0 \vee x = 1 \vee x = 2 \vee x = 3$. In this case whenever $x \in S$ then $T \vdash \phi(x)$, and whenever $x \notin S$, $T \vdash \neg\phi(x)$.

Theorem 6.4. *All recursive sets $S \subset \mathbb{N}^k$ are Σ_1-represented in PA.*

Proof. See [18, 8, 3]. □

Corollary 6.5. *All recursive functions $f : \mathbb{N}^k \to \mathbb{N}$ are Σ_1-represented in PA.*

Proof. Let S_f be the set of ordered pairs $(\overline{x}, y)$ such that $(\overline{x}, y) \in S_f$ just in case $f(\overline{x}) = y$. Then our result follows immediately from the previous theorem. □

6.2. **Gödel Coding.** Gödel coding gives us a way of associating each formula, in an injective manner, with its code, so that each formula has a unique code. Since formulas of $\mathscr{L}_A$ are just strings of symbols, namely: , ( ) $\neg$ $\vee$ $\wedge$ $\forall$ $=$ 0 1 $\leq$ $+$ $\cdot$ x_1 x_2 $\cdots$. We associate with each symbol a unique integer, n_s, from the list $0, 1, 2, 3, 4, \cdots$, where b represents the number of unique symbols. There are many different Gödel codes, what is important is that each code be a recursive, injection $\sigma \to \ulcorner\sigma\urcorner$, where $\ulcorner\sigma\urcorner$ is understood to be the Gödel code for σ. See the appendix for two worked out examples of Gödel coding. As it turns out, how exactly one defines a code is relatively unimportant. Briefly, however, here are two examples of how one might construct such codes:

Example 6.6. One example (taken from Poizat) is to map the symbols: $($ $)$ $\neg$ $\vee$ $\wedge$ $\forall$ $=$ 0 1 $\leq$ $+$ $\cdot$ x_1 x_2 $\cdots$, in order with the integers, starting with 0. We let the Gödel code of a sentence be the number

$$p_1^{a_0+1} \cdots p_n^{a_n+1}$$

Where p_i is the ith prime number. In a sentence with n symbols we represent the ith symbol with $p_i^{a_i+1}$, where a_i is the integer that corresponds to that symbol. Because each unique sentence will pick out a unique prime decomposition of some natural number, we can see that we can uniquely represent any sentence in this manner. Moving from sentences to formulas requires introducing some more theory, but is still possible.

Example 6.7. Another example using a similar technique of such a coding (taken from [8]) is to send the string $s_0 s_1 s_2 \cdots s_k$ to the integer

$$\ulcorner s_0 s_1 s_2 \cdots s_k \urcorner = \sum_{i=0}^{k} b^i \cdot (n_{s_i} + 1).$$

We do so by associating with each letter of the alphabet a number $0, 1, 2, \ldots, 25$, and where b is the number of symbols in our alphabet plus one. Once again, we are guaranteed that each unique string maps to a unique number.

Further, if we understand a proof simply as a finite string of sentences, then we can simply add a new symbol between sentences and produce codes for proofs. We now demonstrate the power of such codes in the following famous result.

Theorem 6.8 (Tarski's Theorem)**.** *Let $\mathscr{L}_A$ be our language fix $M \vDash PA$ as our $\mathscr{L}_A$ structure. There is no formula $V(x)$, such that V is satisfied on M in exactly those cases in which x is the code of a true sentence of a model of arithmetic M.*

Proof. Assume by way of contradiction that there is a such a $V(x)$. Let $\mathscr{L}_A$ be our language fix M as our $\mathscr{L}_A$ structure. Let A be the set of codes of formulas whose only free variable is x. Now, we define $\phi : A \times \mathbb{N} \to \mathbb{N}$ such that $\phi(n, m)$ yields the code obtained by replacing x in the formula that corresponds to n with the number m. ϕ is recursive, and therefore by corollary 6.5 has a Σ_1-representation.

Now, we note that the following formula has only one free variable, x: $\neg V(\phi(x, x))$, which by supposition holds in exactly those cases in which $\phi(x, x)$ is *not* the code of a true sentence of M. Because it has only one free variable its code is in A. Let n_0 be the code, therefore, for $\neg V(\phi(x, x))$. It follow then that $M \vDash V(\phi(n_0, n_0)) \leftrightarrow \neg V(\phi(n_0, n_0))$, which is a contradiction. That is, if $V(\phi(n_0, n_0))$ is true, then $\neg V(\phi n_0, n_0)$ is true, and if $V(\phi n_0, n_0)$ is false, then $\neg V(\phi(n_0, n_0))$ is false. Thus we have our result. □

Tarski's theorem essentially says that "this sentence is false" is a sentence that, though it appears we have the machinery for talking about it in $\mathscr{L}_A$, doing so leads to contradiction. The result being that we cannot speak of the truth of sentences of arithmetic from within arithmetic. Another important take away is that it is often not important exactly how one defines a Gödel code. The specifics of a code rarely come into play in a theorem. Tarski's theorem also introduces another important method that we will need to prove Gödel's incompleteness theorems: diagonalization.

Theorem 6.9 (Diagonalization Theorem)**.** *Given any $\mathscr{L}_A$-theory T that Σ_1 - represents all recursive functions and $\phi(x)$, a formula with one free variable, we can find a sentence S such that $T \vdash S \leftrightarrow \phi(\ulcorner S \urcorner)$. If ϕ is a Π_1 formula, then S may taken as equivalent to a Π_1 formula.*

Proof. Define $d(x)$ as follows

$$d(x) = \begin{cases} \ulcorner \forall y(y = x \to \sigma(y)) \urcorner & \text{if } x = \ulcorner \sigma(x) \urcorner, \text{ for some } \mathscr{L}_A \text{ formula} \\ 0 & \text{otherwise.} \end{cases}$$

