Charles University, Faculty of Arts

Department of Logic

Anna Horská

DÔKAZY BEZESPORNOSTI ARITMETIKY
CUT ELIMINATION AND CONSISTENCY PROOFS

Doctoral thesis

Supervisor: prof. RNDr. Pavel Pudlák, DrSc.

2017

Prohlašuji, že jsem disertační práci napsala samostatně s využitím pouze uvedených a řádně citovaných pramenů a literatury a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

I declare that I carried out this doctoral thesis independently only with the cited sources, literature and other professional sources. The thesis was not used during any other study program to obtain the same or any equivalent degree.

Prague, September 10, 2017

Anna Horská

Abstrakt Táto práca pozostáva z dvoch častí. Prvá časť sa zaoberá Gentzenovým dôkazom bezespornosti Peanovej aritmetiky (PA), ktorý pochádza z roku 1935. Skúmame hlavne Gentzenovu stratégiu eliminácie rezu, ktorá eliminuje rezy, ktorých premisy majú bezrezové odvodenia. Neberie sa pritom ohľad na zložitosť eliminovaného rezu. Naša analýza Gentzenovej stratégie ukázala, že Gentzen vo svojom dôkaze implicitne využíva transfinitnú indukciu po $\Phi_\omega(0)$, kde Φ_ω je Veblenova funkcia s poradovým číslom ω. Jedná sa o horný odhad a hodnota $\Phi_\omega(0)$ je horný odhad na výšku nekonečných bezrezových odvodení, ktoré Gentzen konštruuje pre sekventy dokazateľné v PA. V súčasnosti nemáme výsledky o spodnom odhade. Prvá časť ďalej obsahuje formalizáciu tohto Gentzenovho dôkazu. Na základe nej vidíme, že hore spomínaná transfinitná indukcia je jediný princíp použitý v dôkaze, ktorý nejde formalizovať v PA.

Druhá časť porovnáva Gentzenovu a Taitovu stratégiu eliminácie rezu v klasickej výrokovej logike. Taitova stratégia znižuje tzv. cut-rank odvodenia. Keďže výroková logika nepoužíva odvodzovacie pravidlá s vlastnými premennými, s tzv. eigenvariables, podarilo sa nám nadefinovať elimináciu rezu tak, že obe stratégie dávajú v klasickej výrokovej logike identické bezrezové odvodenia.

Kľúčové slová: eliminácia rezu, bezespornosť Peanovej aritmetiky, Gerhard Gentzen, Veblenova hierarchia, nekonečné kalkuly

Abstract The thesis consists of two parts. The first one deals with Gentzen's consistency proof of 1935, especially with the impact of his cut elimination strategy on the complexity of the proof. Our analysis of Gentzen's cut elimination strategy, which eliminates uppermost cuts regardless of their complexity, yields that, in his proof, Gentzen implicitly applies transfinite induction up to $\Phi_\omega(0)$ where Φ_ω is the ω-th Veblen function. This is an upper bound and $\Phi_\omega(0)$ represents an upper bound on heights of cut-free infinitary derivations which Gentzen constructs for sequents derivable in Peano arithmetic (PA). We currently do not know whether this is a lower bound too. The first part also contains a formalization of Gentzen's proof of 1935. Based on the formalization, we see that the transfinite induction mentioned above is the only principle used in the proof that exceeds PA.

The second part compares the performance of Gentzen's and Tait's cut elimination strategy in classical propositional logic. Tait's strategy reduces the cut-rank of the derivation. Since the propositional logic does not use inference rules with eigenvariables, we managed to organize the cut elimination in the way that both strategies yield identical cut-free derivations in classical propositional logic.

Keywords: cut elimination, consistency of Peano arithmetic, Gerhard Gentzen, Veblen hierarchy, infinitary calculus

Contents

Outline of the thesis

This thesis consists of two parts. Each part is preceded by an abstract that summarizes the topic and the results of the part. Since both parts can be read independently, each of them contains its own comprehensive introduction and a common introduction is redundant.

The first part is called "What is the height of Gentzen's reduction trees?". It deals with Gentzen's consistency proof of 1935 [4] and, especially, with his cut elimination strategy that chooses uppermost cuts for elimination. The aim of Gentzen's proof is to show that reduction trees, cut-free infinitary derivations, can be constructed for all sequents that are derivable in Peano arithmetic (PA). Since a contradictory sequent $\to 0 = 1$ has no reduction tree by definition, Gentzen concludes that PA is consistent.

This part is divided into five sections. The first section, called "Introduction", mentions the history of the proof and explains doubts that led Gentzen to withdraw the proof from publishing in the 1930s. Furthermore, it describes the method of the original Gentzen's proof.

The second section, called "Preliminaries", contains basic definitions, and in particular, it introduces Veblen hierarchy that is used to measure heights of cut-free derivations. It also presents Gentzen's cut elimination strategy in detail. The strategy is later applied to construct reduction trees.

The third section is called "Cut elimination" and it analyzes Gentzen's cut elimination strategy. In particular, we added upper bounds on heights of cut-free derivations to his cut elimination theorem denoted by Hilfssatz.

From the fourth section on, we start to use a sequent calculus for PA besides the infinitary one. We state Gentzen's algorithm that constructs reduction trees for initial sequents of PA. These reduction trees always have a finite height. Second, we present Gentzen's way to construct reduction trees for sequents that are derivable in PA and, with the help of results of the third section, we show that the upper bound on heights of these reduction trees is $\Phi_\omega(0)$.

The last section of the first part contains the formalization of Gentzen's

proof of 1935. We show that it can be formalized in III_3 plus transfinite induction on the height of reduction trees for sequents that are derivable in PA. The transfinite induction uses at most Δ_3 induction formulas.

The second part is called "Comparison between Tait's and Gentzen's cut elimination strategy in classical propositional logic". Gentzen's strategy always chooses a cut for elimination whose premises have cut-free derivations. Tait's strategy reduces the cut-rank by choosing one of the most complex cuts. We know that if Gentzen had applied Tait's strategy in his proof of 1935 [4], the height of reduction trees for sequents derivable in PA would be bounded by ε_0. This fact and the analysis in the previous part raise the question whether both strategies yield different cut-free derivations. We show that this is *not* the case in classical propositional logic where we are able to organize the cut elimination in such a way that we obtain the same cut-free derivations regardless of the strategy. It must be mentioned that methods which we have used for propositional logic do not apply to both the predicate logic and the infinitary calculus, hence, the question is still open for these systems.

The second part is divided into four sections. The first section is called "Introduction" and it clarifies what we mean by a comparison of cut elimination strategies. We discuss aspects that affect the cut elimination and how we want to handle them.

The second section has the usual title "Preliminaries". It contains technical definitions that allow us to speak about particular formulas, sequents etc.

The third section is called "Elimination of a single cut". It introduces an algorithm for elimination of a single cut that is inspired by the algorithm of Buss ([3], pp. 37-40) that makes global changes to the derivation. Furthermore, we discuss some properties of this cut elimination, especially, which parts of the derivation are made redundant during the elimination and why the algorithm does not work in predicate logic.

The last section of the second part defines a cut elimination strategy, called *general cut elimination*, that includes both strategies under consideration, the one of Gentzen and the one of Tait. This strategy has the strong normalization property and we also prove that it has the weak Church-Rosser property. It follows that normals forms, cut-free derivations in classical propositional logic, are given unambiguously.

Part 1

What is the height of Gentzen's reduction trees?

Abstract In his consistency proof of 1935, Gentzen constructs reduction procedures for sequents that are derivable in Peano arithmetic, whereas contradictory sequents have no reduction procedures. Reduction procedures generate reduction trees which we interpret as cut-free infinitary derivations. A cut elimination theorem is used to build reduction trees; Gentzen calls it Hilfssatz in this particular proof. Hilfssatz is interesting because of two reasons: (1) The cut elimination strategy applied there eliminates always an uppermost cut, regardless of its complexity. (2) The proof of Hilfssatz makes use of transfinite induction on the height of reduction trees that have been constructed so far. In this part, we analyse Gentzen's proof and particularly the cut elimination strategy of Hilfssatz to quantify the transfinite induction implicitly applied in the consistency proof. We determine an upper bound for the heights of reduction trees that belong to sequents which are derivable in Peano arithmetic. Namely, the heights of these reduction trees are less than $\Phi_\omega(0)$ where Φ_ω is the ω-th Veblen function. The question stays open what the lower bound is, but this seems to be quite difficult. If Gentzen had applied Tait's cut elimination strategy, which reduces the cut-rank, the heights of reduction trees for sequents derivable in PA would be bounded by ε_0.

The second half deals with the question whether the transfinite induction mentioned above is the only tool in the consistency proof that exceeds PA. We shall show that this is the case by formalizing the proof in Π_3 plus transfinite induction on the height of reduction trees for sequents derivable in PA. The transfinite induction uses at most Δ_3 induction formulas. One can view it as an improvement of Gentzen's original formulation where he implicitly uses induction formulas that contain the notion of well-foundedness, which is a second order property.

1.1 Introduction

In 1935, Gentzen planned to publish the consistency proof for PA for the first time, but he withdrew it later because of the criticism of his contemporaries. Gentzen, who was in Göttingen at that time, had a vigorous correspondence about the proof with Bernays, who was in Princeton. As Bernays later wrote in the introduction to the original Gentzen's paper [4] published not until 1974, the main objection was that Gentzen had used the Fan Theorem implicitly to show the termination of a certain reduction procedure. At the same time, Bernays admitted in the introduction that the objections were false. It is almost certain that Weyl, Gödel and von Neumann read Gentzen's manuscript, too. Whereas we know that Bernays and Weyl criticized the idea of the proof, it is not sure what the opinion of Gödel and von Neumann was. However, it is sure that the opinions among mathematicians were different. For example, in a letter to Kneser dated October 27, 1935 ([8], p. 54), Gentzen mentioned that Van der Waerden had praised his proof.

Gentzen discussed the technical aspects of his proof mainly with Bernays. In a letter dated November 4, 1935 ([8], p. 52), Gentzen was explaining to him why the worries about his proof are groundless. Nevertheless, Bernays obviously had more powerful arguments because Gentzen gave up in December 1935. He decided to remake the proof completely. This resulted in [5], the first published consistency proof. More on history of Gentzen's consistency proofs and his life, but no mathematics, can be found in [8].

From the classical point of view, the Fan Theorem is the contrapositive of König's lemma for binary trees and, in 1987, Kreisel noted that this principle is not sufficient for proving the consistency of PA [14]. Indeed, the principle used in the proof is bar induction [9]. Bar induction operates on (infinitely branching) trees. It uses two predicates, let us call them B and S. The predicate B must be decidable and it stands for "bar" because one of the three assumptions of bar induction is that (1) every branch has a vertex which can be reached from the root within a finite number of steps and which fulfills B. This assumption ensures that we have a well-founded tree whose leaves have the property B. Vertices which follow the nodes with property B in the original tree can be forgotten. Further assumptions are that (2) if a vertex fulfills B, then it must fulfill S and (3) if all children of a vertex v fulfill S, then S is inherited by v, too. As a result, we obtain that the root satisfies S. Such an induction can be replaced by transfinite induction on the height of well-founded trees. When we consider countably branching well-founded trees, as we will do below, this leads, in general, to transfinite induction up to any countable ordinal number.

Gentzen does not explicitly speak about countably branching well-founded trees in [4]. He uses certain *reduction rules* that generate such a tree, *reduction tree*, for every sequent that is valid in the standard model of PA. The intuition behind this is, as Gentzen writes ([4], p. 100), that reduction rules are a syntactic representation of the semantic correctness. If we can construct a reduction tree for every sequent that is derivable in PA, we obtain the consistency of PA because it is easy to see that the contradictory sequent $\rightarrow 0 = 1$ has no reduction tree. The question is, of course, what kind of mathematical principles we apply during the construction. If we, for example, use the predicate of truth, then we are able to construct a reduction tree with a finite height for every sequent that is derivable in PA. But this is not exactly what we want.

A *reduction procedure* for a sequent is a nondeterministic procedure that uses reduction rules to decompose this sequent into, let us say, simpler sequents. One path of this procedure represents one branch of the reduction tree and it must end in a finite number of steps with a sequent that is obviously valid. Hence, reduction rules generate a reduction tree for a sequent from the root to the leaves. However, it is more convenient to consider the reduction rules "upside down", i.e., as a set of derivation rules that proceed from the leaves (axioms) to the root. The sequent to reduce represents the conclusion of such a derivation rule and when we apply a single reduction rule to this sequent, we obtain sequents that represent premises. It is possible to get countably many of them. Under these circumstances, reduction trees can be understood as *deduction trees* and the inductive definition of these deduction trees is in accordance with the transfinite induction on the height of trees that Gentzen implicitly uses in his proof of 1935 [17].

In the analysis below, we shall consider two calculi: (1) The calculus of first order number theory with the ω-rule that is obtained by considering Gentzen's reduction rules as derivation rules. Derivations in this calculus will be called deduction trees. (2) A first order calculus for PA that derives sequents for which we need to construct deduction trees. One can easily see that the contradictory sequent $\rightarrow 0 = 1$ cannot have a deduction tree and since all sequents derivable in PA have one, this gives us the consistency of PA.

1.1.1 Gentzen's proof

Now, we explain Gentzen's method of the proof in [4]. First, he defines reduction rules that generate reduction trees for sequents that are valid in

the standard model of PA. The leaves of reduction trees are sequents in *endform*. The crucial property of sequents in endform is that their validity is easy to see. As we intend to use deduction trees instead of reduction trees, sequents in endform will be our initial sequents (axioms).

