CHAPTER 1

Proof theory of arithmetic

The goal of this chapter is to present some in a sense "most complex" proofs that can be done in first-order arithmetic.

The main tool for proving theorems in arithmetic is clearly the induction schema

$$A(0) \to \forall_x(A(x) \to A(Sx)) \to \forall_x A(x).$$

Here $A(x)$ is an arbitrary formula. An equivalent form of this schema is "course-of-values" or cumulative induction

$$\forall_x(\forall_{y<x} A(y) \to A(x)) \to \forall_x A(x).$$

Both schemes refer to the standard ordering of the natural numbers. Now it is tempting to try to strengthen arithmetic by allowing more general induction schemas, e.g., with respect to the lexicographical ordering of $\mathbb{N}\times\mathbb{N}$. More generally, we might pick an arbitrary well-ordering $\prec$ over $\mathbb{N}$ and use the schema of *transfinite induction*:

$$\forall_x(\forall_{y\prec x} A(y) \to A(x)) \to \forall_x A(x).$$

This can be read as follows. Suppose the property $A(x)$ is "progressive", i.e., from the validity of $A(y)$ for all $y \prec x$ we can always conclude that $A(x)$ holds. Then $A(x)$ holds for all x.

One might wonder for which well-orderings this schema of transfinite induction is actually derivable in arithmetic. We will prove here a classic result of Gentzen (1943) which in a sense answers this question completely. However, in order to state the result we have to be more explicit about the well-orderings used. This is done in the next section.

1.1. Ordinals below ε_0

We want to discuss the derivability of initial cases of transfinite induction in arithmetical systems. In order to do that we shall need some knowledge and notations for ordinals. We do not want to assume set theory here; hence we introduce a certain initial segment of the ordinals (the ordinals $< \varepsilon_0$) in a formal, combinatorial way, i.e., via ordinal notations. Our treatment is

based on the Cantor normal form for ordinals; cf. Bachmann (1955). We also introduce some elementary relations and operations for such ordinal notations, which will be used later. For brevity we from now on use the word "ordinal" instead of "ordinal notation".

1.1.1. Basic definitions. We define the two notions

- α is an ordinal
- $\alpha < \beta$ for ordinals α, β

simultaneously by induction:

(1) If $\alpha_m, \ldots, \alpha_0$ are ordinals, $m \geq -1$ and $\alpha_m \geq \cdots \geq \alpha_0$ (where $\alpha \geq \beta$ means $\alpha > \beta$ or $\alpha = \beta$), then

$$\omega^{\alpha_m} + \cdots + \omega^{\alpha_0}$$

is an ordinal. Note that the empty sum denoted by 0 is allowed.

(2) If $\omega^{\alpha_m} + \cdots + \omega^{\alpha_0}$ and $\omega^{\beta_n} + \cdots + \omega^{\beta_0}$ are ordinals, then

$$\omega^{\alpha_m} + \cdots + \omega^{\alpha_0} < \omega^{\beta_n} + \cdots + \omega^{\beta_0}$$

iff there is an $i \geq 0$ such that $\alpha_{m-i} < \beta_{n-i}$, $\alpha_{m-i+1} = \beta_{n-i+1}$, $\ldots$, $\alpha_m = \beta_n$, or else $m < n$ and $\alpha_m = \beta_n$, $\ldots$, $\alpha_0 = \beta_{n-m}$.

For proofs by induction on ordinals it is convenient to introduce the notion of *level* of an ordinal α by the stipulations (a) if α is the empty sum 0, $\mathrm{lev}(\alpha) = 0$, and (b) if $\alpha = \omega^{\alpha_m} + \ldots + \omega^{\alpha_0}$ with $\alpha_m \geq \ldots \geq \alpha_0$, then $\mathrm{lev}(\alpha) = \mathrm{lev}(\alpha_m) + 1$.

For ordinals of level $k+1$ we have $\omega_k \leq \alpha < \omega_{k+1}$, where $\omega_0 = 0$, $\omega_1 = \omega$, $\omega_{k+1} = \omega^{\omega_k}$.

We shall use the notation 1 for ω^0, k for $\omega^0 + \cdots + \omega^0$ with k copies of ω^0 and $\omega^\alpha k$ for $\omega^\alpha + \cdots + \omega^\alpha$ again with k copies of ω^α.