Now, let $\psi(x) = \forall z(d(x) = z \to \phi(x))$, and let S be the sentence

$$\forall y(y = \ulcorner \phi(x) \urcorner \to \phi(y)).$$

Now, I claim that $T \vdash S \leftrightarrow \psi(\ulcorner \psi(x) \urcorner)$. Which the following calculations show to be true

$$\begin{aligned} T &\vdash S \leftrightarrow S \\ T &\vdash \forall z(d(\ulcorner \phi(x) \urcorner) = z \to \phi(z)) \leftrightarrow \forall z(d(\ulcorner \phi(x) \urcorner) = z \to \phi(z)) \\ T &\vdash \forall y(\ulcorner \phi(x) \urcorner = y \to \forall z(d(\ulcorner \phi(x) \urcorner) = z \to \phi(y)) \leftrightarrow \forall z(d(\ulcorner \phi(x) \urcorner) = z \to \phi(z)) \\ T &\vdash \forall y(\ulcorner \phi(x) \urcorner = y \to \forall z(d(y) = z \to \phi(y)) \leftrightarrow \forall z(d(\ulcorner \phi(x) \urcorner) = z \to \phi(z)) \\ T &\vdash \forall y(\ulcorner \phi(x) \urcorner = y \to \phi(y) \leftrightarrow \forall z(d(\ulcorner \phi(x) \urcorner = z \to \phi(z)) \\ T &\vdash S \leftrightarrow \forall z(d(\ulcorner \phi(x) \urcorner = z \to \phi(z)) \qquad \text{(a)} \\ T &\vdash S \leftrightarrow \psi(\ulcorner \psi(x) \urcorner). \end{aligned}$$

Now, because $\ulcorner S \urcorner = d(\ulcorner \phi(x) \urcorner)$ it follows that $T \vdash \forall z(d(\ulcorner \phi(x) \urcorner) = z \leftrightarrow z = \ulcorner S \urcorner$. From (a) it therefore follows that

$$T \vdash S \leftrightarrow \forall z(z = \ulcorner S \urcorner \to \phi(z)),$$

and thus,

$$T \vdash S \leftrightarrow \phi(\ulcorner S \urcorner).$$

Finally, if $\phi(x)$ is Π_1, then ψ is, since excluding $\phi(x)$, $\psi(x)$ is Δ_0, and likewise for S, since we can just as easily rewrite S as $\psi(\ulcorner \psi(x) \urcorner)$. □

Intuitively, what this says is that we can find, for any formula, a sentence that says "I am true precisely when the code of me is true for that formula". Initially, we might think finding such a sentence will be impossible since any attempt to define it initially will mean a change to its code and then we will have to go back and forth. d says that we can play this game without running into trouble. In Tarski's theorem, the sentence we looked for said "I am true precisely when my code corresponds to a false sentence," which is why we ran into trouble. An interesting result of diagonalization is that for sentences such at ϕ, we have that $T \vdash \phi(\ulcorner \theta \urcorner) \to \phi(\ulcorner \phi(\ulcorner \theta \urcorner) \urcorner$. The significance of this will become clear in the proof of Gödel's second theorem. Now, we move onto Gödel's proofs.

6.3. **Gödel's Incompleteness Theorems.**

Theorem 6.10 (Gödel's First Incompleteness Theorem)**.** *Let T be a consistent recursively axiomatized $\mathscr{L}_A$-theory extending PA. Then there is a Π_1 sentence τ such that neither $T \vdash \tau$ nor $T \vdash \neg\tau$.*

Lemma 6.11. *Let T be a recursive set of Gödel codes of $\mathscr{L}_A$ sentences such that all Σ_1 and Π_1 sentences that PA proves are contained in T and if $\ulcorner \sigma \urcorner \in T$, then $\ulcorner \neg\sigma \urcorner \notin T$ (T is* simply consistent*). T is incomplete meaning that there is a Π_1 sentence such that neither its code or the code of its negation are in T.*

Proof. Because T is recursive, there is a Σ_1 formula $\theta(x)$, such that for all $n \in \mathbb{N}$, $n \in T \leftrightarrow PA \vdash \theta(n)$, and $n \notin T \leftrightarrow PA \vdash \neg\theta(n)$. Now, we use 6.9 to find a sentence S such that $PA \vdash S \leftrightarrow \neg\theta(\ulcorner S \urcorner)$, but then if $\ulcorner S \urcorner \in T$, then $PA \vdash S$ by supposition and so $PA \vdash \neg\theta(\ulcorner S \urcorner)$, which means $\ulcorner S \urcorner \notin T$, i.e. a contradiction. Similarly, if we suppose $\ulcorner S \urcorner \notin T$, then $\ulcorner \neg S \urcorner \in T$ and so $PA \vdash \neg S$, thus $PA \vdash \theta(\ulcorner S \urcorner)$ and so $\ulcorner S \urcorner \in T$. Thus, $\ulcorner S \urcorner \notin T$ and $\ulcorner \neg S \urcorner \notin T$. □

The reader will note a striking similarity between the proof of this method and Tarski's theorem. This is precisely how proofs utilizing diagonalization go. Now we are prepared to prove Gödel's theorem.

Proof. Let $A = \{\ulcorner \sigma \urcorner | \sigma \text{ is a } \mathscr{L}_A \text{ sentence, } \Pi_1 \text{ or } \Sigma_1, \text{ and } T \vdash \sigma\}$. By the above lemma then, we can find a Π_1 sentence S such that $\ulcorner S \urcorner \notin A$ and $\ulcorner \neg S \urcorner \notin A$. Thus, $T \nvdash S$ and $T \nvdash \neg S$. □

Now, we move onto Gödel's second theorem. His second theorem shall not prove all that important for our purposes, but since we have developed all the machinery needed to prove it, we do so here. It essentially says that sufficiently complex first-order theories cannot prove their own consistency.

Definition 6.12. We say that an $\mathscr{L}_A$ formula $\theta(x)$ with one free variable is a *provability predicate* for T, an extension of PA, if for all sentences σ, τ of $\mathscr{L}_A$

(1) if $T \vdash \sigma$, then $T \vdash \theta(\ulcorner \sigma \urcorner)$
(2) $T \vdash \theta(\ulcorner \sigma \to \tau \urcorner) \to (\theta(\ulcorner \sigma \urcorner) \to \theta(\ulcorner \tau \urcorner))$
(3) $T \vdash \theta(\ulcorner \sigma \urcorner) \to \theta(\ulcorner \theta(\ulcorner \sigma \urcorner) \urcorner)$.

