Annika Kanckos

A Possible and Necessary Consistency Proof

Academic dissertation to be publicly discussed, by due permission of the Faculty of Arts at the University of Helsinki in auditorium XII (Main Building, Unioninkatu 34), on the 21st of May, 2011 at 10 o'clock.

Filosofisia tutkimuksia Helsingin yliopistosta
Filosofiska studier från Helsingfors universitet
Philosophical Studies from the University of Helsinki

Publishers:
Theoretical Philosophy and Philosophy (in Swedish)
Department of Philosophy, History, Culture and Art Studies

Social and Moral Philosophy
Department of Political and Economic Studies

P.O. box 24 (Unioninkatu 40 A)
00014 University of Helsinki
Finland

Annika Kanckos

A Possible and Necessary

Consistency Proof

ISBN 978-952-10-6945-1 (paperback)
ISBN 978-952-10-6946-8 (PDF)
ISSN 1458-8331
Helsinki 2011
Helsinki University Print

Contents

Acknowledgments

I am deeply indebted to my advisor Professor Jan von Plato who has encouraged my work and made me a member of his research project at the Academy of Finland. I would also like to thank the logic groups at the Department of Mathematics and the Department of Philosophy at the University of Helsinki.

Finally, I would like to thank my family for their support during the process of writing this thesis.

Helsinki, April 2011 Annika Kanckos

Content of the thesis

The aim of this thesis is to examine the consistency proofs for arithmetic by Gerhard Gentzen from different angles. The first chapter is an introduction to how the problem of consistency proofs relates to the foundational debate of the 20th century. This paper was presented at the Paris-Nancy PhilMath workshop in 2009 and part of this paper has appeared in the Logica Yearbook 2009 [13]. The second chapter examines the different proofs from a more technical aspect. The subject of the third chapter is the extension of logical systems with mathematical rules, a method which will be used throughout the thesis. The fourth chapter gives a consistency proof for an intuitionistic sequent calculus. The result is based on Takeuti's proof in [31]. The proof includes a cut elimination theorem for the calculus and a syntactical study of the purely arithmetical part of the system, resulting in a consistency proof for purely arithmetical derivations that do not contain compound formulas or the induction rule. This chapter will appear in a Gentzen centenary volume. The fifth chapter consists of a consistency proof for Heyting arithmetic in natural deduction. The proof is based on a normalization proof by Howard and assigns vectors to derivations, which are then interpreted as ordinals. The proof appears in Math. Log. Quart. [14].

This thesis is based on a Licensiate thesis approved by the Department of Mathematics and Statistics at the University of Helsinki in 2010.

Chapter 1

Looking for consistency

1.1 The problem

In 1900 Hilbert presented a list of 23 open problems in different fields of mathematics. The second of these problems was to find a consistency proof for the arithmetic of real numbers, that is, analysis. The statement of the problem included the task of presenting an axiomatization, in which all axioms are independent. But according to Hilbert the most important question was to prove that the axioms are not contradictory, that is, that a definite number of logical steps from the axioms cannot lead to contradictory results.

The methods employed in the sought proof should be finitistic, and it is therefore not sufficient to prove the consistency in a stronger theory. The finitistic methods used should not presuppose a completed infinity, but instead rely on constructive methods that are directly accessible even to the man on the street.

The axioms of primitive recursive arithmetic (PRA) are the defining equations of primitive recursive functions and the system consists of a propositional calculus with induction on quantifier-free formulas. PRA is a weaker theory than Peano arithmetic (PA) and it is generally included in, and often identified with, finitistic logic, because unbounded quantification over the domain of natural numbers is not allowed.

Gödel's second incompleteness theorem implies that the methods of PA or PRA are not sufficient for proving even the consistency of PA. Therefore, there is no solution to Hilbert's problem if the methods are restricted to PRA. The consequences of the result of Gödel cannot be questioned with respect to Hilbert's second problem. It proves that the problem is unsolvable and that Hilbert's programme cannot be carried out in full, but a partial realization is possible.

1.2 From the problem to Hilbert's programme

Hilbert's programme was initiated as a consequence of the foundational debate at the turn of the century. During the early 1900's Hilbert developed his views on the foundations of mathematics and presented his views in a succession of papers. He proposed a method for solving the foundational crisis that had emerged after the paradoxes of set theory. In 1921 the aims of Hilbert's programme consisted of formalizing all mathematical theories, and providing 'finitary' consistency proofs for them. Furthermore, the programme included that the questions of mutual independence and completeness of the axioms of the theory were to be answered and possibly a decision method found for the theory.

A narrow description of the programme requires finitary proofs of the formal consistency of formal arithmetic. In broader terms the program asserts that infinitary notions should only be used as abbreviations. The aim of the programme is to give an understanding of existing proofs from a finitistic view. Hilbert's opinions on infinitistic and in particular set theoretic notions in mathematics is that because they are more or less abbreviations for other concepts, they should be possible to eliminate from proofs. Although the programme does not explicitly mention an elimination procedure, we believe that searching for a procedure that eliminates all reference to infinitistic concepts is essential. Such a procedure should not solely consist of restricting the methods of proof, but it should apply to all methods of proof. Furthermore, the process of eliminating infinitary concepts could be regarded as a useful scientific tool, as the elimination at times could increase our conviction in the theorem proved.[1]

Hilbert's main point in his second problem was that it should be possible to make the finiteness of all proofs explicit. This idea developed into his programme. However, in his statement of the problem he had left open the question of exactly which axioms were to be considered and which modes of inference were to be proven free from contradiction.

The consistency of a theory may be proven either semantically or syntactically. A semantical proof consists of proving that the theory is satisfiable by a model. An alternative to Hilbert's programme is to use infinite models and establish not only consistency, but soundness of the axioms for the intended meaning. This means too that the inference rules prove only formulas that are valid with respect to the system's semantics or that the rules 'preserve truth'. A semantical proof is by no means finite if it deals with infinite domains. This means that the consistency of arithmetic, as referred to in Hilbert's programme, should be established without the use of infinite models. A syntactical consistency proof, on the other hand, requires only proof theoretical means, as it concerns provability. Completeness of predicate calculus, however, implies that the semantical and syntactical notions of consistency are equivalent. The difference in these approaches is noticeable in the methods of proof which are accepted. The proof theoretical approach is namely constructive.

1.3 Gentzen's work related to Hilbert's programme

By Gödel's incompleteness theorem from 1931 it was shown that no formalization of elementary arithmetic can be complete and that it is impossible to find a finite consistency proof for PA in the sense that Hilbert's programme required. Therefore, the methods that are proper to the theory, the consistency of which we are proving, do not suffice when proving consistency of the theory. To produce a consistency proof, the consistency of the methods used need to be presupposed. That is, no absolute consistency proof exists and all

[1] Kreisel 1976 [16], p. 98.

proofs merely reduce the question of consistency to that of the other theory used.

Kreisel notes that even if the consistency of the theory ensures the existence of some concept satisfying all the theorems of this system, it does not ensure that the particular concept (of natural numbers) for which axioms of arithmetic is intended, satisfies those theorems.[2] The incompleteness theorem means that in any formal theory, there are always true number-theoretical sentences that are not provable within the theory. Another description of the result is that "sentences can always be found, the proofs of which again always require new modes of inference". ([6, p.357]) This reveals a weakness in the axiomatic method, implying that the consistency proofs must be extended whenever the proof means are extended. In 1937 Gentzen however considers the extensions not relevant in practice, because at that time no Gödel sentence of practical significance had been revealed, except for the sentence expressing consistency. In 1943 he would himself accomplish such an extension, by proving that the principle of transfinite induction up to ϵ_0 is independent of PA.

With broader methods it is still possible to produce a proof, though the finiteness of these methods is debated. Gödel's dialectica interpretation as well as Gentzen's consistency proofs for PA can be seen as a realization of Hilbert's programme, if it is extended to include constructive methods.

1.4 Gentzen's proofs

The earliest proofs of the consistency of Peano arithmetic were presented by Gentzen, who worked out a total of four proofs that were published between 1936 and 1974. Neither Bernays nor Gödel were satisfied with Gentzen's first consistency proof, which is shown in correspondence from Gentzen to Bernays in the fall of 1935.[3] The proof was withdrawn from publication due to the criticism by Bernays for implicit use of the fan theorem, although this assessment was later retracted [4]. However, a galley proof of the article was preserved and

[2]Kreisel 1976 [16], p. 97.
[3]von Plato 2007 [25], p. 392.
[4]Bernays 1970 [1].

excerpts were published posthumously in English translation [9], as well as unabridged in the German original [10].

König's lemma, which states that a finitely branching tree with an infinity of nodes has an infinite branch, is not constructively valid. The contrapositive of König's lemma, called the fan theorem, is however constructive. It states that if all branches of a tree are finite, then the whole tree is also finite.

However, it has been noted by Kreisel in 1987 that this principle is not sufficient for proving the consistency of Peano arithmetic. The principle that was implicitly used to prove termination is bar recursion. Bar recursion is essentially recursion on well-founded trees, it is the contrapositive of a similar classical principle for infinitely branching trees. Gentzen, who had already thought of the objections, reworked his proof and instead relied on the principle of transfinite induction. The result was the published second proof [5], which is contains an ordinal assignment and a constructive proof of the principle of transfinite induction up to the ordinal ϵ_0.

The third proof in sequent calculus was published in 1938. By Gentzen's fourth proof from 1943, it is proven that the consistency of PA can be proven relative to a theory if and only if the proof theoretical ordinal is greater than ϵ_0.

1.5 The principle of transfinite induction

The principle of transfinite induction can be expressed in the following way: Let $P(\beta)$ be a property defined for all ordinals β and let α be an arbitrary ordinal. Then if we assume that for all $\beta < \alpha$, $P(\beta)$ holds, and from this it follows that $P(\alpha)$ holds, then by the principle the property holds for all ordinals.

Gentzen's use of the principle was restricted to primitive recursive predicates. The primitive recursive predicates, $P(n)$, can be verified for an arbitrary number, n, by a bounded computation.

In his proof from 1943 he represented transfinite induction up to ϵ_0 as an arithmetical formula and showed that it is not provable in Peano arithmetic, but that any weaker induction principle is provable. In the proof the natural numbers are extended by what are called constructive ordinals. The induction principle is also extended into a transfinite induction principle.

Schütte and Schwichtenberg [30] note that “the transfinite induction certainly transcended the finite standpoint, as by Gödel is necessary, but it proceeds in a completely constructive way, so that the proof of Gentzen is seen as a testimonial for pure number theory in the sense of the extended Hilbert Programme...”

In general, a set-theoretical proof of the principle of transfinite induction is not acceptable if the methods are to be considered reliable from a constructive point of view. Instead Gentzen proves that each ordinal up to ϵ_0 is accessible. Accessibility means that all descending chains of ordinals are finite, or that the ordinals are well-ordered. This principle is used in order to prove termination in a finite number of steps of the reduction procedure described in Gentzen's proof.

1.6 Crisis and paradoxes

The importance of consistency proofs was debated due to the foundational crisis at the time when Gentzen published his proofs. Gentzen points out that despite the efforts to find a solution for the paradoxes of set theory, that is, to pinpoint the fallacy of the reasoning that leads to antinomies, a clear solution should not be expected. The flaw in the reasoning cannot definitely be pointed at. Gentzen however follows the proponents of intuitionism by claiming that the antinomies of set theory have their origin in the liberal use of the concept of infinity. He claims that "we can only say definitely that the materialisation of the antinomies is connected with the *concept of infinity*, because a purely finite mathematics, as far as anyone can judge, no contradictions can arise, provided the mathematics is correctly constructed." ([6, p. 353])

One simple solution is to draw a clear line between permissible and impermissible modes of inference, thereby blocking the undesired inferences that lead to antinomies. This method has been employed, for example, in axiomatic set theory, by restricting the comprehension schema. However, according to Gentzen [6, p. 353] this solution gives rather arbitrary restrictions if the source of the antinomies has not been properly identified. Furthermore, new antinomies may in the future prove to be derivable with the allowed inferences. Another solution to the paradox is the introduction of a type structure, which is also noted by Gentzen.[5]

In Gentzen's opinion Russell's paradox reveals a fault in the logical inferences involved. He opposes impredicative definitions and regards only constructive definitions as valid. New sets should be defined on the basis of already formed sets, because it is illicit to define an object by means of a totality and then to regard it as belonging to that totality, so that it contributes to its own definition. The definition of a set of all sets is circular, as this set is defined and then concluded to belong to itself.[6] A problem that emerges from invalidating impredicative reasoning is that this form of reasoning is also used in analysis, in the proofs of basic theorems, such as the intermediate value theorem. The definition of the intermediate point is problematic, because the point is included in the intervals defined in the proof of the theorem. This means that the point is defined by referring to the totality of reals and is then concluded to belong to this totality.

A radical standpoint is taken by the intuitionists who do not consider the arguments used in classical analysis to be valid, because the law of trichotomy is not true on the intuitionistic continuum. Thus, they reject the means of proof that allows a division of the reals into two intervals. The intuitionist standpoint is not only taken to avoid possible antinomies, but because the classical statements are considered meaningless.

1.7 The intuitionistic method

Hilbert's programme had not been abandoned by Gentzen. After the finitistic methods had proven to be insufficient for proving the consistency of arithmetic, Gentzen continued the search for a proof. Gentzen's work explores the consequences of the finitistic view of formalist mathematics as stated by Hilbert. In order to prove consistency Gentzen followed the general aims of Hilbert's program, which were to prove consistency "by means of inference that are completely

[5]Gentzen 1969 [9], p. 214.
[6]Gentzen 1969 [9], p. 134.

unimpeachable ('finitist' forms of inference)."[7]

In 1933 the Gödel-Gentzen negative translation showed that classical arithmetic can be reduced to intuitionistic arithmetic, implying that the constructive methods go beyond finitistic reasoning. Therefore, Hilbert's programme could be continued if it was modified to use the broader constructive methods. In the light of this, the negative translation was considered the first consistency proof for arithmetic.[8]

Gentzen's aim was to prove the consistency of classical mathematics, in the first place arithmetic and then analysis, by extending the methods to constructive or intuitionistically acceptable methods.[9] The constructive method used as a foundation for the consistency proofs is similar to, although somewhat broader than, Hilbert's finitistic standpoint. In Gentzen's opinion it provides a secure foundation because it employs the concept of possible infinity, not an actual infinity. The actual infinity is identified as a doubtful element in the methods of proof. The constructive concept of infinity, on the other hand, is not included in the framework of elementary number theory, but is conjectured to be extensible beyond any formal theory.

These methods include a constructive interpretation of quantification over the infinite domain of natural numbers. If the numbers are substituted one by one in a formula that has been universally quantified, then the result is a true formula. An existentially quantified formula, on the other hand, means that a witness to the formula has been found. Even so, Gentzen thinks that some methods encountered in his proof "give cause for concern" from the finitistic standpoint, in particular the principle of transfinite induction.[10] Whether the proof can be regarded as finitistic depends on if the principle of transfinite induction can be accepted as a finitistic method.

1.8 How consistency proofs are possible

In a lecture from 1937 ([6, p. 355]) Gentzen characterized Hilbert's programme as a way to reduce the metamathematical presupposi-

[7] Gentzen, 1969, [9], p.135.
[8] von Plato 2007 [25], p. 392.
[9] Gentzen 2007 [6], and von Plato 2007 [25], p. 383.
[10] Gentzen 1969 [9], p.136.

tions and he followed Hilbert in this respect. In a letter from 1932 Gentzen states that through the formalisation of logical deduction, the task of producing a consistency proof becomes a purely mathematical problem.[11] The purpose was to eliminate all philosophical problems, or at least separate them from scientific practice. This was at a time when he worked on extending the consistency proof for arithmetic to include the rule of induction.

In his paper from 1936 Gentzen clearly states that the purpose of the proof is to reduce the question of consistency of arithmetic to certain general and fundamental principles. He concludes in [5] that a consistency proof is still possible and meaningful if the methods used are more reliable, even if not proper to elementary number theory. It is possible to reduce some parts of arithmetic to other parts, e.g. arithmetic of complex numbers to that of real numbers. But Gentzen concludes that there remains the task of proving the consistency of elementary number theory. His main concern is with the proving of the consistency of the logical reasoning used when proving statements about the natural numbers. This means that the consistency of the system of axioms or the basic relations between numbers is not what he is aiming to prove, because it is the reasoning employed that may produce antinomies. Gentzen discusses to what extent it is possible to carry out a consistency proof and claims that it is both necessary and possible to produce a proof, due to the paradoxes that had emerged in other areas of mathematics.

Kreisel, on the other hand, expresses doubt in the significance of consistency proofs. The question is whether the proofs have epistemological value. "If ordinary mathematics is really so reliable [as Hilbert emphasized] then the value of Hilbert's consistency program cannot possibly consist in increasing significantly the *degree of reliability* (of ordinary mathematics)."[12] It can be noted that Tarski regarded Gentzen's proof as an interesting metamathematical result, but he did not think that the proof made the consistency of arithmetic more evident than by epsilon.[13]

According to Kreisel the analysis of the significance of a consis-

[11] Menzler-Trott 2007 [19], p. 29.
[12] Kreisel 1971 [15], p. 240.
[13] Menzler-Trott 2007 [19], p. 81.

tency proof can be more complicated than the proof itself. The point that Kreisel makes is a criticism of Hilbert's programme; that consistency proofs are sought as a reduction of complex concepts to simpler ones. The elimination of problematic notions is contrary to our intellectual experience. Our experience instead consists of eliminating concepts in practice, not just in theory, or of giving independent meaning to concepts and steps which, originally, occur as mere technical auxiliaries. [14]

1.9 A partial solution

The argument against a relative consistency proof is that it provides only a limited support for the consistency of arithmetic. In the case of the negative translation it proves that an inconsistency cannot stem from the principle of indirect proof. But Gentzen's proofs, however, claim to be absolute in another sense. Their purpose is to provide a secure foundation, taken into consideration the limitations that Gödel's theorem impose. In particular, the additional principle of transfinite induction used in Gentzen's proof makes the finiteness of the proof debatable. By Gentzen's proof it is established that it is possible to prove consistency without relying on intuitionistic logic. The reduction procedure of the proof can be represented in primitive recursive arithmetic and transfinite induction up to the ordinal ϵ_0 restricted to primitive recursive predicates. The claimed finiteness of the principle relies on the fact that the predicates to which it is applied are finite, that is, they do not contain quantification over the whole domain of natural numbers. But, as mentioned, Gentzen also provides a constructive proof for the principle itself.

Thus, there are three methods that can be employed to find consistency proofs; finitism, Gentzen's approach and intuitionism. The unsuccessful finitistic approach prohibits the use of negation over a proposition that has been universally quantified over an infinite domain.[15] The intuitionistic approach, on the other hand bans the law of excluded middle to infinite sets, which is a weaker restriction. Thirdly, the principle TI in Gentzen's approach extends inductive

[14] Kreisel 1971 [15], p. 252.
[15] Mehlberg 2002 [17], p. 74.

reasoning to a transfinite domain. This principle is unprovable in intuitionistic reasoning.

1.10 A partial solution from the empirical sciences

According to Hilbert the metamathematics or the knowledge of how problems are solved should solely be based on finitistic reasoning. The reasoning is therefore more restricted than in proper mathematics, in hope of accomplishing an intuitive line of thought. But according to Mehlberg (2002) too strong restrictions on metamathematics may limit our knowledge. He states that "in particular, the metamathematical problem of consistency may prove to transcend the potentialities of human knowledge if the knowledge of a system's consistency were expected to meet the unrealistic conditions which were inherent in the initial phase of the formalist program." [16]

Our knowledge may be dependable, even if it is not of the infallible deductive kind and this kind of knowledge offers a solution to the consistency proof. In Mehlberg's opinion the consistency of a theory can be dependable if serious and diverse, but unsuccessful, attempts have been made to derive contradictions. The future possibility of a proven contradiction points merely to the fact that the knowledge is not infallible. By a reasonable degree of certainty, as in the empirical sciences, this can be given the epithet knowledge, rather than belief.

With reference to a conversation with Gödel, Mehlberg states that the quest for a set-theoretical foundation for mathematics in Gödel's opinion was mainly for explanatory purpose, not in order to provide a real foundation. The aim is to explain the phenomena, as is done in physics, where phenomena are explained by the theory.

1.11 Following in Gentzen's path

The legacy of Gentzen's work in this field is that through him ordinal analysis became known. This is the method of measuring the proof-theoretic strength of a formal system of mathematics, by the least

[16] Mehlberg 2002 [17], p.72.

ordinal, α, with the property that no recursive well-ordering of ordinal type α may be proven well-ordered in the system in question. That the proof-theoretical ordinal of first-order arithmetic is ϵ_0 was proven by Gentzen in 1943.

After the Second World War some subsystems of classical analysis were proven to be consistent using the methods developed by Gentzen. By restricting the application of the comprehension axiom in second order predicate calculus, subsystems of classical analysis are obtained. For some of these systems it is possible to produce constructive consistency proofs.[17]

[17] Schütte and Schwichtenberg 1990 [30], p. 725.

Chapter 2

Consistency proofs in different calculi

2.1 Natural deduction and sequent calculus

In Gentzen's opinion the object of logic is to study the general structures of proofs. This opinion is a break with the logicist tradition of Frege, Peano, Russell and Hilbert who considered the object of logic to study logical truth.[1]

Gentzen developed the systems of natural deduction and sequent calculus to analyze the structure of proofs. The former was successful for the intuitionistic case and the latter was needed to deal with the classical case. Natural deduction with its hypothetical reasoning was developed to echo better than axiomatic calculi the actual reasoning in mathematical proofs. It can be noted that the system of natural deduction was independently developed by Jaskowski in 1934. His system is presented in linear form and his work does not contain any analysis of the structure of the derivations.[2]

Sequent calculus, on the other hand, proved to be the system in which Gentzen found his main result, the Hauptsatz or cut elimina-

[1] von Plato 2007, [25], p. 384.
[2] Jaskowski 1967 [12].

tion theorem. The system was developed to prove the Hauptsatz for predicate logic. The result can be used to prove consistency for the system of rules. The calculus formalizes the derivability of a formula from other formulas, $\Gamma \to C$, represented by an arrow between a list of assumptions, Γ, as an antecedent and a conclusion, C, as a succedent of the sequent. As a generalization of the notion of sequent a classical multi-succedent calculus is obtained. In the classical calculus the sequents, $\Gamma \to \Delta$, can be interpreted as a number of cases, Δ, under the open assumptions, Γ. It is not necessarily decidable which of the cases hold.

In Gentzen's formalized systems intuitionistic logic gains a strong position because it becomes a special case of the classical calculus. This property may not be as striking in other calculi, in which the rules are chosen differently. The calculi show his intent to use intuitionistic logic as a base for his argumentation.

2.2 From 1936 to 1938

The calculus used in Gentzen's first two proofs is natural deduction in sequent calculus style. The calculus has rules from natural deduction operating on sequents. Instead of left rules, operating on the antecedent, there are elimination rules operating on the succedent.

In the latter proof from 1936 the number of rules is decreased. Initial sequents are used in order to replace logical rules and disjunction, implication and the existential quantifier are eliminated. The new initial sequents replacing the logical rules are among others $A\&B \to A$, $A, B \to A\&B$, $\forall x A(x) \to A(t)$ and $\neg\neg A \to A$. Gentzen regarded structural rules as purely formal modifications of the sequents, except for the rule of weakening. These rules were added to the calculus in order to obtain special features for the formalism.

The proof from 1936 can be explained as a 'reduction procedure' for sequents. Firstly, all free variables are replaced by numerals and then choices are made as the sequent is reduced to less complex sequents. The choices made can be regarded as aiming for the worst possible scenario, in which the sequent is falsified. The reduction ends in a reduced form, which consists of a true sequent. A sequent is true if the antecedent contains a false atomic formula or if the succedent is a true atomic formula. Gentzen shows that initial sequents are reducible and that the rules preserve reducibility. Consistency follows from the fact that the sequent $\rightarrow 0 = 1$ is not in reduced form nor reducible.

The proof also gives an ordinal assignment to prove that the process terminates. This can be compared to the proof from 1938, which also uses an ordinal assignment, but has a standard notation for the ordinal numbers.

As a standard version of the classical consistency proof we consider Takeuti's version [31], which is based on Gentzen's proof [7]. This third proof is the best known of Gentzen's papers on this subject. Gentzen's consistency proof from 1938 can be explained as consisting of a well-ordering of all derivations and a reduction procedure for derivations of the empty sequent. Derivations are ordered by complexity and the reduction decreases the complexity of the derivation. Therefore, if there exists a derivation of the empty sequent, then by a finite number of steps a simple derivation, which does not contain any induction rule, is reached. Consistency then follows by proving that the empty sequent is not derivable without the induction rule.

In the article from 1938 a standard multi-succedent sequent calculus is used. In contrast to the earlier proofs the reduction process resembles cut elimination. In the first step of the reduction procedure free variables in a derivation of the empty sequent are replaced by numerals. Then the 'end-piece' of the derivation is considered. The end-piece consists of structural rules and induction at the end of the derivation. Induction rules and initial sequents are reduced if they occur in the end-piece. Lastly, cuts on compound formulas are reduced. The cuts are not directly reduced to cuts on less complex formulas, but additional cuts on the less complex formulas are introduced. This introduction of additional cuts is called the height-line argument. The ordinal assignment defines a notion of height of a cut and the additional cuts push up the places in the derivation where the heights of the cuts drop. These drops affect the ordinal assigned and the result is a reduction of the ordinal of the derivation.

2.3 Gentzen's consistency proof performed in natural deduction

Since the publication of Gentzen's proof, the conducting of the consistency proof in standard natural deduction has been an open problem. This problem has recently been solved for an intuitionistic calculus by the present author in chapter 5. The result is based on a normalization proof by Howard [11], recommended to the author by Per Martin-Löf. The new consistency proof is performed in the manner of Gentzen, by giving a reduction procedure for derivations of falsity. In contrast to Gentzen's proof, the procedure uses a vector assignment. The reduction reduces the first component of the vector and this component can be interpreted as an ordinal less than ϵ_0, thus ordering the derivations by complexity and proving termination of the process.