Second, Gentzen takes a sequent calculus for PA and shows that every sequent derivable in PA has a reduction tree. He does not actually use the notion reduction tree. He tries to avoid using trees, which are well-founded yet infinite, by using reduction procedures, where one path of the procedure conducts only a finite number of steps. Nevertheless, his treatment of reduction procedures would be more appropriate for trees and thus the use of reduction procedures seems to be artificial. Hence, we shall speak about reduction trees in connection with Gentzen too, even if it is not completely exact.

So, we can say that he constructs reduction trees for sequents derivable in PA and he does it by induction on the complexity of the derivation in PA. In the base step, he shows how to construct reduction trees for initial sequents. The height of these reduction trees is always finite. Then, he assumes that there already exist reduction trees for premises and he shows how to obtain a reduction tree for the conclusion. To find reduction trees for the conclusions of the rules of negation and the induction rule, he uses Hilfssatz. Hilfssatz is a statement essential for the proof. It shows that cut is an admissible rule in the calculus of first order number theory with the ω-rule, i.e., the calculus that we obtain by turning the reduction rules upside down. Gentzen's formulation of Hilfssatz is that one has a reduction tree for $\Gamma, \Delta \to C$ when $\Gamma \to D$ and $D, \Delta \to C$ have reduction trees ([4], p. 108). The proof of Hilfssatz proceeds by induction on the number of logical operations in the cut formula. When we want to prove the statement for cut formula D with n logical operations, we must know that it holds for simpler cut formulas and, furthermore, we must use induction on the height of the reduction tree for the second cut premise. This is the transfinite induction or, if you like, the bar induction discussed above.

Although Gentzen's Hilfssatz proves that the cut elimination works in the calculus of first order number theory with the ω-rule, the proof mentions no bounds on heights of reduction trees. Exactly these bounds are necessary to quantify the transfinite induction used in the consistency proof. The transfinite induction is hidden in Gentzen's proof; it is believed that his motive for this was that he wanted to reason "finitistically". But in fact, the transfinite induction is a part of the proof and it plays a significant role because it is the only principle of the proof that cannot be formalized in PA

unless it would be a transfinite induction up to some ordinal $\alpha < \varepsilon_0$ (see Section 1.5).

We know today that a successive decrease of the cut-rank during the cut elimination leads to a superexponential increase of the height of the derivation. If Gentzen had applied such a cut elimination strategy in his Hilfssatz, he would have obtained reduction trees for sequents derivable in PA with heights bounded by ε_0 [10, 12]. But the point is that his cut elimination strategy is different. He always eliminates an uppermost cut regardless of its complexity. When one tries to do a natural analysis of such a cut elimination strategy, one has to employ Veblen hierarchy and obtains reduction trees for sequents derivable in PA whose heights are less than $\Phi_\omega(0)$, where Φ_ω is the ω-th Veblen function. This is much greater than ε_0, the optimal ordinal for PA, and it is not clear whether Gentzen's elimination strategy is really so inefficient or the analysis can be done better.

In the following text, we prove that $\Phi_\omega(0)$ is an upper bound on the heights of reduction trees for sequents derivable in PA that are constructed with Hilfssatz. We would need to prove the lower bound on the heights of reduction trees to obtain an accurate answer to the question stated in the title of this part. Unfortunately, this seems to be difficult. A possible way to study this is to first analyse Gentzen's cut elimination strategy in finite calculi that do not contain arithmetic, especially classical first order logic.

Let us stress that our aim is not to do some kind of ordinal analysis for PA, but rather investigate the tools that Gentzen's consistency proof of 1935 applies, even if they are implicit in the proof or can be inefficient from the point of view of the current knowledge.

1.2   Preliminaries

In this section, we introduce necessary definitions and describe Gentzen's cut elimination strategy in more detail.

We shall use the language $\mathcal{L} = \{+, \cdot, S, =, 0\}$ of PA, where $+$ and $\cdot$ are binary functional symbols, S is a unary functional symbol, $=$ is a binary relational symbol and 0 is a constant symbol. The symbols have their usual meaning. For now, we will not use free variables. The treatment of them is covered in Section 1.4.

Definition 1. *A* sequent *is an expression of the form* $\Gamma \to B$ *where* $\Gamma = \{A_1, \ldots, A_n\}$ *are antecedent formulas and* B *is a succedent formula. There must always be exactly one succedent formula in every sequent. We view antecedent formulas as a multiset and it can be empty.*

7

Definition 2. *A sequent is said to be in* endform *when the following conditions are met: (1) It does not contain any free variables. (2a) Its succedent formula is a true equation, or (2b) its succedent formula is a false equation and there is at least one false equation among the antecedent formulas.*

Since we do not use free variables, the first condition is satisfied automatically. Sequents in endform will be treated as axioms because they are obviously valid. For the purpose of seeing the validity of a sequent, it would be enough to say that its succedent is a true equation or there is a false equation among the antecedent formulas. Nevertheless, Gentzen's definition of endform is as above. It probably resulted from his definition of reduction rules that start by reducing the succedent formula and go on to look for a false equation among the antecedent formulas only if the reduction rules applied to the succedent have turned it into a false equation.

Definition 3. *A* deduction tree *is a well-founded tree that consists of sequents. Each sequent is in endform or is derived from previous ones using one of the following derivation rules:*

$$\frac{\Gamma \to F(0) \qquad \Gamma \to F(\bar{1}) \qquad \Gamma \to F(\bar{2}) \ \ldots}{\Gamma \to \forall x F(x)} \ \forall R \qquad \frac{\Gamma \to A \qquad \Gamma \to B}{\Gamma \to A \& B} \ \& R \qquad \frac{\Gamma, A \to 0=1}{\Gamma \to \neg A} \ \neg R$$

$$\frac{\Gamma, F(\bar{n}) \to 0=1}{\Gamma, \forall x F(x) \to 0=1} \ \forall L_1 \qquad \frac{\Gamma, \forall x F(x), F(\bar{n}) \to 0=1}{\Gamma, \forall x F(x) \to 0=1} \ \forall L_2$$

$$\frac{\Gamma, A_i \to 0=1}{\Gamma, A_1 \& A_2 \to 0=1} \ \& L_1 \qquad \frac{\Gamma, A_1 \& A_2, A_i \to 0=1}{\Gamma, A_1 \& A_2 \to 0=1} \ \& L_2$$

$$\frac{\Gamma \to A}{\Gamma, \neg A \to 0=1} \ \neg L_1 \qquad \frac{\Gamma, \neg A \to A}{\Gamma, \neg A \to 0=1} \ \neg L_2$$

Equation $0 = 1$ stands for an arbitrary false equation. Similarly, we denote a true equation by $0 = 0$. The root of a deduction tree is called endsequent *and the leaves are called* initial sequents. *We have $i = 1$ or $i = 2$ in the rules of $\& L_1$ and $\& L_2$.*

Deduction trees are derivations in a formalization of first order number theory in the sequent calculus with the ω-rule. The derivation rules of this calculus turned upside down are Gentzen's reduction rules. For example, a sequent of the form $\Gamma \to A \& B$ must be reduced to $\Gamma \to A$ or $\Gamma \to B$, but, viewed as a game, this choice is made by our opponent, so, we must be able to reduce both to endform if necessary. On the other hand, we need to find one right possibility that we are able to reduce to endform when applying reduction rules to the antecedent formulas. So, for example, we

8

can reduce a sequent of the form $\Gamma, A_1 \& A_2 \to 0 = 1$ to $\Gamma, A_1 \to 0 = 1$, but if we are not sure that A_1 is really the right choice, we can reduce it to $\Gamma, A_1 \& A_2, A_1 \to 0 = 1$. This allows us to change our mind and choose A_2 later. Formally, the L-rules have two variants so that there is no need to introduce the rule of contraction. See [9] for more on the reduction rules. We do not use the explicit rule of cut.

Definition 4. *We define the* height *of a tree as follows: Every leaf has height 0. If the children of a node have heights $\alpha_0, \alpha_1, \ldots$, then the node has height $\alpha = sup\{\alpha_0 + 1, \ \alpha_1 + 1, \ldots\}$. The height of a tree is the ordinal of its root.*

We write as $\vdash^\alpha \Gamma \to B$ to mean that we have a deduction tree with endsequent $\Gamma \to B$ whose height is at most α.

Definition 5. *We denote by $|C|$ the* number of logical operations *in formula C.*

Lemma 1. *If we have a deduction tree T with the endsequent $\Delta \to C$ whose height is α, then we also have a deduction tree S with the endsequent $A, \Delta \to C$ whose height is α too.*

Proof. The deduction tree S for $A, \Delta \to C$ is essentially the same as the deduction tree T for $\Delta \to C$. We simply add A to the antecedent of every sequent in T. $\square$

Lemma 2. *If we have a deduction tree T with the endsequent $A, A, \Gamma \to C$ whose height is α, then we also have a deduction tree S with the endsequent $A, \Gamma \to C$ whose height is α too.*

Proof. The deduction tree S with the endsequent $A, \Gamma \to C$ conducts the same derivation rules as T with one exception. The *first* rule of L_1 in every branch of T that operates on an antecedent formula A and that we meet when we proceed from the endsequent $A, A, \Gamma \to C$ towards the leaves is replaced by the rule of L_2 that operates on A too, so that the particular occurrence of A stays preserved in S. $\square$

Lemma 3. *If we have a deduction tree T with the endsequent $0 = 0, \Delta \to C$ whose height is α, then we also have a deduction tree S with the endsequent $\Delta \to C$ whose height is α too.*

Proof. The deduction tree S is constructed from T by deleting the antecedent formulas $0 = 0$ from every sequent. Formula $0 = 0$ is redundant in the antecedent because no rule is able to operate on $0 = 0$ and it does not influence whether a sequent is in endform. $\square$

Definition 6. *A structure tree for an atom A is the tree whose only node is its root called A. Let us denote an arbitrary logical operation by $\circ$. The context makes it clear whether it stands for a binary or a unary operation. Assume that we already have structure trees T_1 and T_2 for formulas A_1 and A_2, respectively. Then, the structure trees for $A_1 \circ A_2$ and $\circ A_1$ are of the form:*

$$
\begin{array}{ccc}
\overset{\circ}{\overbrace{\quad\quad}} & \quad and \quad & \overset{\circ}{\underset{T_1}{|}} \\
T_1 \quad\quad T_2 & &
\end{array}
$$

respectively.

Definition 7. *Assume that D is a formula. We denote by $rk(D)$ the height of the structure tree for D. Assume that S is a sequent $A_1, A_2, \ldots, A_n \to B$. We set*

$$rk(S) = rk(A_1) + \cdots + rk(A_n) + rk(B)$$

Lemma 4. *If T is a deduction tree with the endsequent S where S contains no free variables and no L_2 rules are used in T, then T has height at most $rk(S)$.*

Proof. Assume that sequent S is of the form $A_1, \ldots, A_n \to B$. We proceed by induction on $rk(S)$. If $rk(S) = 0$, then all formulas $A_1, \ldots, A_n, B$ are atomic and since T is a deduction tree for S, sequent S must be in endform. Hence, the height of T is $0 \leq rk(S)$.

Assume that $rk(S) > 0$. Then, there is at least one formula in S, for example A_i, that is not atomic. Assume that B is of the form $0 = 1$ and the rule of L_1 has been applied in T to obtain S:

$$
\frac{
\begin{array}{c}
\vdots \\
A_1, \ldots, A_i', \ldots, A_n \to 0 = 1
\end{array}
}{A_1, \ldots, A_i, \ldots, A_n \to 0 = 1} \; L_1
$$

We have $rk(A_i') < rk(A_i)$, thus,

$$rk(A_1, \ldots, A_i', \ldots, A_n \to 0 = 1) + 1 \leq rk(A_1, \ldots, A_i, \ldots, A_n \to 0 = 1) = rk(S)$$

Since the induction hypothesis gives us that the left hand side of the inequality is the bound on the height of T, we obtain the required result. Other derivation rules are treated similarly. $\square$

1.2.1 Gentzen's elimination strategy in Hilfssatz

In this section, we want to present our interpretation of the proof of Gentzen's Hilfssatz. In order to do this, we temporarily add an explicit rule of cut to the calculus from Definition 3, whereas cut-free derivations in this calculus are still called deduction trees. To distinguish between cut-free derivations and derivations with cuts, we will denote the cut-free derivability of sequent $\Gamma \to A$ by $\vdash_0 \Gamma \to A$.

Hilfssatz says: If $\Gamma \to D$ and $D, \Delta \to C$ are sequents without free variables that have deduction trees T_1 and T_2, respectively, then $\Gamma, \Delta \to C$ has a deduction tree, too. Hence we have a derivation of the form

$$\frac{\vdots\, T_1 \qquad \vdots\, T_2}{\vdash_0\, \Gamma \to D \qquad \vdash_0\, D, \Delta \to C} \vartheta \atop \Gamma, \Delta \to C$$

and we need to eliminate cut ϑ. We apply two induction arguments: (1) Induction on the number of the logical operations in the cut formula and (2) induction on the height of the deduction tree for the second cut premise. Hence, we need two bases of induction: First basis is for the case when the cut formula is an atom. Second basis is for the case when the height of the deduction tree for the second cut premise is 0, i.e., when the second cut premise is in endform. Cuts on atomic sentences can be eliminated without difficulty (Lemma 16). On the other hand, when the second cut premise is in endform and the cut formula is not an atom, the conclusion of the cut is in endform too and no cut is needed anymore (Lemma 17).

Next, we take a cut with an arbitrary cut formula D and distinguish cases according to the last derivation rule in T_2. If this rule operates on the succedent formula or on an antecedent formula other than the cut formula D, we easily move the cut up along T_2, the deduction tree for the second cut premise:

$$\cfrac{\vdash_0 \Gamma \to D \quad \cfrac{\vdash_0 D, \Delta \to A \quad \vdash_0 D, \Delta \to B}{\vdash_0 D, \Delta \to A\&B}\&R}{\Gamma, \Delta \to A\&B}\vartheta \quad \rightsquigarrow \quad \cfrac{\cfrac{\vdash_0 \Gamma \to D \quad \vdash_0 D, \Delta \to A}{\Gamma, \Delta \to A}cut \quad \cfrac{\vdash_0 \Gamma \to D \quad \vdash_0 D, \Delta \to B}{\Gamma, \Delta \to B}cut}{\Gamma, \Delta \to A\&B}\&R$$

This can lead to several new cuts of the same complexity as the old one. But since the heights of the deduction trees for the second cut premises got smaller, we remove these cuts completely by applying the induction hypothesis.