Theorem 6.13 (Gödel's Second Incompleteness Theorem)**.** *Let T be a consistent $\mathscr{L}_A$-theory extending PA. T cannot prove its own consistency. That is, for some provability predicate θ and some contradictory statement σ ($0 \neq 0$ works for this) it is not the case that $T \vdash \neg\theta(\ulcorner \sigma \urcorner)$. Thus we must show $T \nvdash \neg\theta(\ulcorner \sigma \urcorner)$.*

Proof. $\neg\theta(x)$ is a formula with one free variable, thus by the Diagonalization theorem we have that there exists some sentence S such that $T \vdash S \leftrightarrow \neg\theta(\ulcorner S \urcorner)$. If $T \vdash S$ then $T \vdash \neg\theta(\ulcorner S \urcorner)$, i.e. $T \nvdash S$, thus $T \nvdash S$ since if it does it doesn't and if it doesn't it doesn't. Now, because σ is a contradiction, we have that $T \vdash \sigma \to S$, and thus $T \vdash \theta(\ulcorner \sigma \to S \urcorner)$ by 1. It therefore follows that $T \vdash \theta(\ulcorner \sigma \urcorner) \to \theta(\ulcorner S \urcorner)$ by 2, which is equivalent to $T \vdash \neg\theta(\ulcorner S \urcorner) \to \neg\theta(\ulcorner \sigma \urcorner)$, and so we have $T \vdash S \to \neg\theta(\ulcorner \sigma \urcorner)$.

Now, observe that $T \vdash S \leftrightarrow \neg\theta(\ulcorner S \urcorner)$ is equivalent to $T \vdash \theta(\ulcorner S \urcorner) \leftrightarrow \neg S$. We have from 1 therefore that

$$T \vdash \theta(\ulcorner \theta(\ulcorner S \urcorner) \to \neg S \urcorner),$$

and so from 2 and the above

$$T \vdash \theta(\ulcorner \theta(\ulcorner S \urcorner) \urcorner) \to \theta(\ulcorner \neg S \urcorner).$$

Now, we have $T \vdash \theta(\ulcorner S \urcorner) \to \theta(\ulcorner \theta(\ulcorner S \urcorner) \urcorner)$ from 3, and from this and the above we have $T \vdash \theta(\ulcorner S \urcorner) \to \theta(\ulcorner \neg S \urcorner)$. Since $T \vdash \theta(\ulcorner S \urcorner) \to \theta(\ulcorner S \urcorner)$, it follows that

$$T \vdash \theta(\ulcorner S \urcorner) \to \theta(\ulcorner S \wedge \neg S \urcorner).$$

Since $S \wedge \neg S$ is a contradiction it is equivalent to our σ, thus $T \vdash \theta(\ulcorner S \urcorner) \to \theta(\ulcorner \sigma \urcorner)$, which is equivalent to $T \vdash \neg\theta(\ulcorner \sigma \urcorner) \to \neg \vdash \theta(\ulcorner S \urcorner)$. Recall that $T \vdash S \leftrightarrow \neg\theta(\ulcorner S \urcorner)$, thus $T \vdash \neg\theta(\ulcorner \sigma \urcorner) \to S$.

We therefore have $T \vdash S \to \neg\theta(\ulcorner \sigma \urcorner)$ and $T \vdash \neg\theta(\ulcorner \sigma \urcorner) \to S$ and thus

$$T \vdash S \leftrightarrow \neg\theta(\ulcorner \sigma \urcorner).$$

Since $T \nvdash S$, we therefore have that

$$T \nvdash \neg\theta(\ulcorner\sigma\urcorner).$$

□

7. Consequences of incompleteness

Theorem 7.1. *Let $M \vDash PA$ and I be a proper initial segment of M, with the properties that:*[16]

- *$x < y \in I \to x \in I$, and*
- *I is closed under the successor function*

If $a \in M$ and $\phi(x, a)$ is a $\mathscr{L}_A$ formula, such that $M \vDash \phi(b, a)$ for all $b \in I$, then there is a $c > I$ such that $M \vDash \forall x \leq c(\phi(x, a))$.

Proof. Suppose not. It follows then that we can define I in the following way. $I = \{x \in M | M \vDash \phi(x, a)\}$. Because models of PA are closed under the successor function we have that $M \vDash \phi(0, a) \wedge \forall x(\phi(x, a) \to \phi(x+1, a))$, thus $M \vDash \forall x \phi(x, a)$, which means that $I = M$, which contradicts our assumption. □

What this theorem says essentially is that we cannot recursively specify a subset of a nonstandard model. If we suppose we can, by for example saying that a certain formula holds only for members of that initial segment, then it's a consequence that it also holds for elements greater than that initial segment, which is a contradiction.

7.1. **Nonstandard models of PA.** We say that all nonstandard models are *end-extensions* of $\mathbb{N}$, and thus that $\mathbb{N}$ is an initial segment of all nonstandard models. This follows from the fact that all models of PA must contain 0 and must be closed under the successor function. Thus $\mathbb{N}$ is an initial segment of all models of PA. What distinguishes the nonstandard models of PA is the end-extensions of $\mathbb{N}$. As it turns out nonstandard models may have nonstandard initial segments as well, i.e. initial segments that aren't $\mathbb{N}$. In fact, there is a famous theorem, Friedman's Theorem,[17] which says that all nonstandard models of PA will have a proper initial segment to which they are isomorphic, though we will not develop the machinery to prove it here. We adopt the following notation for initial segments/end extensions. We write that $I \subset_e M$, just incase I is an initial segment of M and (thus) M is an end extension of I.

Theorem 7.2. *Let T be a recursively axiomatized extension of PA. T has $2^{\aleph_0}$ complete extensions.*

Proof. By Gödel's first incompleteness theorem, there is a Π_1 sentence τ such that neither $T \vdash \tau$ nor $T \vdash \neg\tau$. Thus, $T + \tau$ and $T + \neg\tau$ are both consistent, and so each is an extension of T. In fact, even if we add τ to T, we will still be able to find another Π_1 sentence τ_1, such that neither $T \cup \{\tau\} \vdash \tau$ nor $T \cup \{\tau\} \vdash \neg\tau$. Nevertheless, each model will either satisfy τ or it will not. Let us enumerate the independent Π_1 sentences that we are guaranteed as $\tau_1, \tau_2, \tau_3, \ldots$. Now, we can specify a model according to which of these Π_1 sentences it satisfies. Following Cantor's diagonal argument, let us represent each model M with with an infinite binary sequence $(b_1, b_2, b_3 \ldots)$ where $b_n = 0$ if $M \nvDash \tau_n$ and $b_n = 1$ if $M \vDash \tau_n$.

[16] Typically, we define such I's as proper cuts, but we do not appeal to this idea often enough to warrant introducing another term.

[17] [8] contains a formal statement and proof of Friedman's theorem.