The assignment uses vectors instead of a direct ordinal assignment because the length of the vector is used as a parameter coding the complexity of the formulas in the derivation. An interesting feature of the proof is that the reduction of induction rules produces non-normalities in the reduced derivation as it introduces an implication, which is directly followed by an elimination of the same implication. This can be compared to Gentzen's proof from 1938, which introduces additional cuts in the derivation. However, if Gentzen's proof were translated into natural deduction, the reduced implication would become a composition of the premises of the induction rule. Gentzen's procedure otherwise resembles cut elimination and the natural deduction proof resembles a normalization proof as standard detour conversions are made after the induction inferences have been reduced.

In the article Zusammenfassung von mehreren vollständigen Induktionen zu einer einzigen, which was published posthumously, Gentzen shows a method of fusing several induction inferences in a derivation into one. A formula $(y = 1 \supset A_1(x))\& \ldots \&(y = n \supset A_n(x))$, denoted $B(x)$, is constructed, which contains the free variables x and y and fuses the induction formulas A_i, where $1 \leq i \leq n$. Then from the formula $[B(0)\&\forall x(B(x) \supset B(x + 1))] \supset \forall x B(x)$, the induction axiom for each separate formula may be derived, by substituting numbers for the free variable y. A consequence of this result is that the number of induction inferences cannot be used as a measure of the derivation's complexity, it is the complexity of the induction formula that counts.

2.4 Consistency without induction

When Gentzen began writing his thesis, Untersuchungen über das logische Schliessen (1934) [9], he intended to provide a consistency proof for arithmetic, by proving the Hauptsatz. However, it turned out not to be possible to treat the rule of induction in this manner. Therefore, a corollary occurring in the thesis is only a consistency proof for the system without induction.

In the thesis Gentzen presents a formal axiomatic system for elementary arithmetic without induction. He concludes that it cannot be proven that the system actually allows us to represent all types of proofs customary in formal arithmetic. It can only be tested that individual proofs are representable. He then proves consistency for this system. A contradiction is derivable if and only if there is a derivation in the logical calculus of a sequent with an empty succedent and arithmetical axioms in the antecedent. The sharpened Hauptsatz is then applied to the derivation and free variables are replaced with a constant. Furthermore, by replacing eigenvariables in subdervations it is concluded that if an inconsistency is derivable, then it is also derivable from numerical propositions using only propositional logic. And such a derivation is not possible, which Gentzen indicates by referring to a soundness proof for the propositional calculus.

The sharpened Hauptsatz states that if each formula in a derived sequent has quantifiers only as outermost connectives, then there is a cut-free derivation, which has only quantifier rules at the end of the derivation. There is a midsequent in the derivation dividing it into an upper part, containing only propositional logic and a lower part containing only quantifier rules.

The aim of the first part of the chapter 4 is to provide a proof analysis of a system of arithmetical rules. As a corollary of the main lemmas together with cut elimination we get a consistency proof for arithmetic without induction by using purely proof-theoretical means. A consistency proof for the full arithmetical system can be obtained by a Gentzen-style proof, such as in [31, ch. 2, §12].

Lemma 4.4.11 proves that the empty sequent is not derivable without the second infinity rule, which states that the successor function is injective. By a combinatorial argument it is then proven that the second infinity rule is admissible if the antecedent is empty.

Gentzen and Takeuti use semantical arguments to prove a lemma (our lemma 4.4.20) stating that there is no so called simple derivation of the empty sequent. Their proof is short, but we shall instead use methods which are coherent with the proof theoretical analysis of Gentzen's consistency proof. It is shown that the lemma can be proved purely proof-theoretically, by formulating the arithmetical axioms as rules instead of initial sequents and by considering all possible combinations of these rules, as in lemma 4.4.14.

2.5 A direct proof in an intutionistic calculus

A study of the papers Gentzen left behind shows that he worked on yet another fifth proof between 1939 and 1943. The aim was to rework the 1938 proof with an intuitionistic sequent calculus, to get a direct proof of the consistency of intuitionistic Heyting arithmetic. Gentzen's attempts are preserved in the form of close to a hundred large pages of stenographic notes, with the signum BTJZ that stands for "Proof theory of intuitionistic number theory". For further reading and description of Gentzen's manuscripts we recommend the thorough discussion of Gentzen's work found in [27].

The aim of the second part of chapter 4 is to give a direct Gentzen-style proof of the consistency of intuitionistic arithmetic or Heyting arithmetic. It is based on Gentzen's classical proof from 1938 formulated by G. Takeuti in [31, ch. 2, §12]. Takeuti's proof can be considered the standard proof for the classical calculus. The proof is carried out by giving a reduction procedure (as in our lemma 4.4.21) for every derivation of the empty sequent that represents a contradiction in the system. By giving every sequent an ordinal it is shown that the reduction procedure terminates.

Another proof of the consistency of Heyting arithmetic is given by B. Scarpellini in [29]. His proof is based on the reductions of the classical calculus. An intuitionistic derivation is reduced by the classical reductions. This results in a classical derivation with multi-succedent sequents. However, as the additional formulas in the succedent have been introduced by weakening, they can be deleted from the derivation, making it an intuitionistic derivation.

Chapter 3

Rules of proof extending a logical calculus

3.1 Axioms as rules

There are four ways of extending sequent calculus by axioms of a mathematical theory. When extending a logical system with formalized axioms for proof analysis, the standard methods lead to the failure of main results, such as Gentzen's Hauptsatz. The first way is to add an axiom A in the form of a sequent $\Rightarrow A$. These sequents can be leafs of a derivation. This way of adding axioms leads to the failure of cut elimination. Gentzen (1938) added mathematical basic sequents $P_1, \ldots, P_m \Rightarrow Q_1, \ldots, Q_n$ to the logical system. In this case the cuts can be limited to cuts on these basic sequents. A third way is to treat axioms as a context Γ and relativizing each theorem to Γ thus proving results of the form $\Gamma \Rightarrow C$. In this case cut elimination applies. The fourth method, which will be examined below, is to extend the logical system by mathematical rules. If a sequent is derivable in one of these four systems, then it is derivable in the other systems as well. Thus, the four systems are equivalent.

The treatment of mathematical rules is developed in [20] and [21]. The method consists of converting axioms into rules of proof extending the logical calculus. By the treatment of axioms as rules, the derivation of a compound formula can be transformed into a derivation, in which the mathematical rules are separated from the logical part. A consequence is that there are logic-free derivations of atomic formulas from atomic assumptions.

Negri and von Plato [21] give a formulation of mathematical rules in a G3 system (both classical and intuitionistic), in which the structural rules (weakening, contraction and cut) are admissible and not explicit rules. In [20] the method is extended to geometric theories[1] which contain existential axioms.

If a set of axioms for a theory are (the universal closure of) quantifier-free formulas, then these axioms can be converted into rules of proof for the construction of formal derivations. Any quantifier-free axiom is equivalent to a conjunction of disjunctions of atoms and negations of atoms, $\neg P_1 \vee \cdots \vee \neg P_m \vee Q_1 \vee \cdots \vee Q_n$. A conjunction of disjunctions of this kind is equivalent to a conjunction of implications of the form: $P_1 \& \ldots \& P_m \supset Q_1 \vee \cdots \vee Q_n$. This implication can be transformed into a rule of sequent calculus in which the active formulas of the rule are on the right side or the left side of the sequent arrow. The right rule scheme, R-RS, that corresponds to the implication is

$$\frac{\Gamma \to \Delta, Q_1, \ldots, Q_n, P_1 \quad \ldots \quad \Gamma \to \Delta, Q_1, \ldots, Q_n, P_m}{\Gamma \to \Delta, Q_1, \ldots, Q_n} \; {\scriptstyle R-RS}$$

The right rule can be interpreted as: if each atomic formula $P_1, \ldots, P_m$ follows as a case under the assumptions Γ, then the cases under Γ are $Q_1, \ldots, Q_n$.

The left rule scheme, L-RS, that corresponds to the implication is

$$\frac{Q_1, P_1, \ldots, P_m, \Gamma \to \Delta \quad \ldots \quad Q_n, P_1, \ldots, P_m, \Gamma \to \Delta}{P_1, \ldots, P_m, \Gamma \to \Delta} \; {\scriptstyle L-RS}$$

The interpretation of the left rule is that if something follows from each of the cases $Q_1, \ldots, Q_n$, then it already follows from just assuming $P_1, \ldots, P_m$.

The conversion of axioms into rules of proof can be extended to geometric theories, which have existential quantifiers included in

[1] Geometric formula is a term from category theory.

the axioms. The axioms of geometric theories belong to the set of geometric implications. Examples of geometric theories are Robinson arithmetic, the theory of nondegenerate ordered fields and the theory of real closed fields.

3.1.1 Definition. A geometric formula contains no implication or universal quantifier. A geometric implication is a sentence of the form $\forall\overline{x}(A \supset B)$, where A and B are geometric formulas. Furthermore, a geometric theory is axiomatized by geometric implications.

Any geometric implication can be reduced to a conjunction of formulas of the form $\forall\overline{x}(P_1\&\dots\&P_m \supset \exists y_1 M_1 \vee \dots \vee \exists y_n M_n)$, where P_i is an atomic formula for $1 \leq i \leq m$ and M_j is a conjunction of atomic formulas for $1 \leq j \leq n$. In the implication the variable y_j does not appear in P_i. We will use a vector notation, $\overline{P}$ for a multiset of formulas $P_1, \dots, P_m$ and $\overline{Q}_j$ for $Q_{j_1}, \dots, Q_{j_{k_j}}$. Let M_j be the conjunction $Q_{j_1}\&\dots\&Q_{j_{k_j}}$ where Q_{j_l} are atomic formulas. A replacement in a vector, $\overline{Q}_j(y_j/x_j)$ denotes the replacement in each of the components, that is $Q_{j_1}(y_j/x_j), \dots, Q_{j_{k_j}}(y_j/x_j)$.

The left geometric rule scheme, L-GRS, that corresponds to the geometric axiom is

$$\frac{\overline{Q}_1(y_1/x_1), \overline{P}, \Gamma \to \Delta \quad \dots \quad \overline{Q}_n(y_n/x_n), \overline{P}, \Gamma \to \Delta}{\overline{P}, \Gamma \to \Delta}\ {\scriptstyle L-GRS}$$

where the variables y_i, $1 \leq i \leq n$, are the eigenvariables of the rule. The eigenvariables must not occur in the conclusion of the rule, that is in $\overline{P}, \Gamma, \Delta$.

The geometric axiom is equivalent to the geometric rule scheme, because assuming admissibility of the structural rules, the axiom is derivable from the geometric rule scheme and the scheme is derivable if the geometric axiom is assumed.

The principal formulas of the rule ($P_1, \dots, P_m$ for the left rule and $Q_1, \dots, Q_n$ for the right rule) have to be repeated in the premises of the rules in order to preserve admissibility of contraction when adding mathematical rules to the logical system. When proving (height-preserving) admissibility of contraction by induction on the length of the derivation the repetition of the formula in the premises makes it possible to permute a contraction on P_i above the mathematical rule, by duplicating the contraction for each premise.

It can be noted that a substitution of formulas in the rule scheme can produce a duplicated principal formula, P_i. Therefore, to ensure that contraction is admissible we also need to add a rule where the formula is not duplicated. In other words, we have a closure condition on the system of rules: If a given geometric theory includes a rule of the form

$$\frac{\overline{Q}_1(y_1/x_1), \overline{P}, P, P, \Gamma \to \Delta \quad \ldots \quad \overline{Q}_n(y_n/x_n), \overline{P}, P, P, \Gamma \to \Delta}{\overline{P}, P, P, \Gamma \to \Delta}$$

then the system should also include the rule

$$\frac{\overline{Q}_1(y_1/x_1), \overline{P}, P, \Gamma \to \Delta \quad \ldots \quad \overline{Q}_n(y_n/x_n), \overline{P}, P, \Gamma \to \Delta}{\overline{P}, P, \Gamma \to \Delta}$$

The vector $\overline{P}$ is $P_1, \ldots, P_{m-2}$.

3.1.2 Example. In the theory of equality the rule of transitivity

$$\frac{a = c, a = b, b = c, \Gamma \to \Delta}{a = b, b = c, \Gamma \to \Delta}\ {\scriptstyle Tr}$$

has a limiting case if all terms are the same. In this case however the rule given by the closure condition,

$$\frac{a = a, a = a, \Gamma \to \Delta}{a = a, \Gamma \to \Delta}$$

is a special case of the reflexivity rule,

$$\frac{a = a, \Gamma \to \Delta}{\Gamma \to \Delta}\ {\scriptstyle Ref}$$

which is already included in the theory.

3.2 Natural deduction

A Harrop formula contains no hidden disjunctions, which implies that if a disjunction is derivable from a set of Harrop formulas, then one of the disjuncts is derivable.

3.2.1 Definition. The class of Harrop formulas is defined by

1. atomic formulas and $\perp$ are Harrop formulas,
2. if A and B are Harrop formulas, then $A\&B$ is a Harrop formula,
3. if B is a Harrop formula, then $A \supset B$ is a Harrop formula.

Any Harrop axiom is equivalent to a conjunction of implications of the form: $P_1\&\ldots\&P_m \supset Q$ for atomic Q and $P_1, \ldots, P_m$. Axioms that are Harrop-formulas may be converted into rules of natural deduction, because the calculus gives a single conclusion in a natural-deduction-style rule. Therefore, the natural deduction rule corresponding to the implication above is

$$\frac{P_1 \quad \ldots \quad P_m}{Q}\ {\scriptstyle Rule}$$

For axioms of a general form the system can be transformed into a multi-conclusion natural deduction. Given an axiom of the general form $P_1\&\ldots\&P_m \supset Q_1 \vee \cdots \vee Q_n$ the corresponding rule of proof is

$$\frac{P_1 \quad \ldots \quad P_m}{Q_1 \ldots Q_n}\ {\scriptstyle MultiRule}$$

The atomic formulas $P_1, \ldots, P_m$ are the premises and the atomic formulas $Q_1, \ldots, Q_n$ are the conclusions of the rule.

For geometric theories the axioms may be converted into rules with eigenvariables resembling the rule for disjunction elimination:

$$\frac{\begin{matrix} \\ \vdots \\ P_1 \end{matrix} \quad \ldots \quad \begin{matrix} \\ \vdots \\ P_m \end{matrix} \quad \begin{matrix} [\overline{Q}_1(y_1/x_1)] \\ \vdots \\ C \end{matrix} \quad \ldots \quad \begin{matrix} [\overline{Q}_n(y_n/x_n)] \\ \vdots \\ C \end{matrix}}{C}$$

The conclusion of the rule, C, is an arbitrary formula and the sets of atomic formulas $\overline{Q}_j$, containing the eigenvariables y_j, are discharged. The eigenvariables must not occur free in the open assumptions (excluding the discharged assumptions), or the conclusion of the rule.

3.3 The subterm property

One feature of mathematical rules is that the active and principal formulas are atomic formulas. If a logical rule occurs above the mathematical rule in a derivation, then the two rules may be permuted. This holds because the active formula of the mathematical rule is not the principal formula of the logical rule, but it is in the context of the conclusion of the logical rule. Thus, there is a derivation in which the all mathematical rules are above the logical rules. As a result of Gentzen's cut elimination in sequent calculus or normalization in natural deduction it is not necessary to consider the logical rules at all if both the assumptions and the conclusion are atomic formulas. Therefore, it is possible to separate the mathematical and the logical part of a derivation.

Because the logical system for sequent calculus, described above, is of type G3 with the structural rules admissible, we may do a root-first proof search. By this search the derivability of a sequent reduces to the derivability of the leafs with mathematical rules.

If a subterm property is proven for a theory, then the terms occurring in the derivation may be restricted to known terms from the assumptions or the conclusion of the derivation. If this is the case and our assumptions are a finite set of atomic formulas, then the possible combinations of terms in atomic formulas can be restricted to a finite number. By combining these formulas we get a finite number of possible derivations and it may be checked if any of these derivations are valid. Therefore, if the subterm property holds, we have a positive solution to the so-called uniform word problem, claiming that the derivability of an atomic formula from a number of atomic formulas is decidable. Thus, it is decidable if a leaf is derivable with mathematical rules.

3.4 Applications of the axioms-as-rules method

The method of axioms-as-rules has been applied to predicate logic with equality, theories of apartness and order, projective and plane affine geometry [28], as well as lattice theory, ordered fields and real closed fields.[2]

The subterm property for lattices was established in [22] by an analysis of the formal derivations of a natural deduction system extended by rules for lattice theory. The result has also been extended to minimal quantum logic or orthologic, which is the study of ortholattices. Meninander [18] presents a positive solution to the uniform word problem for ortholattices. By analysis of the structure of possible derivations it is shown that proof search is bounded and thus that the uniform word problem is solvable.

A consequence of a positive solution to the uniform word problem for a finite Harrop axiom system is the existence of a polynomial-time decision algorithm for the derivability of an atomic formula from a finite number of atomic assumptions. There is only a linear number of known subterms of the assumptions and the conclusion. Then the derivation rules are applied to derive new atomic formulas with these subterms until no new atomic formulas are derivable. This process is a polynomial time decision algorithm for the uniform word problem. The proof of this is analogous to the proof of a polynomial-time algorithm for lattices by Skolem from 1920, which was rediscovered by Cosmadakis [3][3].

Another example of a system extended with rules is found in [24], in which a system of rules in natural deduction for Heyting arithmetic is presented. The logical part of the system includes general elimination rules, which are of the same form as the standard disjunction elimination rule. The induction rule is similarly formulated which makes possible the permutative conversions. A normalization theorem is then proven for Heyting arithmetic and as a consequence the existence property is proven.

3.5 Independence of Euclid's parallel postulate

Von Plato [28] applies the method of axioms-as-rules to projective and affine plane geometry proving that the rules with eigenvariables

[2]Negri 2003 [20]
[3]Burris 1995 [2] and von Plato 2007 [26].

are conservative over the base theory. The method employed consists of proving the subterm property, concluding that the geometrical objects of a derivation may be restricted to terms known from the assumptions or from the conclusion of the theorem. The lowermost occurrences of a new term in a derivation are considered and by permuting rules a standard form of derivation is obtained to which a combinatorial analysis can be applied. By combinatorial analysis of the possible rules the derivation is transformed to a shorter derivation. Thereby, using induction on the length of the derivation the subterm property is proved.

As a consequence of the subterm property consistency proofs for the two geometric theories are obtained, because the empty sequent is not derivable. The stronger statement that any set of atomic formulas is consistent also follows due to the fact that no sequent $\Gamma \rightarrow$, with Γ a set of atomic formulas, can be derived in a system of mathematical right rules. The main corollary however is the independence of Euclid's parallel axiom for affine geometry. The parallel postulate states that given a point, a, outside a line, l, there is no point incident with both the line l and the parallel line through the point a. This can be formalized as $\neg(a \in l) \supset \neg(b \in l \& b \in par(l,a))$ and expressed as a sequent without logical connectives as $b \in l, b \in par(l,a) \rightarrow a \in l$.[4] The rule corresponding to the axiom included in the system of rules is the rule for uniqueness of parallel lines.

$$\frac{a \in l \quad a \in m \quad l \parallel m}{l = m}\ {\scriptstyle Unipar}$$

If we assume $b \in l$ and $b \in par(l,a)$, as well as the additional assumption $l \parallel par(l,a)$, then by application of the rule we get the conclusion $l = par(l,a)$. By line substitution and the affine axiom of incidence, $a \in par(l,a)$, we get the sought conclusion $a \in l$.

The proof of the independence of the parallel postulate comes from the fact that restriction to known terms from the sequent $b \in l, b \in par(l,a) \rightarrow a \in l$ can only produce a few new atomic formulas, but not the sought formula $a \in l$. None of the rules, excluding the rule of the uniqueness of the parallel line, can be applied to the premises $b \in l$ and $b \in par(l,a)$. Using the available rules of the system we can

[4]The construction $par(l,a)$ is the parallel line to l through the point a.

only produce the new atomic formulas $a \in par(l, a)$, $par(l, a) \parallel l$ and $l \parallel par(l, a)$. After that nothing but loops are produced. Therefore, there cannot be a derivation of the parallel axiom if the corresponding rule of the uniqueness of the parallel line is left out.

Thus, it is possible to prove the independence of Euclid's parallel postulate using proof theoretical means as opposed to the standard method of referring to non-Euclidean geometries.

Chapter 4

A direct Gentzen-style consistency proof for Heyting arithmetic

The aim of this chapter is to give a direct Gentzen-style proof of the consistency of intuitionistic arithmetic. The proof is based on Takeuti's proof [31, ch. 2, §12]. The first part of the chapter removes semantical arguments from the proof, by giving a purely proof theoretical analysis of the calculus without induction. A direct proof of cut elimination is included in the analysis of the system. Finally, the consistency proof for the complete system is given in Gentzen-style. The proof is direct, instead of concluding that consistency of the intuitionistic calculus follows from consistency of the classical calculus, as the former is a subsystem of the latter. We shall assume that the reader has basic knowledge of ordinals and refer to [31] for a more detailed treatment of the subject.

4.1 The sequent calculus G0i

A *sequent* is an expression of the form $\Gamma \rightarrow A$ or $\Gamma \rightarrow$, where the *antecedent* Γ is a (possibly empty) multiset. A multiset is a finite list of formulas where the order of the formulas does not matter but the multiplicity of the formulas does, in contrast to ordinary sets. In the *succedent* A is a formula, but the succedent can also be empty. The rules for the intuitionistic sequent calculus G0i, from [23] except that we have no rule of weakening, are as follows.

Initial sequent:

$$A, \Gamma \to A$$

Logical rules:

$$\frac{A, B, \Gamma \to C}{A\&B, \Gamma \to C}\ L\& \qquad \frac{\Gamma \to A \quad \Gamma' \to B}{\Gamma, \Gamma' \to A\&B}\ R\&$$

$$\frac{A, \Gamma \to C \quad B, \Gamma' \to C}{A \vee B, \Gamma, \Gamma' \to C}\ L\vee \qquad \frac{\Gamma \to A}{\Gamma \to A \vee B}\ R\vee \qquad \frac{\Gamma \to B}{\Gamma \to A \vee B}\ R\vee$$

$$\frac{\Gamma \to A}{\sim A, \Gamma \to}\ L\sim \qquad \frac{A, \Gamma \to}{\Gamma \to \sim A}\ R\sim$$

$$\frac{\Gamma \to A \quad B, \Gamma' \to C}{A \supset B, \Gamma, \Gamma' \to C}\ L\supset \qquad \frac{A, \Gamma \to B}{\Gamma \to A \supset B}\ R\supset$$

$$\frac{A(t/x), \Gamma \to C}{\forall x A, \Gamma \to C}\ L\forall \qquad \frac{\Gamma \to A(y/x)}{\Gamma \to \forall x A}\ R\forall$$

$$\frac{A(y/x), \Gamma \to C}{\exists x A, \Gamma \to C}\ L\exists \qquad \frac{\Gamma \to A(t/x)}{\Gamma \to \exists x A}\ R\exists$$

Structural rules:

$$\frac{A, A, \Gamma \to C}{A, \Gamma \to C}\ LC$$

$$\frac{\Gamma \to A \quad A, \Gamma' \to C}{\Gamma, \Gamma' \to C}\ Cut$$

In the quantifier rules the expression $A(t/x)$ means that every free occurrence of x in A is substituted with the term t. In the rules $L\exists$

and $R\forall$ the standard variable restriction holds that y, also called the eigenvariable of the rule, must not be free in the conclusion of the rule. The formula that is introduced in the conclusion of a logical rule, for example $A\&B$ in the conjunction rules, is the *principal formula* of the rule. The formulas that the rule is applied on are the *active* formulas. In the structural rules the principal formula is the formula that the rules are applied on, in this case A. The formula is also called contraction or cut formula. The multiset Γ in the sequents is called the *context* of the rule. We use a calculus with arbitrary contexts in all initial sequents and hence no rule of weakening is needed. We will use the notation Γ_{1-2} as short for Γ_1, Γ_2.

4.2 Heyting arithmetic

The language of Heyting arithmetic consists of the constant 0, the unary functional symbol s, the binary functional symbols $+$ and $\cdot$ and the binary predicate symbol $=$.

4.2.1 Definition. *Terms* are inductively defined. The constant 0 and variables are terms and if t and t' are terms then $s(t), t + t'$ and $t \cdot t'$ are also terms. Terms are *closed* if they do not contain any variable.

Formal expressions for the natural numbers, *numerals*, are inductively defined: 0 is a numeral and if $\overline{m}$ is a numeral, then $s(\overline{m})$ is also a numeral. The numeral $\overline{m}$ is m copies of s followed by a 0.

The axioms of Heyting arithmetic can be formulated as rules of natural deduction, expanding the logical calculus. Together with an induction rule the logical and arithmetical rules constitute the system of Heyting arithmetic (HA). Negri and von Plato [21] developed the general method for converting mathematical axioms into rules for the primary purpose of proving cut elimination in systems of sequent calculus. The specific system for arithmetic was first used by von Plato [24] to prove the disjunction and existential properties. These rules act on the succedent part of the sequents and have arbitrary contexts. As a special case we get rules without premises.

Rules for the equality relation:

$$\frac{}{\Gamma \to t = t}\ Ref \qquad \frac{\Gamma \to t = t'}{\Gamma \to t' = t}\ Sym$$

$$\frac{\Gamma_1 \to t = t' \quad \Gamma_2 \to t' = t''}{\Gamma_{1-2} \to t = t''}\ Tr$$

Recursion rules:

$$\frac{}{\Gamma \to t + 0 = t}\ {+Rec0} \qquad \frac{}{\Gamma \to t + s(t') = s(t + t')}\ {+Recs}$$

$$\frac{}{\Gamma \to t \cdot 0 = 0}\ {\cdot Rec0} \qquad \frac{}{\Gamma \to t \cdot s(t') = t \cdot t' + t}\ {\cdot Recs}$$

Replacement rules:

$$\frac{\Gamma \to t = t'}{\Gamma \to s(t) = s(t')}\ sRep$$

$$\frac{\Gamma \to t = t'}{\Gamma \to t + t'' = t' + t''}\ {+Rep_1} \qquad \frac{\Gamma \to t' = t''}{\Gamma \to t + t' = t + t''}\ {+Rep_2}$$

$$\frac{\Gamma \to t = t'}{\Gamma \to t \cdot t'' = t' \cdot t''}\ {\cdot Rep_1} \qquad \frac{\Gamma \to t' = t''}{\Gamma \to t \cdot t' = t \cdot t''}\ {\cdot Rep_2}$$

Infinity rules:

$$\frac{\Gamma \to s(t) = 0}{\Gamma \to}\ Inf_1 \qquad \frac{\Gamma \to s(t) = s(t')}{\Gamma \to t = t'}\ Inf_2$$

Induction rule:

$$\frac{\Gamma_1 \to A(0/x) \quad A(y/x), \Gamma_2 \to A(sy/x) \quad A(t/x), \Gamma_3 \to D}{\Gamma_{1-3} \to D}\ Ind$$

In the arithmetical rules t, t' and t'' are terms. In the induction rule y is the eigenvariable of the rule and it should not occur free in the conclusion. The induction formula A is arbitrary.