It is easy to see (by induction on the levels) that $<$ is a linear order with 0 being the smallest element.

We define addition for ordinals by

$$\omega^{\alpha_m} + \cdots + \omega^{\alpha_0} + \omega^{\beta_n} + \cdots + \omega^{\beta_0} := \omega^{\alpha_m} + \cdots + \omega^{\alpha_i} + \omega^{\beta_n} + \cdots + \omega^{\beta_0}$$

where i is minimal such that $\alpha_i \geq \beta_n$.

It is easy to see that $+$ is an associative operation which is strictly monotonic in the second argument and weakly monotonic in the first argument. Note that $+$ is not commutative: $1 + \omega = \omega \neq \omega + 1$.

There is also a commutative version on addition. The *natural* (or Hessenberg) sum of two ordinals is defined by

$$(\omega^{\alpha_m} + \cdots + \omega^{\alpha_0}) \# (\omega^{\beta_n} + \cdots + \omega^{\beta_0}) := \omega^{\gamma_{m+n}} + \cdots + \omega^{\gamma_0},$$

where $\gamma_{m+n}, \dots, \gamma_0$ is a decreasing permutation of $\alpha_m, \dots, \alpha_0, \beta_n, \dots, \beta_0$. It is easy to see that $\#$ is associative, commutative and strictly monotonic in both arguments.

We will also need to know how ordinals of the form $\beta + \omega^\alpha$ can be approximated from below. First note that

$$\delta < \alpha \to \beta + \omega^\delta k < \beta + \omega^\alpha.$$

Furthermore, for any $\gamma < \beta + \omega^\alpha$ we can find a $\delta < \alpha$ and a k such that

$$\gamma < \beta + \omega^\delta k.$$

1.1.2. Enumerating ordinals. In order to work with ordinals in a purely arithmetical system we set up some effective bijection between our ordinals $< \varepsilon_0$ and non-negative integers (i.e., a Gödel numbering). For its definition it is useful to refer to ordinals in the form

$$\omega^{\alpha_m} k_m + \dots + \omega^{\alpha_0} k_0 \quad \text{with } \alpha_m > \dots > \alpha_0 \text{ and } k_i \neq 0 \ (m \geq -1).$$

(By convention, $m = -1$ corresponds to the empty sum.)

For every ordinal α we define its Gödel number $\ulcorner \alpha \urcorner$ inductively by

$$\ulcorner \omega^{\alpha_m} k_m + \dots + \omega^{\alpha_0} k_0 \urcorner := \Big(\prod_{i \leq m} p_{\ulcorner \alpha_i \urcorner}^{k_i} \Big) - 1,$$

where p_n is the n-th prime number starting with $p_0 := 2$. For every non-negative integer x we define its corresponding ordinal notation $\mathrm{o}(x)$ inductively by

$$\mathrm{o}\Big((\prod_{i \leq l} p_i^{q_i}) - 1 \Big) := \sum_{i \leq l} \omega^{\mathrm{o}(i)} q_i,$$

where the sum is to be understood as the natural sum.

LEMMA. (a) $\mathrm{o}(\ulcorner \alpha \urcorner) = \alpha$,
(b) $\ulcorner \mathrm{o}(x) \urcorner = x$.

PROOF. This can be proved easily by induction. □

Hence we have a simple bijection between ordinals and non-negative integers. Using this bijection we can transfer our relations and operations on ordinals to computable relations and operations on non-negative integers. We use the following abbreviations.

$$\begin{aligned} x \prec y &:= \mathrm{o}(x) < \mathrm{o}(y), \\ \omega^x &:= \ulcorner \omega^{\mathrm{o}(x)} \urcorner, \\ x \oplus y &:= \ulcorner \mathrm{o}(x) + \mathrm{o}(y) \urcorner, \end{aligned}$$

$$\begin{aligned} xk &:= \ulcorner \mathrm{o}(x)k \urcorner, \\ \omega_k &:= \ulcorner \omega_k \urcorner. \end{aligned}$$

We leave it to the reader to verify that $\prec$, $\lambda_x \omega^x$, $\lambda_{x,y}(x \oplus y)$, $\lambda_{x,k}(xk)$ and $\lambda_k \ulcorner \omega_k \urcorner$ are all elementary.