Now suppose that there are countably many models of PA. Then we ought to be able to create a correspondence between our infinite binary sequences and $\mathbb{N}$. So therefore we can enumerate our binary sequences as follows

$$\begin{aligned} s_1 &= (0,0,0,0,0,0,\ldots) \\ s_2 &= (0,1,0,1,0,1,\ldots) \\ s_3 &= (0,0,1,1,0,0,\ldots) \\ s_4 &= (1,1,1,1,1,1,\ldots) \\ &\vdots \end{aligned}$$

where we associate each natural number n with the sequence s_n. Now, we let $s_m^{b_n}$ be the nth term of the mth sequence. In the above, for example $s_1^{b_1} = 0$, and $s_2^{b_2} = 1$ and so on. Now, we define a sequence $s_0 = (b_1, b_2, b_3, \ldots)$ in the following way. We let $b_n \neq s_n^{b_n}$, i.e. if $s_n^{b_n} = 0$, then $b_n = 1$, and if $s_n^{b_n} = 1$, then $b_n = 0$. It follows then that s_0 is not a sequence in our list, precisely because we have defined it to be different from every other element in the list, because it varies from each element in at least one place. In other words, for arbitrary m, $s_0 \neq s_m$ because we have defined s_0 such that $s_0^{b_m} \neq s_m^{b_m}$, and so they must be different sequences. This contradicts our supposition that there are countably many models of PA. Thus there are $2^{\aleph_0}$ countable nonelementary-equivalent models. □

Further, by completeness, each of the above theories have a model, and thus there are $2^{\aleph_0}$ countable nonelementarily-equivalent extensions of PA. By Theorem 3.18 there are nonelementarily equivalent models of all cardinalities. An interesting question that remains is what a countable nonstandard model would look like, especially since there are $2^{\aleph_0}$ of them.

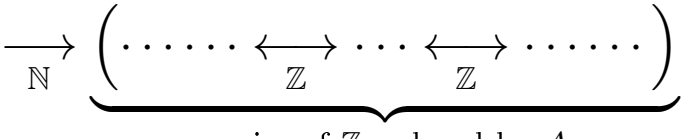
Theorem 7.3. *The order type of countable models of PA is $\mathbb{N} + \mathbb{Z} \cdot \mathbb{Q}$.*

Proof. $PA \restriction <$ is the reduct of the theory of arithmetic to the theory of the order. Recall above at definition 1.2, where various order types are defined. $PA \restriction <$ is a discrete linear order with first, but not last element; PA realizes $Th(DIS)$. Briefly,

- 2 is PA9.
- 4 is PA8.
- 5 is PA10
- 6 follows from PA11, PA14.
- 7 follows from PA11, PA14.
- 8 follows from PA15

While such a claim seems trivial, there are alternative axiomatization schemes for which this claim is not as obvious.

Now, I claim that $Th(DIS)$ has models of the form $\mathbb{N} + \mathbb{Z} \cdot A$, where A orders the copies of $\mathbb{Z}$ that are placed after $\mathbb{N}$. This picture looks like



If A is empty, then we just get $\mathbb{N}$. If A is a single element, then we get $\mathbb{N}+\mathbb{Z}$, etc. I claim such a model satisfies $Th(DIS)$ since $\mathbb{N}$ and $\mathbb{Z}$ are both discrete linear orders, there is no right endpoint, and only $\mathbb{N}$ is bounded on the left. Thus, $\mathbb{N}+\mathbb{Z}\cdot A$ satisfies $Th(DIS)$.

Now, my claim is that if we are looking at a model of PA, then A must be a dense linear order. A satisfies 2, 4, and 5 (in definition 1.2). Thus we must only prove that A realizes 9 and 10, i.e. that A is dense and unbounded.

Suppose then we have a nonstandard model of PA, M. That A has more than one element is given by the following. Suppose, by way of contradiction, that A has only one element: Z_0, and that a is nonstandard (i.e. $a > \mathbb{N}$). $a+a$ then must also be nonstandard and by supposition in Z_0. But because $a+a$ and a are both in Z_0 the distance between them must be finite. This means that $(a+a)-a=b$, for some $b\in\mathbb{N}$, but then $a=b$, which is a contradiction since b is not nonstandard. Thus, A is at least not bounded on the right.

To see that it is also not bounded on the left, we define $[a/b]$ to be the integer component of a divided by b. That is, $[a/b]=c$, just in case $a=b\cdot c+d$, for some $d<b$, which we can make sense of from the axioms of PA. Thus, let us suppose again, by way of contradiction, that A is bounded on the left. Call the smallest element of A, Z_0. Now, let $c=[a/2]$ for nonstandard $a\in Z_0$. Either c is nonstandard or it is not. If $c\in\mathbb{N}$, then we have that either $2\cdot c=a$ or $2\cdot c+1=a$, but since $2,c,1\in\mathbb{N}$, a must also be in $\mathbb{N}$, which is a contradiction. Likewise, suppose c is nonstandard. Because we have that either $2\cdot c=a$ or $2\cdot c+1=a$, it follows from PA12 that $c<a$, meaning that $c\in Z_0$. Now, we rewrite c as $c+c$, but by our above proof then $a\notin Z_0$, which is a contradiction. Thus A is not bounded on the left. Thus A satisfies 10 (unboundedness).

To see that A is dense, we use a similar argument. Suppose that A is not dense. It follows then, that for all $Z_n\in A$, Z_n has an immediate successor Z_{n+1}. Now, suppose that Z_b is the immediate successor of Z_a and that $a\in Z_a$ and $b\in Z_b$ are nonstandard. Without loss of generality, suppose $a<b$, and let $c=b-a$. Now, let $[c/2]=d$, i.e. $d\cdot 2=c$ or $d\cdot 2+1=c$ (we can ignore this second disjunct henceforth, when doing so results in no loss of rigor). Because $c=b-a$, we have $a+c=b$, and it follows that

$$a+2\cdot d=b.$$

Now, either $a+d\in Z_a$ or $a+d\in Z_b$. In the first case we have that $(a+d)-a=f$ for some $f\in\mathbb{N}$, but then $2\cdot f\in\mathbb{N}$, and thus $a+2\cdot d\in Z_a$, which is a contradiction. The second case follows similarly (instead, we notice that $a=b-2\cdot d$). Thus A must be dense, and so A must realize $Th(DLO)$.