4.2.2 Definition. A *valid derivation* in HA is defined inductively. An initial sequent or an arithmetical rule without premises is a valid derivation and a valid derivation is obtained by applying a rule on valid derivations of the premises of the rule.

The *end-piece* of a derivation is defined in the following way: the end-sequent belongs to the end-piece. Furthermore, if the conclusion of a structural rule or Ind is included in the end-piece, then the premises of the rule are also included in the end-piece. An arithmetical or logical rule borders on the end-piece if the conclusion of the rule is included in the end-piece.

A formula A is a *descendant* of a formula B if A is in the context of the conclusion of a rule and B is an identical formula in the context of a premise or if A is the principal formula of the rule and B is an active formula in a premise. Furthermore, if A a descendant of B and B is a descendant of C, then A is a descendant of C. If A is a descendant of B, then B is a *predecessor* of A.

4.3 The ordinal of a derivation

We define the height of a sequent in a derivation.

4.3.1 Definition. (i) The *grade* of a formula is the number of logical symbols in the formula. The grade of a Cut or an Ind is the grade of the cut or the induction formula.

(ii) The *height* of a sequent S in a derivation P, denoted $h(S;P)$ or $h(S)$, is the maximum of the grades of the cuts and inductions below S in P.

Note that the height of the end-sequent is 0 and that the premises of a rule all have the same height. If S_1 is a sequent under another sequent S_2, then $h(S_1) \leqslant h(S_2)$.

To be able to calculate with ordinals we need to define a suitable sum operation.

4.3.2 Definition. If two ordinals μ and ν are expressed in normal form $\mu = \omega^{\mu_1} + \omega^{\mu_2} + \cdots + \omega^{\mu_m}$ and $\nu = \omega^{\nu_1} + \omega^{\nu_2} + \cdots + \omega^{\nu_n}$, where $\mu_1 \geqslant \mu_2 \geqslant \cdots \geqslant \mu_m$ and $\nu_1 \geqslant \nu_2 \geqslant \cdots \geqslant \nu_n$, then the *natural sum*, denoted $\mu\#\nu$, is equal to $\omega^{\lambda_1} + \omega^{\lambda_2} + \cdots + \omega^{\lambda_{m+n}}$, where

$\{\lambda_1, \lambda_2, \ldots, \lambda_{m+n}\} = \{\mu_1, \mu_2, \ldots, \nu_1, \nu_2, \ldots\}$ are equal multisets and $\lambda_1 \geqslant \cdots \geqslant \lambda_{n+m}$.

We shall also use the following notation: for an ordinal α and a natural number $n, \omega_n(\alpha)$ is inductively defined as $\omega_0(\alpha) \equiv \alpha$ and $\omega_{n+1}(\alpha) \equiv \omega^{\omega_n(\alpha)}$. Thus, we have

$$\omega_n(\alpha) \equiv \left.\omega^{\cdot^{\cdot^{\cdot^{\omega^\alpha}}}}\right\} n \text{ times } \omega.$$

The limit of $\omega_n(0)$ when n approaches infinity is ϵ_0, an ordinal which in some ways characterizes the strength of derivability of arithmetic. We can conclude that the following property holds for the ordinal ϵ_0.

4.3.3 Definition. An ordinal α is *accessible* if it has been shown that every decreasing sequence beginning with α is finite.

4.3.4 Lemma. ϵ_0 *is accessible.*

Takeuti [31] proves the lemma by defining eliminators, which are operations on concretely given decreasing sequences of ordinal numbers. An argument with the standard well-ordering is also given to convince the reader that it is indeed a well-ordering. The notion of accessibility is only considered when it has been constructively demonstrated that a sequence is finite. The aim of the proof is to avoid abstract notions, except for concepts which are reduced to concrete operations. This, makes the proof an extension of the finitistic standpoint of Hilbert.

We can now give every derivation in HA an ordinal.

4.3.5 Definition. The *ordinal of a sequent* S in a derivation P, denoted $o(S;P)$ or $o(S)$, is defined inductively as follows:

1. An initial sequent has the ordinal 1.
2. The conclusion of an arithmetical rule without premises has the ordinal 1.
3. If S is the conclusion of a contraction then the ordinal is the same as the ordinal of the premise.
4. If S is the conclusion of a one-premise arithmetical or logical rule, where the ordinal of the premise is μ, then $o(S) = \mu + 1$.

5. If S is the conclusion of a two-premise arithmetical or logical rule, where the ordinals of the premises are μ and ν respectively, then $o(S) = \mu \# \nu$.

6. If S is the conclusion of a cut where the premises have the ordinals μ and ν, then $o(S) = \omega_{k-l}(\mu \# \nu)$, or

$$\left.\omega^{\cdot^{\cdot^{\omega^{\mu\#\nu}}}}\right\} k-l \text{ times } \omega,$$

   where k is the height of the premises and l is the height of the conclusion.

7. If S is the conclusion of an induction and the premises have the ordinals μ_1, μ_2 and μ_3 and the height k and the conclusion has the height l, then the ordinal of the conclusion is $o(S) = \omega_{k-l+1}(\mu_1 \# \mu_2 \# \mu_3)$.

The ordinal of a derivation P, denoted $o(P)$, is the ordinal of the end-sequent. Thus, every derivation has an ordinal less than ϵ_0.

If the height remains unchanged in a cut the ordinal of the conclusion in point 6 is $\mu \# \nu$, whereas the ordinal of the corresponding case in point 7 is $\omega^{\mu_1 \# \mu_2 \# \mu_3}$.

4.4 The consistency of Heyting arithmetic

4.4.1 The consistency theorem

4.4.1 Definition. A system is said to be inconsistent if the empty sequent $\rightarrow$ is derivable. If the system is not inconsistent it is *consistent.*

4.4.2 Theorem (The consistency of Heyting arithmetic)**.** *The empty sequent $\rightarrow$ is not derivable in HA, that is, HA is consistent.*

To prove this theorem we give a reduction procedure for derivations. Assume that there is a derivation of the empty sequent. Furthermore, we may assume that the arithmetical rules are applied before logical and structural rules in the derivation. If needed, it is possible to change the order of the rules according to lemma 4.4.5, even though this may increase the ordinal of the derivation. The permutation only has to be performed once before the reduction procedure. By the reduction procedure we conclude that if there is a derivation of the empty sequent, then there is a reduced derivation with a lower ordinal and another reduced derivation and so on. Then we would have an infinite succession of decreasing ordinals all less than ϵ_0, but this is impossible and the reduction procedure must terminate. This is a contradiction and we therefore cannot have a derivation of the empty sequent. Thus, the system of Heyting arithmetic, HA, is consistent.

The reduction procedure for derivations is described in lemma 4.4.21, but before we give the proof we need some additional results.

4.4.2 Properties of derivations

4.4.3 Definition. A *thread* in a derivation is a sequence of sequents in a derivation, for which the following holds:

1. It begins with an initial sequent or the conclusion of an arithmetical rule without premises.
2. Every sequent but the last one is a premise of a rule and the sequent is followed by the conclusion of that rule.

4.4.4 Lemma. *Assume that S_1 is a sequent in a derivation P. Let P_1 be the subderivation ending with S_1 and let P_1' be another derivation ending with S_1. Now let P' be the derivation that results from the process of substituting P_1' for P_1 in P.*

If $o(S_1;P') < o(S_1;P)$, then $o(P') < o(P)$.

Proof. For every thread in P passing through S_1 we show that the following holds: If S is a sequent in a thread at or below S_1 and if S' is the corresponding sequent to S in P', then $o(S';P') < o(S;P)$. According to the assumption the proposition holds if $S = S_1$. The heights of the sequents below S in P and S' in P' are the same and for every ordinal α, β and γ that satisfy $\alpha < \beta$, we have $\alpha \# \gamma < \beta \# \gamma$. Thus, the inequality is retained for every rule applied. If we then let S be the end-sequent of the derivation we obtain the inequality for the derivations. □

4.4.5 Lemma. *In a derivation we can permute the order of the rules and first apply the arithmetical rules and then Ind and the logical and structural rules.*

Proof. If we have a logical rule followed by an arithmetical rule, then the arithmetical rule is not applied on the principal formula of the logical rule, since this formula is compound. Hence, we can permute the order of the rules and apply the arithmetical rule first.

Assume that we have an instance of contraction followed by an arithmetical rule. If the arithmetical rule is not applied on the contraction formula, then we can permute the order of the rules. We now consider the case that the arithmetical rule is applied on the contraction formula. If the arithmetical rule is a one-premise rule, then we can apply the arithmetical rule on each copy of the formula followed by an instance of contraction. If on the other hand the arithmetical rule has two premises, that is, if the rule is an instance of transitivity, then we can apply transitivity on each copy of the formula, multiplying the derivation of the other transitivity premise, and then apply contraction on the principal formula of the transitivity and also on possible formulas in the context of the multiplied premise.

If we have an instance of Cut followed by an arithmetical rule, then we can permute the order of the rules and the same holds for an instance of Ind followed by an arithmetical rule. □

Note that this change in the order of the rules can increase the ordinal of the derivation.

4.4.6 Lemma. *(i) For an arbitrary closed term t there exists a numeral $\overline{n}$ such that $\to t = \overline{n}$ can be derived without Ind or Cut.*

(ii) Let t and t' be closed terms for which $\to t = t'$ can be derived without Ind or Cut and let q be an arbitrary term. Now the sequent $\to q(t/x) = q(t'/x)$ is derivable without Ind or Cut.

(iii) Let t and t' be closed terms for which $\to t = t'$ can be derived without Ind or Cut and let q and r be terms. Then the sequent $q(t/x) = r(t/x) \to q(t'/x) = r(t'/x)$ can be derived without Ind or Cut.

(iv) *Let t and t' be closed terms for which $\rightarrow t = t'$ can be derived without Ind or Cut. Then for an arbitrary formula A the sequent $A(t/x) \rightarrow A(t'/x)$ can be derived without Ind or Cut.*

Proof. (i) For the constant 0 the proposition holds. Assume that the proposition holds for the closed terms t and t', that is there are n and m, such that $\rightarrow t = \overline{n}$ and $\rightarrow t' = \overline{m}$ can be derived without Ind or Cut. Then the sequent $\rightarrow s(t) = s(\overline{n})$ is derivable with $sRep$ where $s(\overline{n}) \equiv \overline{n+1}$.

The sequent $\rightarrow t + t' = \overline{n+m}$ can be derived as follows. First we get a derivation of $\rightarrow t + t' = \overline{n} + \overline{m}$.

$$\dfrac{\dfrac{\rightarrow t = \overline{n}}{\rightarrow t + t' = \overline{n} + t'}\,{\scriptstyle +Rep_1} \qquad \dfrac{\rightarrow t' = \overline{m}}{\rightarrow \overline{n} + t' = \overline{n} + \overline{m}}\,{\scriptstyle +Rep_2}}{\rightarrow t + t' = \overline{n} + \overline{m}}\,{\scriptstyle Tr}$$

Furthermore, if $m = 0$ we have $\rightarrow \overline{n} + \overline{0} = \overline{n+0}$ with $+Rec0$ since $\overline{n+0} \equiv \overline{n}$. If $m > 0$, that is $m = sm'$ for some m', then we have as induction hypothesis a derivation of $\rightarrow \overline{n} + \overline{m'} = \overline{n+m'}$.

$$\dfrac{\dfrac{}{\rightarrow \overline{n} + s\overline{m'} = s(\overline{n} + \overline{m'})}\,{\scriptstyle +Recs} \qquad \dfrac{\rightarrow \overline{n} + \overline{m'} = \overline{n+m'}}{\rightarrow s(\overline{n} + \overline{m'}) = s(\overline{n+m'})}\,{\scriptstyle sRep}}{\rightarrow \overline{n} + s\overline{m'} = s(\overline{n+m'})}\,{\scriptstyle Tr}$$

We now have $\rightarrow \overline{n} + \overline{m} = \overline{n+m}$ for every m. With transitivity on the conclusions of these derivations we get the result $\rightarrow t + t' = \overline{n+m}$.

The sequent $\rightarrow t \cdot t' = \overline{n \cdot m}$ is derivable in a similar manner.

(ii) If q is the constant 0 or a variable different from x, then the sequent is derivable with Ref. If q is the variable x, then we already have the derivation according to the assumption. Now assume that $q \equiv s(q')$ and as induction hypothesis we have a derivation of $\rightarrow q'(t/x) = q'(t'/x)$ that fulfills the requirements. Then we get $\rightarrow s(q'(t/x)) = s(q'(t'/x))$ with $sRep$. If $q \equiv q' + q''$ we get the following derivation where we write $q'(t)$ and $q''(t)$ instead of $q'(t/x)$ and $q''(t/x)$ and the sequent arrow is left out.

$$\dfrac{\dfrac{q'(t) = q'(t')}{q'(t) + q''(t) = q'(t') + q''(t)}\,{\scriptstyle +Rep_1} \qquad \dfrac{q''(t) = q''(t')}{q'(t') + q''(t) = q'(t') + q''(t')}\,{\scriptstyle +Rep_2}}{q'(t) + q''(t) = q'(t') + q''(t')}\,{\scriptstyle Tr}$$

If $q \equiv q' \cdot q''$ the derivation is similar.

(iii) According to point (ii) we have derivations of $\rightarrow q(t) = q(t')$ and $\rightarrow r(t) = r(t')$ that fulfill the requirements. We can now construct the derivation:

$$\dfrac{\dfrac{\dfrac{\rightarrow q(t) = q(t')}{\rightarrow q(t') = q(t)}\,Sym \qquad q(t) = r(t) \rightarrow q(t) = r(t)}{q(t) = r(t) \rightarrow q(t') = r(t)}\,Tr \qquad \rightarrow r(t) = r(t')}{q(t) = r(t) \rightarrow q(t') = r(t')}\,Tr$$

(iv) The proof is carried out by induction on the complexity of the formula. If A is an atomic formula, then the proposition is proved in case (iii).

If $A \equiv B\&C$ and we as induction hypothesis have that $B(t/x) \rightarrow B(t'/x)$ and $C(t/x) \rightarrow C(t'/x)$ are derivable without Ind or Cut, then we get the derivation:

$$\dfrac{\dfrac{B(t/x) \rightarrow B(t'/x) \quad C(t/x) \rightarrow C(t'/x)}{B(t/x), C(t/x) \rightarrow B(t'/x)\&C(t'/x)}\,R\&}{B(t/x)\&C(t/x) \rightarrow B(t'/x)\&C(t'/x)}\,L\&$$

Assume that $A \equiv \forall yB$. If $x \equiv y$ then x is not free in A and $A(t/x) \rightarrow A(t'/x)$ is an initial sequent. On the other hand if x is not y, then we have by the induction hypothesis that $(B(z/y))(t/x) \rightarrow (B(z/y))(t'/x)$, where $x \neq z$, can be derived without Ind or Cut. Because t and t' are closed terms, they do not contain y and we may change the order of the substitutions, that is $(B(z/y))(t/x) = (B(t/x))(z/y)$ and $(B(z/y))(t'/x) = (B(t'/x))(z/y)$. We now get the derivation:

$$\dfrac{\dfrac{(B(t/x))(z/y) \rightarrow (B(t'/x))(z/y)}{\forall yB(t/x) \rightarrow (B(t'/x))(z/y)}\,L\forall}{\forall yB(t/x) \rightarrow \forall yB(t'/x)}\,R\forall$$

The other cases are similar.

□

In point (i) of the lemma, we only state the existence of a numeral that equals the closed term, not that this numeral is unique. The uniqueness of the numeral is equivalent to the consistency of simple derivations proved in lemma 4.4.20.

4.4.3 Cut elimination in Heyting arithmetic

We shall give a direct proof of cut elimination in the system HA. Note that the Cut rule is a special case of our induction rule, if the induction formula has no occurrence of the variable x. In this case the second premise of the induction is an initial sequent and we have a form of vacuous induction. Thus, cuts can be eliminated by replacing them with inductions. But as the cut elimination theorem 4.4.8 shows, we can also properly eliminate Cut.

4.4.7 Definition. The *length* of a derivation in HA is defined inductively.
An initial sequent has the length 1.
The length of the conclusion of an arithmetical rule without premises is 1.
The length of the conclusion of the rule *Sym* is the same as the length of the premise.
The length of the conclusion of a one-premise rule (except *Sym*), where the premise has the length α is $\alpha + 1$.
The length of the conclusion of a two-premise rule, where the premises have the lengths α and β is $\alpha + \beta$.
The length of the conclusion of *Ind*, where the premises have the lengths α, β and γ is $\alpha + \beta + \gamma$.

4.4.8 Theorem (Cut elimination in HA)**.** *If there is a derivation of the sequent* $\Gamma \to D$ *in HA, then we can transform the derivation into a derivation of the same sequent without Cut or additional inductions.*

Proof. The proof is by induction on the grade of the cut formula with a subinduction on the length of the derivation. We assume that there are no instances of Cut above the cut we consider.

We assume that the right cut premise has been derived with $n-1$ instances of contraction on the cut formula, where $n \geqslant 1$. We consider the premise of the first contraction.

1. Firstly, we consider the case that the premise is an initial sequent.

$$\dfrac{\Gamma_1 \to A \qquad \dfrac{A^n, \Gamma_2 \to A}{A, \Gamma_2 \to A}\, {\scriptstyle LC^{n-1}}}{\Gamma_{1-2} \to A}\, {\scriptstyle Cut}$$

In this case we can add the missing context Γ_2 in the derivation of the left cut premise and get the sought derivation without Cut.

We now assume that the premise of the contraction has been derived by a rule R.

$$\dfrac{\Gamma_1 \to A \qquad \dfrac{\dfrac{}{A^n, \Gamma_2 \to D}\, {\scriptstyle R}}{A, \Gamma_2 \to D}\, {\scriptstyle LC^{n-1}}}{\Gamma_{1-2} \to D}\, {\scriptstyle Cut}$$

If rule R is an instance of Sym we can permute the contractions and the cut above the Sym. The length of the cut remains unchanged. Thus, we may assume that R is not Sym.

2. If rule R is an arithmetical rule without premises, then also the conclusion of the cut is an instance of the same rule.

3. If rule R is an arithmetical one-premise rule, then A is not principal in the rule. We can then permute the contractions and the cut above the arithmetical rule, diminishing the length of the cut.

4. Suppose rule R is Tr.

$$\dfrac{\Gamma_1 \to A \qquad \dfrac{\dfrac{A^k, \Gamma_1' \to t = t' \quad A^l, \Gamma_2' \to t' = t''}{A^n, \Gamma_2 \to t = t''}\, {\scriptstyle Tr}}{A, \Gamma_2 \to t = t''}\, {\scriptstyle LC^{n-1}}}{\Gamma_{1-2} \to t = t''}\, {\scriptstyle Cut}$$

where $\Gamma_2 = \Gamma_{1-2}'$ and $n = k + l$. We then transform the derivation diminishing the length of the cuts on A.

$$\dfrac{\dfrac{\Gamma_1 \to A \qquad \dfrac{A^k, \Gamma_1' \to t = t'}{A, \Gamma_1' \to t = t'}\, {\scriptstyle LC^{k-1}}}{\Gamma_1, \Gamma_1' \to t = t'}\, {\scriptstyle Cut} \qquad \dfrac{\Gamma_1 \to A \qquad \dfrac{A^l, \Gamma_2' \to t' = t''}{A, \Gamma_2' \to t' = t''}\, {\scriptstyle LC^{l-1}}}{\Gamma_1, \Gamma_2' \to t' = t''}\, {\scriptstyle Cut}}{\Gamma_1^2, \Gamma_2 \to t = t''}\, {\scriptstyle Tr}$$

$$\vdots \; \textit{contractions}$$

$$\Gamma_{1-2} \to t = t''$$

5. If rule R is a logical one-premise rule where A is not principal, then we can permute the contractions and the cut above the rule, diminishing the length of the cut.

6. If rule R is a logical two-premise rule where A is not principal, then we transform the derivation as in case 4, diminishing the length of the cuts.

7. Suppose rule R is a logical rule where A is principal. We consider the rule with which the left premise of the cut has been derived.

7.1 If the left cut premise is an initial sequent, then the formula A is in Γ_1. Thus, we can get the conclusion of the cut by adding the missing context Γ_1 without A in the derivation of the right cut premise.

7.2 The left cut premise has not been derived by an arithmetical rule, since the formula A has logical structure.

7.3 If the left cut premise has been derived by a logical one-premise rule where A is not principal, then we can permute the cut above the rule.

7.4 If the left cut premise has been derived by a logical two-premise rule where A is not principal, that is $L\supset$ or $L\vee$, then we can in the case of $L\vee$ apply Cut twice, once on each premise of the logical rule and then apply the logical rule and in the case of $L\supset$ apply Cut before the rule.

7.5 If the left cut premise has been derived by a logical rule where A is principal, then we consider the derivation according to the form of the formula. We consider the case where A is a conjunction $B\&C$.

$$\dfrac{\dfrac{\Gamma_1' \to B \quad \Gamma_1'' \to C}{\Gamma_1 \to B\&C}{\scriptstyle R\&} \quad \dfrac{\dfrac{B, C, (B\&C)^{n-1}, \Gamma_2 \to D}{(B\&C)^n, \Gamma_2 \to D}{\scriptstyle L\&}}{B\&C, \Gamma_2 \to D}{\scriptstyle LC^{n-1}}}{\Gamma_{1-2} \to D}{\scriptstyle Cut}$$

In the derivation

$$\dfrac{\dfrac{\Gamma_1' \to B \quad \Gamma_1'' \to C}{\Gamma_1 \to B\&C}{\scriptstyle R\&} \quad \dfrac{B, C, (B\&C)^{n-1}, \Gamma_2 \to D}{B, C, B\&C, \Gamma_2 \to D}{\scriptstyle LC^{n-2}}}{B, C, \Gamma_{1-2} \to D}{\scriptstyle Cut}$$

the cut length is shorter. Thus, we have by the induction hypothesis a derivation of the sequent $B, C, \Gamma_{1-2} \to D$ without Cut. We now construct the following derivation, where the grades of the cut formulas are less.

$$\dfrac{\Gamma_1' \to B \qquad \dfrac{\Gamma_1'' \to C \quad B, C, \Gamma_{1-2} \to D}{B, \Gamma_1'', \Gamma_{1-2} \to D}\,{\scriptstyle Cut}}{\dfrac{\Gamma_1^2, \Gamma_2 \to D}{\Gamma_{1-2} \to D}\,{\scriptstyle contractions}}\,{\scriptstyle Cut}$$

The other cases of cut formula are treated in a similar manner.

7.6 If the left cut premise has been derived by a contraction, then we can permute the cut above the rule.

7.7 If the left cut premise has been derived by Ind, then we can permute the cut above the rule.

8. If rule R is an instance of contraction, where A is not principal, then we can permute the contractions and the cut above the rule, diminishing the length of the cut.

9. Suppose rule R is an instance of Ind.

$$\dfrac{\Gamma_1 \to A \qquad \dfrac{\dfrac{A^k, \Gamma_1' \to B(0) \quad A^l, B(y), \Gamma_2' \to B(sy) \quad A^m, B(t), \Gamma_3' \to D}{A^n, \Gamma_2 \to D}\,{\scriptstyle Ind}}{A, \Gamma_2 \to D}\,{\scriptstyle LC^{n-1}}}{\Gamma_{1-2} \to D}\,{\scriptstyle Cut}$$

Here we have $\Gamma_2 = \Gamma_{1-3}'$ and $n = k + l + m$. We transform the derivation as in case 4, diminishing the length of the cuts on A. $\square$

This direct proof of cut elimination in Heyting arithmetic is an extension of the proof given in [23]. Note that unlike Gentzen's original proof of cut elimination for sequent calculus in his thesis of 1933, our proof is carried out without introducing any rule of multicut.

4.4.4 Consistency proof for simple derivations

4.4.9 Definition. A *simple derivation* is a derivation without free variables, without Ind and that contains only atomic formulas.

Thus, in a simple derivation we have only initial sequents, arithmetical and structural rules, and in addition there are no compound formulas in the contexts.

Our aim is now to show that there is no simple derivation of the empty sequent, but first we consider only the case that the derivation does not contain rule Inf_2.

4.4.10 Definition. We inductively define if the *value* of a closed term is 0 or 1. The constant 0 has value 0. A term of the form $s(t)$ has value 1. A term of the form $t + t'$ has value 0 if both t and t' have value 0 and otherwise it has value 1. A term of the form $t \cdot t'$ has value 0 if t or t' has value 0 and otherwise it has value 1.

According to the definition a closed term has value 0 if it equals 0 and value 1 if it is greater than 0.

4.4.11 Lemma. *There is no simple derivation of the empty sequent without rule* Inf_2.

Proof. Assume that there is a derivation of the empty sequent without rule Inf_2. According to theorem 4.4.8 there is then a derivation of the empty sequent without Cut (and this new derivation without Cut is also without Inf_2 and Ind). Furthermore, we note that in a cut-free simple derivation of the empty sequent all sequents have an empty antecedent, since formulas in the antecedent can only disappear through cut. Therefore, there are no initial sequents or instances of contraction in the derivation, but only arithmetical rules.

Now, the last rule of the derivation must be Inf_1, because all other rules give as a conclusion a sequent with a formula in the succedent. Thus, we have a derivation of the sequent $\rightarrow s(t) = 0$ for some term t.