1.2. Provability of initial cases of transfinite induction

We now derive initial cases of the principle of transfinite induction in arithmetic, i.e., of

$$\forall_x(\forall_{y \prec x} Py \to Px) \to \forall_{x \prec a} Px$$

for some number a and a predicate symbol P, where $\prec$ is the standard order of order type ε_0 defined in the preceding section. One can show that our results here are optimal in the sense that for the full system of ordinals $< \varepsilon_0$ the principle

$$\forall_x(\forall_{y \prec x} Py \to Px) \to \forall_x Px$$

of transfinite induction is underivable. All these results are due to Gentzen (1943).

1.2.1. Arithmetical systems. By an *arithmetical system* **Z** we mean a theory based on minimal logic in the $\forall\to$-language (including equality axioms), with the following properties. The language of **Z** consists of a fixed (possibly countably infinite) supply of function and relation constants which are assumed to denote fixed functions and relations on the non-negative integers for which a computation procedure is known. Among the function constants there must be a constant S for the successor function and 0 for (the 0-place function) zero. Among the relation constants there must be a constant $=$ for equality and $\prec$ for the ordering of type ε_0 of the natural numbers, as introduced in section 1.1. In order to formulate the general principle of transfinite induction we also assume that a unary relation symbol P is present, which acts like a free set variable.

Terms are built up from object variables x, y, z by means of $f(t_1, \ldots, t_m)$, where f is a function constant. We identify closed terms which have the same value; this is a convenient way to express in our formal systems the assumption that for each function constant a computation procedure is known. Terms of the form $\mathrm{S}(\mathrm{S}(\ldots \mathrm{S0} \ldots))$ are called *numerals*. We use the notation $\mathrm{S}^n 0$ or $\overline{n}$ or (only in this chapter) even n for them. *Formulas* are built up from atomic formulas $R(t_1, \ldots, t_m)$, with R a relation constant or a relation symbol, by means of $A \to B$ and $\forall_x A$.

The *axioms* of **Z** include compatibility of equality

$$x = y \to A(x) \to A(y),$$

the *Peano axioms*, i.e., the universal closures of

$$\mathrm{S}x = \mathrm{S}y \to x = y, \tag{1.1}$$

$$\mathrm{S}x = 0 \to A, \tag{1.2}$$

$$A(0) \to \forall_x(A(x) \to A(\mathrm{S}x)) \to \forall_x A(x), \tag{1.3}$$

with $A(x)$ an arbitrary formula. We express our assumption that for every relation constant R a decision procedure is known by adding the axiom $R\vec{n}$ whenever $R\vec{n}$ is true. Concerning $\prec$ we require as axioms irreflexivity and transitivity for $\prec$

$$x \prec x \to A,$$

$$x \prec y \to y \prec z \to x \prec z$$

and also – following Schütte – the universal closures of

$$x \prec 0 \to A, \tag{1.4}$$

$$z \prec y \oplus \omega^0 \to (z \prec y \to A) \to (z = y \to A) \to A, \tag{1.5}$$

$$x \oplus 0 = x, \tag{1.6}$$

$$x \oplus (y \oplus z) = (x \oplus y) \oplus z, \tag{1.7}$$

$$0 \oplus x = x, \tag{1.8}$$

$$\omega^x 0 = 0, \tag{1.9}$$

$$\omega^x(\mathrm{S}y) = \omega^x y \oplus \omega^x, \tag{1.10}$$

$$z \prec y \oplus \omega^{\mathrm{S}x} \to z \prec y \oplus \omega^{e(x,y,z)} m(x,y,z), \tag{1.11}$$

$$z \prec y \oplus \omega^{\mathrm{S}x} \to e(x,y,z) \prec \mathrm{S}x, \tag{1.12}$$

where $\oplus$, $\lambda_{x,y}(\omega^x y)$, e and m denote the appropriate function constants and A is any formula. (The reader should check that e, m can be taken to be elementary.) These axioms are formal counterparts to the properties of the ordinal notations observed in the preceding section.

1.2.2. Gentzen's proof.

THEOREM (Provable initial cases of transfinite induction in **Z**). *Transfinite induction up to ω_n, i.e., for arbitrary $A(x)$ the formula*

$$\forall_x(\forall_{y \prec x} A(y) \to A(x)) \to \forall_{x \prec \omega_n} A(x),$$

is derivable in **Z**.