Therefore, we conclude that nonstandards models of PA are ordered by $\mathbb{N}+\mathbb{Z}\cdot\mathbb{Q}$. □

Theorem 7.4 (Tennenbaum's Theorem)**.** *Let $M\vDash PA$ be nonstandard. Then M is not recursive. In fact if $M\cong(\mathbb{N},+,\cdot,<,n_1,n_2)$, then neither $+$ nor $\cdot$ is recursive.*

Proof. See [8, 153]. □

Unfortunately a formal proof of this requires developing more machinery than I have room for in this paper. What's important to note though is that while nonstandard models exists, it is extremely difficult to specify them because $+$ and $\cdot$ are not computable on them. In fact, we have an extremely weird picture in that in the past two theorems we have established that there is a continuum of nonstandard models, the order that they must have, but yet we cannot recursively specify *any* of them in first-order logic.

8. Indicator Theory

Next we move onto indicators which will prove central both for proving the independence of Goodstein's theorem and for classification.

Definition 8.1. Let $M \vDash PA$ be nonstandard and T a $\mathscr{L}_A$ theory, and let $Y : M \times M \to M$ be definable in M. We call Y an indicator for T, if whenever $a, b \in M$, $Y(a,b) > \mathbb{N}$ if and only if $\exists I \subset_e M$ such that $a \in I < b$ (meaning $a \in I$ and $\forall x \in I(x < b)$) and $I \vDash T$.

Intuitively $Y(a,b)$ is an indicator for some property that initial segments of models of PA may have if and only if whenever $Y(a,b)$ is nonstandard, there is some initial segment I between a and b, such that I has that property. We define this by saying that I satisfied some formula. We may on occasion relax and say that something is an indicator for a function, or a set of sentences, or a theory, but we should just take this as shorthand for some sentence that represents this function, set of sentences (i.e. the conjunction of the sentences), or that the initial segments satisfies some theory, and so on. In the following example I discuss the indicator of a function, for example. I am explicit in what the corresponding formula would look like, but we need not be so explicit in every case.

Example 8.2. $Y(a,b) = b - a$ is an indicator for the successor function, or to be more formal, it is an indicator for the sentence which says that our universe is closed under the successor and every element save 0 is a successor. That is, $\phi = \forall x(x = 0 \vee \exists y(s(y) = x))$. Suppose $Y(a,b,) > \mathbb{N}$. It follows that $a < b, S(a) < b, S(S(a)) < b, \cdots$. Choose $I = \{x \in M | x < a+n$ for some $n \in \mathbb{N}\}$.

See [8] for more (easy) examples involving indicators.

Definition 8.3. We say an indicator is *well-behaved for* $\mathscr{L}_A$ theory Q *in models of* PA if

(1) $Y(x,y) = z$ is a Σ_1 $\mathscr{L}_A$ formula with only the free variables x,y,z
(2) $PA \vdash \forall x, y \exists! z Y(x,y) = z$
(3) $PA \vdash \forall x, y Y(x,y) \leq y$
(4) $PA \vdash \forall x, y, x', y'((x' \leq x \wedge y \leq y') \to Y(x,y,) \leq Y(x',y'))$
(5) for any nonstandard model M and for $a, b \in M$, $(Ya, b) \geq \mathbb{N}$ if and only if there is a $I \subset_e M$ such that $a \in I < b$ and $I \vDash Q$.

8.1. **All recursive $\mathscr{L}_A$-theories have well-behaved indicators.** With these definitions, we now introduce the most important theorem in this section. As a quick warning, I should say that as with any mathematical subject time and space only make it possible to delve so deep. In particular, in this section certain aspects of my proof I will not be able to cover, particularly as they pertain to satisfaction. While we have proven above that we cannot talk about truth within PA satisfaction gives us a weaker notion of truth that we can talk about with the language of arithmetic. As I said, I avoid going into detail here and where there are gaps in my proof I will point the reader to places in the literature.

Theorem 8.4. *Any consistent recursively axiomatized $\mathscr{L}_A$ theory T extending PA has a well-behaved indicator, $Y(x,y)$ for T in models of PA.*

Before I begin I should say the proof will have the following sort of structure. First, it suffices to show that there exists Y, such that $M \vDash PA$ with $a, b \in M$ and $Y(a,b) > \mathbb{N}$ if and only if there is $I \subset_e M$ with $a \in I < b$ and $I \vDash T$, such that Y is well-behaved. My central claim will be that the following functions as such an indicator. First define $\Psi(x, y, \overline{z})$

to hold just in case x is the Gödel number of a proof from $T + PA$ of a formula whose Gödel number is y. Note that $\overline{z}$ can be a 0-tuple, in which case Ψ can be treated as having only two free variables. It follows then that x is the Gödel number of a proof of a formula whose Gödel number is y only if $PA \vdash \exists \overline{z} \Psi(x, y, \overline{z})$, otherwise $PA \vdash \neg \exists \overline{z} \Psi(x, y, \overline{z})$. Finally, we will need to introduce two ideas very briefly from satisfaction: form_{Δ_0} and Sat_{Δ_0}. I will not be able to give an indepth treatment of these formulas, meaning that I cannot give a rigorous proof of their existence nor can I give a very formal definition of what they would look like. Intuitively they mean the following:

- $\text{form}_{\Delta_0}(\ulcorner x \urcorner)$ holds whenever x is a Δ_0 formula.
- $\text{Sat}_{\Delta_0}(\ulcorner \phi(x) \urcorner, [\overline{a}])$ holds whenever ϕ is a Δ_0 formula such that $PA \vdash \phi(\overline{a})$.

To see how we can define these formulas, we fix a Gödel code and note that we can recursively enumerate all Δ_0 formulas of our language by complexity on the formula. Here $[\overline{x}]$ is understood to be the Gödel code of the sequence formed by $\overline{x}$. We can actually define both of these formulas for Π_n and Σ_n formulas, but we don't here. The reason that their existence might be questionable is that the first says that we can recursively define the set of all Gödel numbers of Δ_0 formulas. The second seems even more questionable in that it says we can recursively define the set of all Gödel numbers of Δ_0 formulas, such that PA proves them. As I say, I do not have the time or space to justify the existence of either or to show what they look like. The reader can find such information in [8, 104ff.].