In a simple derivation there are only closed terms and every term therefore has a value. We now prove by induction on the length of the derivation that every sequent in the derivation of $\rightarrow s(t) = 0$ has the property that the succedent is a formula $t = t'$ where t and t' have the same value.

Base case of the induction. As stated we have no initial sequents in the derivation and thus, we only consider the conclusions of the arithmetical rules without premises as the base case. We want to prove that the terms of the principal formula in the succedent have the same value.

In Ref both terms of the principal formula, $t = t$, have the same value. In $+Rec0$ the terms $t + 0$ and t of the principal formula,

$t + 0 = t$, have the same value. In $+Recs$ the principal formula is of the form $t + s(t') = s(t + t')$. Both $t + s(t')$ and $s(t + t')$ in $+Recs$ have the value 1. In $\cdot Rec0$ the principal formula is of the form $t \cdot 0 = 0$. The constant 0 has the value 0 and the term $t \cdot 0$ therefore also has the same value. In $\cdot Recs$ the principal formula is of the form $t \cdot s(t') = t \cdot t' + t$. If the term t has the value 1, then both terms $t \cdot s(t')$ and $t \cdot t' + t$ have the value 1. If t on the other hand has the value 0, then both terms have the value 0.

Induction step. Assume as induction hypothesis that the proposition holds for the premises of an arithmetical rule, that is, that the terms of the formulas in the succedents of the premises have the same value.

In *Sym* we can conclude that if the terms t and t' in the formula $t = t'$ have the same value, then the same applies for the formula $t' = t$ in the conclusion. In Tr we can see that if the terms of the formula $t = t'$ and $t' = t''$ have the same values, then the terms of the formula $t = t''$ have the same value. In *sRep* both terms of the formula $s(t) = s(t')$ in the conclusion have the value 1. In $+Rep_1$, if the terms of the formula $t = t'$ in the premise have the same value, then also the terms of the formula $t + t'' = t' + t''$ in the conclusion have the same value. The same holds for rule $+Rep_2$ and the $\cdot Rep$-rules.

Because all sequents in the derivation have an empty antecedent, rule Inf_1 gives the empty sequent as the conclusion and thus it can occur only as the last rule in the derivation.

Thus, we have completed the induction and have proved that in a simple derivation of the sequent $\rightarrow s(t) = 0$, all sequents have in the succedent an equation where the terms have the same value. On the other hand the terms $s(t)$ and 0 have different values. This is a contradiction and therefore there cannot exist any simple derivation of the empty sequent. □

4.4.12 Lemma. *If we have a derivation of a sequent $\Gamma \rightarrow D$, then there is a derivation of the same length of the sequent where all instances of Sym come directly after arithmetical rules without premises or after initial sequents.*

Proof. Suppose that we have a premise of *Sym* derived by a rule that is not an arithmetical rule without premises. If the rule is a one-premise arithmetical rule, that is *sRep*, *+Rep*, *·Rep*, or Inf_2, we can permute the instance of *Sym* above the other rule. If we have two instances of *Sym*, we have a loop and can delete both rules. If the rule is logical (except $L\vee$), structural or an instance of *Inf*, we can also permute *Sym* above the other rule.

If the rule is an instance of *Tr*, then the derivation is:

$$\dfrac{\dfrac{\Gamma_1 \to t = t' \quad \Gamma_2 \to t' = t''}{\Gamma_{1-2} \to t = t''}\,{\scriptstyle Tr}}{\Gamma_{1-2} \to t'' = t}\,{\scriptstyle Sym}$$

We can then instead apply *Sym* on each premise followed by *Tr*.

$$\dfrac{\dfrac{\Gamma_2 \to t' = t''}{\Gamma_2 \to t'' = t'}\,{\scriptstyle Sym} \quad \dfrac{\Gamma_1 \to t = t'}{\Gamma_1 \to t' = t}\,{\scriptstyle Sym}}{\Gamma_{1-2} \to t'' = t}\,{\scriptstyle Tr}$$

This does not alter the length of the derivation. The case of $L\vee$ is similar. □

4.4.13 Lemma. *There is a derivation of the sequent* $\to 0 \cdot c = 0$ *(without* Inf_2*) for every closed term* c*.*

Proof. Firstly we show by induction that for every numeral $\overline{m}$ we have a derivation of the sequent $\to 0 \cdot \overline{m} = 0$. We can derive $\to 0 \cdot 0 = 0$ with *·Recs*. Now assume that $\overline{m}$ is $s\overline{n}$ for some numeral $\overline{n}$ and we have a derivation of $\to 0 \cdot \overline{n} = 0$. We then get the derivation

$$\dfrac{\dfrac{}{\to 0 \cdot s(\overline{n}) = 0 \cdot \overline{n} + 0}\,{\scriptstyle \cdot Recs} \quad \dfrac{\dfrac{}{\to 0 \cdot \overline{n} + 0 = 0 \cdot \overline{n}}\,{\scriptstyle +Rec0} \quad \to 0 \cdot \overline{n} = 0}{\to 0 \cdot \overline{n} + 0 = 0}\,{\scriptstyle Tr}}{\to 0 \cdot s(\overline{n}) = 0}\,{\scriptstyle Tr}$$

Thus, the proposition holds for every numeral.

For every closed term c there is a numeral $\overline{m}$ for which the sequent $\to c = \overline{m}$ is derivable (without Inf_2), this according to lemma 4.4.5(i). We then get the sought derivation

$$\dfrac{\dfrac{\to c = \overline{m}}{\to 0 \cdot c = 0 \cdot \overline{m}}\,{\scriptstyle \cdot Rep_2} \quad \to 0 \cdot \overline{m} = 0}{\to 0 \cdot c = 0}\,{\scriptstyle Tr}$$

□

4.4.14 Lemma. *If there is a simple derivation of the sequent* $\to s(t) = s(t')$ *without the rule* Inf_2*, then there is a simple derivation of the sequent* $\to t = t'$ *without* Inf_2*.*

Proof. The proof is by induction on the length of the derivation. We assume that if there is a shorter derivation of some sequent $\to s(a) = s(b)$, then we have a derivation of $\to a = b$ without rule Inf_2.

Assume that we have a simple derivation of a sequent $\to s(t) = s(t')$ without Inf_2. We can by theorem 4.4.8 assume that the derivation is cut free. Thus, every sequent in the derivation has an empty antecedent. By lemma 4.4.12 we can assume that all instances of Sym come directly after arithmetical rules without premises (note that there are no initial sequents in the derivation because the antecedents are empty).

We consider the form of the derivation. The last rule can be *sRep*, Ref, Sym, or Tr.

1. Assume that the last rule of the derivation is *sRep*. The premise of the rule is $\to t = t'$ and we can remove the rule and get the sought derivation.

2. Assume that the last rule is Ref. Then $t \equiv t'$ and the sequent $\to t = t'$ is also derivable with Ref.

3. Assume that the last rule is Sym. Since the premise of Sym is derived by an arithmetical rule without premises the only possibility is that this rule is Ref. The case is treated as in case 2.

4. The remaining possibility is that the last rule is derived by Tr. We trace up in the derivation along the left premise until we reach a sequent not derived by Tr. The derivation is of the form

$$\dfrac{\dfrac{\dfrac{\to s(t) = a_1 \quad \to a_1 = a_2}{\to s(t) = a_2}\,Tr}{\vdots\; Tr - rules}}{} \qquad \dfrac{\to s(t) = a_n \qquad\qquad \to a_n = s(t')}{\to s(t) = s(t')}\,Tr \tag{4.4.15}$$

where $n \geqslant 1$ and the sequent $\to s(t) = a_1$ is not derived by Tr.

If one of the other Tr-premises $\to a_i = a_{i+1}$ is derived by Tr

$$\dfrac{\to s(t) = a_i \qquad \dfrac{\to a_i = a \quad \to a = a_{i+1}}{\to a_i = a_{i+1}}}{\to s(t) = a_{i+1}}$$

we can change the order of the Tr-rules without altering the length of the derivation.

$$\dfrac{\dfrac{\to s(t) = a_i \quad \to a_i = a}{\to s(t) = a} \quad \to a = a_{i+1}}{\to s(t) = a_{i+1}}$$

Hence, we can assume that the derivation is of the form 4.4.15 and that none of the premises $\to a_i = a_{i+1}$ have been derived by Tr. When a derivation has the form of derivation 4.4.15, then the right premises of the two consecutive Tr rules are called *adjecent*.

If some term a_i is of the form $s(t'')$, then the sequent $\to s(t) = a_i$ is the sequent $\to s(t) = s(t'')$. We can then alter the order of the Tr-rules and get a derivation of the same length.

$$\dfrac{\to s(t) = s(t'') \quad \dfrac{\begin{array}{c}\to s(t'') = a_{i+1} \\ \vdots \\ \to s(t'') = a_n \end{array} \quad \to a_n = s(t')}{\to s(t'') = s(t')}}{\to s(t) = s(t')}$$

The derivations of the sequents $\to s(t) = s(t'')$ and $\to s(t'') = s(t')$ are shorter and we therefore have derivations of the sequents $\to t = t''$ and $\to t'' = t'$. By Tr we get the sought derivation of $\to t = t'$.

We can now assume that the derivation has the form 4.4.15 and that no term a_i has the form $s(t'')$. We consider the different possibilities to derive the Tr-premises.

4.1 Assume that one of the premises has been derived by Ref. We now have a loop in the derivation since the conclusion of the following Tr is the same as the other premise. We can delete the rule Tr and get a shorter derivation. Thus, we may assume that no premise has been derived by Ref.

4.2 Assume that two adjacent Tr-premises have been derived by the same replacement rule $+Rep_1, +Rep_2, \cdot Rep_1$, or $\cdot Rep_2$ or that three adjacent Tr-premises have been derived by two instances of the same replacement rule with one instance of the other replacement rule in between. As an example we consider the following derivation.

$$\dfrac{\dfrac{\to s(t)=a+b \qquad \dfrac{\to b=c}{\to a+b=a+c}\,{\scriptstyle +Rep_2}}{\to s(t)=a+c}\,{\scriptstyle Tr} \qquad \dfrac{\to c=d}{\to a+c=a+d}\,{\scriptstyle +Rep_2}}{\to s(t)=a+d}\,{\scriptstyle Tr}$$

We can then apply Tr on the premises of the replacement rules and get a shorter derivation.

$$\dfrac{\to s(t)=a+b \qquad \dfrac{\dfrac{\to b=c \quad \to c=d}{\to b=d}\,{\scriptstyle Tr}}{\to a+b=a+d}\,{\scriptstyle +Rep_2}}{\to s(t)=a+d}\,{\scriptstyle Tr}$$

Thus, we can assume that we at most have two adjacent Tr-premises derived by $+Rep$ or $\cdot Rep$ and that these rules have different indexes.

4.3 Assume that some of the Tr-premises have been derived by Sym and $+Rec0$. We consider the rightmost premise derived in this way. It cannot be the last Tr-premise $\to a_n = s(t)$ since the sequent is of the form $\to a_i = a_i + 0$. Thus, the derivation is of the form

$$\dfrac{\dfrac{\dfrac{\dfrac{}{\to a_i+0=a_i}\,{\scriptstyle +Rec0}}{\to a_i=a_i+0}\,{\scriptstyle Sym}}{\to s(t)=a_i+0}\,{\scriptstyle Tr?} \qquad \dfrac{}{\to a_i+0=b}\,{\scriptstyle R}}{\to s(t)=b}\,{\scriptstyle Tr} \tag{4.4.16}$$

where $Tr?$ indicates that if $a_i \equiv s(t)$ we have no rule there, but if $a_i \not\equiv s(t)$ we have a Tr-rule there.

Rule R can according to the form of the term be Sym, $+Rec0$, $+Rep_1$, or $+Rep_2$ and if the rule is Sym, then the premise can be derived by $\cdot Recs$. We consider the different alternatives.

4.3.1 Assume that R is $+Rec0$. Then $b \equiv a_i$. If $a_i \equiv s(t)$, then we have derived an instance of Ref and if $a_i \not\equiv s(t)$, then we have a loop in the derivation with the sequent $\to s(t) = a_i$ two times. By eliminating the loop we get a shorter derivation.

4.3.2 Assume that R is $+Rep_1$. Now $b \equiv c+0$ and the derivation

4.4.16 is

$$\dfrac{\dfrac{\dfrac{\dfrac{}{\to a_i + 0 = a_i}\,{+Rec0}}{\to a_i = a_i + 0}\,Sym}{\to s(t) = a_i + 0}\,Tr? \qquad \dfrac{\to a_i = c}{\to a_i + 0 = c + 0}\,{+Rep_1}}{\to s(t) = c + 0}\,Tr$$

We can transform the derivation into a shorter derivation.

$$\dfrac{\dfrac{\to a_i = c}{\to s(t) = c}\,Tr? \qquad \dfrac{\dfrac{}{\to c + 0 = c}\,{+Rec0}}{\to c = c + 0}\,Sym}{\to s(t) = c + 0}\,Tr$$

4.3.3 Assume that R is Sym and that the premise of this rule is derived by $\cdot Recs$. Now $a_i \equiv 0 \cdot c$, $b \equiv 0 \cdot s(c)$ and the derivation 4.4.16 is

$$\dfrac{\dfrac{\to s(t) = 0 \cdot c \qquad \dfrac{\dfrac{}{\to 0 \cdot c + 0 = 0 \cdot c}\,{+Rec0}}{\to 0 \cdot c = 0 \cdot c + 0}\,Sym}{\to s(t) = 0 \cdot c + 0}\,Tr \qquad \dfrac{\dfrac{}{\to 0 \cdot s(c) = 0 \cdot c + 0}\,{\cdot Recs}}{\to 0 \cdot c + 0 = 0 \cdot s(c)}\,Sym}{\to s(t) = 0 \cdot s(c)}\,Tr$$

According to lemma 4.4.13 there is a derivation of the sequent $\to 0 \cdot s(c) = 0$ (without rule Inf_2). With Tr we get a derivation of the sequent $\to s(t) = 0$ without Inf_2. Thus, applying Inf_1 we get a derivation of the empty sequent without Inf_2. This is a contradiction according to lemma 4.4.11.

4.3.4 Assume that R is $+Rep_2$. Then $b \equiv a_i + c$ and we have another Tr-premise to the right derived by a rule R'. The derivation 4.4.16 is

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{\to a_i + 0 = a_i}\,{+Rec0}}{\to a_i = a_i + 0}\,Sym}{\to s(t) = a_i + 0}\,Tr? \qquad \dfrac{\to 0 = c}{\to a_i + 0 = a_i + c}\,{+Rep_2}}{\to s(t) = a_i + c}\,Tr \qquad \dfrac{}{\to a_i + c = d}\,R'}{\to s(t) = d}\,Tr \tag{4.4.17}$$

Considering the form of the formula $a_i + c = d$ the rule R' can be $Sym, +Rec0, +Recs$, or $+Rep_1$ (note that according to 4.2 the rule cannot be $+Rep_2$) and if it is Sym, then the Sym-premise can only be derived by $\cdot Recs$. We consider the different possibilities.

4.3.4.1 Assume that R' is $+Rec0$. The derivation is treated as in case 4.3.1.

4.3.4.2 Assume that R' is $+Recs$. Now $c \equiv s(e)$ and $d \equiv s(a_i + e)$. The sequent $\to 0 = c$ is then $\to 0 = s(e)$. This gives a contradiction as in case 4.3.3.

4.3.4.3 Assume that R' is $+Rep_1$. Now $d \equiv a_i + e$ and the derivation 4.4.17 is

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{\to a_i + 0 = a_i}\;{\scriptstyle +Rec0}}{\to a_i = a_i + 0}\;{\scriptstyle Sym}}{\to s(t) = a_i + 0}\;{\scriptstyle Tr?} \qquad \dfrac{\to 0 = c}{\to a_i + 0 = a_i + c}\;{\scriptstyle +Rep_2}}{\to s(t) = a_i + c}\;{\scriptstyle Tr} \qquad \dfrac{\to a_i = e}{\to a_i + c = e + c}\;{\scriptstyle +Rep_1}}{\to s(t) = e + c}\;{\scriptstyle Tr}$$

We can transform the derivation into a shorter derivation.

$$\dfrac{\dfrac{\dfrac{\to a_i = e \qquad \dfrac{\dfrac{}{\to e + 0 = e}\;{\scriptstyle +Rec0}}{\to e = e + 0}\;{\scriptstyle Sym}}{\to a_i = e + 0}\;{\scriptstyle Tr} \qquad \dfrac{\to 0 = c}{\to e + 0 = e + c}\;{\scriptstyle +Rep_2}}{\to a_i = e + c}\;{\scriptstyle Tr}}{\to s(t) = e + c}\;{\scriptstyle Tr?}$$

4.3.4.4 Assume that R' is Sym and that the Sym-premise has been derived by $\cdot Recs$. Now $a_i \equiv c \cdot e$, $d \equiv c \cdot s(e)$ and the conclusion of derivation 4.4.17 is $\to s(t) = c \cdot s(e)$. We get a simple derivation of the sequent $\to s(t) = 0$ without Inf_2, since according to lemma 4.4.13 we have a simple derivation of the sequent $\to 0 \cdot s(e) = 0$.

$$\dfrac{\dfrac{\to s(t) = c \cdot s(e) \qquad \dfrac{\dfrac{\to 0 = c}{\to c = 0}\;{\scriptstyle Sym}}{\to c \cdot s(e) = 0 \cdot s(e)}\;{\scriptstyle \cdot Rep_1}}{\to s(t) = 0 \cdot s(e)}\;{\scriptstyle Tr} \qquad \to 0 \cdot s(e) = 0}{\to s(t) = 0}\;{\scriptstyle Tr}$$

This is a contradiction as in case 4.3.3.

We have now treated all the possibilities of rule R' and case 4.3.4 is finished. We have also treated all cases in 4.3 and thus, we can assume that no Tr-premise in derivation 4.4.15 has been derived by Sym and $+Rec0$.

4.4 We consider derivation 4.4.15. The leftmost Tr-premise $\to s(t) = a_1$ can only be derived by Sym and the premise of Sym by

$+Recs$. The following Tr-premise can be derived by $+Rep_1, +Rep_2$, Sym, or $+Recs$ and if it is derived by Sym, then the Sym-premise is derived by $\cdot Recs$. We treat the different cases simultaneously, since the derivation will ultimately have the same form disregarding some Rep-rules and possible instances of $\cdot Recs$. According to case 4.2 we can only have two adjacent Tr-premises derived by the $+Rep$-rules. We assume that we have one premise derived by $+Rep_1$ and one by $+Rep_2$. The following Tr-premise can be derived by $+Recs, +Rec0$, or Sym and $\cdot Recs$. If it is derived by $+Rec0$ we get a contradiction as in case 4.3.3. We assume that the premise is derived by Sym and $\cdot Recs$. The following two premises can be derived by $\cdot Rep_1$ and $\cdot Rep_2$ and the next only by $\cdot Recs$, because if it is derived by $\cdot Rec0$ we have a contradiction as in case 4.3.3. Again we can have two $+Rep$-rules and a number of repetitions of the rules $\cdot Recs, \cdot Rep_1, \cdot Rep_2, \cdot Recs, +Rep_1$ and $+Rep_2$. The last Tr-premise is derived by $+Recs$.

Hence, the derivation has the following form (where we have left out the sequent arrow and unnecessary parentheses):

$$\dfrac{\dfrac{\dfrac{\dfrac{}{a+sb=s(a+b)}\,{\scriptstyle +Recs}}{s(a+b)=a+sb}\,{\scriptstyle Sym} \qquad \dfrac{a=c}{a+sb=c+sb}\,{\scriptstyle +Rep_1}}{s(a+b)=c+sb}\,{\scriptstyle Tr} \qquad \dfrac{sb=d}{c+sb=c+d}\,{\scriptstyle +Rep_2}}{s(a+b)=c+d}\,{\scriptstyle Tr}$$

From the rule $\cdot Recs$ we have $c \equiv d \cdot e$.

$$\dfrac{\dfrac{\begin{matrix}\vdots\\ s(a+b)=c+d\end{matrix} \qquad \dfrac{\dfrac{}{d\cdot se=c+d}\,{\scriptstyle \cdot Recs}}{c+d=d\cdot se}\,{\scriptstyle Sym}}{s(a+b)=d\cdot se}\,{\scriptstyle Tr} \qquad \dfrac{d=f}{d\cdot se=f\cdot se}\,{\scriptstyle \cdot Rep_1}}{s(a+b)=f\cdot se}\,{\scriptstyle Tr}$$

From the rule $\cdot Recs$ we have $g \equiv sh$.

$$\dfrac{\dfrac{\begin{matrix}\vdots\\ s(a+b)=f\cdot se\end{matrix} \qquad \dfrac{se=g}{f\cdot se=f\cdot g}\,{\scriptstyle \cdot Rep_2}}{s(a+b)=f\cdot g}\,{\scriptstyle Tr} \qquad \dfrac{}{f\cdot g=f\cdot h+f}\,{\scriptstyle \cdot Recs}}{s(a+b)=f\cdot h+f}\,{\scriptstyle Tr}$$

$$\dfrac{\dfrac{\begin{matrix}\vdots\\ s(a+b)=f\cdot h+f\end{matrix} \quad \dfrac{f\cdot h=c_2}{f\cdot h+f=c_2+f}\,{\scriptstyle +Rep_1}}{s(a+b)=c_2+f}\,{\scriptstyle Tr} \qquad \dfrac{f=d_2}{c_2+f=c_2+d_2}\,{\scriptstyle +Rep_2}}{s(a+b)=c_2+d_2}\,{\scriptstyle Tr}$$

From the formula $s(a+b) = c_2 + d_2$ we can have a repetition of $\cdot Recs$ and Rep-rules. If we have $n-1$ repetitions, where $n \geqslant 1$, then the end of the derivation is

$$\dfrac{\begin{matrix}\vdots\\ s(a+b)=c_n+d_n\end{matrix} \quad \dfrac{}{c_n+d_n=s(a_2+b_2)}\,{\scriptstyle +Recs}}{s(a+b)=s(a_2+b_2)}\,{\scriptstyle Tr} \tag{4.4.18}$$

Here we have $c_n \equiv a_2$ and $d_n \equiv sb_2$ and also $a_2 + b_2 \equiv t'$.

If in the derivation we have at least one row of the specified rules, that is, if $n > 1$, then we show that we can derive $c_i = c_{i+1}$ and $d_i = d_{i+1}$. If we don't have all Rep-rules in the derivation, then we have identities instead of equations and the derivation is shorter.

In the derivation we have subderivations of the formulas $d_i = f_i$ and $f_i \cdot h_i = c_{i+1}$ and we also have the identity $c_i \equiv d_i \cdot e_i$. Since $g_i \equiv sh_i$ and we have a subderivation of $se_i = g_i$, that is $se_i = sh_i$, we have by the induction hypothesis a derivation of $e_i = h_i$. Thus, we can construct a derivation of $c_i = c_{i+1}$.

$$\dfrac{\dfrac{\dfrac{d_i=f_i}{d_i\cdot e_i=f_i\cdot e_i}\,{\scriptstyle \cdot Rep_1} \quad \dfrac{e_i=h_i}{f_i\cdot e_i=f_i\cdot h_i}\,{\scriptstyle \cdot Rep_2}}{d_i\cdot e_i=f_i\cdot h_i}\,{\scriptstyle Tr} \qquad f_i\cdot h_i=c_{i+1}}{d_i\cdot e_i=c_{i+1}}\,{\scriptstyle Tr}$$

On the other hand we get $d_i = d_{i+1}$ with Tr from $d_i = f_i$ and $f_i = d_{i+1}$.

With Tr we get derivations of $c = c_n$ and $d = d_n$. We now construct a derivation of $t = t'$, that is $a + b = a_2 + b_2$. From the subderivation of $a = c$ and the derivation of $c = c_n$, we get with Tr a derivation of $a = c_n$. Since $c_n \equiv a_2$ we now have a derivation of $a = a_2$.

From the subderivation of $sb = d$ and the derivation of $d = d_n$ we get with Tr a derivation of $sb = d_n$. Since $d_n \equiv sb_2$ we have a derivation of $sb = sb_n$ and this derivation is shorter. According to the induction hypothesis we have a derivation of $b = b_n$. We now get the sought derivation

$$\dfrac{\dfrac{a = a_2}{a + b = a_2 + b}\,{+Rep_1} \qquad \dfrac{b = b_2}{a_2 + b = a_2 + b_2}\,{+Rep_2}}{a + b = a_2 + b_2}\,Tr$$

Hence, we have treated case 4.4 and also case 4 is finished. □

4.4.19 Lemma. *If there is a simple derivation of the sequent $\rightarrow t = t'$, then there is a derivation of the same sequent without rule Inf_2*

Proof. Assume that the sequent $\rightarrow t = t'$ is derivable with at least one instance of Inf_2 in the derivation. Then take an uppermost instance of Inf_2. The premise of this rule is $\rightarrow s(u) = s(v)$. According to lemma 4.4.14 the conclusion of the rule $\rightarrow u = v$ is derivable without Inf_2. Thus, we can replace the subderivation with this derivation without Inf_2. In this way we can remove every instance of Inf_2 in the derivation. □

4.4.20 Lemma. *There is no simple derivation of the empty sequent.*

Proof. Assume that we have a simple derivation of the empty sequent. According to theorem 4.4.8 there is a cut-free derivation of the sequent. The last rule of this derivation must be Inf_1 with a premise $\rightarrow s(t) = 0$ because all other rules give as the conclusion a sequent with a formula in the succedent. According to lemma 4.4.19 the premise is derivable without Inf_2. Therefore, we also have a derivation of the empty sequent without Inf_2. This is a contradiction according to lemma 4.4.11 and thus, there cannot be any simple derivation of the empty sequent. □

Gentzen and Takeuti use semantical arguments in their proofs of this lemma, while we managed to complete the proof using purely proof-theoretical means. Takeuti proves that there is either a false formula in the antecedent of a sequent in a simple proof or a true formula in the succedent. He needs these semantical arguments because he has arbitrary initial sequents in his system only specified by the requirement that they have a true atomic formula with closed terms in the succedent or a false formula in the antecedent. We managed to abolish the semantical arguments of the lemma through our formulation of the system HA.