PROOF. To every formula $A(x)$ we assign a formula $A^+(x)$ (with respect to a fixed variable x) by

$$A^+(x) := \forall_y(\forall_{z\prec y} A(z) \to \forall_{z\prec y\oplus\omega^x} A(z)).$$

We first show

$$\text{If } A(x) \text{ is progressive, then } A^+(x) \text{ is progressive,}$$

where "$B(x)$ is *progressive*" means $\forall_x(\forall_{y\prec x}B(y) \to B(x))$. So assume that $A(x)$ is progressive and

$$\forall_{y\prec x}A^+(y). \tag{1.13}$$

We have to show $A^+(x)$. So assume further

$$\forall_{z\prec y}A(z) \tag{1.14}$$

and $z \prec y \oplus \omega^x$. We have to show $A(z)$.

Case $x = 0$. Then $z \prec y \oplus \omega^0$. By (1.5) it suffices to derive $A(z)$ from $z \prec y$ as well as from $z = y$. If $z \prec y$, then $A(z)$ follows from (1.14), and if $z = y$, then $A(z)$ follows from (1.14) and the progressiveness of $A(x)$.

Case $\mathrm{S}x$. From $z \prec y \oplus \omega^{\mathrm{S}x}$ we obtain $z \prec y \oplus \omega^{e(x,y,z)}m(x,y,z)$ by (1.11) and $e(x,y,z) \prec \mathrm{S}x$ by (1.12). From (1.13) we obtain $A^+(e(x,y,z))$. By the definition of $A^+(x)$ we get

$$\forall_{u\prec y\oplus\omega^{e(x,y,z)}v}A(u) \to \forall_{u\prec(y\oplus\omega^{e(x,y,z)}v)\oplus\omega^{e(x,y,z)}}A(u)$$

and hence, using (1.7) and (1.10)

$$\forall_{u\prec y\oplus\omega^{e(x,y,z)}v}A(u) \to \forall_{u\prec y\oplus\omega^{e(x,y,z)}(\mathrm{S}v)}A(u).$$

Also from (1.14) and (1.9), (1.6) we obtain

$$\forall_{u\prec y\oplus\omega^{e(x,y,z)}0}A(u).$$

Using an appropriate instance of the induction schema we can conclude

$$\forall_{u\prec y\oplus\omega^{e(x,y,z)}m(x,y,z)}A(u)$$

and hence $A(z)$.

We now show, by induction on n, how for an arbitrary formula $A(x)$ we can obtain a derivation of

$$\forall_x(\forall_{y\prec x}A(y) \to A(x)) \to \forall_{x\prec\omega_n}A(x).$$

So assume the left hand side, i.e., assume that $A(x)$ is progressive.

Case 0. Then $x \prec \omega^0$ and hence $x \prec 0 \oplus \omega^0$ by (1.8). By (1.5) it suffices to derive $A(x)$ from $x \prec 0$ as well as from $x = 0$. Now $x \prec 0 \to A(x)$ holds by (1.4), and $A(0)$ then follows from the progressiveness of $A(x)$.

Case $n+1$. Since $A(x)$ is progressive, by what we have shown above $A^+(x)$ is also progressive. Applying the induction hypothesis to $A^+(x)$ yields $\forall_{x\prec\omega_n} A^+(x)$, and hence $A^+(\omega_n)$ by the progressiveness of $A^+(x)$. Now the definition of $A^+(x)$ (together with (1.4) and (1.8)) yields $\forall_{z\prec\omega^{\omega_n}} A(z)$. □

Note that in the induction step of this proof we have derived transfinite induction up to ω_{n+1} for $A(x)$ from transfinite induction up to ω_n for a formula of level higher than the level of $A(x)$. The *level* of a formula A is defined by

$$\begin{aligned}
\mathrm{lev}(R\vec{t}\,) &:= 0,\\
\mathrm{lev}(A\to B) &:= \max(\mathrm{lev}(A)+1, \mathrm{lev}(B)),\\
\mathrm{lev}(\forall_x A) &:= \max(1, \mathrm{lev}(A)).
\end{aligned}$$