11

The situation is different when the last derivation rule in T_2 works on the cut formula D. The cases for cut formulas D with a conjunction or a universal quantifier as an outermost logical operation are essentially the same, hence, let us take D of the form $A\&B$ and $\neg A$. When the rule in question is the rule of L_1, we obtain one new cut with simpler cut formula:

$$\&R\cfrac{\cfrac{\vdash_0 \Gamma\to A \qquad \vdash_0 \Gamma\to B}{\vdash_0 \Gamma\to A\&B} \qquad \cfrac{\vdash_0 A,\Delta\to 0=1}{\vdash_0 A\&B,\Delta\to 0=1}\&L_1}{\Gamma,\Delta\to 0=1}\vartheta \quad\rightsquigarrow\quad \cfrac{\vdash_0 \Gamma\to A \qquad \vdash_0 A,\Delta\to 0=1}{\Gamma,\Delta\to 0=1}cut$$

$$\neg R\cfrac{\cfrac{\vdash_0 A,\Gamma\to 0=1}{\vdash_0 \Gamma\to\neg A} \qquad \cfrac{\vdash_0 \Delta\to A}{\vdash_0 \neg A,\Delta\to 0=1}\neg L_1}{\Gamma,\Delta\to 0=1}\vartheta \quad\rightsquigarrow\quad \cfrac{\vdash_0 \Delta\to A \qquad \vdash_0 A,\Gamma\to 0=1}{\Gamma,\Delta\to 0=1}cut$$

The definition of derivation rules ensures that the last rule in T_1 must be the rule of $\&R$ or $\neg R$, respectively. One of the premises of this R-rule becomes one of the premises of the new cut. As expected, we now apply the induction hypothesis for simpler cut formulas. This removes the cut on A.

Finally, assume that the last derivation rule in T_2 is the rule of L_2 that operates on the cut formula $A\&B$ or $\neg A$, respectively. Now we obtain two new cuts and both induction hypotheses are needed:

$$\&R\cfrac{\cfrac{\vdash_0 \Gamma\to A \quad \vdash_0 \Gamma\to B}{\vdash_0 \Gamma\to A\&B} \quad \cfrac{\vdash_0 A\&B,A,\Delta\to 0=1}{\vdash_0 A\&B,\Delta\to 0=1}\&L_2}{\Gamma,\Delta\to 0=1}\vartheta \rightsquigarrow \cfrac{\vdash_0 \Gamma\to A \quad \cfrac{\vdash_0 \Gamma\to A\&B \quad \vdash_0 A\&B,A,\Delta\to 0=1}{\Gamma,A,\Delta\to 0=1}cut}{\Gamma,\Gamma,\Delta\to 0=1}cut$$

$$\neg R\cfrac{\cfrac{\vdash_0 A,\Gamma\to 0=1}{\vdash_0 \Gamma\to\neg A} \quad \cfrac{\vdash_0 \neg A,\Delta\to A}{\vdash_0 \neg A,\Delta\to 0=1}\neg L_2}{\Gamma,\Delta\to 0=1}\vartheta \rightsquigarrow \cfrac{\cfrac{\vdash_0 \Gamma\to\neg A \quad \vdash_0 \neg A,\Delta\to A}{\Gamma,\Delta\to A}cut \quad \vdash_0 A,\Gamma\to 0=1}{\Gamma,\Gamma,\Delta\to 0=1}cut$$

The new cuts on $A\&B$ and $\neg A$, respectively, are removed because the deduction tree of the second cut premise has a smaller height. Thus, we obtain $\vdash_0 \Gamma, A, \Delta \to 0 = 1$ and $\vdash_0 \Gamma, \Delta \to A$, respectively. Since both premises of the new cut on A have deduction trees and the cut formula A is simpler than $A\&B$ and $\neg A$, we can use the induction hypothesis for simpler cut formulas. The result is $\vdash_0 \Gamma, \Gamma, \Delta \to 0 = 1$. Lemma 2 changes it to a deduction tree for $\Gamma, \Delta \to 0 = 1$.

Assume now that we do not apply the induction hypotheses to say that

there exist deduction trees for certain sequents, but we rather unfold the induction and construct the deduction trees explicitly. This means that we have to move the original cut ϑ up along the deduction tree T_2. Every time when we encounter the rule of L_2 that operates on the cut formula D, the algorithm produces two cuts above each other. The cut at the lower position is simpler than ϑ. The cut above this simpler one is of the same complexity as ϑ. Since our cut elimination algorithm knows only how to eliminate cuts whose premises have deduction trees, we go on to eliminate the cut of the same complexity as ϑ. Hence, we continue moving it up along the deduction tree for the second cut premise, whereas the simpler cut stays in the place where it was created. It waits there until it is its turn to be eliminated, i.e., until there are no other cuts above it. Since there can be an arbitrary finite number of the rules of L_2 that work on the cut formula D above each other in every branch of T_2 and all these rules are transformed into cuts on subformulas of D during the elimination of ϑ, we can obtain an arbitrary finite number of simpler cuts above each other in every branch of T_2. All these cuts have to be eliminated in the same way as ϑ, i.e., when any of these cuts becomes an uppermost cut, we move it up and every rule of L_2 that works on the cut formula gives us a new simpler cut.

As we know how to eliminate cuts with atomic cut formulas and cuts with the sequent in endform as the second cut premise, the process described above must give us a deduction tree with the endsequent $\Gamma, \Delta \rightarrow C$ in the end, possibly with the use of Lemma 2.

1.2.2   Veblen hierarchy

In this section, we introduce Veblen functions that were originally developed by Oswald Veblen in [18]. Our elaboration is based on ([11], pp. 73–84) that contains a very good explanation of this topic. We shall work only with countable ordinals whose union will be denoted by $\mathbb{O}$. It follows that $\mathbb{O}$ itself is not countable.

Definition 8. *We say that $A \subseteq \mathbb{O}$ is an $\mathbb{O}$-segment if the following holds:*

$$\forall \xi < \eta \quad (\eta \in A \;\supset\; \xi \in A)$$

Definition 9. *We say that $f : A \rightarrow B$ is an ordering function of a set $B \subseteq \mathbb{O}$ if the following holds:*

   1. Set A is an $\mathbb{O}$-segment.

   2. Function f is strictly monotone: $\forall \eta, \xi \in A$ we have $\eta < \xi \;\supset\; f(\eta) < f(\xi)$.

3. The image of f is the whole set B.

Note that an ordering function $f : A \to B$ is a bijection. It is onto by property 3 and it is also one-to-one since otherwise we obtain a contradiction with property 2. It is easy to see that every set $B \subseteq \mathbb{O}$ has exactly one ordering function. If $B = \emptyset$, then the empty function $\varnothing \to \varnothing$ is the ordering function of B.

Lemma 5. *If $f : A \to B$ is an ordering function of B, then $\alpha \leq f(\alpha)$ for all $\alpha \in A$.*

Proof. Assume that there exists $\alpha \in A$ such that $f(\alpha) < \alpha$. All numbers $\beta < \alpha$ belong to A because A is an $\mathbb{O}$-segment. Set $C = \{\alpha, f(\alpha) < \alpha\}$ is non-empty by assumption and it has the smallest member $\alpha_0 \in A$. Assumption $f(\alpha_0) < \alpha_0$ together with monotonicity give us (1) $f(f(\alpha_0)) < f(\alpha_0)$. Since α_0 is the smallest in C and $f(\alpha_0) < \alpha_0$, value $f(\alpha_0)$ is not in C and we have (2) $f(\alpha_0) \leq f(f(\alpha_0))$. Transitivity applied to (1) and (2) gives us $f(\alpha_0) < f(\alpha_0)$. This is a contradiction. $\square$

Definition 10. *We say that $B \subseteq \mathbb{O}$ is* unbounded *when for every $\alpha \in \mathbb{O}$ there exists $\beta \in B$ such that $\alpha < \beta$.*

Definition 11. *We say that $B \subseteq \mathbb{O}$ is* closed *when for every non-empty countable set $M \subseteq B$ we also have $\sup M \in B$.*

Definition 12. *We say that an ordering function $f : A \to B$ is* continuous *if A is closed and for every non-empty countable set $U \subseteq A$ we have $f(\sup U) = \sup f(U)$.*

Set A has to be closed so that f can make the image of $\sup U$.

Definition 13. *We say that an ordering function $f : A \to B$ is* normal *if it is continuous and $A = \mathbb{O}$.*

Lemma 6. *The ordering function $f : A \to B$ of a set $B \subseteq \mathbb{O}$ is continuous if and only if B is closed.*

Proof. Assume that f is continuous. We want to prove that B is closed. Hence, let $M \subseteq B$ be non-empty and countable. We want to show that $\sup M \in B$. Function f is bijective, so, for set M there exists $U \subseteq A$ non-empty and countable such that $f(U) = M$. We have

$$\sup M = \sup f(U) = f(\sup U)$$

The second equality holds because f is continuous. We see that $\sup M \in \mathrm{Rng}\,(f) = B$.

On the other hand, assume that B is closed. We aim to prove that f is continuous. Assume that $U \subseteq A$ is non-empty and countable. There are two properties that have to be shown: (1) $\sup U \in A$ and (2) $f(\sup U) = \sup f(U)$.

Set U has the image $f(U) \subseteq B$ that is non-empty and countable. Since B is closed, we have $\sup f(U) \in B$. Then, there is $\alpha \in A$ such that $f(\alpha) = \sup f(U)$. It follows that $\forall \xi \in U : f(\xi) \leq f(\alpha)$. Function f is strictly monotone, hence, $\forall \xi \in U : \xi \leq \alpha$. Clearly, $\sup U \leq \alpha$. Since $\alpha \in A$ and A is an $\mathbb{O}$-segment, $\sup U \in A$.

Inequality $\sup U \leq \alpha$ gives us $f(\sup U) \leq f(\alpha)$. By definition, $f(\alpha) = \sup f(U)$, thus, (i) $f(\sup U) \leq \sup f(U)$. Further, $\forall \xi \in U : \xi \leq \sup U$ and strict monotonicity yields $f(\xi) \leq f(\sup U)$. Hence, (ii) $\sup f(U) \leq f(\sup U)$. Observations (i) and (ii) give us $f(\sup U) = \sup f(U)$. $\qquad \square$

Lemma 7. *The ordering function $f : A \to B$ of a set $B \subseteq \mathbb{O}$ is normal if and only if B is closed and unbounded.*

Proof. Assume that f is continuous and $A = \mathbb{O}$. Lemma 6 gives us that B is closed. We proceed to prove that it is also unbounded. Assume that it is not the case and we have $\alpha \in \mathbb{O}$ such that $\forall \xi \in B : \xi < \alpha$. Since f is the ordering function of B, it follows that $A \subset \mathbb{O}$ and we obtain a contradiction.

Assume now that B is closed and unbounded. According to Lemma 6, it suffices to show that $A = \mathbb{O}$. Since B is unbounded, it is not countable because a countable set cannot be cofinal in $\mathbb{O}$. As f is a bijection, it follows that A must be equal to $\mathbb{O}$. $\qquad \square$

Lemma 8. *Every ordinal number $\beta \neq 0$ can be written in Cantor normal form*

$$\beta = \sum_{i=1}^{n} \omega^{\delta_i} \cdot m_i,$$

where $n, m_i \neq 0$; $n, m_i \in \mathrm{N}$; $\delta_1 > \delta_2 > \ldots > \delta_n$.

Proof. We proceed by induction on β. Assume that all $\gamma < \beta$ have their Cantor normal forms. We want to find one for β. Since the function $F(\alpha) = \omega^\alpha$ is strictly monotone, set $\{\alpha, \ \beta < \omega^\alpha\}$ is non-empty and it has the smallest member. Let us denote it by α_0. Ordinal α_0 cannot be limit. If it was limit, then the definition of exponentiation gives us $\omega^{\alpha_0} = \sup \{\omega^\eta, \ \eta < \alpha_0\}$ and there would exist $\eta < \alpha_0$ such that $\beta < \omega^\eta < \omega^{\alpha_0}$. This contradicts the choice of α_0. Hence, α_0 is a successor ordinal. There exists β_0 such that $\alpha_0 = \beta_0 + 1$ and $\omega^{\beta_0} \leq \beta$. If $\beta = \omega^{\beta_0}$, we have the required form. Thus, assume $\omega^{\beta_0} < \beta$. We are able to divide ordinals

$$\exists! \mu \exists! \nu \ (\beta = \omega^{\beta_0} \cdot \mu + \nu \quad \& \quad \nu < \omega^{\beta_0})$$

where ν is the remainder and μ must be a natural number. If it was not, then $\beta \geq \omega^{\beta_0} \cdot \omega + \nu = \omega^{\beta_0+1} + \nu = \omega^{\alpha_0} + \nu$ and this contradicts the choice of α_0. Hence,

$$\beta = \omega^{\beta_0} \cdot m + \nu$$

where $m \in \mathrm{N}$. Since $\nu < \beta$, the induction hypothesis gives us Cantor normal form for ν. The construction reveals that the greatest exponent in Cantor normal form for ν is smaller than β_0. $\qquad \square$

Cantor normal form for any ordinal $\beta \neq 0$ is given unambiguously, but it can also be expressed as

$$\beta = \sum_{i=1}^{n} \omega^{\delta_i},$$

where $n \neq 0$; $n \in \mathrm{N}$; $\delta_1 \geq \delta_2 \geq \ldots \geq \delta_n$. This is obviously equivalent to the form above. Nevertheless, the current form has an advantage over that one from Lemma 8 when we compare ordinals in Cantor normal form: α is greater than β when the first exponent from the left in which α and β differ is greater in α.