4.4.5 The reduction procedure for derivations

We can now begin to describe the actual reduction procedure for derivations of the empty sequent. The main idea of the proof is that we first substitute free variables in the derivation. Then according to the form of the derivation we convert inductions or cuts on compound formulas with predecessors in arithmetical rules or initial sequents. If this is not possible, then we have a so-called suitable cut. If we have a suitable cut, then we can introduce cuts on formulas of a lower grade. The problematic case is that if there are contractions on the cut formula we cannot directly convert the suitable cut into cuts on formulas of lower grade. The problem is solved by the so-called height lines that are permuted up in the derivation by introducing additional cuts on formulas of lower grade, lowering the ordinal of the derivation.

4.4.21 Lemma (Reduction Procedure)**.** *If P is a derivation of the empty sequent $\rightarrow$ in which the arithmetical rules are applied before the logical and structural rules, then there exists a derivation P' of the empty sequent such that $o(P') < o(P)$.*

Proof. The proof describes a reduction procedure where a derivation P is transformed into a derivation P' with a lower ordinal. The reduction consists of several steps, which are performed as many times as possible before proceeding to the next step and the reduction ends when a derivation with a lower ordinal is reached.

Let P be a derivation of the empty sequent $\rightarrow$. We may assume that the eigenvariables of the rules are different and that an eigenvariable occurs only above the rule in the derivation.

Step 1. If there are any free variables in the derivation that are not eigenvariables, we substitute them with the constant 0. The derivation that results from this process is also a valid derivation of the empty sequent and it has the same ordinal as P.

Step 2. If the end-piece of P contains an induction, then we perform the following reduction. Assume I to be one of the last inductions of the derivation.

$$\dfrac{\begin{array}{c}\vdots\\ \Gamma_1 \overset{\mu_1}{\to} A(0)\end{array} \quad \begin{array}{c}P_0(x)\\ \vdots\\ A(x),\Gamma_2 \overset{\mu_2}{\to} A(sx)\end{array} \quad \begin{array}{c}\vdots\\ A(t),\Gamma_3 \overset{\mu_3}{\to} D \quad (l)\end{array}}{\begin{array}{c}\Gamma_{1-3} \to D \quad (k)\\ \vdots\\ \to\end{array}}\; I$$

Here $P_0(x)$ is the subderivation ending with $A(x), \Gamma_2 \to A(sx)$ and S is the sequent $\Gamma_{1-3} \to D$. The premises of I all have the same height, l. Let k be the height of the conclusion of the rule and let μ_i, where $i = 1, 2, 3$, be the ordinals of the premises. Now the conclusion has the ordinal $o(\Gamma_{1-3} \to D; P) = \omega_{l-k+1}(\mu_1 \# \mu_2 \# \mu_3)$.

The term t in the third premise of the rule does not contain any free variable since they were substituted in step 1. Neither does t contain any eigenvariables because I is the last rule with an eigenvariable in the derivation. Thus, t is a closed term and there exists a number n, for which the sequent $\to t = \overline{n}$ is derivable without inductions or cuts (this according to lemma 4.4.6(i)). Therefore, we have a derivation, Q, of the sequent $A(\overline{n}) \to A(t)$ also without inductions or cuts. This according to lemma 4.4.6(iv).

The derivation P can now be reduced to P' according to the following principle if $n > 0$. If n equals 0 the corresponding reduction is used but no contractions are needed and instead the missing context Γ_2 is added in the derivation.) Let $P_0(\overline{m})$ be the derivation that results from $P_0(x)$ when every occurrence of x is substituted with $\overline{m}$ and let Π be the derivation:

$$\dfrac{\dfrac{\begin{array}{c}\vdots\\ \Gamma_1 \to A(0)\end{array} \quad \begin{array}{c}P_0(\overline{0})\\ \vdots\\ A(0),\Gamma_2 \to A(s0)\end{array}}{\Gamma_1,\Gamma_2 \to A(s0)}\;Cut \quad \begin{array}{c}P_0(\overline{1})\\ \vdots\\ A(s0),\Gamma_2 \to A(ss0)\end{array}}{\begin{array}{c}\Gamma_1,\Gamma_2^2 \to A(ss0)\\ \vdots\\ \Gamma_1,\Gamma_2^n \to A(\overline{n})\end{array}}\;Cut$$

We reduce P to the following derivation P' where Π is a subderivation:

$$\dfrac{\dfrac{\begin{array}{c}\Pi\\ \vdots\\ \Gamma_1, \Gamma_2^n \to A(\overline{n})\end{array} \quad \begin{array}{c}Q\\ \vdots\\ A(\overline{n}) \to A(t)\end{array}}{\Gamma_1, \Gamma_2^n \to A(t)}\,{\scriptstyle Cut} \qquad \begin{array}{c}\vdots\\ A(t), \Gamma_3 \to D\end{array}}{\Gamma_1, \Gamma_2^n, \Gamma_3 \to D}\,{\scriptstyle Cut}$$

$$\begin{array}{c} \vdots \;\; \text{Contractions} \\ \Gamma_{1-3} \to D \\ \vdots \\ \to \end{array}$$

All cuts shown in Π and P' are on formulas of the same grade, so all cut premises have the same height l. Therefore, the ordinals of the premises of the first cut in Π are $o(\Gamma_1 \to A(0); P') = \mu_1$ and $o(A(0), \Gamma_2 \to A(s0); P') = \mu_2$. The ordinal of the conclusion, S_1', is then $o(S_1') = \omega_{l-l}(\mu_1 \# \mu_2) = \mu_1 \# \mu_2$. The conclusion of the second cut, S_2', then has the ordinal $o(S_3') = \mu_1 \# \mu_2 \# \mu_2$ and so on. If we write $\mu * m = \mu \# \mu \# \ldots \# \mu (m \text{ times})$ we get $o(S_m') = \mu_1 \# (\mu_2 * m)$ for every $m = 1, \ldots, n$. If we denote the ordinal of Q by q, we have $o(A(\overline{n}) \to A(t)) = q < \omega$ because Q does not contain any inductions or cuts. Because each of the ordinals $\mu_1, \mu_2 * n, q$, and μ_3 are less than $\omega^{\mu_1 \# \mu_2 \# \mu_3}$, the sum is also less, that is we have the inequality $\mu_1 \# (\mu_2 * n) \# q \# \mu_3 < \omega^{\mu_1 \# \mu_2 \# \mu_3}$. From this follows that $o(S; P') = \omega_{k-l}(\mu_1 \# (\mu_2 * n) \# q \# \mu_3) < \omega_{l-k+1}(\mu_1 \# \mu_2 \# \mu_3) = o(S; P)$, that is $o(S; P') < o(S; P)$. According to lemma 4.4.4 we then have $o(P') < o(P)$.

Thus, if there is an induction in the end-piece we have reduced the derivation. Otherwise we can assume that the end-piece is free from inductions.

Step 3. Assume that there is a compound formula E in the end-piece of the derivation. Let I be the cut in the end-piece where the formula disappears. No predecessor of the formula in the left cut premise can be derived by an arithmetical rule that borders on the end-piece since the formula E has logical structure. Now assume that a predecessor of the formula in the right cut premise has been derived by an arithmetical rule that borders on the end-piece.

$$
\begin{array}{c}
\dfrac{\begin{array}{cc} & \overline{E,\Gamma_2' \to D'}\ {\scriptstyle Arithm.} \\ \vdots & \vdots \\ \Gamma_1 \to E & E,\Gamma_2 \to D \quad (k) \end{array}}{\Gamma_{1-2} \to D \quad (l)}\ {\scriptstyle I} \\ \vdots \\ \to
\end{array}
$$

Above the arithmetical rule that borders on the end-piece we have only other arithmetical rules and initial sequents (this according to the assumption made in the beginning of the proof.) The formula E is therefore not principal in any rule above the arithmetical rule and it cannot be introduced in an initial sequent as the formula on both sides either, since no succedent of a sequent above the arithmetical rule can be compound.

Hence, the formula E has been introduced in the context of an arithmetical rule without premises or in an initial sequent and we can eliminate the formula and trace down in the derivation deleting the formula in the context of every arithmetical rule. Thus, the derivation that results from this process is a derivation of the sequent $\Gamma_2' \to D'$ that is otherwise similar to the derivation of $E, \Gamma_2' \to D'$.

We now divide the reduction into two cases depending on whether we have any contractions on the formula E between the arithmetical rule that borders on the end-piece and the cut I where the formula disappears.

Case 1. Assume that there are no contractions on the formula E between the arithmetical rule and I. We now continue deleting every occurrence of E and also the cut I, instead adding the missing context Γ_1 in the antecedent. Thus, we have a valid derivation of the sequent $\Gamma_{1-2} \to D$ and the derivation P' is as follows:

$$
\begin{array}{c}
\overline{\Gamma_1, \Gamma_2' \to D'}\ {\scriptstyle Arithm.} \\ \vdots \\ \Gamma_{1-2} \to D \\ \vdots \\ \to
\end{array}
$$

Now in order to calculate the ordinal of the new derivation let S

be a sequent in P above $E, \Gamma_2 \to D$ and let S' be the corresponding sequent in P'. We then show by induction on the number of inferences up to $E, \Gamma_2 \to D$ that the following inequality holds

$$\omega_{k_1-k_2}(o(S;P)) \geqslant o(S';P'), \tag{4.4.22}$$

where $k_1 = h(S;P)$ and $k_2 = h(S';P')$ and thus $k_1 \geqslant k_2$.

If S is an initial sequent or the conclusion of an arithmetical rule without premises, then $o(S;P) = o(S';P') = 1$ and the proposition holds. Now assume that the sequent S has been derived with a rule and that the claim holds for its premises. If S has been derived with contraction, the heights and the ordinals of the conclusions S and S' are the same as for the premises and the proposition holds.

If S has been derived with an arithmetical or logical one-premise rule then the heights of the conclusions are the same as for the premises. If we let the ordinals of the premises be α and α' we get $\omega_{k_1-k_2}(o(S;P)) = \omega_{k_1-k_2}(\alpha+1) > \omega_{k_1-k_2}(\alpha)$. Since the claim holds for the premises, that is $\omega_{k_1-k_2}(\alpha) \geqslant \alpha'$, we get $\omega_{k_1-k_2}(\alpha+1) > \alpha'$ and furthermore, $\omega_{k_1-k_2}(\alpha+1) \geqslant \alpha'+1$ and the proposition holds.

If S has been derived with an arithmetical or logical two-premise rule then again the heights of the conclusions are the same as for the premises. If we let the ordinals of the premises be α, β and α', β' we have the following inequalities for the premises of the rules $\omega_{k_1-k_2}(\alpha) \geqslant \alpha'$ and $\omega_{k_1-k_2}(\beta) \geqslant \beta'$. If $k_1 = k_2$, then we get from the inequalities of the premises $\alpha \geqslant \alpha'$ and $\beta \geqslant \beta'$ the inequality $\omega_{k_1-k_2}(o(S;P)) = o(S;P) = \alpha \# \beta \geqslant \alpha' \# \beta'$. On the other hand if $k_1 > k_2$, then we get $\omega_{k_1-k_2}(\alpha \# \beta) > \omega_{k_1-k_2}(\alpha) \geqslant \alpha'$ and $\omega_{k_1-k_2}(\alpha \# \beta) > \omega_{k_1-k_2}(\beta) \geqslant \beta'$. This gives $\omega_{k_1-k_2}(\alpha \# \beta) > \alpha' \# \beta'$ and the proposition holds.

If S has been derived with a cut the premises of which have the height m_1 and the ordinals α and β and S' has been derived with a cut the premises of which have the height m_2 and the ordinals α' and β', then we have the following inequalities for the premises $\omega_{m_1-m_2}(\alpha) \geqslant \alpha'$ and $\omega_{m_1-m_2}(\beta) \geqslant \beta'$. We then get $\omega_{k_1-k_2}(o(S;P)) = \omega_{k_1-k_2}(\omega_{m_1-k_1}(\alpha \# \beta)) = \omega_{m_1-k_2}(\alpha \# \beta) = \omega_{m_2-k_2}(\omega_{m_1-m_2}(\alpha \# \beta))$. If $m_1 = m_2$ then from the inequalities of the premises $\alpha \geqslant \alpha'$ and $\beta \geqslant \beta'$ we get the inequality $\omega_{m_2-k_2}(\omega_{m_1-m_2}(\alpha \# \beta)) = \omega_{m_2-k_2}(\alpha \# \beta) \geqslant \omega_{m_2-k_2}(\alpha' \# \beta')$. If $m_1 > m_2$ then we get $\omega_{m_1-m_2}(\alpha \# \beta) > \omega_{m_1-m_2}(\alpha)$

$\geqslant \alpha'$ and $\omega_{m_1-m_2}(\alpha\#\beta) > \omega_{m_1-m_2}(\beta) \geqslant \beta'$. Thus, we get $\omega_{m_1-m_2}(\alpha\#\beta) > \alpha'\#\beta'$ and from this follows that $\omega_{m_2-k_2}(\omega_{m_1-m_2}(\alpha\#\beta)) > \omega_{m_2-k_2}(\alpha'\#\beta')$, that is the proposition holds.

If S has been derived with an Ind the premises of which have the height m_1 and the ordinals α, β and γ and S' has been derived with an Ind the premises of which have the height m_2 and the ordinals α', β' and γ' then we have the following inequalities for the premises $\omega_{m_1-m_2}(\alpha) \geqslant \alpha'$, $\omega_{m_1-m_2}(\beta) \geqslant \beta'$ and $\omega_{m_1-m_2}(\gamma) \geqslant \gamma'$. We then have

$$\begin{aligned}\omega_{k_1-k_2}(o(S;P)) &= \omega_{k_1-k_2}(\omega_{m_1-k_1+1}(\alpha\#\beta\#\gamma)) \\ &= \omega_{m_1-k_2+1}(\alpha\#\beta\#\gamma) \\ &= \omega_{m_2-k_2+1}(\omega_{m_1-m_2}(\alpha\#\beta\#\gamma)) \\ &\geqslant \omega_{m_2-k_2+1}(\alpha'\#\beta'\#\gamma') = o(S';P')\end{aligned}$$

Thus, it has been proved that inequality 4.4.22 holds.

Now let S be the sequent $E, \Gamma_2 \rightarrow D$ and S' the corresponding sequent $\Gamma_{1-2} \rightarrow D$. If we let $o(\Gamma_1 \rightarrow E; P) = \mu_1, o(E, \Gamma_2 \rightarrow D; P) = \mu_2, o(\Gamma_{1-2} \rightarrow D; P) = \nu$ and $o(\Gamma_{1-2} \rightarrow D; P') = \nu'$ and also let $h(\Gamma_{1-2} \rightarrow D; P) = l$ and $h(E, \Gamma_2 \rightarrow D; P) = k$, then we have $l \leqslant k$ and $h(\Gamma_{1-2} \rightarrow D; P') = l$. From the inequality we get

$$\omega_{k-l}(\mu_2) \geqslant \nu'$$

and from this follows the inequality

$$\nu = \omega_{k-l}(\mu_1\#\mu_2) > \omega_{k-l}(\mu_2) \geqslant \nu'.$$

According to lemma 4.4.4 we can conclude that $o(P) > o(P')$.

Case 2. Assume that there is at least one contraction on the formula E between the arithmetical rule and I. Let the uppermost contraction be I'. Recall that we have a derivation of the sequent $\Gamma_2' \rightarrow D'$ that is otherwise similar to the derivation of $E, \Gamma_2' \rightarrow D'$. We can now reduce the derivation on the left into the one on the right by eliminating the contraction.

$$\dfrac{\dfrac{\dfrac{}{E,\Gamma_2'\rightarrow D'}\ \textit{Arithm.}}{\vdots}}{} \qquad \dfrac{\begin{array}{c}E,E,\Gamma_2''\rightarrow D''\end{array}}{E,\Gamma_2''\rightarrow D''}\ I' \qquad \begin{array}{c}\vdots\\ E,\Gamma_2\rightarrow D\\ \vdots\end{array} \quad\rightsquigarrow\quad \begin{array}{c}\overline{\Gamma_2'\rightarrow D'}\ \textit{Arithm.}\\ \vdots\\ E,\Gamma_2''\rightarrow D''\\ \vdots\\ E,\Gamma_2\rightarrow D\\ \vdots\end{array}$$

In this reduction the ordinal is preserved and $o(P) = o(P')$. We now repeat step 3 if we can or continue with step 4 and assume that compound formulas in the end-piece of P do not have predecessors in arithmetical rules that border on the end-piece. Therefore, these formulas must have predecessors in initial sequents or logical rules that border on the end-piece.

Step 4. Assume that the end-piece contains an initial sequent $D, \Gamma \rightarrow D$. Since the end-sequent is empty both formulas D (or rather descendants of both formulas) must disappear through cuts. Assume that the D in the antecedent is the first formula to disappear in a cut (the other case is similar). The derivation P now has the form

$$\dfrac{\begin{array}{c}\vdots\\ \Gamma_1\rightarrow D\end{array} \quad \begin{array}{c}D,\Gamma\rightarrow D\\ \vdots\\ D,\Gamma_2\rightarrow D\end{array}}{\Gamma_{1-2}\rightarrow D}\ \textit{Cut}$$
$$\begin{array}{c}\vdots\\ \rightarrow\end{array}$$

We can reduce P into a derivation P' where the cut has been eliminated by adding the missing context Γ_2 in the antecedent of the derivation of the left premise.

Since both D's from the sequent $D, \Gamma \rightarrow D$ disappear through cuts, we have a cut on the other D in the succedent below the sequent $\Gamma_{1-2} \rightarrow D$. Therefore, the heights of the sequents remain unchanged, while the ordinal of the subderivation ending with $\Gamma_{1-2} \rightarrow D$ decreases. Thus, we get $o(P') < o(P)$ by lemma 4.4.4.

We can now proceed to step 5 and can assume that the end-piece does not contain any initial sequents but only cuts and contractions.

Step 5. To continue the reduction procedure we consider the compound cut formulas of the end-piece. We want to diminish the ordinal of the derivation by introducing cuts on shorter formulas. For this we need a suitable cut in the end-piece.

4.4.23 Definition. A cut in the end-piece of a derivation is a *suitable cut* if both copies of the cut formula have predecessors that are principal in logical rules that border on the end-piece.

4.4.24 Sublemma. *Assume that a derivation P fulfills the following requirements:*

1. *the end-piece of P contains at least one cut on a compound formula.*

2. *In every cut on a compound formula in the end-piece each copy of the cut formula has a predecessor in the conclusion of a logical rule that borders on the end-piece.*

3. *The principal formula of the logical rule mentioned in point (2) has a descendant that disappears through a cut in the end-piece.*

Then P has a suitable cut.

Proof. The proof is an induction on the number of cuts on compound cut formulas in the end-piece.

In the end-piece of P there is at least one cut on a compound formula according to point (1). If there is only one cut, then the cut formulas of both premises have a predecessor in a logical rule bordering on the end-piece according to point (2). If the principal formula of the rule was not the predecessor of the cut formula, then, according to point (3), it would have to disappear through another cut in the end-piece. Thus, the principal formula has to be the predecessor of the only cut and we have a suitable cut.

Now assume that P has n cuts on compound formulas in the end-piece. As induction hypothesis we have that any derivation with fewer such cuts has a suitable cut, provided that the derivation fulfils the stipulated requirements. Let I be the last of the cuts on some compound formula, D.

$$\dfrac{\begin{matrix} P_1 & & P_2 \\ \vdots & & \vdots \\ \Gamma_1 \to D & & D, \Gamma_2 \to E \end{matrix}}{\Gamma_{1-2} \to E}\ I$$

If I is a suitable cut the proposition is proved. Therefore, we assume that I is not a suitable cut. Both cut formulas of the premises have, according to point (2) a predecessor in the conclusion of a logical rule bordering on the end-piece. Since the cut is not a suitable cut a predecessor of one D is not principal in one of the logical rules. We may assume that this is the case for the D in the premise $\Gamma_1 \to D$. According to point (3) a descendant of the principal formula in the logical rule disappears through a cut. If this cut was I, then the principal formula would be D, but then I would be a suitable cut. Therefore, there must be another cut on a compound formula and this cut is above I in P_1 since I was the last cut. Thus, P_1 satisfies point (1). P_1 also inherits property (2) from P. None of the principal formulas in the logical rules bordering on the end-piece can disappear through the cut I, since that would make I a suitable cut, therefore the cuts must be in P_1 and P_1 fulfills criterion (3). With that, the subderivation P_1 fulfills all three requirements and according to the induction hypothesis has a suitable cut. This is also a suitable cut of the derivation P. □

We now continue to consider the derivation P of the empty sequent. If the derivation P contained only atomic formulas, then any instances of Ind would be in the end-piece, but this is not possible since these were reduced in step 2. Hence, the derivation P contains a compound formula, for otherwise the derivation would be simple which is impossible according to lemma 4.4.20. Since the end-sequent is empty and the end-piece does not contain any instances of Ind all formulas in the end-piece must disappear through cuts. At least one of these formulas has logical structure. The derivation P therefore satisfies the first criterion in sublemma 4.4.24. Assume that D is a compound formula that disappears though a cut in the end-piece. The formula D cannot have a predecessor in an arithmetical rule that borders on the end-piece, since these were treated in step 3. Neither can a predecessor of D have been introduced in an initial sequent in the end-piece, since these were treated in step 4. The only remaining possibility is that the formula has a predecessor in the conclusion of a logical rule bordering on the end-piece. This means that P satisfies the second criterion in sublemma 4.4.24. From the fact that the end-sequent is empty and that there are no inductions in the end-piece we draw the conclusion that P satisfies the third criterion in lemma 4.4.24. Therefore, P fulfills all the requirements of the sublemma and P contains a suitable cut.

Now consider the lowermost suitable cut I and perform the following reduction according to the form of the cut formula.

Case 1. Assume that the cut formula of the last suitable cut is a conjunction $B\&C$. Now P has the form

$$
\dfrac{\dfrac{\dfrac{\dfrac{\vdots}{\Gamma_1'' \to B} \quad \dfrac{\vdots}{\Gamma_1''' \to C}}{\Gamma_1' \to B\&C}\,R\&}{\Gamma_1 \xrightarrow{\mu} B\&C} \qquad \dfrac{\dfrac{\dfrac{\vdots}{B, C, \Gamma_2' \to D'}}{B\&C, \Gamma_2' \to D'}\,L\&}{B\&C, \Gamma_2 \xrightarrow{\nu} D \quad (l)}}{\dfrac{\dfrac{\Gamma_{1-2} \to D}{\Theta \xrightarrow{\lambda} E \quad (k)}}{\to}}\,I
$$

where $\Gamma_1' = \Gamma_1'', \Gamma_1'''$ and $\Theta \to E$ is the first sequent below I that has a lower height than the premises of the cut. Such a sequent exists because the height of the end-sequent is 0 while the cut premises have a height of at least 1. Let l be the height of the premises of the cut I and let $h(\Theta \to E; P) = k$. Then we have $k < l$. The sequent $\Theta \to E$ must be the conclusion of a cut since the end-piece only contains contractions and cuts and the conclusion of a contraction has the same height as the premise. Furthermore, we let $o(\Gamma_1 \to B\&C) = \mu, o(B\&C, \Gamma_2 \to D) = \nu$ and $o(\Theta \to E) = \lambda$.

In the derivation of $B, C, \Gamma_2' \to D'$ we can add the formula $B\&C$ in the context and get a derivation of the sequent $B\&C, B, C, \Gamma_2' \to D'$.

Now let P_3 be the following derivation:

$$
\dfrac{\begin{array}{c}\vdots \\ \Gamma_1 \overset{\mu_3}{\to} B\&C\end{array} \quad \begin{array}{c}\vdots \\ B\&C, B, C, \Gamma_2' \to D' \\ \vdots \\ B\&C, B, C, \Gamma_2 \overset{\nu_3}{\to} D\end{array}}{\begin{array}{c}B, C, \Gamma_{1-2} \to D \\ \vdots \\ B, C, \Theta \to E\end{array}} \; J_3
$$

We take the derivation of $\Gamma_1'' \to B$ and instead of applying a right conjunction rule we add the missing formulas Γ_1''' in the context and get a derivation of the sequent $\Gamma_1' \to B$. Then we apply the cuts and contractions above the left premise of the cut J_3 shown in P_3 (this is possible because the descendant of the conjunction in the succedent disappears through the cut J_3 and therefore cannot be principal in another rule above the cut.) Hence, we have constructed a derivation of the sequent $\Gamma_1 \to B$. We again instead of applying the cut J_3 add the missing context Γ_2 and get a derivation of $\Gamma_{1-2} \to B$. After this we continue with the same rules as below P_3 applying the same rules on the same formulas if we have a contraction or a cut on formulas in the antecedent. If we on the other hand in P_3 have a cut on the formula in the succedent (that is a cut on the formula in P_3 that has been replaced by the formula B in the constructed derivation) we instead of applying the cut add the missing context in the antecedent of the sequent. Thus, we get a valid derivation of the sequent $\Theta \to B$ and we call this derivation P_1. Correspondingly we construct a derivation of the sequent $\Theta \to C$ from the derivation P_3 and call this derivation P_2.

We now compose the three derivations into the derivation P':

$$\dfrac{\begin{array}{c} P_1 \\ \vdots \\ \Theta \overset{\lambda_1}{\to} B \end{array} \qquad \dfrac{\begin{array}{c} P_2 \\ \vdots \\ \Theta \overset{\lambda_2}{\to} C \end{array} \quad \begin{array}{c} P_3 \\ \vdots \\ B, C, \Theta \overset{\lambda_3}{\to} E \quad (m_2) \end{array}}{B, \Theta^2 \to E \quad (m_1)}\, Cut}{\begin{array}{c} \Theta^3 \overset{\lambda_0}{\to} E \quad (k) \\ \vdots \quad \text{contractions} \\ \Theta \to E \\ \vdots \\ \to \end{array}}\, Cut$$

Let m_1 be the height of the premises of the cut on the formula B and let m_2 be the height of the premises of the cut on the formula C. The premises of the cut J_3 in P' have the height l because all cuts below the premises of the cut I also occur below J_3. And both added cuts have a lower grade than the cut formula $B\&C$. Furthermore, we have that $h(\Theta^3 \to E; P') = k$.