Now, I put forward what I claim is an indicator for arbitrary T:

(8.1)

$$Y(a,b) = \max\{c \leq b | \forall x < c, y < c, d < c, \overline{z} < c \Big(\big[\Psi(x, y, \overline{z}) \wedge y = \ulcorner \forall v_0 \exists v_1 \phi_d(v_0, v_1) \urcorner \wedge \text{form}_{\Delta_0}(d)\big] \rightarrow \big[\forall u \leq a \ \exists v \leq b \ \text{Sat}_{\Delta_0}(d, [u, v])\big] \Big)\}.$$

Where $\ulcorner \phi_d(x, y) \urcorner = d$, i.e. $\phi_d(x)$ is the formula whose Gödel code is d. Essentially what this equation says is the following. Whenever x is the Gödel number of a proof from $T + PA$ of some Π_2 formula, $\forall v_0 \exists v_1 d$, PA proves the corresponding bounded formula, where a and b supply the bounds for v_0, v_1 respectively; then $Y(a, b)$ will be nonstandard just in case x, $\ulcorner d \urcorner$, and $\ulcorner \forall v_0 \exists v_1 \phi_d(v_0, v_1) \urcorner$ are all less than b in all cases. d are taken to be sentences from the arbitrary theory T under consideration.

Intuitively this indicator says something about the complexity of formulas with relation to a theory extending PA. It says that however constrained we are with respect to formulas we can represent with a Gödel code, we must also be constrained with respect to valid proofs of those formulas. That $c > \mathbb{N}$ means that we can consider all the formulas of the language which have proofs from our theory T extending PA. With this rather bulky Y picked out, we now move onto the proof where we must show that it is in fact an indicator for T and that it is well-behaved.

Proof. Suppose $M \vDash PA$ with $a, b \in M$, $Y(a, b)$ as in equation 8.1 and that $Y(a, b) > \mathbb{N}$. We must show that $Y(a, b)$ satisfies (1) - (5) of definition 8.3.

(1) That x, y, z are the only free variables in the formula $Y(x, y) = z$ follows from our choice of Y. Now we must show that $Y(a, b) = c$ is Σ_1. Notice how $Y(a, b) = c$ is the

conjunction of the following three formulas:

(8.2) $$c \leq b$$

(8.3)
$$\forall x < c, y < c, d < c, \overline{z} < c \Bigg(\big[\Psi(x, y, \overline{z}) \wedge y = \ulcorner \forall v_0 \exists v_1 d \urcorner \wedge \mathrm{form}_{\Delta_0}(d)\big] \rightarrow \big[\forall u \leq a \exists v \leq b \mathrm{Sat}_{\Delta_0}(d, [u, v])\big] \Bigg)$$

(8.4)
$$\Bigg[\neg \forall x < c, y < c, d < c, \overline{z} < c+1 \Bigg(\big[\Psi(x, y, \overline{z}) \wedge y = \ulcorner \forall v_0 \exists v_1 d \urcorner \wedge \mathrm{form}_{\Delta_0}(d)\big] \rightarrow \big[\forall u \leq a \exists v \leq b \mathrm{Sat}_{\Delta_0}(d, [u, v])\big] \Bigg) \Bigg] \vee c + 1 > b.$$

Briefly, 8.2 says that $c \leq b$ which is a stipulation of our Y, 8.3 says that c is as defined, and 8.4 says that either $c + 1 > b$ (disqualifying it by definition) or that $c + 1$ does not work for our definition; i.e. 8.4 says that c is maximal. It follows therefore that $Y(a, b) = c$ is a Σ_1 formula.

(2). We have just shown that $Y(a, b) = c$ is a Σ_1-formula, so c does exist in each case since $Y(a, b) = c$ is a well-formed $\Sigma - 1$-formula. Uniqueness follows, since if we had c_1, c_2 such that $Y(a, b) = c_1$ and $Y(a, b) = c_2$, but $c_1 \neq c_2$, then either c_1 is the maximum, in which case $\neg(Y(a, b) = c_2)$, or c_2 is the maximum, in which case $\neg(Y(a, b) = c_2)$.

(3) follows from how we have defined $Y(a, b)$. That is, we have defined $Y(a, b) = c$ such that it is always the case that $c \leq b$.

(4). For arbitrary $a, a', b, b' \in M$ suppose that $a' \leq a$ and $b \leq b'$. We must show that $Y(a, b) = c \leq Y(a', b') = c'$. Now, since b serves as a bound for c, we have that

$$Y(a, b) \leq Y(a, b').$$

For the same reason $Y(a', b) \geq Y(a, b)$. Thus it follows that

$$Y(a', b') \geq Y(a, b),$$

and so we are done.

(5) Let $M \vDash PA$ with $a, b \in M$, and suppose $Y(a, b) > \mathbb{N}$. We must show that there is an $I \subset_e M$ with $a \in I < b$ such that $I \vDash T$. Unfortunately proving 5 requires an extensive survey of the results of satisfaction, for which we do not have the space. A formal proof can be found in [8, 201ff.]. □

Now, we demonstrate a proof that a formula is a indicator for models of PA. First we must define ω_c, such that $\omega_0 = \omega$, and $\omega_{\alpha+1} = \omega^{\omega_\alpha}$. ω_c is well-defined for $c \in M \vDash PA$ since all $c \in M$, save 0, are the successor of some number; however, it is impossible to recursively specify how it will act for nonstandard c, since by Tennenbaum's theorem multiplication and addition (and by extension the successor function) are not computable in nonstandard models. Nevertheless, we can see that it is recursively definable on the standard model via the following definition.

Define d_c such that:

- $d_0 = 1$
- $d_{a+1} = d^{d_a}$

d_c is recursive since we can find a representation $f(x,y,z)$, such that $M \vDash f(d,a,v)$ if and only if $v = d_a$.

Theorem 8.5. *$Y(a,b) = \max\{c \in M | [a,b]$ is ω_c-large$\}$ is an indicator for models of PA.*

Proof. Suppose that $M \vDash PA$, $a,b \in M$, $I \subset_e M$, $a \in I < b$ with $I \vDash PA$. We must show that $Y(a,b) > \mathbb{N}$. Suppose that $Y(a,b) \not> \mathbb{N}$. Thus, there is some $n \in \mathbb{N}$ such that $Y(a,b) = n$. Thus, $[a,b]$ is $\alpha = \omega^{\omega^{\cdot^{\cdot^{\omega}}}}$ (iterated n times)-large. Thus $\{\alpha\}(a, a+1, \ldots, b-1, b) = 0$. It therefore follows that the sequence $\{\alpha\}(a), \{\alpha\}(a, a+1), \ldots \{\alpha\}(a, a+1, \ldots, b-1), \{\alpha\}(a, a+1, \ldots, b-1, b) = 0$ is finite. It follows, therefore, that the cardinality of $[a,b]$ is in $\mathbb{N}$. Thus $a + j = b$ for some $j \in \mathbb{N}$. But because $I \vDash PA$, it must follow that $b \in I$, a contradiction.