Definition 14. *We say that ordinal α is an* additive principal number *if $\alpha \neq 0$ and $\forall \xi < \alpha : \ (\xi + \alpha = \alpha)$.*

The smallest additive principal number is 1 immediately followed by ω.

Lemma 9. *If α is an additive principal number, then*

$$\forall \xi, \chi < \alpha : \ (\xi + \chi < \alpha)$$

Proof. Our assumption $\chi < \alpha$ yields $\xi + \chi < \xi + \alpha$. Further, the assumption that α is an additive principal number gives us $\xi + \chi < \xi + \alpha = \alpha$. $\qquad \square$

Lemma 10. *The set of additive principal numbers is closed and unbounded.*

Proof. First, we prove the unboundedness and then the closure.

(1) Assume that $\alpha \in \mathbb{O}$. We want to find an additive principal number greater than α. Let us define the following sequence of ordinals:

$$\begin{aligned}
\beta_0 &= \alpha + 1 \\
\beta_{n+1} &= \beta_n + \beta_n \\
M &= \{\beta_n; \ n \in \mathrm{N}\} \\
\beta &= \sup M
\end{aligned}$$

We show that $\beta \neq 0$ is an additive principal number greater than α. Choose an arbitrary $\xi < \beta$. Then, there exists $n \in \mathrm{N}$ such that $\xi < \beta_n$. It follows for

all $m \geq n : (\xi + \beta_m \leq \beta_m + \beta_m = \beta_{m+1} < \beta)$. Hence sup $(\xi + M) \leq \beta$. This observation and the properties of addition give us: $\beta \leq \xi + \beta = \sup (\xi + M) \leq \beta$. We obtain that $\xi + \beta = \beta$ for all $\xi < \beta$.

(2) Let U be a non-empty and countable set of additive principal numbers. We want to show that sup U is an additive principal number. By now, we know that $0 \neq \sup U$. Choose an arbitrary $\xi < \sup U$. Then, there exists $\alpha \in U$ such that $\xi < \alpha$. If we take $\beta \in U$ with $\beta \geq \alpha$, we obtain $\xi + \beta = \beta$ because β is an additive principal number by assumption. Hence, we have $\xi + \sup U = \sup \{\xi + \beta; \beta \in U\} = \sup U$. Clearly, sup U is an additive principal number. $\square$

Lemmas 7 and 10 give us that the ordering function of the set of all additive principal numbers is normal. We shall investigate now what these numbers look like.

Lemma 11. *For all $\delta < \alpha$ we have $\omega^\delta + \omega^\alpha = \omega^\alpha$.*

Proof. We proceed by induction on α. Assume that $\alpha = \gamma + 1$ is a successor ordinal and $\delta \leq \gamma$. Then

$$\omega^\delta + \omega^{\gamma+1} \leq \omega^\gamma + \omega^{\gamma+1} = \omega^\gamma + \omega^\gamma \cdot \omega = \omega^\gamma(1 + \omega) = \omega^\gamma \cdot \omega = \omega^{\gamma+1}$$

Since we know that $\omega^{\gamma+1} \leq \omega^\delta + \omega^{\gamma+1}$, we obtain the required equality.

Assume that $\alpha = \lambda$ is a limit ordinal and $\omega^\delta < \omega^\lambda = \sup \{\omega^\xi, \xi < \lambda\}$. Hence, there exists $\xi_0 < \lambda$ such that $\omega^\delta < \omega^{\xi_0}$. The induction hypothesis gives us then for all $\chi < \lambda$ such that $\chi \geq \xi_0$: $(\omega^\delta + \omega^\chi = \omega^\chi)$. We can now calculate:

$$\omega^\delta + \omega^\lambda = \omega^\delta + \sup\{\omega^\xi, \xi < \lambda\} = \sup\{\omega^\delta + \omega^\xi, \xi < \lambda\} = \sup\{\omega^\xi, \xi < \lambda\} = \omega^\lambda \quad \square$$

Lemma 12. *Ordinal β is an additive principal number if and only if there exists α such that $\beta = \omega^\alpha$.*

Proof. Assume that β has more than one summand in its Cantor normal form, i.e., $\beta = \omega^{\delta_1} + \cdots + \omega^{\delta_n}$ where $\delta_1 \geq \cdots \geq \delta_n$ and $n \in \mathbb{N}$. Now, it is easy to find ordinal $\xi < \beta$ which witnesses that β is not an additive principal number. We can take for example $\xi = \omega^{\delta_1}$.

Assume that $\beta = \omega^\alpha$ and $\xi < \beta$. Ordinal ξ has its Cantor normal form $\xi = \omega^{\chi_1} + \cdots + \omega^{\chi_k}$; $\chi_1 \geq \cdots \geq \chi_k$. Since $\xi < \beta$, we have $\chi_1 < \alpha$ and, thus, $\chi_1, \ldots, \chi_k < \alpha$. Now, we are ready to calculate:

$$\xi + \beta = \omega^{\chi_1} + \cdots + \omega^{\chi_k} + \omega^\alpha = \omega^\alpha = \beta$$

We used associativity of addition and Lemma 11. $\square$

Definition 15. *We define a set $Cr(\alpha) \subset \mathbb{O}$ and a function Φ_α for each $\alpha \in \mathbb{O}$ inductively as follows:*

1. *$Cr(0)$ is the set of all additive principal numbers.*

2. *Function $\Phi_\alpha : A_\alpha \to Cr(\alpha)$ is the ordering function of $Cr(\alpha)$.*

3. *If $\alpha \neq 0$, then it holds that $\eta \in Cr(\alpha)$ if and only if $\forall \xi < \alpha : (\eta \in A_\xi \,\&\, \Phi_\xi(\eta) = \eta)$.*

We call the members of $Cr(\alpha)$ the α-critical ordinals and the ordering functions Φ_α of sets $Cr(\alpha)$ are called Veblen functions.

The third item says that $\mathrm{Cr}(\alpha)$ is the set of all common fixed points of all functions Φ_ξ where $\xi < \alpha$. The following properties of α-critical ordinals and Veblen functions can be easily seen:

- $\alpha < \beta \,\supset\, \mathrm{Cr}(\beta) \subseteq \mathrm{Cr}(\alpha)$

- Every ordinal $\Phi_\alpha(\beta)$ is an additive principal number.

- $\Phi_0(\alpha) = \omega^\alpha$

Lemma 13. *The set $Cr(\alpha)$ is closed and unbounded.*

Proof. We proceed by transfinite induction on α. Set $\mathrm{Cr}(0)$ of all additive principal numbers is closed and unbounded according to Lemma 10. Assume that $\xi < \alpha$. Set $\mathrm{Cr}(\xi)$ is closed and unbounded by the induction hypothesis. Hence, the ordering function $\Phi_\xi : A_\xi \to \mathrm{Cr}(\xi)$ is normal according to Lemma 7 and particularly $A_\xi = \mathbb{O}$. First, we show that $\mathrm{Cr}(\alpha)$ is unbounded and then that it is also closed.

Assume that $\beta \in \mathbb{O}$. We need to find an element in $\mathrm{Cr}(\alpha)$ that is greater than β. Let us define:

$$
\begin{aligned}
\gamma_0 &= \beta + 1 \\
\gamma_{n+1} &= \sup\{\Phi_\xi(\gamma_n);\ \xi < \alpha\} \\
U &= \{\gamma_n;\ n \in \mathbb{N}\} \\
\gamma &= \sup U
\end{aligned}
$$

We want to argue that $\gamma > \beta$ is an α-critical ordinal. This means that for all $\xi < \alpha$, we need $\Phi_\xi(\gamma) = \gamma$. Choose an arbitrary $\xi < \alpha$. We have:

$$\Phi_\xi(\gamma_n) \leq \gamma_{n+1} \leq \gamma$$

Therefore, $\gamma \geq \sup\{\Phi_\xi(\gamma_n); \ \gamma_n \in U\}$. Since Φ_ξ is continuous by the induction hypothesis, it holds:

$$\gamma \geq \sup\{\Phi_\xi(\gamma_n); \ \gamma_n \in U\} = \Phi_\xi(\gamma)$$

The second inequality $\gamma \leq \Phi_\xi(\gamma)$ follows from Lemma 5.

Assume now that $M \subseteq \mathrm{Cr}(\alpha)$ is non-empty and countable. We want to show that $\sup M \in \mathrm{Cr}(\alpha)$. For this, $\sup M$ must be a fixed point of all Φ_ξ, $\xi < \alpha$. Choose an arbitrary $\xi < \alpha$. The domain A_ξ of Φ_ξ is equal to $\mathbb{O}$, hence we can calculate:

$$\Phi_\xi(\sup M) = \sup\{\Phi_\xi(\mu); \ \mu \in M\} = \sup\{\mu; \ \mu \in M\} = \sup M$$

We obtain this because Φ_ξ is continuous by the induction hypothesis and M contains only fixed points of functions Φ_ξ, $\xi < \alpha$. $\qquad\square$

Lemmas 7 and 13 yield that every Veblen function is normal. The last issue is to investigate how numbers represented by Veblen functions can be compared.

Lemma 14. *We have $\Phi_{\alpha_1}(\beta_1) = \Phi_{\alpha_2}(\beta_2)$ if and only if one of the following conditions is met:*

- *$\alpha_1 < \alpha_2$ and $\beta_1 = \Phi_{\alpha_2}(\beta_2)$*

- *$\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$*

- *$\alpha_2 < \alpha_1$ and $\Phi_{\alpha_1}(\beta_1) = \beta_2$*

Proof. By definition, $\Phi_{\alpha_2}(\beta_2)$ is an element of $\mathrm{Cr}(\alpha_2)$, i.e., it is a fixed point of all Φ_ξ, $\xi < \alpha_2$. Since α_1, α_2 are ordinal numbers, exactly one of the following must occur: $\alpha_1 < \alpha_2$, $\alpha_1 = \alpha_2$, $\alpha_2 < \alpha_1$. If $\alpha_1 < \alpha_2$, then $\Phi_{\alpha_1}(\Phi_{\alpha_2}(\beta_2)) = \Phi_{\alpha_2}(\beta_2)$ and we have that $\Phi_{\alpha_1}(\beta_1) = \Phi_{\alpha_2}(\beta_2)$ if and only if $\beta_1 = \Phi_{\alpha_2}(\beta_2)$. The case for $\alpha_2 < \alpha_1$ is similar and the case for $\alpha_1 = \alpha_2$ is obvious. $\qquad\square$

Lemma 15. *We have $\Phi_{\alpha_1}(\beta_1) < \Phi_{\alpha_2}(\beta_2)$ if and only if one of the following conditions is met:*

- *$\alpha_1 < \alpha_2$ and $\beta_1 < \Phi_{\alpha_2}(\beta_2)$*

- *$\alpha_1 = \alpha_2$ and $\beta_1 < \beta_2$*

- *$\alpha_2 < \alpha_1$ and $\Phi_{\alpha_1}(\beta_1) < \beta_2$*

Proof. By definition, $\Phi_{\alpha_2}(\beta_2)$ is an element of $\mathrm{Cr}\,(\alpha_2)$ and one of these possibilities must occur: $\alpha_1 < \alpha_2$, $\alpha_1 = \alpha_2$, $\alpha_2 < \alpha_1$. If $\alpha_1 < \alpha_2$, then $\Phi_{\alpha_1}(\Phi_{\alpha_2}(\beta_2)) = \Phi_{\alpha_2}(\beta_2)$ and we obtain that $\Phi_{\alpha_1}(\beta_1) < \Phi_{\alpha_2}(\beta_2)$ if and only if $\beta_1 < \Phi_{\alpha_2}(\beta_2)$ since Φ_{α_1} is strictly monotone. The case for $\alpha_2 < \alpha_1$ is similar and the case for $\alpha_1 = \alpha_2$ follows again from the monotonicity. $\square$

As an example, we show that $\omega < \Phi_\omega(0)$. We use obvious equalities $\omega = \Phi_0(1)$ and $1 = \Phi_0(0)$:

$$
\begin{aligned}
\Phi_0(1) < \Phi_\omega(0) &\equiv \\
0 < \omega \;\&\; \Phi_0(0) < \Phi_\omega(0) &\equiv \\
0 < \omega \;\&\; 0 < \Phi_\omega(0) &
\end{aligned}
$$

This holds true because any additive principal number is greater than 0.

Let us summarize the properties of Veblen functions that we have shown in this section and that we will use below:

(V1) $\beta_1 < \beta_2 \quad \supset \quad \Phi_\alpha(\beta_1) < \Phi_\alpha(\beta_2)$

(V2) $\beta_1, \beta_2 < \Phi_\alpha(\beta) \quad \supset \quad \beta_1 + \beta_2 < \Phi_\alpha(\beta)$

(V3) $\alpha_1 < \alpha_2 \quad \supset \quad \Phi_{\alpha_1}(\Phi_{\alpha_2}(\beta)) = \Phi_{\alpha_2}(\beta)$

(V4) $\alpha_1 < \alpha_2 \;\&\; \beta_1 < \Phi_{\alpha_2}(\beta_2) \quad \supset \quad \Phi_{\alpha_1}(\beta_1) < \Phi_{\alpha_2}(\beta_2)$

(V5) $\Phi_0(\beta) = \omega^\beta$

(V6) $\omega < \Phi_\omega(0)$

(V7) $\sup\{\Phi_\alpha(\beta_1), \Phi_\alpha(\beta_2), \Phi_\alpha(\beta_3), \ldots\} = \Phi_\alpha(\sup\{\beta_1, \beta_2, \beta_3, \ldots\})$

(V8) $\beta \leq \Phi_\alpha(\beta)$

1.3   Cut elimination

Lemma 16. *Assume that sequents $\Gamma \to D$ and $D, \Delta \to C$ have deduction trees T_1 and T_2 with heights α_1 and α_2, respectively. Further assume that D is an atomic sentence. Then, sequent $\Gamma, \Delta \to C$ has a deduction tree whose height is at most $\alpha_1 + \alpha_2$.*

Proof. If D is a true atomic sentence, then $\Gamma \to D$ is in endform and $\alpha_1 = 0$. Deduction tree T_2 and Lemma 3 give us a deduction tree with the endsequent $\Delta \to C$ whose height is α_2. When we apply Lemma 1, we obtain a deduction tree with the endsequent $\Gamma, \Delta \to C$ whose height is $\alpha_2 = \alpha_1 + \alpha_2$.