Assume that the grade of B is higher than or equal to the grade of C (otherwise we may exchange the order of the two cuts). Now we have $m_1 = m_2$. If k is higher than the grade of B (and the grade of C), then we have that $k = m_1 = m_2$ and if not m_1 equals the grade of B. In both cases we have $k \leqslant m_1$.

Let

$$\lambda_0 = o(\Theta^3 \to E; P')$$
$$\lambda_1 = o(\Theta \to B; P')$$
$$\lambda_2 = o(\Theta \to C; P')$$
$$\lambda_3 = o(B, C, \Theta \to E; P')$$
$$\mu_3 = o(\Gamma_1 \to B\&C; P')$$
$$\nu_3 = o(B\&C, B, C, \Gamma_2 \to D; P')$$

Then we have that $\nu_3 < \nu$ since the heights of the sequents above remain unchanged and a logical rule has been removed. Furthermore, we have that $\mu_3 = \mu$.

Now let

$$\dfrac{S_1' \quad S_2'}{S'}\, J'$$

be an arbitrary rule between J_3 and the sequent $B, C, \Theta \to E$ in the subderivation P_3 of P' and let

$$\frac{S_1 \quad S_2}{S} \; J$$

be the corresponding rule between I and $\Theta \to E$ in P. Let

$\alpha'_1 = o(S'_1; P')$ $\quad \alpha'_2 = o(S'_2; P')$ $\quad \alpha' = o(S'; P')$
$\alpha_1 = o(S_1; P)$ $\quad \alpha_2 = o(S_2; P)$ $\quad \alpha = o(S; P)$
$k_1 = h(S'_1; P) = h(S'_2; P')$ $\quad k_2 = h(S'; P')$

Then we have that $\alpha = \alpha_1 \# \alpha_2$ if S' is not the sequent $B, C, \Theta \to E$ and $\alpha = \omega_{l-k}(\alpha_1 \# \alpha_2)$ if S' is the sequent $B, C, \Theta \to E$. On the other hand we have that $\alpha' = \omega_{k_1 - k_2}(\alpha'_1 \# \alpha'_2)$.

We show by induction on the number of inferences between J_3 and S' that

$$\alpha' < \omega_{l-k_2}(\alpha) \tag{4.4.25}$$

if S' is not the sequent $B, C, \Theta \to E$.

If J' is J_3 then we have that

$$\alpha' = \omega_{l-k_2}(\mu_3 \# \nu_3) < \omega_{l-k_2}(\mu \# \nu) = \omega_{l-k_2}(\alpha)$$

because $\mu_3 = \mu$ and $\nu_3 < \nu$.

If we assume that the inequality holds for the premises of J', that is $\alpha'_1 < \omega_{l-k_1}(\alpha_1)$ and $\alpha'_2 < \omega_{l-k_1}(\alpha_2)$ then we get that $\alpha'_1 \# \alpha'_2$ is less than $\omega_{l-k_1}(\alpha_1) \# \omega_{l-k_1}(\alpha_2)$, this implies that $\alpha'_1 \# \alpha'_2 < \omega_{l-k_1}(\alpha_1 \# \alpha_2)$. From this follows that the inequality holds for the conclusion, because we have

$$\begin{aligned} \alpha' &= \omega_{k_1-k_2}(\alpha'_1 \# \alpha'_2) < \omega_{k_1-k_2}(\omega_{l-k_1}(\alpha_1 \# \alpha_2)) \\ &= \omega_{l-k_2}(\alpha_1 \# \alpha_2) = \omega_{l-k_2}(\alpha). \end{aligned}$$

Thus, it is proved that the inequality 4.4.25 holds.

The inequality 4.4.25 holds for the premises of the cut that gives the sequent $B, C, \Theta \to E$. The premises have the height $l = k_2$ and if we denote the ordinals of the premises α'_1 and α'_2 and for the corresponding premises in P, α_1 and α_2, we get from the inequalities of the premises that $\alpha'_1 < \omega_{l-l}(\alpha_1) = \alpha_1$ and $\alpha'_2 < \omega_{l-l}(\alpha_2) = \alpha_2$ hold. From this follows that $\lambda_3 = \omega_{l-m_2}(\alpha'_1 \# \alpha'_2) < \omega_{l-m_2}(\alpha_1 \# \alpha_2) = \omega_{l-m_2}(\kappa)$, if we let $\lambda = \omega_{l-k}(\kappa)$.

Then remains to calculate corresponding inequalities for the ordinals of the other subderivations P_1 and P_2. We consider the derivation P_1. There are two possibilities to consider, namely, that the last cut above the sequent $\Theta \to E$ in P_3 has been eliminated in the construction of P_1 and the possibility that there is a corresponding cut above the sequent $\Theta \to B$ in P_1. We show that in both cases $\lambda_1 \leqslant \lambda_3$.

Assume that there is a corresponding cut in P_1. The conclusion of the cut in P_3 has the height m_2, the premises have the height $l > m_2$ and the cut formula has the grade l. The cut formula of the cuts between J_3 and the cut in question have a grade lower or equal to l. Thus, all heights remain unchanged when the cuts are removed in P_1. And we conclude that $\lambda_1 \leqslant \lambda_3$.

Now, assume for the other case that the last cut above the sequent $\Theta \to B$ has been eliminated. This means that the heights of the corresponding sequents in P_1 and P_3 are no longer equal. We define the notion height difference to be able to inductively prove the inequality we want.

4.4.26 Definition. Let the premises of a cut or an induction have the height g and the conclusion the height h. The *height difference* of the cut or the induction is $g - h$ for the cut and $g - h + 1$ for the induction. The height difference between two sequents in a derivation is the sum of the height differences for all cuts and inductions between the two sequents.

The height difference between two sequents is equal to the height of the uppermost sequent, minus the height of the lowermost sequent, plus the number of inductions between the sequents.

Let S be a sequent in P_3 with the ordinal α and S' the corresponding sequent in P_1 with the ordinal α'. We show by induction that

$$\alpha' \leqslant \omega_{h-h'}(\alpha) \tag{4.4.27}$$

where h is the height difference between S and the conclusion of the subderivation P_3, that is $B, C, \Theta \to E$ and h' is the height difference between S' and the conclusion of the subderivation P_1, that is $\Theta \to B$.

The expression is well defined if $h \geqslant h'$. The sequents $B, C, \Theta \to E$ and $\Theta \to B$ have the same height $m_1 = m_2$ and the number of inductions between S and $B, C, \Theta \to E$ and between S' and $\Theta \to B$ is the same. Since the cut formulas below S' also occur below S we have that the height of S is greater or equal to the height of S'. This means that $h \geqslant h'$ and the expression is well defined. We can now proceed to proving the inequality 4.4.27.

If S is an initial sequent or the conclusion of an arithmetical rule without premises, then $\alpha' = \alpha = 1$ and the inequality holds regardless of the size of $h - h'$.

Assume that the inequality holds for the premise of a one-premise rule. Let the height difference under the premise in P_3 be h and in P_1 h' and let the ordinals of the premises be α and α' respectively. The height differences under the conclusions are the same. If the rule is a contraction the inequality of the premises is preserved. If the rule is logical or arithmetical then we get $\alpha' \leqslant \omega_{h-h'}(\alpha) < \omega_{h-h'}(\alpha + 1)$ and from this $\alpha' + 1 \leqslant \omega_{h-h'}(\alpha + 1)$.

Assume that the inequality holds for the premises of a two-premise arithmetical or logical rule, that is $\alpha'_1 \leqslant \omega_{h-h'}(\alpha_1)$ and $\alpha'_2 \leqslant \omega_{h-h'}(\alpha_2)$ hold. Here α_1 and α_2 are the ordinals of the premises in P_3 and α'_1 and α'_2 are the ordinals of the premises in P_1. The height differences under the premises, h and h', are the same as under the conclusion. We then get $\alpha' = \alpha'_1 \# \alpha'_2 \leqslant \omega_{h-h'}(\alpha_1) \# \omega_{h-h'}(\alpha_2) \leqslant \omega_{h-h'}(\alpha_1 \# \alpha_2) = \omega_{h-h'}(\alpha)$.

Assume that the inequality holds for the premises of a cut in P_3, that is $\alpha'_1 \leqslant \omega_{h-h'}(\alpha_1)$ and $\alpha'_2 \leqslant \omega_{h-h'}(\alpha_2)$ hold. If the cut has been eliminated in P_1, then S' has the ordinal α'_1. Let the height difference of the cut be g in P_3. Now the height difference under S is $h - g$ and we get the inequality $\alpha' = \alpha'_1 \leqslant \omega_{h-h'}(\alpha_1) < \omega_{h-h'}(\alpha_1 \# \alpha_2) = \omega_{(h-g)-h'}(\omega_g(\alpha_1 \# \alpha_2)) = \omega_{(h-g)-h'}(\alpha)$. On the other hand if the cut also occurs in P_1, in other words if it has not been eliminated, we let the height difference in P_1 be g'. Now the height difference under S is $h - g$ and under S' $h' - g'$ and we get the inequality for the conclusion $\alpha' = \omega_{g'}(\alpha'_1 \# \alpha'_2) \leqslant \omega_{g'}(\omega_{h-h'}(\alpha_1) \# \omega_{h-h'}(\alpha_2)) \leqslant \omega_{g'}(\omega_{h-h'}(\alpha_1 \# \alpha_2)) = \omega_{g'+h-h'-g}(\omega_g(\alpha_1 \# \alpha_2)) = \omega_{(h-g)-(h'-g')}(\alpha)$.

Lastly assume that the inequality holds for the premises of an instance of Ind, that is $\alpha'_1 \leqslant \omega_{h-h'}(\alpha_1)$, $\alpha'_2 \leqslant \omega_{h-h'}(\alpha_2)$ and $\alpha'_3 \leqslant \omega_{h-h'}(\alpha_3)$. Let the height difference for the induction in P_1 be g' and in P_3 g. Now the height difference under S is $h - g$ and under S' $h' - g'$ and we get the inequality $\alpha' = \omega_{g'}(\alpha_1' \# \alpha_2' \# \alpha_3') \leqslant \omega_{g'}(\omega_{h-h'}(\alpha_1) \# \omega_{h-h'}(\alpha_2) \# \omega_{h-h'}(\alpha_3)) \leqslant \omega_{g'}(\omega_{h-h'}(\alpha_1 \# \alpha_2 \# \alpha_3)) = \omega_{g'+h-h'-g}(\omega_g(\alpha_1 \# \alpha_2 \# \alpha_3)) = \omega_{(h-g)-(h'-g')}(\alpha)$.

Thus, it has been proved that the inequality holds. Now let S be the sequent $B, C, \Theta \to E$ and S' the sequent $\Theta \to B$. Then the height differences h and h' are 0 and we get $\lambda_1 \leqslant \omega_{h-h'}(\lambda_3) = \lambda_3$.

Regardless of the last cut has been eliminated, we thus have $\lambda_1 \leqslant \lambda_3$. Correspondingly we get $\lambda_2 \leqslant \lambda_3$. Using the inequality $\lambda_3 < \omega_{l-m_2}(\kappa)$ and the fact that $m_1 = m_2$ we then get $\lambda_1 \# \lambda_2 \# \lambda_3 < \omega_{l-m_1}(\kappa)$, since $l > m_1$. Furthermore, we get that $\lambda_0 = \omega_{m_1-k}(\lambda_1 \# (\omega_{m_2-m_1}(\lambda_2 \# \lambda_3))) = \omega_{m_1-k}(\lambda_1 \# \lambda_2 \# \lambda_3) < \omega_{m_1-k}(\omega_{l-m_1}(\kappa)) = \omega_{l-k}(\kappa) = \lambda$.

From the inequality $\lambda_0 < \lambda$ we get, according to lemma 4.4.4, that $o(P) > o(P')$.

Case 2. Assume that the cut formula of the last suitable cut is $\forall x B(x)$. The derivation P then has the form

$$
\dfrac{\dfrac{\dfrac{\vdots}{\Gamma_1' \to B(y/x)}}{\Gamma_1' \to \forall x B(x)}\, R\forall \quad \dfrac{\dfrac{\vdots}{B(t/x), \Gamma_2' \to D'}}{\forall x B(x), \Gamma_2' \to D'}\, L\forall}{\dfrac{\Gamma_1 \to \forall x B(x) \qquad \forall x B(x), \Gamma_2 \to D}{\Gamma_{1-2} \to D}\, I}
$$

$$
\begin{array}{c}
\Gamma_1' \to \forall x B(x) \qquad \forall x B(x), \Gamma_2' \to D' \\
\vdots \qquad\qquad\qquad \vdots \\
\dfrac{\Gamma_1 \to \forall x B(x) \qquad \forall x B(x), \Gamma_2 \to D}{\Gamma_{1-2} \to D}\, I \\
\vdots \\
\Theta \to E \\
\vdots \\
\to
\end{array}
$$

where the sequent $\Theta \to E$ is defined in the same way as in case 1.

From the derivation of the sequent $B(t/x), \Gamma_2' \to D_2'$ we get a derivation of the sequent $\forall x B(x), B(t/x), \Gamma_2' \to D_2'$ by adding a formula in the context. Let P_2 be the following derivation:

$$
\begin{array}{c}
\begin{array}{c}
\vdots \\
\forall x B(x), B(t/x), \Gamma_2' \to D' \\
\vdots
\end{array} \\[4pt]
\dfrac{\Gamma_1 \to \forall x B(x) \quad \forall x B(x), B(t/x), \Gamma_2 \to D}{B(t/x), \Gamma_{1-2} \to D}\ {\scriptstyle J_2} \\
\vdots \\
B(t/x), \Theta \to E
\end{array}
$$

We can get a derivation of the sequent $\Gamma_1' \to B(t/x)$ from the derivation of $\Gamma_1' \to B(y/x)$ by substituting y with t. We then apply the rules between the logical rule and J_2 in P_2 to the sequent $\Gamma_1' \to B(t/x)$ (this is possible because the quantified formula in the succedent of the sequents in P_2 is not principal in any rule above the cut J_2). We now have a derivation of the sequent $\Gamma_1 \to B(t/x)$ and can instead of applying the cut add the missing context in the antecedent and get a derivation of $\Gamma_{1-2} \to B(t/x)$. Then we apply the cuts and contractions below the cut J_2 on formulas in the antecedent. If we have a cut on the succedent, that is on the formula that has been replaced with $B(t/x)$, we just add the missing context in the antecedent and eliminate the cut. Thus, we obtain a valid derivation of the sequent $\Theta \to B(t/x)$ and we call this derivation P_1.

Now we can join the two derivations together into one derivation P'

$$
\begin{array}{c}
\dfrac{\begin{array}{c} P_1 \\ \vdots \\ \Theta \to B(t/x) \end{array} \quad \begin{array}{c} P_2 \\ \vdots \\ B(t/x), \Theta \to E \end{array}}{\Theta^2 \to E}\ {\scriptstyle Cut} \\
\vdots \ \text{contractions} \\
\Theta \to E \\
\vdots \\
\to
\end{array}
$$

The ordinal calculations are similar to the ones in case 1 and for the other cases of cut formulas the proofs are also similar.

Thus, we have reduced the derivation P into a derivation P' with a lower ordinal and the proof of lemma 4.4.21 is finished. We can conclude that the derivation P' also fulfills the requirement that all arithmetical rules are applied before the logical and structural rules. This makes it possible to repeat the reduction and get a sequence of decreasing ordinals. □

With the proof of the reduction procedure finished we also have a proof of the consistency theorem 4.4.2.

Some of the essential features of our proof are: Cut elimination is proved directly, without Gentzen's rule of multicut; The arithmetical axioms are treated purely syntactically; all rules with several premises have independent contexts and no rule of weakening is used. It is hoped that a comparison of our proof with Gentzen's notes in his series BTJZ will be helpful in the understanding of Gentzen's work on the consistency of intuitionistic arithmetic. There are still only fragmentary translations of these notes, but the problem seems to have been contraction and thereby the multiplication of certain parts of the derivations in the reductions.

Finally, because of the close connection between an intuitionistic sequent calculus with independent contexts and natural deduction, proof presented in this chapter is a useful step towards the proof in natural deduction. However, it will be shown in chapter 5 that the ordinal assignment of the natural deduction system requires special treatment. The ordinal assignment differs completely, due to the fact that the cut rule places sequents beside each other, whereas the composition of derivations in natural deduction stores subderivations on top of each other.

Chapter 5

Consistency of Heyting arithmetic in natural deduction

This is the pre-peer reviewed version of the following article: A. Kanckos, Consistency of Heyting Arithmetic in Natural Deduction. *Mathematical Logic Quarterly*, vol. 56 (2010), no. 6, pp. 611– 624.

The earliest proofs of the consistency of Peano arithmetic were presented by Gentzen. Since the publication of Gentzen's proof in sequent calculus, the conducting of the consistency proof in standard natural deduction has been an open problem. The aim of this chapter is to solve this problem by giving a consistency proof in natural deduction for Heyting arithmetic. The result is based on a normalization proof by Howard [11].

The present consistency proof is performed in the manner of Gentzen, by giving a reduction procedure for derivations of falsity. The procedure is appended with the assignment of a vector to each derivation and it is shown that the reduction reduces the first component. This component can be interpreted as an ordinal less than ϵ_0, thus ordering the derivations by complexity and proving termination of the process. To prove consistency it needs to be established that no derivation of the simplest kind exists. An important initial task is then to examine how natural deduction can be extended into an arithmetical system.

5.1 Logical calculus

The following rules constitute the standard calculus of intuitionistic natural deduction. Negation is a defined concept with $\neg A \equiv A \supset \bot$.

$$\frac{\bot}{C}\,{\scriptstyle \bot E}$$

$$\frac{A \quad B}{A\&B}\,{\scriptstyle \&I} \qquad \frac{A\&B}{A}\,{\scriptstyle \&E} \qquad \frac{A\&B}{B}\,{\scriptstyle \&E}$$

$$\frac{A}{A \vee B}\,{\scriptstyle \vee I} \qquad \frac{B}{A \vee B}\,{\scriptstyle \vee I} \qquad \frac{A \vee B \quad \begin{array}{c}[A]\\ \vdots \\ C\end{array} \quad \begin{array}{c}[B]\\ \vdots \\ C\end{array}}{C}\,{\scriptstyle \vee E}$$

$$\frac{\begin{array}{c}[A]\\ \vdots \\ B\end{array}}{A \supset B}\,{\scriptstyle \supset I} \qquad \frac{A \supset B \quad A}{B}\,{\scriptstyle \supset E}$$

$$\frac{A(y/x)}{\forall x A}\,{\scriptstyle \forall I} \qquad \frac{\forall x A}{A(t/x)}\,{\scriptstyle \forall E}$$

$$\frac{A(t/x)}{\exists x A(x)}\,{\scriptstyle \exists I} \qquad \frac{\exists x A(x) \quad \begin{array}{c}[A(y/x)]\\ \vdots \\ C\end{array}}{C}\,{\scriptstyle \exists E}$$

In the formula denoted by $A(t/x)$ every occurrence of the variable x in $A(x)$ has been substituted with the term t. The standard variable restriction holds in the rules $\forall I$ and $\exists E$: the eigenvariable y must not be free in the conclusion of the rule, nor in any assumption that the conclusion depends on, except for the discarded assumption $A(y/x)$ in the existential rule. The premise with logical structure, which is eliminated in the conclusion of an elimination rule, is the *major premise* of the rule. The other premises are *minor premises.*

5.2 Arithmetical rules and induction

As mentioned in the paper by Negri and von Plato [21], it is possible to convert the rules of sequent calculus into non-logical introduction and elimination rules of natural deduction.

Rules for the equality relation:

$$\frac{}{t = t}\ Ref \qquad \frac{t = t'}{t' = t}\ Sym$$

$$\frac{t = t' \quad t' = t''}{t = t''}\ Tr$$

Recursion rules:

$$\frac{}{t + 0 = t}\ +Rec0 \qquad \frac{}{t + s(t') = s(t + t')}\ +Recs$$

$$\frac{}{t \cdot 0 = 0}\ \cdot Rec0 \qquad \frac{}{t \cdot s(t') = t \cdot t' + t}\ \cdot Recs$$

Replacement rules:

$$\frac{t = t'}{s(t) = s(t')}\ sRep$$

$$\frac{t = t'}{t + t'' = t' + t''}\ +Rep \qquad \frac{t' = t''}{t + t' = t + t''}\ +Rep$$

$$\frac{t = t'}{t \cdot t'' = t' \cdot t''}\ \cdot Rep \qquad \frac{t' = t''}{t \cdot t' = t \cdot t''}\ \cdot Rep$$

Infinity rules:

$$\frac{s(t) = 0}{\perp}\ Inf_1 \qquad \frac{s(t) = s(t')}{t = t'}\ Inf_2$$

Induction rule:

$$\frac{A(0/x) \quad \begin{matrix}[A(y/x)] \\ \vdots \\ A(sy/x)\end{matrix}}{A(t/x)}\ {\scriptstyle Ind}$$

The eigenvariable y of the induction rule obeys the standard variable restriction and the induction formula A is arbitrary.

Although the system is intuitionistic it is possible to derive the law of excluded middle for atomic formulas, $\forall x \forall y(x = y \vee \neg x = y)$, by the induction rule. Therefore, the formula $A \vee \neg A$ can be derived for an arbitrary quantifier-free formula A.

5.2.1 Definition. A *purely arithmetical* derivation is a derivation where only arithmetical rules occur. (Induction is not included among the arithmetical rules.)

5.3 Properties of arithmetical derivations

The overall aim of Gentzen-style consistency proofs is to reduce complex derivations of a contradiction to simpler derivations. Therefore, a natural starting point is to consider the most elementary kind of derivations. The primary goal is to prove consistency for these purely arithmetical derivations.

5.3.1 Lemma. *(i) For a closed term t there exists a unique numeral $\overline{m}$, for which there is a purely arithmetical derivation of $t = \overline{m}$.*

(ii) Let t and t' be closed terms and assume that there is a purely arithmetical derivation of the formula $t = t'$. Then there is a purely arithmetical derivation of the formula $q(t/x) = q(t'/x)$ for an arbitrary term $q(x)$.

(iii) Let t and t' be closed terms and assume that there is a purely arithmetical derivation of the formula $t = t'$. Then for arbitrary terms $q(x)$ and $r(x)$ there is a purely arithmetical derivation of $q(t'/x) = r(t'/x)$ from the open assumption $q(t/x) = r(t/x)$.

(iv) *Let t and t' be closed terms and assume that there is a purely arithmetical derivation of the formula $t = t'$. Then for an arbitrary formula A the formula $A(t/x) \supset A(t'/x)$ can be derived without Ind.*

Proof. The proof of (i)-(iii) is by induction on the complexity of the term and the proof of (iv) is by induction on the complexity of the formula A. By proving that there is a derivation of $A(t')$ from the open assumption $A(t)$ a proof of (iv) is obtained through a final implication introduction. If A is an atomic formula, then the claim is proved in (iii). For the inductive step it is assumed that formula A is a compound formula. Assuming that A is a conjunction $B\&C$, the following derivation may be constructed:

$$\dfrac{\dfrac{\dfrac{[B(t)\&C(t)]}{B(t)}\,\&E \quad \dfrac{[B(t)\&C(t)]}{C(t)}\,\&E}{\dfrac{B(t') \qquad\qquad C(t')}{B(t')\&C(t')}\,\&I}}{(B(t)\&C(t)) \supset (B(t')\&C(t'))}\,\supset I$$

The other cases are similar. □

5.3.2 Lemma. *There is no purely arithmetical derivation of falsity.*

Proof. By the uniqueness of the numeral equal to a term, the premise of rule Inf_1 cannot be derived without open assumptions. Therefore, it is not possible to derive falsity.

□

5.4 Assignment of vectors to derivations

The normalization proof of Howard [11] provides a unique ordinal assignment up to ϵ_0 to terms of Gödel's theory T of primitive recursive functionals and proves that restricted reductions of the terms reduce the ordinals. In addition, a non-unique assignment is given for general reductions. By the well-ordering of ϵ_0, each reduction sequence terminates into a normal form, thereby proving strong normalization.

The unique assignment can be adapted to derivations in natural deduction. If each derivation is assigned an ordinal number as a measure of its complexity, then derivations can be ordered aiming for a proof of termination. The assignment of ordinal numbers to terms in T is indirect through a vector assignment and an interpretation of vectors as ordinals. A detour by vectors is needed, because the length of the vector provides an additional parameter for the calculations. This parameter is used in the definition of two operations (also originating in Howard's paper) that will provide desired properties of vectors. The length of the vector assigned to a formula in a derivation will depend on the complexity of the formula.

5.4.1 Definition. The *level* of a formula A, denoted $l(A)$, is inductively defined.

1. The level of an atomic formula and falsity is 0.
2. The level of a conjunction $A\&B$ is $max\{l(A), l(B)\}$.
3. The level of a disjunction $A \vee B$ is $max\{l(A), l(B)\}$.
4. The level of an implication $A \supset B$ is $max\{l(A) + 1, l(B)\}$.
5. The level of a universally quantified formula $\forall x A$ is $l(A)$.
6. The level of an existentially quantified formula $\exists x A$ is $l(A)$.

5.4.1 The theory $\mathcal{E}$

The vectors that will be assigned to derivations and formulas of derivations are vectors of expressions. These expressions are defined by introducing an axiomatic theory $\mathcal{E}$ of an order relation $\prec$ on expressions. In section 5.4.3 the expressions of this theory will be interpreted as ordinals.

5.4.2 Definition. *Expressions* are inductively defined.

(i) The constants $0, 1$ and ω are expressions.

(ii) For all formulas A and all i in $\mathbb{N}$, the variable x_i^A is an expression.

(iii) If f and g are expressions, then $f+g$ and (f,g) are also expressions.

Equality between expressions is treated axiomatically and obeys reflexivity and the replacement axiom. The weak order relation $f \preccurlyeq g$ is defined as $f \prec g$ or $f = g$ and the relation $f \succ g$ as $g \prec f$. The axioms of the theory $\mathcal{E}$ are listed below.