For the other direction, suppose $Y(a,b) > \mathbb{N}$, and now we must show that there is some initial segment $I \subset_e M$ such that $a \in I < b$ and $I \vDash PA$. Suppose not. It therefore follows that *all* proper initial segments of $I \subset_e M$, which realize PA and which contain a must also contain b. That is, the following is something that holds of all initial segments $I \subset_e M$ (regardless of whether they contain a or are proper):

$$I \vDash a \in I \to b \in I.$$

Now of course, that is not yet a first-order statement. However, we can take $a \in I$ to be short for the statement which says that a exists (and simiilarly for $b \in I$). These statements would be defined with respect to the model M, of which I is an initial segment. It is again difficult to specify how this would go for nonstandard elements, but certainly for standard elements, we can write say that 2, for example is an element of our model, by saying that $\exists x(1+1 = x)$. Now, if all initial segments say it though (including the intended model), then not only is it provable from PA, but $b - a = d$ is a statement which we can likewise express in all such initial segments (again, including the intended model). It therefore follows that $b - a$ is not nonstandard. By the above then, $Y(a,b)$ cannot be nonstandard either, which contradicts our assumption. It follows then that we must be able to find $I \subset_e M$, such that $I \vDash PA$ and $a \in I < b$. □

In fact, this indicator is well-behaved, though I do not prove this here. In definition 8.3, 1 and 2 are straightforward. We have just shown 5. 3 and 4 take some work, but should be plausible.

9. Independence Proof

Theorem 9.1 (Independence of Goodstein's Theorem)**.** *$\mathbb{N} \vDash \forall m \exists k (m_k = 0)$, but $PA \nvdash \forall m \exists k (m_k = 0)$.*

That $\mathbb{N} \vDash \forall m \exists k (m_k = 0)$ was the result of the first part of this paper. Thus we must show that we cannot prove Goodstein's theorem from PA. In order to do this, we must develop the following two lemmas.

Lemma 9.2. *We can find $M \vDash PA$ and nonstandard $c \in M$ such that*

$$M \vDash \neg \exists y ([1,y] \text{ is } \omega_c\text{-large}).$$

Proof. $K \vDash PA$ be nonstandard. Let Y be as in theorem 8.5. Now, because $Y(a,b)$ is well-behaved we choose nonstandard $a, b \in K$ such that $Y(a,b) > \mathbb{N}$ and $Y(a,b) < a - 1$. Because, by Theorem 8.5, $Y(a,b)$ is an indicator for PA, this entails that there exists $I \subset_e K$ such that $a \in I < b$. Call this I, M.

We can choose such a K and $a, b \in K$ in the following manner. First we note that $\forall a \forall c \exists b([a, b]$ is ω_c-large) is independent of PA (see [9, 10, 15]). Since it is independent of PA we can find nonstandard models on which it is true and on which it is false. Thus let K be such that $K \vDash \forall a \forall c \exists b([a, b]$ is ω_c-large). Next we choose $a, c \in K$ such that $a - 1 > c > \mathbb{N}$. We are guaranteed a b then such that $Y(a, b) < a - 1$.

such that $a - 1 > Y(a, b) > \mathbb{N}$, by taking nonstandard $c \in K$ and a such that $c < a$. Now if assume we cannot. If particular then, regardless of our choice of K and $a, b \in K$, we cannot have that both $Y(a, b) > \mathbb{N}$ and $Y(a, b) < a - 1$. Therefore, assume that

Now, suppose that $M \nvDash \neg \exists y([1, y]$ is ω_c-large). Let $[1, d]$ be ω_c-large. Recall, however, that $c < a - 1$. It follows then that that there is an initial segment $I \subset M$, such that $I \vDash PA$ and $1 \in I < d$. But this is a contradiction for we can find $[d/n], n \in I$, which would mean that $[d/n] \cdot \notin I$, meaning that I is not closed under multiplication, and thus I cannot be a model of PA. Thus we have our result. □

The reader may notice something strange about our last result and wonder whether we really found a contradiction in our supposition or if this merely exposes a larger contradiction hidden elsewhere. The weird feeling is that it seems like we ought to be able to make the same argument with b, and if not, why not? As it turns out, what this really reveals is something about initial segments. Recall from Theorem 7.3 that nonstandard models of PA have the order type $\mathbb{N} + \mathbb{Z} \cdot \mathbb{Q}$. What we have revealed is that initial segments that pick out models of PA must be Dedekind cuts, in that we can always find more and more "larger" (in terms of the DLO that orders the copies of $\mathbb{Z}$) copies of $\mathbb{Z}$.

Lemma 9.3. *Let $b_0, b_1, b_2, \ldots$ be the Goodstein sequence for m starting at n and let k be minimal such that $b_k = 0$. It follows that $[n - 1, n + k]$ is $o_n(m)$-large.*

Proof. To begin consider the corresponding sequence of ordinals as in Goodstein's theorem above. That is, let

$$\begin{aligned} o_n(m) = o_n(b_0) &= \alpha \\ o_{n+1}(b_1) &= \langle \alpha \rangle(n) \\ o_{n+2}(b_2) &= \langle \alpha \rangle(n, n+1) \\ &\vdots \\ o_{n+k}(b_k) = o_{n+k}(0) = 0 &= \langle \alpha \rangle(n, n+1, \ldots, n+k). \end{aligned}$$

Now, it follows from Theorem 1.29 that

$$\langle \alpha \rangle(n, n+1, \ldots, n+k) \underset{j}{\to} \{\alpha\}(n, n+1, \ldots, n+k),$$

and it follows from this and Theorem 1.28 that

$$\langle \alpha \rangle(n, n+1, \ldots, n+k) \geq \{\alpha\}(n, n+1, \ldots, n+k).$$

Thus, $[n - 1, n + k]$ is α-large if $\{\alpha\}(n - 1, \ldots, n + k) = 0$, but this precisely what we have just shown. It therefore follows that $[n - 1, n + k]$ is $o_n(m)$-large.

□

Now, we turn to the independence proof for Goodstein's theorem.