If D is a false atomic sentence, we decompose the succedent formula C from $\Gamma, \Delta \to C$ the same way as deduction tree T_2 does. This turns the succedent formula either into a true atomic sentence or into a false atomic sentence:

$$
\cfrac{
\cfrac{\vdash^0 \ \Gamma, \Delta \to 0=0}{\vdots}
\qquad
\cfrac{\cfrac{\vdots \ T_1}{\vdash^{\alpha_1} \Gamma, \Delta \to 0=1}}{\vdots}
\qquad
\cfrac{\cfrac{\cfrac{\vdots \ T_1}{\vdash^{\alpha_1} \Gamma, \Delta, E \to 0=1}}{\vdash \Gamma, \Delta \to \neg E} \ \neg R}{\vdots}
}{\vdash^{\alpha_1 + \alpha_2} \ \Gamma, \Delta \to C} \ R\text{-rules}
$$

Lemma 1 allows us to attach T_1 to the nodes where C has turned into a false atomic sentence. Hence, the deduction tree for $\Gamma, \Delta \to C$ can be viewed as T_2 whose leaves are assigned α_1 instead of 0. It is easy to prove by induction on α_2 that the deduction tree for $\Gamma, \Delta \to C$ has height at most $\alpha_1 + \alpha_2$. $\quad\square$

Lemma 17. *Assume that sequent $D, \Delta \to C$ is in endform and D is not atomic. Then, sequent $\Gamma, \Delta \to C$ is in endform too.*

Proof. If $D, \Delta \to C$ is in endform and D is not atomic, then C is either of the form $0 = 0$ or C is of the form $0 = 1$ and there must be a false atomic sentence in Δ. Hence, sequent $\Gamma, \Delta \to C$ is obviously in endform. $\quad\square$

It is easy to see that if we take monotone cut formula D, i.e., the one that contains only conjunctions and universal quantifiers but no negations, then Gentzen's elimination procedure gives us a deduction tree for $\Gamma, \Delta \to C$ whose height is at most $\alpha_1 + \alpha_2$. However, things are more complicated when we include negation.

Theorem 1. *Assume that sequents $\Gamma \to D$ and $D, \Delta \to C$ have deduction trees T_1 and T_2 with heights α_1 and α_2, respectively, and $|D| = n$. Then, sequent $\Gamma, \Delta \to C$ has a deduction tree whose height is at most $\Phi_{n-1}(\alpha_1 + \alpha_2)$, where $\Phi_{-1} = Id$.*

Proof. We apply induction on the number of the logical operations in cut formula D and induction on the height of the deduction tree T_2. The base steps are accomplished in Lemmas 16 and 17.

Assume that $n > 0$. Now, we examine the last derivation rule in T_2, i.e., the rule that was used to obtain the sequent $D, \Delta \to C$. If this rule operates on any formula except cut formula D, for example:

$$\frac{\overset{\vdots\, T_2'}{\vdash^{\gamma < \alpha_2} D, A \& B, A, \Delta' \to\, 0=1}}{\vdash^{\alpha_2}\, D, A \& B, \Delta' \to\, 0=1}\,\&L_2$$

where $\Delta = (A \& B, \Delta')$ and $C = (0 = 1)$, then, by the induction hypothesis, we obtain $\vdash^{\Phi_{n-1}(\alpha_1+\gamma)} \Gamma, A \& B, A, \Delta' \to 0 = 1$. When we apply the last rule from T_2 to this sequent, the rule of $\&L_2$ in this case, we have

$$\vdash^{\Phi_{n-1}(\alpha_1+\gamma)+1} \Gamma, A \& B, \Delta' \to 0 = 1$$

Properties (V1) and (V2) of Veblen hierarchy give us

$$\Phi_{n-1}(\alpha_1 + \gamma) + 1 < \Phi_{n-1}(\alpha_1 + \alpha_2)$$

The other cases are similar. When C is of the form $\forall x F(x)$ and the last derivation rule in T_2 is the rule of $\forall R$, use (V7).

Assume now that the last derivation rule in T_2 operates on formula D. This rule is either the rule of L_1 or L_2. If it is L_1, we have:

$$\frac{\overset{\vdots\, T_2'}{\vdash^{\gamma_1 < \alpha_2} A, \Delta \to 0=1}}{\vdash^{\alpha_2} A \& B, \Delta \to 0=1}\,\&L_1 \qquad\qquad \frac{\overset{\vdots\, T_2'}{\vdash^{\delta_2 < \alpha_2} \Delta \to A}}{\vdash^{\alpha_2} \neg A, \Delta \to 0=1}\,\neg L_1$$

where the formula D is of the form $A \& B$ or $\neg A$, respectively. The case when D is of the form $\forall x F(x)$ is treated the same way. Since Definition 3 gives us that deduction trees T_1 are of the form

$$\frac{\overset{\vdots}{\vdash^{\delta_1 < \alpha_1} \Gamma \to A} \quad \overset{\vdots}{\Gamma \to B}}{\vdash^{\alpha_1} \Gamma \to A \& B}\,\&R \qquad\qquad \frac{\overset{\vdots}{\vdash^{\gamma_2 < \alpha_1} A, \Gamma \to 0=1}}{\vdash^{\alpha_1} \Gamma \to \neg A}\,\neg R$$

respectively, the induction hypothesis for simpler cut formulas gives us

$$\vdash^{\Phi_{n-2}(\delta_i+\gamma_i)} \Gamma, \Delta \to 0 = 1$$

for both $i = 1$ or $i = 2$. Property (V8) yields that $\delta_i, \gamma_i < \Phi_{n-1}(\alpha_1 + \alpha_2)$. Further, property (V2) gives us $\delta_i + \gamma_i < \Phi_{n-1}(\alpha_1 + \alpha_2)$. When we now apply (V4), we have our desired result:

$$\Phi_{n-2}(\delta_i + \gamma_i) < \Phi_{n-1}(\alpha_1 + \alpha_2)$$

If the last derivation rule in T_2 is the rule of L_2 and it operates on cut formula D, we have the following:

$$
\frac{\vdash^{\gamma < \alpha_2} A\&B, A, \Delta \to 0 = 1}{\vdash^{\alpha_2} A\&B, \Delta \to 0 = 1} \&L_2
\qquad
\frac{\vdash^{\gamma < \alpha_2} \neg A, \Delta \to A}{\vdash^{\alpha_2} \neg A, \Delta \to 0 = 1} \neg L_2
$$

with $\vdots\, T_2'$ above each.

Since $\Gamma \to A\&B$ and $A\&B, A, \Delta \to 0 = 1$ have deduction trees T_1 and T_2', respectively, and the height of T_2' is less than α_2, we can use the induction hypothesis and we obtain

$$\vdash^{\Phi_{n-1}(\alpha_1 + \gamma)} \Gamma, A, \Delta \to 0 = 1$$

Similarly for $\neg L_2$:

$$\vdash^{\Phi_{n-1}(\alpha_1 + \gamma)} \Gamma, \Delta \to A$$

Now, we have to remove the formula A. For $D = A\&B$, we have $\vdash^{\Phi_{n-1}(\alpha_1 + \gamma)} \Gamma, A, \Delta \to 0 = 1$ and $\vdash^{\varrho < \alpha_1} \Gamma \to A$ as a subtree of T_1. Hence, we apply the induction hypothesis for simpler cut formulas and we obtain

$$\vdash^{\Phi_{n-2}(\varrho + \Phi_{n-1}(\alpha_1 + \gamma))} \Gamma, \Gamma, \Delta \to 0 = 1$$

For $D = \neg A$, we have $\vdash^{\Phi_{n-1}(\alpha_1 + \gamma)} \Gamma, \Delta \to A$ and $\vdash^{\varrho < \alpha_1} A, \Gamma \to 0 = 1$ as a subtree of T_1. Thus, the induction hypothesis gives us

$$\vdash^{\Phi_{n-2}(\Phi_{n-1}(\alpha_1 + \gamma) + \varrho)} \Gamma, \Gamma, \Delta \to 0 = 1$$

The last step is to show that both numbers $\Phi_{n-2}(\varrho + \Phi_{n-1}(\alpha_1 + \gamma))$ and $\Phi_{n-2}(\Phi_{n-1}(\alpha_1 + \gamma) + \varrho)$ can be bounded by $\Phi_{n-1}(\alpha_1 + \alpha_2)$. When we apply (V8), we obtain $\varrho < \Phi_{n-1}(\alpha_1 + \alpha_2)$. Property (V1) gives us $\Phi_{n-1}(\alpha_1 + \gamma) < \Phi_{n-1}(\alpha_1 + \alpha_2)$. These two observations and property (V2) give us

$$
\begin{aligned}
\varrho + \Phi_{n-1}(\alpha_1 + \gamma) &< \Phi_{n-1}(\alpha_1 + \alpha_2) \\
\Phi_{n-1}(\alpha_1 + \gamma) + \varrho &< \Phi_{n-1}(\alpha_1 + \alpha_2)
\end{aligned}
$$

Next, we use (V1) again and we have

$$
\begin{aligned}
\Phi_{n-2}(\varrho + \Phi_{n-1}(\alpha_1 + \gamma)) &< \Phi_{n-2}(\Phi_{n-1}(\alpha_1 + \alpha_2)) \\
\Phi_{n-2}(\Phi_{n-1}(\alpha_1 + \gamma) + \varrho) &< \Phi_{n-2}(\Phi_{n-1}(\alpha_1 + \alpha_2))
\end{aligned}
$$

According to (V3), value $\Phi_{n-1}(\alpha_1 + \alpha_2)$ is a fixed point of Φ_{n-2}. Hence there is the value $\Phi_{n-1}(\alpha_1 + \alpha_2)$ itself on the right hand side of the inequalities. Finally, we use Lemma 2. This gives us our desired result. $\qquad\square$

1.4 An upper bound on the height of deduction trees for sequents that are derivable in PA

We have analysed Gentzen's reduction rules that are viewed as a calculus with the ω-rule. Now, we move on to PA. Unlike in the previous sections, here we will use free variables and, therefore, we add two new rules to the calculus with the ω-rule:

$$(1) \quad \frac{\Gamma(x/0) \to C(x/0) \quad \Gamma(x/\bar{1}) \to C(x/\bar{1}) \quad \Gamma(x/\bar{2}) \to C(x/\bar{2}) \ \ldots}{\Gamma \to C} \ var$$

Variable x is a free variable somewhere in Γ or C and the rule replaces all occurrences of x in $\Gamma \to C$ by every possible numeral. We do not want to apply this rule consecutively, so, let us assume that we are able to replace all free variables $x_1, \ldots, x_n$ in $\Gamma \to C$ by all possible n-tuples of numerals as a result of a single application. This preserves countable branching. The second new rule is of the form:

$$(2) \quad \frac{\Gamma(t_1,..,t_k/\bar{n}_1,..,\bar{n}_k) \to C(t_1,..,t_k/\bar{n}_1,..,\bar{n}_k)}{\Gamma \to C} \ term$$

Symbols $t_1, \ldots, t_k$ are terms without free variables so that their values can be calculated and the rule replaces them by their corresponding values. The use of this rule will be implicit.

We allow these rules to occur in a deduction tree only immediately before the endsequent, hence, they do not harm our previous result: We take a sequent whose deduction tree we want to find, for example $D(x, y) \to C(x, y)$ where x, y are free variables, and we replace all free variables in this sequent by all possible n-tuples of numerals. Then, we are interested in deduction trees for all instances of the sequent and when we find them, we say that the original sequent has a deduction tree too:

$$\frac{\vdash^\alpha D(0,0) \to C(0,0) \quad \vdash^\beta D(1,0) \to C(1,0) \quad \vdash^\gamma D(0,1) \to C(0,1) \quad \vdash^\delta D(0,2) \to C(0,2) \quad \vdash^\mu D(1,1) \to C(1,1) \quad \vdash^\nu D(2,0) \to C(2,0)...}{D(x,y) \to C(x,y)} \ var$$

The height of the deduction tree for $D(x, y) \to C(x, y)$ is $\sup\{\alpha + 1, \beta + 1, \gamma + 1, \delta + 1, \mu + 1, \nu + 1, \ldots\}$.

1.4.1 Peano arithmetic

We will now introduce the calculus for PA – natural deduction in sequent calculus style. Our sequents are defined as above (Definition 1). The calculus includes two kinds of initial sequents: logical and mathematical. Logical initial sequents are of the form $D \to D$, where D is an arbitrary formula in $\mathcal{L}$ (Section 1.2). Mathematical initial sequents are of the form $\to C$, where C is an equality axiom or a Robinson arithmetic axiom. Just to remind the reader, we provide the list of these axioms. The scheme of induction is implemented as a rule.

Equality axioms:

- $\forall x(x = x)$

- $\forall x \forall y(x = y \supset y = x)$

- $\forall x \forall y \forall z(x = y \ \& \ y = z \supset x = z)$

- $\forall x_1..\forall x_n \forall y_1..\forall y_n(x_1 = y_1 \ \& \ .. \ \& \ x_n = y_n \supset F(x_1,..,x_n) = F(y_1,..,y_n))$
  where F is a n-ary functional symbol in $\mathcal{L}$.