1. If $f \prec g$ and $g \prec h$, then $f \prec h$.
2. If $f \prec g$, then $\neg f = g$.
3. $f + g = g + f$ and $(f+g)+h = f + (g+h)$.
4. If $f \prec g$, then $f + h \prec g + h$.
5. $f + g = f$ if and only if $g = 0$.
6. $0 \preccurlyeq f$ and $0 \prec 1 \prec \omega$.
7. If $f \prec \omega$ and $g \prec \omega$, then $f + g \prec \omega$.
8. $(f, g+h) = (f,g) + (f,h)$.
9. If $g \prec c$ and $h \prec c$, then $(g,f) + (h,f) \preccurlyeq (c,f)$.
10. If $f \prec g$, then $(h,f) \prec (h,g)$.
11. If $f \prec g$ and $\neg h = 0$, then $(f,h) \prec (g,h)$.
12. $(0,f) = f$.
13. $(f,(g,h)) = (f+g,h)$

If $f \succ 0$ and $h \succ 0$, then by axioms 12, 11 and 8 the following inequality holds:

$$(f,g) + h \prec (f, g+h) \text{ if } f \succ 0 \text{ and } h \succ 0. \qquad (5.4.3)$$

5.4.2 Vectors of expressions

Expressions can be divided into classes, C_i, with the property that each expression in a class contains no variable that has a lower index than the class. A class of vectors $\mathbf{C}$ can be defined by presupposing that each component of the vector belongs to the corresponding class of expressions. The vectors in $\mathbf{C}$ will be the ones assigned to formulas in a derivation and two operations, the box- and the delta-operation acting upon these vectors will shortly be defined.

If f_i is an expression for each $0 \leqslant i \leqslant n$, then the $n+1$-tuple $\mathbf{f} = \langle f_0, \ldots, f_n \rangle$ is a vector of length n. For $0 \leqslant i \leqslant n$ the expression f_i, also denoted $(\mathbf{f})_i$, is called the ith component of $\mathbf{f}$. For $i > length(\mathbf{f})$, the component $(\mathbf{f})_i$ is defined to be 0. Addition of vectors $\mathbf{f}$ and $\mathbf{g}$ is done component by component, $\mathbf{f} + \mathbf{g} = \langle f_0 + g_0, \ldots, f_n + g_n \rangle$ where $n = max\{length(\mathbf{f}), length(\mathbf{g})\}$.

Finally, a vector of variables is defined for every formula A with the level $l(A) = n$, $\mathbf{x^A} = \langle x_0^A, \ldots, x_n^A \rangle$.

5.4.4 Definition. The *classes* C_i of expressions are defined by four clauses.

(i) If the expression h contains no variables, then h is in C_i.

(ii) For every formula A, the variable x_i^A is in C_i.

(iii) If the expressions f and g are in C_i, then so is $f + g$.

(iv) If the expression f is in C_{i+1} and the expression g is in C_i, then (f, g) is in C_i.

The class $\mathbf{C}$ consists of all vectors $\mathbf{h}$, such that h_i is in C_i for $0 \leqslant i \leqslant length(\mathbf{h})$.

5.4.5 Definition. The *box-operation* of two vectors $\mathbf{f}\Box\mathbf{g}$ is defined to be the vector $h = \langle h_0, \ldots, h_n \rangle$, where $n = max\{length(\mathbf{f}), length(\mathbf{g})\}$, such that

$$h_n = f_n + g_n \text{ and}$$

$$h_i = (h_{i+1}, f_i + g_i) \text{ for } 0 \leqslant i < n. \tag{5.4.6}$$

Note that equation 5.4.6 in fact holds for all $0 \leqslant i \leqslant n$, because recalling that $h_i = 0$ for all $i > max\{length(\mathbf{f}), length(\mathbf{g})\}$, and relying on axiom 12 the component h_n can be written as a pair and a simple calculation gives $h_i = f_n + g_n = (0, f_n + g_n) = (h_{n+1}, f_n + g_n)$. Another noticeable fact is the commutativity of the box-operation that follows from axiom 3, which states commutativity of addition on the expressions.

5.4.7 Definition. The *delta-operation* on a formula A of an expression h in $\cup C_i$, denoted $\delta^A h$, is a vector in $\mathbf{C}$ of length $l(A)+1$ that does not contain any component of the vector $\mathbf{x^A}$. The vector is defined when the C_i, to which h belongs, is specified.

1. If h is in C_i and contains no component of $\mathbf{x^A}$, then $\delta^A h$ is the vector of length $l(A)+1$, defined by $(\delta^A h)_i = h+1$ and $(\delta^A h)_j = 1$, when $j \neq i$ and $0 \leqslant j \leqslant l(A)+1$.
2. If h is x_i^A, then $(\delta^A h)_j = 1$ for $0 \leqslant j \leqslant l(A)+1$.
3. If h contains a component of $\mathbf{x^A}$ and $h = f + g$, where f and g are in C_i, then $\delta^A h = \delta^A f + \delta^A g$.
4. If h contains a component of $\mathbf{x^A}$ and $h = (f, g)$, where f is in C_{i+1} and g is in C_i, then

$$(\delta^A h)_j = (\delta^A f)_j + (\delta^A g)_j \text{ if } 0 \leqslant j \leqslant l(A) \text{ and}$$

$$(\delta^A h)_j = 2(\delta^A f)_j + 2(\delta^A g)_j + 1 \text{ if } j = l(A)+1.$$

The delta-operation is also defined for vectors $\mathbf{h} = \langle h_0, \ldots, h_n \rangle$ in $\mathbf{C}$.

$$(\delta^A \mathbf{h})_j = (\delta^A h_0)_j + \cdots + (\delta^A h_n)_j \text{ if } 0 \leqslant j \leqslant l(A)+1$$

and if $n > l(A)+1$, then we define

$$(\delta^A \mathbf{h})_j = h_j + 1 \text{ for } l(A)+1 < j \leqslant n.$$

The vector $\delta^A \mathbf{h}$ has the length $max\{l(A)+1, n\}$.

By the definitions of the operations both are well-defined operations on vectors in $\mathbf{C}$.

5.4.8 Lemma. *If* $\mathbf{f}$ *and* $\mathbf{g}$ *are in* $\mathbf{C}$*, then* $\mathbf{f} \Box \mathbf{g}$ *and* $\delta^A \mathbf{f}$ *are in* $\mathbf{C}$.

5.4.3 Interpretation of $\mathcal{E}$

To finalize the argument of the consistency proof, expressions of the theory $\mathcal{E}$ are to be interpreted as ordinals. In general, expressions can be interpreted as functions of the variables contained in them. From the first component of the vector assigned to a derivation a function is obtained, which applied to a suitable constant, say 0, will give an ordinal.

The relation $a \succ b$ is interpreted as $a > b$ and $a + b$ as the natural sum $a\#b$. The natural sum of two ordinals a and b represented in Cantor normal form $a = \omega^{a_1} + \cdots + \omega^{a_n}$ and $b = \omega^{b_1} + \cdots + \omega^{b_m}$, where $a_1 \geqslant \cdots \geqslant a_n$ and $b_1 \geqslant \cdots \geqslant b_m$, is defined as $a\#b = \omega^{c_1} + \cdots + \omega^{c_{n+m}}$, where $c_1 \geqslant \cdots \geqslant c_{n+m}$ is a rearrangement of the sequence $a_1, \ldots, a_n, b_1, \ldots, b_m$. The definition of (a, b) is separated into two cases depending on whether $b = 0$. The pair $(a, 0)$ is interpreted as 0. On the other hand, assume that $b \succ 0$ and represent b in Cantor normal form to the base 2, $b = 2^{b_1} + \cdots + 2^{b_n}$, where $b_1 > \cdots > b_n$. Then (a, b) is $2^{c_1} + \cdots + 2^{c_n}$, where $c_i = a\#b_i$ and $1 \leqslant i \leqslant n$. The described interpretation satisfies axioms 1 to 13 of the theory $\mathcal{E}$ given in section 5.4.1.

5.4.4 The vector assignment

After all preparations it is now possible to assign a vector, $\mathbf{f}$, to each formula, A, in a derivation such that $length(\mathbf{f}) = l(A)$ and $\mathbf{f}$ is in $\mathbf{C}$. To increase readability the following notation is introduced: if $\mathbf{f} = \langle f_0, \ldots, f_n \rangle$, and $\mathbf{g} = \langle f_0, \ldots, f_m \rangle$ where $m \leqslant n$, then $\mathbf{g} = (\mathbf{f}) \restriction_m$ is the restricted vector.

5.4.9 Definition. The *vector assigned to a formula* in a derivation is inductively defined as follows:

1. An assumption A is assigned the vector $\mathbf{x^A} = \langle x_0^A, \ldots, x_n^A \rangle$, where $n = l(A)$.
2. The conclusion of an arithmetical rule without premises is assigned the vector $\langle 0 \rangle$.
3. The conclusion of a one-premise arithmetical rule has the same vector as the premise.

4. If the premises of an instance of Tr or $\& I$ are assigned the vectors $\mathbf{f}$ and $\mathbf{g}$, then the conclusion of the rule is assigned the vector $\mathbf{f} + \mathbf{g}$.

5. If the premise of an instance of $\& E$ is assigned the vector $\mathbf{f}$, then the conclusion of the rule is assigned a vector $\mathbf{g}$, such that $g_i = f_i + 1$, for $0 \leqslant i \leqslant n$, where n is the level of the formula in the conclusion.

6. If the premise of $\vee I$ is assigned the vector $\mathbf{f}$, then the conclusion of the rule is assigned the vector $\mathbf{g}$, such that $g_i = f_i + 1$, for $0 \leqslant i \leqslant n$, where n is the level of the formula in the conclusion.

7. If the premises of $\vee E$ are assigned the vectors $\mathbf{f}$, $\mathbf{g}$ and $\mathbf{h}$, then the conclusion of the rule is assigned the vector $\mathbf{e}$, such that $e_i = (\mathbf{f}\Box(\delta^A\mathbf{g} + \delta^B\mathbf{h}))_i$ for $0 \leqslant i \leqslant n$, where n is the level of the formula in the conclusion and A and B are the discarded assumptions of the rule.

8. If the premise of $\supset I$ is assigned the vector $\mathbf{f}$, then the conclusion of the rule is assigned the vector $\delta^A\mathbf{f}$, where A is the discarded assumption of the rule.

9. If the premises of $\supset E$ are assigned the vectors $\mathbf{f}$ and $\mathbf{g}$, then the conclusion of the rule is assigned the vector $\mathbf{h}$, such that $h_i = (\mathbf{f}\Box\mathbf{g})_i$ for $0 \leqslant i \leqslant n$, where n is the level of the formula in the conclusion.

10. If the premise of $\forall I$ is assigned the vector $\mathbf{f}$, then the conclusion of the rule has the same vector.

11. If the premise of $\forall E$ is assigned the vector $\mathbf{f}$, then the conclusion of the rule is assigned the vector $\mathbf{g}$, such that $g_i = f_i + 1$, for $0 \leqslant i \leqslant \mathit{length}(\mathbf{f})$.

12. If the premise of $\exists I$ is assigned the vector $\mathbf{f}$, then the conclusion of the rule has the same vector.

13. If the premises of $\exists E$ are assigned the vectors $\mathbf{f}$ and $\mathbf{g}$, then the conclusion of the rule is assigned the vector $\mathbf{h}$, such that $h_i = (\mathbf{f}\Box\delta^{A(x)}\mathbf{g})_i$, for $0 \leqslant i \leqslant n$, where n is the level of the

formula in the conclusion and $A(x)$ is the discarded assumption of the rule.

14. If the premise of $\perp E$ is assigned the vector $\mathbf{f}$, then the conclusion of the rule is assigned the vector $\mathbf{g}$, such that $g_i = f_i + 1$, for $0 \leqslant i \leqslant n$, where n is the level of the formula in the conclusion.

15. If the formula concluded by an instance of Ind is $A(t)$, then the vector assigned to this formula depends on the term t. Let $\mathbf{f} = \langle f_0, \ldots, f_{n+1} \rangle$, where $n = l(A)$, be the vector assigned to the derivation of $A(\overline{m'}) \supset A(t')$ described in lemma 5.3.1(iv) for some closed term t' for which $t' = \overline{m'}$ is derivable.

    (a) If t is a closed term, then there is a derivation of $t = \overline{m}$ for some unique numeral $\overline{m}$ according to lemma 5.3.1. If the vectors assigned to the premises of the Ind-rule are $\mathbf{h}$ and $\mathbf{g}$, then the vector of the conclusion of the induction is

    $$(((\langle f_0, \ldots, f_{n+1}, 2(m+1) \rangle \Box \delta^{A(x)} \mathbf{g}) \restriction_{n+1} \Box \mathbf{h}) \restriction_n,$$

    where the length of the vector $\langle f_0, \ldots, f_{n+1}, 2(m+1) \rangle$ is $n + 2 = l(A) + 2$.

    (b) If on the other hand the term t contains a variable, then the vector of the conclusion of the induction is

    $$(((\langle f_0, \ldots, f_{n+1}, \omega \rangle \Box \delta^{A(x)} \mathbf{g}) \restriction_{n+1} \Box \mathbf{h}) \restriction_n,$$

    where the length of $\langle f_0, \ldots, f_{n+1}, \omega \rangle$ is $n + 2 = l(A) + 2$.

The vector assigned to the conclusion of a derivation is the vector assigned to the whole derivation.

Note that the delta-operation is always performed on a vector when an assumption is discharged in the subderivation to which the vector is assigned. The performed operation gives a vector not containing the variables assigned to the discarded assumption. Therefore, a formula derived without open assumptions (i.e. a theorem) must have a vector assigned to it which do not contain any variables.

In particular, the vector $\mathbf{f}$ used in the assignment of vectors to inductions does not contain any variables. Furthermore, since the vector $\langle 0 \rangle$ is assigned to a purely arithmetical derivation without open assumptions, the vector $\mathbf{f}$ does not depend on the term t', but only on the logical structure of A and the vector is well-defined.

5.5 Reduction procedure

The restricted reductions of [11] with a unique ordinal assignment correspond to a limitation in choice of the considered reducibility in the HA-derivation. The reducibility may not be a part of a subderivation that has open assumptions. Since all open assumptions must be discharged to derive a theorem, there would be an application of the delta-operation on the corresponding vector if there were open assumptions. The problem that arises with these general reductions is that order preservation is not necessarily provable for the delta-operation. If $f \prec g$, then $(\delta^A f)_i \prec (\delta^A g)_i$ does not follow in any obvious way when the expressions f and g differ in structure and fall under separate clauses in the definition of the delta-operation. However, even if general reductions cannot be treated, a suitable reducibility can be chosen in a derivation of falsity.

5.5.1 Theorem. *If there is a derivation of $\perp$ to which the vector $\mathbf{f}$ is assigned, then there is a derivation of $\perp$ to which the vector $\mathbf{g}$ is assigned and $f_0 \succ g_0$.*

Proof. Assume that there exists a derivation of $\perp$. Reduction steps are performed on the derivation in a specific order, each step is performed as many times as possible before proceeding to the next step. First step 1 is applied as many times as possible, then step 2 if possible. If step 2 is not possible, then step 3 may apply and finally if no other reduction is possible step 4 is performed.

Step 1. All free variables in the derivation, which are not eigenvariables, are replaced with the constant 0.

Step 2. If there is an instance of falsity elimination in the derivation below all instances of introduction rules and inductions and below which there are only arithmetical rules and major premises of elimination rules, then it is possible to eliminate the rule and the rest of the derivation below the rule. The new derivation is also a derivation of falsity, with no open assumptions.

Step 3. Assume that there is at least one induction below which there are no introduction rules, only major premises of elimination rules and arithmetical rules. Consider the lowermost (or rather one of the lowermost) of these inductions.

$$\dfrac{\begin{matrix} & [A(x)] \\ \vdots & \vdots \\ A(0) & A(sx) \end{matrix}}{A(t)}\ {\scriptstyle Ind}$$

Because there are no introduction rules, and in particular no universal introduction below the induction and only major premises of elimination rules, in particular no minor premise of existential elimination, the formula $A(t)$ cannot contain eigenvariables. Furthermore, all free variables were replaced in step 1. Therefore, the term t must be closed and there exists a derivation of $t = \overline{m}$ for some numeral $\overline{m}$ according to lemma 5.3.1(i). The reduction now performed depends on the numeral $\overline{m}$.

Case 1. If $\overline{m} \equiv 0$, then according to lemma 5.3.1(iv) there is a derivation of $A(0) \supset A(t)$ without inductions. The reduced derivation is composed by implication elimination with the first premise of the induction as minor premise.

$$\dfrac{\begin{matrix} \vdots & \vdots \\ A(0) \supset A(t) & A(0) \end{matrix}}{A(t)}\ {\scriptstyle \supset E}$$

Case 2. If $\overline{m} \equiv s(\overline{m'})$ for some numeral $\overline{m'}$, then according to lemma 5.3.1(iv) there is a derivation of $A(s\overline{m'}) \supset A(t)$ without inductions and the following reduction on the derivation is performed:

$$\dfrac{\begin{matrix} \vdots \\ A(s\overline{m'}) \supset A(t) \end{matrix} \qquad \dfrac{\dfrac{\begin{matrix} [A(\overline{m'})] \\ \vdots \\ A(s\overline{m'}) \end{matrix}}{A(\overline{m'}) \supset A(s\overline{m'})}\ {\scriptstyle \supset I} \qquad \dfrac{\begin{matrix} & [A(x)] \\ \vdots & \vdots \\ A(0) & A(sx) \end{matrix}}{A(\overline{m'})}\ {\scriptstyle Ind}}{A(s\overline{m'})}\ {\scriptstyle \supset E}}{A(t)}\ {\scriptstyle \supset E}$$

The derivation of $A(s\overline{m'})$ from $A(\overline{m'})$ is the second premise of the original induction with $\overline{m'}$ substituted for x.

Step 4. According to lemma 5.3.2 there is no purely arithmetical derivation of falsity. Therefore, the derivation must contain an instance of induction, falsity elimination, or another logical rule. Consider a lowermost instance of such a rule. If it is falsity elimination or induction, then either step 2 or step 3 applies. Now assume that we have a logical rule in the derivation. Because the conclusion of the derivation is not compound, there must be an elimination rule below each introduction. Hence, it may be assumed that the rule is an elimination rule. If the major premise of the elimination rule is the conclusion of another elimination rule, then it is possible to trace up through the major premises of elimination rules, until a formula concluded by some other rule is reached. The major premise under consideration is a compound formula and therefore not concluded by an arithmetical rule, neither can the formula be a discharged assumption because no rule discharges assumptions above major premises of elimination rules. Three possibilities remain, each suitable for reduction: the formula is concluded by falsity elimination, induction or an introduction rule. In the first two cases step 2 or step 3 applies. In the third case an operational reduction is performed depending on the outermost logical connective of the formula.

Case 1. If the formula is an implication, then the derivation has the form:

$$\dfrac{\dfrac{\begin{matrix}[A]\\ \vdots \\ B\end{matrix}}{A \supset B}\,{\scriptstyle \supset I} \qquad \begin{matrix}\vdots \\ A\end{matrix}}{B}\,{\scriptstyle \supset E}$$

and this is reduced into the following derivation:

$$\begin{matrix}\vdots \\ A \\ \vdots \\ B\end{matrix}$$

Case 2 to case 5 on conjunctions, disjunctions, universally and existentially quantified formulas respectively are similar standard detour conversions.

Thus, in all cases the derivation has been reduced. The vector calculations associated with the reductions are performed in lemma 5.6.12. □

5.6 Vector calculations

Lemmas 5.6.1-5.6.7, stating properties of the box- and delta-operations, are all proven or simply stated in [11]. The lemmas 5.6.1-5.6.4 are proven by downward induction on i from the axioms of the theory $\mathcal{E}$. The expression denoted by $f[\mathbf{g}/\mathbf{x^A}]$ is obtained from the expression f by substitution of each occurrence of x_i^A with $(\mathbf{g})_i$. The notation also applies to vectors $\mathbf{f}[\mathbf{g}/\mathbf{x^A}]$, where variable occurrences are replaced in each component.

5.6.1 Lemma. $(\mathbf{f}\Box\mathbf{g})_i \succcurlyeq f_i$ *for all* i.

5.6.2 Lemma. *Assume that* $length(\mathbf{f}) = length(\mathbf{g}) = n$ *and* $f_i \succcurlyeq g_i$ *for* $0 \leqslant i \leqslant n$, *then* $(\mathbf{f}\Box\mathbf{h})_i \succcurlyeq (\mathbf{g}\Box\mathbf{h})_i$ *for all* i.

5.6.3 Lemma. *Under the assumption of lemma 5.6.2 and the additional assumption* $f_i \succ g_i$ *for* $0 \leqslant i \leqslant k$ *and some* $k \leqslant n$, *the inequality* $(\mathbf{f}\Box\mathbf{h})_i \succ (\mathbf{g}\Box\mathbf{h})_i$ *holds for* $0 \leqslant i \leqslant k$.

Lemma 5.6.3 can be generalized for vectors that differ in length. Because if $length(\mathbf{f}) = n$, $length(\mathbf{g}) = m$, $n > m$ and $f_i \succ g_i$ for all $i \leqslant m$, then $(\mathbf{f}\Box\mathbf{h})_i \succ (\langle g_0, \ldots, g_m, 0, \ldots, 0\rangle\Box\mathbf{h})_i = (\mathbf{g}\Box\mathbf{h})_i$ for all $i \leqslant m$.

5.6.4 Lemma. *Assume that* $length(\mathbf{f}) = length(\mathbf{g}) = n+1 > length(\mathbf{h})$ *and* $f_i \succ 0$ *and* $g_i \succ 0$ *for* $0 \leqslant i \leqslant n+1$. *Let* $\mathbf{c}$ *be a vector such that* $2f_{n+1} + 2g_{n+1} \prec c_{n+1}$ *and* $f_i + g_i \preccurlyeq c_i$ *for all* $i \leqslant n$. *Then*

$$2((\mathbf{f}\Box\mathbf{h})\Box(\mathbf{g}\Box\mathbf{h}))_i \prec (\mathbf{c}\Box\mathbf{h})_i \text{ for all } i \leqslant n+1.$$

5.6.5 Lemma. *Let* $\mathbf{e}$ *be a vector of length* $l(A)$ *and assume* h *is in* C_i. *Then* $((\delta^A h)\Box\mathbf{e})_i \succ h[\mathbf{e}/\mathbf{x^A}]$.

Proof. The lemma is proven by induction on the number of times clauses 3 and 4 in the definition of the delta-operation are applied in $\delta^A h$. A complete proof is found in [11, Lemma 2.11]. □

5.6.6 Corollary. *If* $\mathbf{h}$ *is in* $\mathbf{C}$ *and* $\mathbf{e}$ *has length* $l(A)$*, then* $((\delta^A\mathbf{h})\Box\mathbf{e})_i \succ (\mathbf{h}[\mathbf{e}/\mathbf{x^A}])_i$ *for all* $i \leqslant length(\mathbf{h})$*.*

The next lemma is proven by induction on the length of the derivation.

5.6.7 Lemma. *Assume that there is a derivation of* A *to which the vector* $\mathbf{f}$ *is assigned and a derivation of* B *to which the vector* $\mathbf{g}$ *is assigned and that* A *is an open assumption in the latter derivation. Assume furthermore that no open assumption in the derivation of* A *becomes discarded in the derivation of* B*, where all assumptions* A *have been replaced with the derivation of* A*, then the vector assigned to this derivation is* $\mathbf{g}[\mathbf{f}/\mathbf{x^A}]$*.*

Now the aim becomes to prove that the reduction performed in step 1, substituting a constant for each free variable, does not increase the components of the vector.

5.6.8 Lemma. *If there is a derivation to which the vector* $\mathbf{h}$ *is assigned and another derivation, to which the vector* $\mathbf{g}$ *is assigned, and the latter derivation is obtained from the former by substituting a term for a free variable, then* $h_i \succcurlyeq g_i$ *for* $0 \leqslant i \leqslant length(\mathbf{h})$*.*

Proof. The vector assignment is otherwise the same, but for the fact that some inductions concluding a term with a variable may now have become inductions with a closed term, which fall under the first clause of the vector assignments to inductions. Assume that this is the case for some induction. Now, $(\langle f_0, \ldots, f_{n+1}, \omega\rangle)_i \succcurlyeq (\langle f_0, \ldots, f_{n+1}, 2(m+1)\rangle)_i$ for all i. By induction on the number of times the box- and delta-operations are applied, it is possible to show that the inequalities are preserved for the components of the vectors. The induction hypothesis is that $h_i \succcurlyeq h'_i$ and $g_i \succcurlyeq g'_i$ for all i, and the inductive step gives $(\mathbf{h}\Box\mathbf{g})_i \succcurlyeq (\mathbf{h}\Box\mathbf{g}')_i \succcurlyeq (\mathbf{h}'\Box\mathbf{g}')_i$ by lemma 5.6.2.

What remains to be shown is that a similar inequality holds for the delta-operation, $(\delta^A\mathbf{h})_i \succcurlyeq (\delta^{A'}\mathbf{h}')_i$, where A' comes from A by substitution of the term for the free variable.

1. If h_j contains no component of $\mathbf{x^A}$, then h'_j contains no component of $\mathbf{x^{A'}}$ and $(\delta^A h_j)_j = h_j + 1 \succcurlyeq h'_j + 1 = (\delta^{A'} h'_j)_j$, where

the inequality follows from the induction hypothesis. The other components of the vectors $\delta^A h_j$ and $\delta^{A'} h'_j$ are 1.

2. If h_j is x_j^A, then h'_j is $x_j^{A'}$ and the components of the vectors $\delta^A h_j$ and $\delta^{A'} h'_j$ are 1.

3. If h_j contains a component of $\mathbf{x^A}$ and $h_j = u + v$, then h'_j contains a component of $\mathbf{x^{A'}}$ and $h'_j = u' + v'$. Then $\delta^A h_j = \delta^A u + \delta^A v \succcurlyeq \delta^{A'} u' + \delta^A v \succcurlyeq \delta^{A'} u' + \delta^{A'} v' = \delta^{A'} h'_j$.

4. If h_j contains a component of $\mathbf{x^A}$ and $h_j = (u, v)$, then h'_j contains a component of $\mathbf{x^{A'}}$ and $h'_j = (u', v')$. The calculations of the inequalities are similar to those in case 3.