Proof. Suppose, by way of contradiction, $PA \vdash \forall m \exists k (m_k = 0)$. Now, let us find M and c as dictated by Lemma 9.2. In M, let $d = 2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}$ (with c iterated exponents). We can define this with respect to ω_c, i.e. let $d = f_{\omega_c,\omega}(2)$. If we take d to be written in hereditary base-2 notation, then this is equivalent to saying that $o_2(d) = \omega_c$. Now, by hypothesis, we have that there exists $k \in M$ such that $d_k = 0$. Let k be minimal; it follows from Lemma 9.3 that $[1, 2 + k]$ is $o_2(d)$-large, and thus ω_c-large. This is to say that

$$M \vDash \exists y([1, y] \text{ is } \omega_c\text{-large}),$$

namely, $(2 + k)$, but this contradicts Lemma 9.2. It therefore follows that

$$PA \nvdash \forall m \exists k (m_k = 0),$$

i.e. Goodstein's theorem is independent of PA. □

Conclusion

Thus far, we have gone through some introductory set theory in order to show that Goodstein's theorem is in fact true on some models of PA, in particular the intended model, $\mathbb{N}$. From there we went into more depth, exploring model theory and looking at Gödel's incompleteness proofs, before returning to a proof of the independence of Goodstein's theorem from PA. In the section on indicator theory, we developed some results, which are useful for showing that we can in fact find an indicator for models that realize Goodstein's theorem. In closing, I'd like to say a few words about concerning how it is possible therefore to classify the nonstandard models of PA by Goodstein's theorem. Unfortunately, specifying an exact indicator for Goodstein's theorem will be beyond the scope of this paper.

Classification. First we note that by Theorem 8.1 any recursive $\mathscr{L}_A$ theory has a well-behaved indicator for initial segments satisfying T in PA. It follows then, that we can classify the nonstandard models of PA using such an indicator. That is, let $\mathrm{Mod}_{\mathscr{L}_A}(PA)$ be the class of all countable models of PA, let $\mathcal{G}$ be the theory of Goodstein's theorem, i.e. $\mathrm{Cn}_{\mathscr{L}_A}(\forall m \exists k(m_k = 0)$. $\mathrm{Mod}_{\mathscr{L}_A}(\mathcal{G})$ is the class of all models of PA $\mathfrak{A}$ such that $\mathfrak{A} \vDash \mathcal{G}$.

Theorem 9.4. *$\mathcal{G}$, the theory of Goodstein's theorem, is a recursive $\mathscr{L}_A$ theory.*

Proof. Recall again that Goodstein's theorem says $\forall m \exists k(m_k = 0)$. It suffices to show that $f_g(m) = \exists k(m_k = 0)$ is recursive. f_g is Σ_1, and so we are done. That we can construct a Turing machine to find after how many steps any particular Goodstein sequence terminates is sufficient for this purpose. □

It follows from this previous result and Theorem 8.4 that Goodstein's theorem has a well-behaved indicator $Y_{\mathcal{G}}(a, b)$. We therefore define $\mathrm{Mod}_{\mathscr{L}_A}(\mathcal{G})$ as the class of models $\mathrm{Mod}_{\mathscr{L}_A}(\mathcal{G}) = \{M \in \mathrm{Mod}_{\mathscr{L}_A}(PA) | M$ is an initial segment of some model of PA such that $Y_{\mathcal{G}}(a, b) > \mathbb{N}$ if and only if $M \vDash \mathcal{G}$ and $a \in M < b.\}$.

Acknowledgments

I would like to thank my adviser on this project, Justin Brody, for taking the time to guide me through much of the background necessary for this paper over the course of three semesters, for helping me locate this topic in particular, and for many, many helpful comments on various drafts. I would also like to thank my honors committee: Professors Michael McCooey, Wendell Ressler, and Glenn Ross for joining on in this project and their helpful comments on various drafts of this paper. I would also like to thank all of the wonderful math professors and teachers I have had over the years, without whom this project would not have been possible. Finally, I should like to thank a fellow student and non-mathematician, Alex Yfraimov, for suffering through several attempts of mine at explaining ordinal arithmetic.

References

1. A. Caicedo, *Goodstein's function*, Revista Colombiana de Matemáticas **41** (2007), 381–391.
2. E. Cichon, *A short proof of two recently discovered independence results using recursion theoretic methods.*, Proc. American Math. Soc. **87** (1983), 704–706.
3. Nigel Cutland, *Computability: an introduction to recursive function theory*, Cambridge University Press, 1980.
4. R. Goodstein, *On the restricted ordinal theorem.*, Journal of Symbolic Logic **9** (1944), 33–41.
5. Ronald Graham, *Ramsey theory*, Ney York: Wiley, 1990.
6. Wilfrid Hodges, *Model theory*, Cambridge University Press, 1993.
7. _____, *A shorter model theory*, Cambridge University Press, 1997.
8. Richard Kaye, *Models of peano arithmetic*, Oxford: Clarendon Press, 1991.
9. J. Ketonen and R. Solovay, *Rapidly growing ramsey functions*, Annals of Mathematics **113** (1981), 267–314.
10. L. Kirby and J. Paris, *Accessible independence results for peano arithmetic*, Bull. London Math. Soc. **14** (1982), 285–293.
11. David W. Kueker, *Notes on mathematical logic*, Unpublished Lecture Notes.
12. Kenneth Kunen, *Set theory. an introduction to independence proofs.*, Amsterdam: Elsevier Science Publishers B.V., 1980.
13. Justin Miller, *On the independence of goodstein's theorem*, Thesis, University of Arizona, 2001.
14. J. Paris, *Some independence results for peano arithmetic*, Journal of Symbolic Logic **43** (1978), 725–731.
15. _____, *A hierachy of cuts in models of arithmetic*, Model Theory of Algebra and Arithmetic (1979), 312–337.
16. J. Paris and L. Harrington, *A mathematical incompleteness in peano arithmetic*, Handbook of Mathematical Logic (J. Barwise, ed.), North-Holland Publishing Company, 1977, pp. 1133–42.
17. J. Paris and R. Tavakol, *Goodstein algorithm as a super-transient dynamical system*, Physics Letters A **180** (1993), 83–86.
18. Bruno Poizat, *A course in model theory: An introduction to contemporary mathematical logic*, New York: Springer, 2000.
19. C. Smorynski, *The incompleteness theorems*, Handbook of Mathematical Logic, North-Holland Publishing Company, 1977.

20. J. Spencer, *Large numbers and unprovable theorems*, American Mathematical Monthly **90** (1983), 669–75.
21. Andreas Weiermann, *A classification of rapidly growing ramsey functions*, Proc. American Math. Soc. **132** (2003), 553–561.

Department of Mathematics, Franklin and Marshall College, Lancaster, Pennsylvania, 17603

E-mail address: dkaplan@fandm.edu