- We do not need the general axiom for n-ary relational symbols because the only relational symbol that we have is equation.

Robinson arithmetic axioms:

- $\forall x \forall y(\mathrm{S}(x) = \mathrm{S}(y) \supset x = y)$

- $\forall x \neg(\mathrm{S}(x) = 0)$

- $\forall x(\neg x = 0 \supset \exists y(\mathrm{S}(y) = x))$

- $\forall x(x + 0 = x)$

- $\forall x \forall y(x + \mathrm{S}(y) = \mathrm{S}(x + y))$

- $\forall x(x \cdot 0 = 0)$

- $\forall x \forall y(x \cdot \mathrm{S}(y) = x \cdot y + x)$

The calculus includes tree kinds of inference rules: structural, logical and an induction rule.

Structural rules:

$$\frac{\Gamma, A, A \to B}{\Gamma, A \to B} \ Ct \qquad \frac{\Gamma \to B}{\Gamma, A \to B} \ Wk$$

Gentzen allows to rename bound variables at any time and counts it as an additional structural rule. He also uses the rule of exchange, but we do not need it, since we work with multisets.

Logical rules:

$$\forall I \; \frac{\Gamma \to F(a)}{\Gamma \to \forall x F(x)} \qquad\qquad \frac{\Gamma \to \forall x F(x)}{\Gamma \to F(t)} \; \forall E$$

$$\& I \; \frac{\Gamma \to A_1 \qquad \Sigma \to A_2}{\Gamma, \Sigma \to A_1 \& A_2} \qquad \frac{\Gamma \to A_1 \& A_2}{\Gamma \to A_i} \; \& E$$

$$\neg I \; \frac{A, \Gamma \to B \qquad A, \Sigma \to \neg B}{\Gamma, \Sigma \to \neg A} \qquad \frac{\Gamma \to \neg\neg A}{\Gamma \to A} \; \neg E$$

Induction rule:

$$\frac{\Gamma \to F(0) \qquad F(a), \Sigma \to F(S(a))}{\Gamma, \Sigma \to F(t)} \; Ind$$

The variable a is an eigenvariable and must, therefore, not occur in Γ, $\forall x F(x)$ and Γ, Σ, $F(0)$, $F(t)$, respectively. The symbol t stands for a term in $\mathcal{L}$ that can be substituted for x in $F(x)$. In the rule of $\& E$ we have $i = 1$ or $i = 2$. We do not need rules for all logical operations as we work in classical predicate logic. There is no problem to replace the mathematical initial sequents with formulas that are equivalent to them and that use only $\forall, \&, \neg$.

Definition 16. *A derivation in PA is a finite tree that consists of sequents. Each sequent is either a logical or a mathematical initial sequent or is derived from previous ones using one of the inference rules that are described in this section.*

1.4.2 Deduction trees for initial sequents of PA

We will show that every initial sequent of PA has a deduction tree. Gentzen assumes this for the mathematical initial sequents and there is no problem indeed. Hence, we concentrate on the logical initial sequents. Gentzen's argument is as follows.

Consider a sequent $D \to D$ in which we have replaced all free variables by numerals according to *var* and *term*. The following scheme explains how to construct a deduction tree for $D \to D$:

$$\uparrow \quad \text{IH, } |C| < |D|$$

$$\text{endform} \quad 0 = 1 \rightarrow 0 = 1 \qquad \frac{C \rightarrow C}{\neg C, C \rightarrow 0 = 1} \ \neg L_1$$

$$\forall L_1, \& L_1 \ \underline{\qquad\qquad} \uparrow \qquad\qquad \uparrow \underline{\qquad\qquad} \ \forall L_1, \& L_1$$

$$D \rightarrow 0 = 1 \qquad \frac{D, C \rightarrow 0 = 1}{D \rightarrow \neg C} \ \neg R$$

$$\text{endform} \quad D \rightarrow 0 = 0 \qquad \nwarrow \quad \uparrow \quad \nearrow$$

$$\underline{\qquad\qquad\qquad\qquad\qquad} \ \forall R, \& R$$

$$D \rightarrow D$$

$$\uparrow \quad var, term$$

We start by decomposing the succedent formula D with the help of $\forall R, \& R$ until it turns into an atomic sentence or a negated sentence. Branches where D has turned into a true atomic sentence are in endform. If D has turned into a negated formula, we apply $\neg R$ to obtain $0 = 1$ in the succedent. Then, we apply $\forall L_1, \& L_1$ to the antecedent formula D in all branches which are not in endform yet. We have to follow the choices that were made along every particular branch while the succedent formula D was reduced. This ensures that the antecedent formula D acquires the same form as the succedent formula D has acquired after steps $\forall R, \& R$ had been applied. Either we reach endform, or we obtain a simpler logical initial sequent $C \rightarrow C$ to which we can apply the induction hypothesis that tells us that $C \rightarrow C$ already has a deduction tree. Note that no L_2 rules are used. For more detailed discussion of deduction trees for initial sequents see ([5], §13) or [7].

We provide now an example of a deduction tree for the initial sequent of the form

$$\neg(\forall x E_1(x) \& E_2) \rightarrow \neg(\forall x E_1(x) \& E_2)$$

where E_2 is an atomic sentence and E_1 is also atomic and does not contain free variables other than x. The application of the rules marked by $(\star)$ is unnecessary when the succedent formula is a true atomic sentence. The deduction tree is the following:

$$\cfrac{\cfrac{\forall L_1(\star) \cfrac{E_1(0) \rightarrow E_1(0)}{\forall x E_1(x) \rightarrow E_1(0)}}{\& L_1(\star) \cfrac{}{\forall x E_1(x)\& E_2 \rightarrow E_1(0)}} \quad \cfrac{\cfrac{E_1(\bar{1}) \rightarrow E_1(\bar{1})}{\forall x E_1(x) \rightarrow E_1(\bar{1})} \forall L_1(\star)}{\forall x E_1(x)\& E_2 \rightarrow E_1(\bar{1})} \& L_1(\star) \quad \cdots \qquad \cfrac{E_2 \rightarrow E_2}{\forall x E_1(x)\& E_2 \rightarrow E_2} \& L_1(\star)}{\cfrac{\cfrac{\forall x E_1(x)\& E_2 \rightarrow \forall x E_1(x)}{\forall x E_1(x)\& E_2 \rightarrow \forall x E_1(x)\& E_2} \& R}{\cfrac{\neg(\forall x E_1(x)\& E_2), \ \forall x E_1(x)\& E_2 \rightarrow 0=1}{\neg(\forall x E_1(x)\& E_2) \rightarrow \neg(\forall x E_1(x)\& E_2)} \neg R} \neg L_1}$$
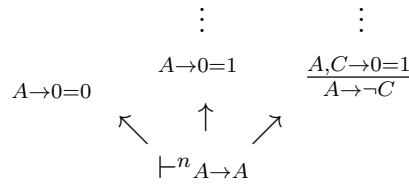
After we apply the rule of $\neg R$ to the initial sequent, we reach the state denoted by $\neg C, C \to 0 = 1$ in the general scheme where $D = \neg C$. The next step is to apply the rule of $\neg L_1$. This creates a simpler initial sequent for which we build a deduction tree according to the same algorithm. So, we apply $\& R$, which is the only rule that is allowed, and we obtain finite branching. The right hand branch may be in endform depending on E_2. If it is not in endform, we have to apply the L_1-variant of $\& R$ and choose E_2 because this is the branch where the rule of $\& R$ has placed E_2 into the succedent.

The left hand branch has $\forall x E_1(x)$ in the succedent. We apply the rule of $\forall R$ and obtain countable branching. Sequents that are not in endform are transformed further. There were the rules of $\& R$ and $\forall R$ applied to our initial sequent $\forall x E_1(x) \& E_2 \to \forall x E_1(x) \& E_2$, hence, $\& L_1$ and $\forall L_1$ must follow. Since $\& R$ has placed $\forall x E_1(x)$ into the succedent, the rule of $\& L_1$ chooses the same formula from the conjunction $\forall x E_1(x) \& E_2$. Every branch created by $\forall R$ is characterized by the numeral that the rule of $\forall R$ has substituted for x in $E_1(x)$. The rule of $\forall L_1$ chooses now the same numeral in every particular branch. Thus, all branches are in endform by now.

The most important property of deduction trees for the logical as well as the mathematical initial sequents follows from Lemma 4: We can always build deduction trees for them that make no use of L_2-rules and, therefore, they have a finite height.

Lemma 18. *Sequent $\neg\neg A \to A$ has a deduction tree with a finite height.*

Proof. We can assume that A contains no free variables. Furthermore, we know that sequent $A \to A$ has a deduction tree with a finite height at most $n = rk(A \to A)$ that can be schematically represented as follows:

$$
\begin{array}{ccc}
 & \vdots & \vdots \\
 & A \to 0=1 & \dfrac{A, C \to 0=1}{A \to \neg C} \\
A \to 0=0 & & \\
 & \nwarrow \quad \uparrow \quad \nearrow & \\
 & \vdash^n A \to A &
\end{array}
$$

The deduction tree for $\neg\neg A \to A$ is similar to the one above except for the length of the branches which may be at most two vertices longer:

$$
\begin{array}{cc}
& \vdots \\
\vdots & \dfrac{A,C\to0=1}{\dfrac{C\to\neg A}{\dfrac{\neg\neg A,C\to0=1}{\neg\neg A\to\neg C}}}
\end{array}
$$

$$
\dfrac{A\to0=1}{\dfrac{\to\neg A}{\neg\neg A\to0=1}}
$$

$$
\neg\neg A\to0=0 \qquad\qquad \nwarrow \quad \uparrow \quad \nearrow
$$

$$
\neg\neg A\to A
$$

Anyway, no L_2-rules are used and Lemma 4 gives us that the deduction tree for $\neg\neg A \to A$ has a finite height bounded by $n + 2 = rk(\neg\neg A \to A)$. $\qquad\square$

1.4.3 Deduction trees for derivable sequents of PA

We will now construct a deduction tree for an arbitrary sequent that is derivable in PA and estimate the height of such a tree.

Theorem 2. *Following Gentzen's procedure, we can construct for every sequent that is derivable in PA a deduction tree whose height is less than $\Phi_\omega(0)$.*

Proof. Assume that we have a derivation ϑ in PA. We proceed by induction on the complexity of ϑ. If ϑ is just one initial sequent, then it has a deduction tree with a finite height (even if there are free variables in the sequent and we have to use the rule of *var*).

Conclusions of the rules of weakening and contraction have deduction trees with heights bounded by $\Phi_\omega(0)$ when the premises have one. This is handled in Lemmas 1 and 2, respectively.

The most important case is when we need to construct a deduction tree for the conclusion $\Gamma, \Sigma \to F(t)$ of the induction rule. Assume that the premises $\Gamma \to F(0)$ and $F(a), \Sigma \to F(\mathrm{S}(a))$ of the induction rule have deduction trees with heights $\alpha, \beta < \Phi_\omega(0)$, respectively, and $\alpha \le \beta$. Formulas and multisets with superscripts denote themselves after the replacement of free variables:

$$
\dfrac{\overset{\vdots}{\Gamma'\to F(0)'} \quad \overset{\vdots}{\Gamma''\to F(0)''} \quad \overset{\vdots}{\Gamma'''\to F(0)'''} \quad \ldots}{\vdash^\alpha \ \Gamma\to F(0)}\ var
$$

$$
\dfrac{\overset{\vdots}{F(0)',\Sigma'\to F(\bar1)'} \quad \overset{\vdots}{F(\bar1)',\Sigma'\to F(\bar2)'} \quad \ldots \quad \overset{\vdots}{F(0)'',\Sigma''\to F(\bar1)''} \quad \overset{\vdots}{F(\bar1)'',\Sigma''\to F(\bar2)''} \quad \ldots}{\vdash^\beta \ F(a),\Sigma\to F(\mathrm{S}(a))}\ var
$$

29

The first step in finding the deduction tree for $\Gamma, \Sigma \to F(t)$ must be var which replaces all free variables by all possible n-tuples of numerals. Term t may contain free variables, too:

$$\frac{\vdash^{\alpha} \Gamma', \Sigma' \to F(0)' \quad \Gamma', \Sigma' \to F(\bar{1})' \quad \Gamma', \Sigma' \to F(\bar{2})' \quad \Gamma', \Sigma' \to F(\bar{3})' \quad ...}{\Gamma, \Sigma \to F(t)} \; var$$

Sequents $\Gamma', \Sigma' \to F(\bar{n})'$ and $\Gamma'', \Sigma'' \to F(\bar{n})''$ with different n-tuple of numerals instead of free variables but with the same value of term t will have deduction trees whose heights will be bounded by the same number. Now, we need to find deduction trees for sequents of the form $\Gamma', \Sigma' \to F(\bar{m})'$ where m is a natural number. We apply an easy induction on m to show the following:

$$\vdash^{\Phi_{|F(0)|}(\beta + m)} \Gamma', \Sigma' \to F(\bar{m})'$$

When this is accomplished, property (V7) gives us the following estimate of the height of the deduction tree with the endsequent $\Gamma, \Sigma \to F(t)$:

$$\sup\{\Phi_{|F(0)|}(\beta), \Phi_{|F(0)|}(\beta + 1), \Phi_{|F(0)|}(\beta + 2), \ldots\} = \Phi_{|F(0)|}(\beta + \omega)$$

If we had $\beta < \alpha$, then we would obtain $\Phi_{|F(0)|}(\alpha + \omega)$. Values α, β are strictly less than $\Phi_{\omega}(0)$ by assumption and ω is less than $\Phi_{\omega}(0)$ by (V6). Hence, the sums $\alpha + \omega$ and $\beta + \omega$ are, according to (V2), also less than $\Phi_{\omega}(0)$. Finally, (V4) gives us that both values $\Phi_{|F(0)|}(\alpha + \omega)$ and $\Phi_{|F(0)|}(\beta + \omega)$ are strictly less than $\Phi_{\omega}(0)$.