Thus, in all cases $(\delta^A h_j)_i \succcurlyeq (\delta^{A'} h'_j)_i$ and therefore $(\delta^A \mathbf{h})_i \succcurlyeq (\delta^{A'} \mathbf{h}')_i$. □

The following lemma calculates the vectors of the reduction performed in step 3 (case 2) dealing with the inductions.

5.6.9 Lemma. *Let* $\mathbf{e} = (((\langle f_0, \ldots, f_{n+1}, 2m+4\rangle \Box \mathbf{g}) \restriction_{n+1} \Box \mathbf{h}) \restriction_n$, *and let* $\mathbf{e}' = (((((\langle f_0, \ldots, f_{n+1}, 2m+2\rangle \Box \mathbf{g}) \restriction_{n+1} \Box \mathbf{h}) \restriction_n \Box \mathbf{g}) \restriction_n \Box \mathbf{f}) \restriction_n$, *where* $\mathbf{f} = \delta^A \mathbf{f}'$, $\mathbf{g} = \delta^A \mathbf{g}'$ *and furthermore* $length(\mathbf{f}) = length(\mathbf{g}) = n+1$ *and* $length(\mathbf{h}) = n$. *Then* $e_i \succ e'_i$ *for* $0 \leqslant i \leqslant n$.

Proof. First, the vectors $\mathbf{r}$, $\mathbf{t}$ and $\mathbf{b}$ are defined.

$$\begin{aligned}
\mathbf{r} &= ((((\langle f_0, \ldots, f_{n+1}, 2m+2\rangle \Box \mathbf{g}) \restriction_{n+1} \Box \mathbf{h}) \restriction_n \Box \mathbf{g}) \restriction_n \\
\mathbf{t} &= (((\langle f_0, \ldots, f_{n+1}, 2m+2\rangle \Box \mathbf{g}) \restriction_{n+1} \Box \mathbf{h}) \restriction_n \\
\mathbf{b} &= ((\langle f_0, \ldots, f_{n+1}, 2m+2\rangle \Box \mathbf{g}) \restriction_{n+1}
\end{aligned}$$

Let $\mathbf{a}$ be the vector of length $n+1$ defined as follows: $a_{n+1} = g_{n+1} + f_{n+1} + 1$ and $a_i = (a_{i+1}, g_i + f_i)$ for $0 \leqslant i \leqslant n$. Relying on these definitions, it is possible to prove that $e'_i \preccurlyeq (\mathbf{a} \Box \mathbf{t})_i$, by downward induction on $i \leqslant n$. In fact, the stronger claim $2e'_i \prec (\mathbf{a} \Box \mathbf{t})_i$ is proven.

For $i = n$ the equation $2e'_n = 2(f_{n+1}, r_n + f_n) = 2(f_{n+1}, (\mathbf{t} \Box \mathbf{g})_n + f_n) = 2(f_{n+1}, (g_{n+1}, t_n + g_n) + f_n)$ holds. Furthermore, because $\mathbf{f} = \delta^A \mathbf{f}'$ for some vector $\mathbf{f}'$, the strict inequality $f_j \succcurlyeq 1 \succ 0$ holds for all $j \leqslant n+1$ by the definition of the delta-operation. By a similar argument: $g_j \succcurlyeq 1 \succ 0$ for all $j \leqslant n+1$.

Thus, by inequality 5.4.3, $2(f_{n+1}, (g_{n+1}, t_n + g_n) + f_n) \prec 2(f_{n+1}, (g_{n+1}, t_n + g_n + f_n))$. By axiom 13 this equals $2(f_{n+1} + g_{n+1}, t_n + g_n + f_n)$. Because also the inequality $g_n + f_n = (0, g_n + f_n) \preccurlyeq (a_{n+1}, g_n + f_n) = a_n$ holds, $2(f_{n+1} + g_{n+1}, t_n + g_n + f_n) \preccurlyeq 2(f_{n+1} + g_{n+1}, t_n + a_n)$. Using axiom 9 and the fact that $f_{n+1} + g_{n+1} \prec a_{n+1}$ the proof of the base case is completed by the calculation $2(f_{n+1} + g_{n+1}, t_n + a_n) \preccurlyeq (a_{n+1}, t_n + a_n) = (\mathbf{a}\square\mathbf{t})_n$.

For the inductive step $i < n$ the vector $\mathbf{e}'$ can be analyzed as follows: $2e'_i = 2(e'_{i+1}, r_i + f_i) = 2(e'_{i+1}, (r_{i+1}, t_i + g_i) + f_i)$. Since $i < n$, $r_{i+1} = (\mathbf{t}\square\mathbf{g})_{i+1} \succcurlyeq g_{i+1} \succcurlyeq 1$ by lemma 5.6.1. Therefore, it is possible to use inequality 5.4.3 followed by axiom 13 to get $2(e'_{i+1}, (r_{i+1}, t_i + g_i) + f_i) \prec 2(e'_{i+1}, (r_{i+1}, t_i + g_i + f_i)) = 2(e'_{i+1} + r_{i+1}, t_i + g_i + f_i)$. By lemma 5.6.1 $r_{i+1} \preccurlyeq (\mathbf{r}\square\mathbf{f})_{i+1} = e'_{i+1}$, which gives $r_{i+1} + e'_{i+1} \preccurlyeq 2e'_{i+1}$. By axiom 9 and the induction hypothesis the inductive step is completed $2(e'_{i+1} + r_{i+1}, t_i + g_i + f_i) \preccurlyeq ((\mathbf{a}\square\mathbf{t})_{i+1}, t_i + g_i + f_i) \preccurlyeq ((\mathbf{a}\square\mathbf{t})_{i+1}, t_i + a_i) = (\mathbf{a}\square\mathbf{t})_i$. Thus, $e'_i \preccurlyeq (\mathbf{a}\square\mathbf{t})_i \preccurlyeq (\mathbf{a}\square(\mathbf{b}\square\mathbf{h}))_i$ for $i \leqslant n+1$.

By lemma 5.6.1 $a_i \preccurlyeq (\mathbf{a}\square\mathbf{h})_i$ for $i \leqslant n+1$ and hence $e'_i \preccurlyeq ((\mathbf{a}\square\mathbf{h})\square(\mathbf{b}\square\mathbf{h}))_i$. It is therefore sufficient to prove that $((\mathbf{a}\square\mathbf{h})\square(\mathbf{b}\square\mathbf{h}))_i \prec (\mathbf{c}\square\mathbf{h})_i$ for all $i \leqslant n+1$ where

$$\mathbf{c} = (\langle f_0, \ldots, f_{n+1}, 2m+4\rangle\square\mathbf{g}) \restriction_{n+1} .$$

Clearly, $a_i \succ 0$ for $i \leqslant n+1$ and as stated above $g_i \succcurlyeq 1 \succ 0$ for $i \leqslant n+1$. Therefore, by lemma 5.6.1, $b_i = (\langle f_0, \ldots, f_{n+1}, 2m+2\rangle\square\mathbf{g})_i \succcurlyeq g_i \succcurlyeq 1$ for $i \leqslant n+1$ and also $b_i \succ 0$ holds. By lemma 5.6.4 it is sufficient to prove

$$2a_{n+1} + 2b_{n+1} \prec c_{n+1}, \tag{5.6.10}$$

$$a_i + b_i \preccurlyeq c_i, \text{ for } i < n+1. \tag{5.6.11}$$

The first goal is to prove inequality 5.6.10. From $f_{n+1} \succcurlyeq 1$ and $g_{n+1} \succcurlyeq 1$ follow that $2a_{n+1} = 2(f_{n+1} + g_{n+1} + 1) \preccurlyeq 3(f_{n+1} + g_{n+1})$. By axioms 12, 9 and 11, $3(f_{n+1} + g_{n+1}) = 3(0, f_{n+1} + g_{n+1}) \preccurlyeq (2, f_{n+1} + g_{n+1}) \prec (2m+3, f_{n+1} + g_{n+1})$. On the other hand by axiom 9 $2b_{n+1} = 2(2m+2, f_{n+1} + g_{n+1}) \preccurlyeq (2m+3, f_{n+1} + g_{n+1})$. Thus,

$2a_{n+1} + 2b_{n+1} \prec 2(2m+3, f_{n+1} + g_{n+1})$ and by axiom 9 $2a_{n+1} + 2b_{n+1} \prec (2m+4, f_{n+1} + g_{n+1}) = c_{n+1}$.

The remaining goal is to prove inequality 5.6.11. Because it was proved above, that $a_{i+1} \succ 0$ and $b_{i+1} \succ 0$, it follows that $a_{i+1}+b_{i+1} \succ a_{i+1}$ and $a_{i+1} + b_{i+1} \succ b_{i+1}$. Therefore, by axiom 9, $a_i + b_i = (a_{i+1}, f_i + g_i) + (b_{i+1}, f_i + g_i) \preccurlyeq (a_{i+1} + b_{i+1}, f_i + g_i)$. Hence, by induction hypothesis $a_i + b_i \preccurlyeq (c_{i+1}, f_i + g_i) = c_i$, which proves the claim. □

The proofs above are sufficient preparation for calculating the vectors of the derivations in the reduction procedure.

5.6.12 Lemma. *Let P be a derivation of $\perp$ to which the vector* $\mathbf{f}$ *is assigned. If P' from P by performing the reduction described in theorem 5.5.1 and* $\mathbf{g}$ *is the vector assigned to P', then $f_0 \succ g_0$.*

Proof. **Step 1.** According to lemma 5.6.8 the expressions of the vector are not increased, by the procedure of substituting a constant for free variables in the derivation.

Step 2. The level of falsity is 0, so the vector assigned to the premise of the falsity elimination has one component, $\mathbf{f} = \langle f_0 \rangle$. By induction on the number of rules below the falsity elimination it can be shown that the first component of the vector assigned to the derivation P is greater than f_0. For the base case of the induction it can be concluded that the rule of falsity elimination increases the vector, because $f_0 + 1 \succ f_0$. Now assume as the induction hypothesis that $g_0 \succ f_0$ for some vector $\mathbf{g}$ assigned to a formula below the falsity elimination. Below the falsity elimination rule there are no rules that discharge assumptions that falsity depends on, so there are no delta-operations on the vector $\mathbf{g}$, but only box-operations and additions of 1. For the case of the box-operation a simple calculation gives $(\mathbf{g}\Box\mathbf{h})_0 \succcurlyeq g_0 \succ f_0$ by lemma 5.6.1 and the induction hypothesis. If the elimination rule only adds 1 to the components of the vector, then the statement is clear. This proves the claim.

Step 3. In this step, where an induction is reduced, there are two cases.

Case 1. The first case considered is when the term in the conclusion of the induction is equal to 0. Let $\mathbf{h}$ and $\mathbf{g}'$ be the vectors assigned to the premises of the induction in P and let $\mathbf{f}$ be the vector assigned to the derivation of $A(\overline{m}) \supset A(t)$. Furthermore, let $\mathbf{g} = \delta^{A(x)}\mathbf{g'}$ and denote $\mathbf{e} = (\langle f_0, \ldots, f_{n+1}, 2\rangle \Box \mathbf{g}) \restriction_{n+1}$. Then the vector assigned to the conclusion of the induction rule in P is $(\mathbf{e}\Box\mathbf{h}) \restriction_n$. The vector assigned to the reduced derivation is $(\mathbf{f}\Box\mathbf{h}) \restriction_n$.

By downward induction on i, it can be proved that $e_i \succ f_i$. For $i = n{+}1$ the component of $\mathbf{e}$ is $e_{n+1} = (2, g_{n+1}{+}f_{n+1})$. Since $\mathbf{f} = \delta^A\mathbf{f'}$ for some vector $\mathbf{f'}$, the components of the vector are positive $f_i \succcurlyeq 1 \succ 0$ for $0 \leqslant i \leqslant n{+}1$. Thus, with axiom 11 of the theory $\mathcal{E}$, the following calculation holds $e_{n+1} \succ (0, g_{n+1} + f_{n+1}) = g_{n+1} + f_{n+1} \succcurlyeq f_{n+1}$.

For the inductive step the component of $\mathbf{e}$ is $e_i = (e_{i+1}, g_i{+}f_i)$. By the induction hypothesis $e_{i+1} \succ f_{i+1} \succ 0$. Again by axiom 11 and the fact that $f_i \succ 0$ it can be concluded that $e_i \succ (0, g_i{+}f_i) = g_i{+}f_i \succcurlyeq f_i$. This concludes the inductive proof.

By lemma 5.6.3 and the claim proved above $(\mathbf{e}\Box\mathbf{h})_i \succ (\mathbf{f}\Box\mathbf{h})_i$ for $0 \leqslant i \leqslant n + 1$ and the vectors of the reduced part of the derivation have been calculated.

What remains to be shown is that the rules below the reduced part of the derivation preserve the inequality. This claim is proved by induction on the number of rules. As in step 2, it can be concluded that there are no rules below discharging assumptions that the conclusion of the induction rule depends on, so there are only box-operations and additions of 1. The base case, that the inequality holds if there are no rules below, is already proved above. Now assume as the induction hypothesis, that $a_0 \succ b_0$ and $a_i \succcurlyeq b_i$ for $i \leqslant \mathit{length}(\mathbf{a}) = \mathit{length}(\mathbf{b})$. Then lemma 5.6.3 can be used to get $(\mathbf{a}\Box\mathbf{c})_0 \succ (\mathbf{b}\Box\mathbf{c})_0$ for some vector $\mathbf{c}$ and $(\mathbf{a}\Box\mathbf{c})_i \succcurlyeq (\mathbf{b}\Box\mathbf{c})_i$ for $i > 0$. On the other hand, also $a_0 + 1 \succ b_0 + 1$ follows from the induction hypothesis as well as $a_i + 1 \succcurlyeq b_i + 1$ for $i > 0$. This proves the claim.

Case 2. The second case of step 3, considers an induction, for which the term in the conclusion equals a successor. Let $\mathbf{h}$, $\mathbf{g}$ and $\mathbf{f}$ be as in case 1. Then the vector assigned to the conclusion of the induction rule in P is $((\langle f_0, \ldots, f_{n+1}, 2m + 4\rangle \Box \mathbf{g}) \restriction_{n+1} \Box \mathbf{h}) \restriction_n$ and this is reduced to $(((((\langle f_0, \ldots, f_{n+1}, 2m + 2\rangle \Box \mathbf{g}) \restriction_{n+1} \Box \mathbf{h}) \restriction_n \Box \mathbf{g}) \restriction_n \Box \mathbf{f}) \restriction_n$ in P'. The claim that the components of the vector in P are greater than the components of the vector in P' for $i \leqslant n$ is proven by lemma 5.6.9. As in case 1 the rules below preserve the inequality.

Step 4. This step is divided into cases according to the logical rules in the detour conversion.

Case 1. Let the vector assigned to the premise of the implication introduction rule in P be $\mathbf{f}$ and the vector assigned to the minor premise of the implication elimination rule be $\mathbf{g}$. Then the vector assigned to the conclusion of the elimination is $(\delta^A\mathbf{f}\Box\mathbf{g})\upharpoonright_{l(B)}$. This vector is reduced to the vector $\mathbf{f}[\mathbf{g}/\mathbf{x^A}]$ according to lemma 5.6.7. Corollary 5.6.6 gives the desired result for these two vectors. As in step 3 (case 1) the rules below preserve the inequality.

Case 2. Let the vectors assigned to the premises of the conjunction introduction rule in P be $\mathbf{f}$ and $\mathbf{g}$. Assume that A is the conclusion of the elimination rule. Then the vector assigned to the conclusion of the elimination is $(\langle f_0 + g_0 + 1, \ldots, f_n + g_n + 1\rangle)\upharpoonright_{l(A)}$, where $n = max\{length(\mathbf{f}), length(\mathbf{g})\}$. This vector is reduced to the vector $\mathbf{f}$.

By the axioms of $\mathcal{E}$ it is easily concluded that $f_i + g_i + 1 \succ f_i$ for all $i \leqslant n$ and as in step 3 (case 1) the rules below preserve the inequality.

Case 3. Let the vector assigned to the premise of the disjunction introduction rule in P be $\mathbf{f}$ and let $\mathbf{g}$ and $\mathbf{h}$ be the vectors assigned to the minor premises of the elimination rule. Assume that the premise of the introduction rule is A, since the other case is dual. Then the vector assigned to the conclusion of the elimination is $(\langle f_0{+}1, \ldots, f_n{+}1\rangle\Box(\delta^A\mathbf{g} + \delta^B\mathbf{h}))\upharpoonright_{l(C)}$, where $n = l(A \vee B) \geqslant length(\mathbf{f})$, and this is reduced to $\mathbf{g}[\mathbf{f}/\mathbf{x^A}]$ according to lemma 5.6.7.

Now, $(\delta^A\mathbf{g} + \delta^B\mathbf{h})_i \succcurlyeq (\delta^A\mathbf{g})_i$ for all i by the axioms of $\mathcal{E}$ and thus $(\langle f_0 + 1, \ldots, f_n + 1\rangle\Box(\delta^A\mathbf{g} + \delta^B\mathbf{h}))_i \succcurlyeq (\langle f_0 + 1, \ldots, f_n + 1\rangle\Box\delta^A\mathbf{g})_i$ for all i by lemma 5.6.2. Furthermore, $(\langle f_0 + 1, \ldots, f_n + 1\rangle\Box\delta^A\mathbf{g})_i \succcurlyeq (\langle f_0, \ldots, f_n\rangle\Box\delta^A\mathbf{g})_i = (\mathbf{f}\Box\delta^A\mathbf{g})_i$ for all i. According to corollary 5.6.6 the desired result $(\mathbf{f}\Box\delta^A\mathbf{g})_i \succ (\mathbf{g}[\mathbf{f}/\mathbf{x^A}])_i$ holds for all $i \leqslant l(C)$. As in step 3 (case 1) the rules below preserve the inequality.

Case 4. Let the vector assigned to the premise of the universal introduction rule in P be $\mathbf{f}$, then the vector assigned to the conclusion of the elimination rule is $\langle f_0 + 1, \ldots, f_n + 1\rangle$. In the derivation of the premise of the introduction rule, $A(y/x)$, the term t can be substituted for x. If t contains a variable, then the vector, $\mathbf{f'}$, of the derivation that results from the reduction procedure remains unchanged and equal to $\mathbf{f}$ and if t is closed, then by lemma 5.6.8 the components of the vector are not increased. Thus, $f_i+1 \succ f_i \succcurlyeq f'_i$ for all $i \leqslant n$. As in step 3 (case 1) the rules below preserve the inequality.

Case 5. Let the vector assigned to the premise of the existential introduction rule in P be $\mathbf{f}$ and let $\mathbf{g}$ be the vector assigned to the minor premise, C, of the elimination rule. Then the vector assigned to the conclusion of the elimination is $(\mathbf{f}\Box\delta^A\mathbf{g}) \restriction_{l(C)}$. Let $\mathbf{g}'$ be the vector of the derivation that results from the process of substituting the term t in the premise of the introduction rule for x in the derivation of the minor premise of the elimination rule. Then, $(\delta^A\mathbf{g})_i \succcurlyeq (\delta^A\mathbf{g}')_i$ and the vector assigned to the reduced derivation is $\mathbf{g}'[\mathbf{f}/\mathbf{x^A}]$. By lemma 5.6.2 follows that $(\mathbf{f}\Box\delta^A\mathbf{g})_i \succcurlyeq (\mathbf{f}\Box\delta^A\mathbf{g}')_i$ for all i. From corollary 5.6.6 the desired result $(\mathbf{f}\Box\delta^A\mathbf{g}')_i \succ (\mathbf{g}'[\mathbf{f}/\mathbf{x^A}])_i$ follows for $i \leqslant l(C)$. As in step 3 (case 1) the rules below preserve the inequality.

□

5.7 The consistency theorem

5.7.1 Theorem (The consistency of Heyting arithmetic)**.** *Falsity is not derivable in the system HA, that is, it is consistent.*

Proof. Assume that the system HA is inconsistent and that there is a derivation of falsity. According to theorem 5.5.1 there is a reduced derivation with a lower ordinal and another reduced derivation and so on. This produces an infinite succession of decreasing ordinals all less than ϵ_0, but this is impossible because the well-ordering of ϵ_0 implies that the reduction procedure must terminate. Thus, there cannot exist a derivation of falsity and the system of Heyting arithmetic is consistent. □

Bibliography

[1] P. Bernays: On the original Gentzen consistency proof for number theory, in *Intuitionism and proof theory*, ed. A. Kino, J. Myhill, and R. Vesley, North-Holland, Amsterdam, 1970.

[2] S. Burris: Polynomial Time Uniform Word Problems, *Mathematical Logic Quarterly*, vol. 41 (1995), Issue 2, pp. 173–182.

[3] S. Cosmadakis: The word and generator problem for lattices, *Information and Computation*, vol. 77 (1988), issue 3, pp. 192–217.

[4] D. Føllesdal, L. Walløe, J. Elster: *Argumentationsteori, språk och vetenskapsfilosofi*, 3rd ed., Thales, Stockholm, 2001.

[5] G. Gentzen: Die Widerspruchsfreiheit der reinen Zahlentheorie, *Mathematische Annalen*, vol. 112 (1936), pp. 493–565.

[6] G. Gentzen: The current situation in research in the foundations of Mathematics, in *Logic's lost genius: The life of Gerhard Gentzen*, E. Menzler-Trott, American mathematical society, 2007.

[7] G. Gentzen: Neue Fassung des Widerspruchsfreiheitsbeweises für die reine Zahlentheorie, *Forschungen zur Logik und zur Grundlegung der exakten Wissenschaften*, vol. 4 (1938), pp. 19–44.

[8] G. Gentzen: Beweisbarkeit und Unbeweisbarkeit von Anfangsfällen der transfiniten Induktion in der reinen Zahlentheorie, *Mathematische Annalen*, vol. 119 (1943), pp. 140–161.

[9] G. Gentzen: *The Collected Papers of Gerhard Gentzen*, ed. M.E. Szabo, North-Holland, Amsterdam, 1969.

[10] G. Gentzen: Der erste Widerspruchsfreiheitsbeweis für die klassische Zahlentheorie, *Archiv für mathematische Logik*, vol. 16 (1974), pp. 97–118.

[11] W. A. Howard: Assignment of ordinals to terms for primitive recursive functionals of finite type, in *Intuitionism and Proof Theory*, ed. A. Kino, J. Myhill, and R. Vesley, North-Holland, Amsterdam, 1970.

[12] S. Jaskowski: On the rules of supposition in formal logic, (translation of Polish original) in *Polish Logic 1920-1939*, ed. S. McCall, pp. 232–258, Oxford, 1967.

[13] A. Kanckos: Gentzen's consistency proofs for arithmetic, in *The Logica Yearbook 2009*, ed. M. Peliš, pp. 109–119, College Publications, London, 2010.

[14] A. Kanckos: Consistency of Heyting Arithmetic in Natural Deduction. *Mathematical Logic Quarterly*, vol. 56 (2010), no. 6, pp. 611– 624.

[15] G. Kreisel: Review of the Collected Papers of Gerhard Gentzen, *The Journal of Philosophy*, vol. 68 (1971), pp. 238–265.

[16] G. Kreisel: What have we learnt from Hilbert's second problem?, in *Mathematical developments arising from Hilbert problems* (Proc. Sympos. Pure Math., Northern Illinois Univ., De Kalb, Ill.,): pp. 93–130, Providence, R. I.: Amer. Math. Soc., 1976.

[17] H. Mehlberg: The present situation in the philosophy of mathematics, in *Philosophy of mathematics, an anthology*, ed. D. Jaquette, Blackwell Publishers Ltd. , 2002

[18] A. MEINANDER: A solution for the uniform word problem for ortholattices, *Mathematical Structures in Computer Science*, vol. 20 (2010), issue 04, pp. 625–638.

[19] E. MENZLER-TROTT, *Logic's lost genius: The life of Gerhard Gentzen*, American mathematical society, 2007.

[20] S. NEGRI: Contraction-free sequent calculi for geometric theories with an application to Barr's theorem, *Archive for Mathematical Logic*, vol. 42 (2003), pp. 389–401.

[21] S. NEGRI, J. VON PLATO: Cut elimination in the presence of axioms, *The Bulletin of Symbolic Logic*, vol. 4 (1998), pp. 418–435.

[22] S. NEGRI, J. VON PLATO: Permutability of rules in lattice theory, *Algebra Universalis*, vol. 48 (2002), pp. 473–477.

[23] J. VON PLATO: A proof of Gentzen's Hauptsatz without multicut, *Archive for Mathematical Logic*, vol. 40 (2001), pp. 9–18.

[24] J. VON PLATO: Normal form and existence property for derivations in Heyting arithmetic, *Acta Philosophica Fennica*, vol. 78 (2006), pp. 159–163.

[25] J. VON PLATO: From Hilbert's Programme to Gentzen's Programme, in *Logic's lost genius: The life of Gerhard Gentzen*, E. Menzler-Trott, American mathematical society, 2007.

[26] J. VON PLATO: In the shadows of the Löwenheim-Skolem theorem: Early combinatorial analyses of mathematical proofs, The Bulletin of Symbolic Logic, vol. 13 (2007), no. 2, pp. 189-225.

[27] J. VON PLATO: Gentzen's Logic, *Handbook of the History of Logic vol. 5*, North-Holland, pp. 667–721, 2009.

[28] J. VON PLATO: Combinatorial analysis of proofs in projective and affine geometry, *Annals of Pure and Applied Logic*, vol. 162 (2010), pp. 144–161.

[29] B. Scarpellini: Some applications of Gentzen's second consistency proof, *Mathematishce Annalen*, vol. 181 (1969), pp. 325–344.

[30] K. Schütte and H. Schwichtenberg: Mathematische Logik, in *Ein Jahrhundert Mathematik 1890-1990. Festschrift zum Jubiläum der DMV*, volume 6 of *Dokumente zur Geschichte der Mathematik,* ed. W. Scharlau, G. Fischer, F. Hirzebruch and W. Törnig, Vieweg, Brauschweig/Wiesbaden, 1990.

[31] G. Takeuti: *Proof Theory*, North-Holland, 1987.