Let us return to the embedded induction on m. Assume that $m = 0$. The deduction tree $\vdash^{\alpha} \Gamma', \Sigma' \to F(0)'$ exists according to the induction hypothesis of the superior induction on the complexity of ϑ. Moreover, assumption $\alpha \leq \beta$ and property (V8) claim $\alpha \leq \Phi_{|F(0)|}(\beta)$. Assume now that $m > 0$. The induction hypothesis gives us the deduction tree

$$\vdash^{\Phi_{|F(0)|}(\beta + m - 1)} \Gamma', \Sigma' \to F(\overline{m-1})'$$

and since we also have

$$\vdash^{\beta} F(\overline{m-1})', \Sigma' \to F(\bar{m})'$$

Theorem 1 gives us

$$\vdash^{\Phi_{|F(0)|-1}(\Phi_{|F(0)|}(\beta + m - 1) + \beta)} \Gamma', \Sigma', \Sigma' \to F(\bar{m})'$$

Properties (V1) and (V8) imply that both values β and $\Phi_{|F(0)|}(\beta + m - 1)$ are strictly less than $\Phi_{|F(0)|}(\beta + m)$. Further, this observation and property (V2) give us

$$\Phi_{|F(0)|}(\beta + m - 1) + \beta < \Phi_{|F(0)|}(\beta + m)$$

An application of (V1) to the inequality above yields

$$\Phi_{|F(0)|-1}(\Phi_{|F(0)|}(\beta + m - 1) + \beta) < \Phi_{|F(0)|-1}(\Phi_{|F(0)|}(\beta + m))$$

Since the property (V3) tells us that $\Phi_{|F(0)|}(\beta+m)$ is a fixed point of $\Phi_{|F(0)|-1}$, we obtain

$$\vdash^{\Phi_{|F(0)|}(\beta+m)} \Gamma', \Sigma', \Sigma' \to F(\bar{m})'$$

The deduction tree with the endsequent $\Gamma', \Sigma' \to F(\bar{m})'$ and with the same height is obtained by Lemma 2.

Hilfssatz is also applied to obtain deduction trees for conclusions of the rules of negation. The deduction tree for the conclusion $\Gamma \to A$ of the rule of $\neg E$ is built by combining the deduction tree for $\neg\neg A \to A$, whose height is finite according to Lemma 18, with the deduction tree for the premise $\Gamma \to \neg\neg A$ of $\neg E$ that exists by the induction hypothesis and whose height is $\alpha < \Phi_\omega(0)$.

Finally, it is straightforward to find deduction trees with heights strictly less than $\Phi_\omega(0)$ for the conclusions of the rules of $\forall I, \forall E, \&I, \&E, \neg I$ when their premises have deduction trees with heights strictly less than $\Phi_\omega(0)$. $\square$

1.5 Formalization of Gentzen's proof

We want to show in this section that Gentzen's consistency proof of 1935 can be formulated in the way that the only tool that exceeds PA and that is used in the proof is the transfinite induction on the height of deduction trees applied to Δ_3 formulas. Roughly speaking, Gentzen himself applies transfinite induction on the height of deduction trees in [4]. Moreover, he uses the notion of well-foundedness in induction formulas, namely, the notion of deduction tree appears in his induction formulas and we know that a deduction tree must be well-founded by definition. This is not very convenient for a consistency proof of PA since well-foundedness is a second order property. Here, we will see that we can do without the general well-foundedness of deduction trees.

1.5.1 Notation system for ordinals represented by Veblen functions

We continue to follow Schütte ([11], pp. 83–92) and introduce a notation system that represents ordinals less than a particular ordinal denoted by Γ_0.

First, we define *strongly critical* ordinals the least of which is Γ_0. Second, we define *maximal α-critical* ordinals whose ordering function called Ψ_α is needed to establish the notation system. The properties of strongly critical ordinals and ordering functions Ψ_α make clear why our notation system will work only for ordinals that are strictly less than Γ_0.

We saw at the end of Section 1.2.2 that there exist ordinals α such that $\alpha < \Phi_\alpha(0)$. For example, ordinal ω is one of them. A natural question is whether there are ordinals α too such that $\alpha > \Phi_\alpha(0)$ or $\alpha = \Phi_\alpha(0)$.

The assumption $\alpha > \Phi_\alpha(0)$ leads to a contradiction: If there were some ordinals with this property, then we would take β, the least of them. Hence, $\forall \xi < \beta \colon \xi \leq \Phi_\xi(0)$ and further

$$\xi \leq \Phi_\xi(0) < \Phi_\xi(\Phi_\beta(0)) = \Phi_\beta(0)$$

This yields $\forall \xi < \beta \colon \xi < \Phi_\beta(0)$. Since $\Phi_\beta(0) < \beta$ by assumption, we have $\Phi_\beta(0) < \Phi_\beta(0)$. Thus, there are no ordinals $\alpha > \Phi_\alpha(0)$.

Lemma 19. *We always have $\alpha \leq \Phi_\alpha(0)$.*

Nevertheless, ordinals α such that $\alpha = \Phi_\alpha(0)$ do exist and they will be called strongly critical.

Definition 17. *We say that α is* strongly critical *if $\alpha \in Cr(\alpha)$.*

Lemma 20. *Ordinal α is strongly critical if and only if $\alpha = \Phi_\alpha(0)$.*

Proof. If $\alpha = \Phi_\alpha(0)$, then α belongs to $\mathrm{Cr}(\alpha)$, the range of Φ_α. If $\alpha \in \mathrm{Cr}(\alpha)$, then there is β such that $\alpha = \Phi_\alpha(\beta)$. Now, we know that $\alpha \leq \Phi_\alpha(0)$. Since Φ_α is strictly monotone, $\Phi_\alpha(0)$ is the least element of $\mathrm{Cr}(\alpha)$ and we obtain $\beta = 0$. $\square$

Strongly critical ordinals exist because one can prove that the set of them is closed and unbounded ([11], p. 83–84).

Definition 18. *The least strongly critical ordinal is Γ_0.*

Definition 19. *We say that ordinal γ is* maximal α-critical *when the following conditions are met:*

1. $\gamma \in Cr(\alpha)$

2. $\forall \xi > \alpha : \gamma \notin Cr(\xi)$

A maximal α-critical ordinal γ belongs to $\mathrm{Cr}\,(\alpha)$, hence, we have $\Phi_\nu(\gamma) = \gamma$ for all $\nu < \alpha$. Further, there must exist β such that $\Phi_\alpha(\beta) = \gamma$ and since γ does not belong to the range of any Φ_ξ where $\xi > \alpha$, it follows that $\beta \neq \gamma$. Moreover, $\beta < \gamma$ because Φ_α is strictly monotone. Let us show that for every additive principal number γ there exists α such that γ is maximal α-critical:

Lemma 21. *For every additive principal number γ there exist unique α and unique $\beta < \gamma$ such that $\gamma = \Phi_\alpha(\beta)$.*

Proof. The least additive principal number is 1, hence, $\gamma \leq \Phi_\gamma(0) < \Phi_\gamma(\gamma)$. The inequality $\gamma < \Phi_\gamma(\gamma)$ tells us that there must exist the least α such that $\gamma \neq \Phi_\alpha(\gamma)$. If $\alpha > 0$, the choice of α gives us $\gamma = \Phi_\xi(\gamma)$ for $\xi < \alpha$ and thus, $\gamma \in \mathrm{Cr}\,(\alpha)$. If $\alpha = 0$, $\gamma \in \mathrm{Cr}\,(\alpha)$ since γ is an additive principal number. Anyway, $\gamma \in \mathrm{Cr}\,(\alpha)$. Therefore, there is β such that $\gamma = \Phi_\alpha(\beta)$. We obtain:

$$\gamma \leq \Phi_\alpha(\gamma) \,\&\, \gamma \neq \Phi_\alpha(\gamma) \;\supset\; \gamma < \Phi_\alpha(\gamma) \;\supset\; \Phi_\alpha(\beta) < \Phi_\alpha(\gamma) \;\supset\; \beta < \gamma$$

The uniqueness of α and β follows from Lemma 14. $\qquad\square$

We know that additive principal numbers can be expressed as values of Veblen functions. Unfortunately, this representation is not unambiguous. On the other hand, Lemma 21 tells us that for every additive principal number there is a unique stage where it leaves the hierarchy in the sense that it will not belong to the range of any further Veblen function. In other words, the stage α at which an additive principal number γ becomes a maximal α-critical ordinal is given unambiguously. It is convenient to have an ordering function Ψ_α of maximal α-critical ordinals:

Definition 20. *The ordering function Ψ_α of maximal α-critical ordinals is defined as follows:*

1. *If there is β_0 and $n \in \mathrm{N}$ such that $\Phi_\alpha(\beta_0) = \beta_0$ and $\beta = \beta_0 + n$, we set $\Psi_\alpha(\beta) := \Phi_\alpha(\beta + 1)$.*

2. *Otherwise, we set $\Psi_\alpha(\beta) := \Phi_\alpha(\beta)$.*

The intuition behind this definition is that if we encounter a fixed point of Φ_α, function Ψ_α skips it and continues enumerating values that follow.

Now, let us prove that Ψ_α is really the ordering function of the maximal α-critical ordinals. According to Definition 9, the domain of Ψ_α must be an $\mathbb{O}$-segment. The function must be strictly monotone and the range must cover all maximal α-critical ordinals. There is no problem with the domain since we see that Ψ_α is defined for any $\beta \in \mathbb{O}$.

Lemma 22. *We have $\Psi_\alpha(\beta_1) < \Psi_\alpha(\beta_2)$ for every $\beta_1 < \beta_2$.*

Proof. Assume that $\beta_1 < \beta_2$. We only show that the possibility

$$
\begin{aligned}
\beta_1 + 1 &= \beta_2 \\
\Psi_\alpha(\beta_1) &= \Phi_\alpha(\beta_1 + 1) \\
\Psi_\alpha(\beta_2) &= \Phi_\alpha(\beta_2)
\end{aligned}
$$

cannot occur. Other possibilities give us trivially the required result because Φ_α is strictly monotone. When $\Psi_\alpha(\beta_1) = \Phi_\alpha(\beta_1 + 1)$, there must be β_0 and $n \in \mathrm{N}$ such that $\Phi_\alpha(\beta_0) = \beta_0$ and $\beta_1 = \beta_0 + n$. Then $\beta_2 = \beta_1 + 1 = \beta_0 + n + 1$ and, hence, we must have $\Psi_\alpha(\beta_2) = \Phi_\alpha(\beta_2 + 1)$. $\qquad\square$

Lemma 23. *The value of $\Psi_\alpha(\beta)$ is a maximal α-critical ordinal.*

Proof. By definition, it is clear that $\Psi_\alpha(\beta) \in \mathrm{Cr}(\alpha)$. Assume that we have $\Psi_\alpha(\beta) = \Phi_\alpha(\beta+1)$. Then $\beta \leq \Phi_\alpha(\beta) < \Phi_\alpha(\beta+1)$ and $1 \leq \Phi_\alpha(\beta) < \Phi_\alpha(\beta+1)$. Since $\Phi_\alpha(\beta+1)$ is an additive principal number, it holds $\beta+1 < \Phi_\alpha(\beta+1) = \Psi_\alpha(\beta)$.

Assume now that we have $\Psi_\alpha(\beta) = \Phi_\alpha(\beta)$. In general $\beta \leq \Phi_\alpha(\beta)$, but in this situation it must be $\beta < \Phi_\alpha(\beta)$ because, otherwise, the other case of the definition of $\Psi_\alpha(\beta)$ would apply.

Hence, we obtained that $\Psi_\alpha(\beta)$ is not a fixed point of Φ_α and, therefore, $\Psi_\alpha(\beta) \notin \mathrm{Cr}(\xi)$ where $\xi > \alpha$. $\qquad\square$

Lemma 24. *If γ is maximal α-critical, then there exists β such that $\gamma = \Psi_\alpha(\beta)$.*

Proof. Assume that $\gamma \in \mathrm{Cr}(\alpha)$ and $\gamma \notin \mathrm{Cr}(\xi)$ where $\xi > \alpha$, thus, there exists $\beta_1 < \gamma$ such that $\gamma = \Phi_\alpha(\beta_1)$. Let us investigate the value of $\Psi_\alpha(\beta_1)$. If $\Psi_\alpha(\beta_1) = \Phi_\alpha(\beta_1) = \gamma$, then the assertion is proved by taking $\beta := \beta_1$.

Assume now that $\Psi_\alpha(\beta_1) = \Phi_\alpha(\beta_1 + 1)$. This means that we have β_0 and $n \in \mathrm{N}$ such that $\beta_0 = \Phi_\alpha(\beta_0)$ and $\beta_1 = \beta_0 + n$. The assumption $\beta_1 < \gamma = \Phi_\alpha(\beta_1)$ yields that $n \neq 0$ and we can consider $n - 1$. If we set $\beta := \beta_0 + (n - 1)$, we obtain $\Psi_\alpha(\beta) = \Phi_\alpha(\beta + 1) = \Phi_\alpha(\beta_1) = \gamma$. $\qquad\square$

We have shown that Ψ_α, as introduced in Definition 20, satisfies the properties of the ordering function for maximal α-critical ordinals. If we combine the results of Lemmas 21 and 24, we obtain the following important observation:

Lemma 25. *For every additive principal number γ there exists unique α and unique β such that $\gamma = \Psi_\alpha(\beta)$.*

Proof. Lemma 21 gives us α such that γ is maximal α-critical and this α is given unambiguously. Further, Lemma 24 gives us β such that $\gamma = \Psi_\alpha(\beta)$. Since Ψ_α is strictly monotone, β is unique too. $\qquad\square$

Now, we would appreciate if we can express additive principal numbers γ with the help of $\Psi_\alpha(\beta)$ where α, β are strictly less than γ itself. This will succeed only for additive principal ordinals that are not strongly critical:

Lemma 26. *(1) Assume that $\gamma = \Psi_\alpha(\beta)$. Then, $\alpha < \gamma$ if and only if γ is not a strongly critical ordinal.*
(2) We have $\beta < \Psi_\alpha(\beta)$ for all $\beta \in \mathbb{O}$.

Proof. (1) First, assume that $\gamma \in \mathrm{Cr}\,(\gamma)$. Therefore $\gamma = \Phi_\gamma(0)$. We further have

$$\Psi_\alpha(\beta) = \gamma = \Phi_\gamma(0) = \Psi_\gamma(0)$$

because there is no fixed point β_0 of Φ_γ such that $0 = \beta_0 + n$ where $n \in \mathrm{N}$. Ordinal γ is additive principal, thus, Lemma 25 claims that α and γ denote the same ordinal: $\alpha = \gamma$.

Second, assume that $\gamma \notin \mathrm{Cr}\,(\gamma)$. It follows that $\gamma < \Phi_\gamma(0)$. Hence, we have

$$\Phi_\alpha(\beta^\star) = \Psi_\alpha(\beta) = \gamma < \Phi_\gamma(0)$$

where $\beta^\star = \beta$ or $\beta^\star = \beta + 1$. According to Lemma 15, there are three options that imply $\Phi_\alpha(\beta^\star) < \Phi_\gamma(0)$, but two of them require $\beta^\star < 0$ or $\Phi_\alpha(\beta^\star) < 0$. Hence, the only possible option is $\alpha < \gamma$ and $\beta^\star < \Phi_\gamma(0)$.

(2) Either we have $\beta < \beta + 1 \leq \Phi_\alpha(\beta + 1) = \Psi_\alpha(\beta)$ or we have $\beta \leq \Phi_\alpha(\beta) = \Psi_\alpha(\beta)$. The first case is clear. The second case actually rules out that $\beta = \Phi_\alpha(\beta)$ since otherwise there would be a fixed point of Φ_α from which we can reach β within a finite number of steps. Consequently, the other branch of the definition of $\Psi_\alpha(\beta)$ would apply. $\qquad\square$

Note that functions Ψ_α do not have any fixed points. This means that they are either not continuous or they are not strictly monotone. Lemma 22 yields that they are not continuous.

Before we move on to the definition of the notation system, let us explain how numbers represented by functions Ψ_α can be compared.

Lemma 27. *We have $\Psi_{\alpha_1}(\beta_1) < \Psi_{\alpha_2}(\beta_2)$ if and only if one of the following conditions is met:*

- $\alpha_1 < \alpha_2$ *and* $\beta_1 < \Psi_{\alpha_2}(\beta_2)$

- $\alpha_1 = \alpha_2$ *and* $\beta_1 < \beta_2$

- $\alpha_1 > \alpha_2$ *and* $\Psi_{\alpha_1}(\beta_1) \leq \beta_2$

Proof. Assume that $\alpha_1 < \alpha_2$ and $\beta_1 < \Psi_{\alpha_2}(\beta_2)$. Hence, $\beta_1 + 1 < \Psi_{\alpha_2}(\beta_2)$ because $\Psi_{\alpha_2}(\beta_2)$ is a limit ordinal. Altogether, we have

$$\Psi_{\alpha_1}(\beta_1) \leq \Phi_{\alpha_1}(\beta_1 + 1) < \Phi_{\alpha_1}(\Psi_{\alpha_2}(\beta_2)) = \Psi_{\alpha_2}(\beta_2)$$

where the first inequality follows from the definition of Ψ_{α_1} and the last equality holds since $\Psi_{\alpha_2}(\beta_2) \in \operatorname{Cr}(\alpha_2)$ and $\alpha_1 < \alpha_2$.

The second possibility is clear, so, we can proceed to the third one. Assume that $\Psi_{\alpha_1}(\beta_1) \leq \beta_2$. Lemma 26 claims $\beta_2 < \Psi_{\alpha_2}(\beta_2)$. The required result is obtained by transitivity.

The other direction is proved by assuming that none of the three possibilities is met. Then, the analysis of all available cases leads to $\Psi_{\alpha_2}(\beta_2) \leq \Psi_{\alpha_1}(\beta_1)$. $\qquad\square$

Let us now summarize the main knowledge that the current stage of this exposition gives us. We know that every ordinal $\gamma \neq 0$ has the Cantor normal form (Lemma 8):

$$\gamma = \omega^{\delta_1} + \cdots + \omega^{\delta_n} \qquad n \in \mathrm{N}; \ \delta_1 \geq \cdots \geq \delta_n$$

that is the sum of additive principal numbers ω^{δ_i}. For every additive principal number ω^{δ_i}, there exist unique α_i and β_i such that $\Psi_{\alpha_i}(\beta_i) = \omega^{\delta_i}$ (Lemma 25). If we take $\gamma < \Gamma_0$, which means that γ is not strongly critical, then all summands ω^{δ_i}, where $i \leq n$, are below Γ_0 too and Lemma 26 guarantees that $\alpha_i, \beta_i < \omega^{\delta_i} \leq \gamma$. Hence, $\gamma \neq 0$ has the normal form:

$$\gamma = \Psi_{\alpha_1}(\beta_1) + \cdots + \Psi_{\alpha_n}(\beta_n) \qquad n \in \mathrm{N}; \ \Psi_{\alpha_1}(\beta_1) \geq \cdots \geq \Psi_{\alpha_n}(\beta_n)$$

with parameters $\alpha_i, \beta_i < \gamma$ and these parameters have normal forms with yet again smaller parameters and so on. As this process must terminate, ordinal numbers $< \Gamma_0$ can be coded by natural numbers. (We assume that we are able to code and decode finite sequences primitive recursively.) Further, we see that any $\gamma \in \mathbb{O}$ can be expressed with the help of 0, $+$ and Ψ.

We will now define *ordinal terms* that correspond to ordinal numbers below Γ_0. They will be composed of "simpler" ordinal terms and since $\Gamma_0 = \Phi_{\Gamma_0}(0) = \Psi_{\Gamma_0}(0)$, it is clear the we cannot reach Γ_0 in this way. Besides the ordinal terms themselves, we also define their length and $<$-relation which allows us to compare them. Further, we will need the addition of ordinal terms and we introduce term $\Phi\alpha\beta$ which, as expected, stands for the ordinal denoted by $\Phi_\alpha(\beta)$.

Definition 21. *We define the set* OT *of ordinal terms as follows:*

- *The symbol 0 is an ordinal term of length 0.*

- *If α, β are ordinal terms of lengths $|\alpha|, |\beta|$, then (α, β) is an ordinal term of length*
$$2 \cdot max\left\{\,|\alpha|, |\beta|\,\right\} + 1$$

- *If $\alpha_1, \ldots, \alpha_n, \ \beta_1, \ldots, \beta_n$ where $n \geq 2$ are ordinal terms of lengths $|\alpha_1|, \ldots, |\alpha_n|, |\beta_1|, \ldots, |\beta_n|$ respectively and $(\alpha_1, \beta_1) \geq \cdots \geq (\alpha_n, \beta_n)$, then $(\alpha_1, \beta_1) \ldots (\alpha_n, \beta_n)$ is an ordinal term of length*
$$2 \cdot max\left\{\,|\alpha_1|, \ldots, |\alpha_n|, |\beta_1|, \ldots, |\beta_n|\,\right\} + n$$

The $<$-relation on OT *is defined as follows:*

- *If $\alpha \in$ OT, then $\alpha < 0$ is never valid.*

- *If $\beta \in$ OT, then $0 < \beta$ if and only if $\beta \neq 0$.*

- *If (α_1, β_1) and $(\alpha_2, \beta_2) \in$ OT, then $(\alpha_1, \beta_1) < (\alpha_2, \beta_2)$ if and only if one of the following conditions is met:*

    - *$\alpha_1 < \alpha_2$ and $\beta_1 < (\alpha_2, \beta_2)$*
    - *$\alpha_1 = \alpha_2$ and $\beta_1 < \beta_2$*
    - *$\alpha_1 > \alpha_2$ and $(\alpha_1, \beta_1) \leq \beta_2$*

- *If $\gamma = (\mu_1, \nu_1) \ldots (\mu_m, \nu_m)$ and $\delta = (\sigma_1, \tau_1) \ldots (\sigma_k, \tau_k) \in$ OT where $m, k \geq 1$ and $m + k > 2$, then $\gamma < \delta$ if and only if one of the following conditions is met:*

    - *$m < k$ and $(\mu_i, \nu_i) = (\sigma_i, \tau_i)$ for all $1 \leq i \leq m$*
    - *There is $j \leq min\{m, k\}$ such that $(\mu_j, \nu_j) < (\sigma_j, \tau_j)$ and $(\mu_i, \nu_i) = (\sigma_i, \tau_i)$ for all $1 \leq i < j$.*

The equality $\alpha = \beta$ between terms α and β means that we deal with identical terms.

Even though we cannot use the intuitive interpretation of ordinal terms while proving statements about them, it is useful to keep the interpretation in mind. The term 0 stands for the ordinal 0. If terms α_t and β_t stand for ordinals α and β, then the term (α_t, β_t) stands for the ordinal denoted by $\Psi_\alpha(\beta)$. Since we know that $\Psi_\alpha(\beta)$ is an additive principal number, we

will call terms of the form (α_t, β_t) *principal terms*. If terms $\alpha_{t_1}, .., \alpha_{t_n}$ and $\beta_{t_1}, .., \beta_{t_n}$ stand for ordinals $\alpha_1, .., \alpha_n$ and $\beta_1, .., \beta_n$ respectively, then the term $(\alpha_{t_1}, \beta_{t_1}) \ldots (\alpha_{t_n}, \beta_{t_n})$, which is a term only if $(\alpha_{t_i}, \beta_{t_i}) \geq (\alpha_{t_{i+1}}, \beta_{t_{i+1}})$, stands for the ordinal denoted by $\Psi_{\alpha_1}(\beta_1) + \cdots + \Psi_{\alpha_n}(\beta_n)$.

All ordinal terms are composed only of brackets and zeros. For example, ordinal terms that denote natural numbers are of the form $(0, 0) \ldots (0, 0)$. The supremum of them is ω denoted by $(0, (0, 0))$.

The $<$-relation on OT is obviously defined according to the analogous relation on $\mathbb{O}$. Nevertheless, the ordinal terms must be understood as formal sequences of symbols without the reference to the interpretation and their properties must be proved using only the inductive definition above.

Lemma 28. *There is a primitive recursive function that tells us whether $\alpha = \beta$, $\alpha < \beta$ or $\alpha > \beta$ for all $\alpha, \beta \in$ OT.*

Proof. The construction of the function proceeds by induction on the sum $|\alpha| + |\beta|$ where $\alpha, \beta \in$ OT and it pursues the inductive definition of the $<$-relation introduced in Definition 21. $\qquad\square$

It can be further shown in $\mathrm{I}\Sigma_1$ by induction on the lengths of ordinal terms that they are linearly ordered by the $<$-relation ([11], p. 88).

Lemma 29. *We have (i) $\beta < (\alpha, \beta)$ and (ii) $\alpha < (\alpha, \beta)$.*

Proof. (i) We proceed by induction on $|\beta|$. If $\beta = 0$, then the assertion holds. Assume that $\beta = (\beta_1, \beta_2)$. We have:

$$\beta_2 < (\alpha, \beta_2) < (\alpha, (\beta_1, \beta_2)) = (\alpha, \beta)$$

The first inequality is valid according to the induction hypothesis. The second inequality also used the induction hypothesis and, moreover, the definition of $<$-relation. Now, we consider three cases. (1) If $\beta_1 < \alpha$, we obtain $\beta = (\beta_1, \beta_2) < (\alpha, \beta)$ since we also have $\beta_2 < (\alpha, \beta)$. (2) If $\beta_1 = \alpha$, we have:

$$\beta = (\beta_1, \beta_2) < (\beta_1, (\beta_1, \beta_2)) = (\alpha, \beta)$$

(3) If $\beta_1 > \alpha$, we use the following equivalence:

$$\beta = (\beta_1, \beta_2) < (\alpha, \beta) \quad \equiv \quad \beta_1 > \alpha \ \& \ \beta \leq \beta$$

Clearly, the right hand side of the equivalence is valid.

Finally, assume that $\beta = \beta_1 \ldots \beta_n$ where $\beta_1, .., \beta_n$ are principal terms and $n \geq 2$. We have:

$$\beta_1 < (\alpha, \beta_1) < (\alpha, \beta_1 \ldots \beta_n) = (\alpha, \beta)$$

Since $\beta_1 < (\alpha, \beta)$, we also obtain $\beta = \beta_1 \ldots \beta_n < (\alpha, \beta)$ because terms β and (α, β) differ already in the first principal term.

(ii) We proceed by induction on $|\alpha|$. Assume that $\alpha = (\alpha_1, \alpha_2)$. It suffices to show $\alpha_1 < \alpha$ and $\alpha_2 < (\alpha, \beta)$. The first inequality is handled by the induction hypothesis. To show the second one, we apply (i): $\alpha_2 < (\alpha_1, \alpha_2) = \alpha$ and $\beta < (\alpha, \beta)$. These two inequalities and the definition of $<$-relation yield $(\alpha_2, \beta) < (\alpha, \beta)$. When we use the induction hypothesis again, we obtain $\alpha_2 < (\alpha_2, \beta) < (\alpha, \beta)$ and this yields the required result.

Finally, assume that $\alpha = \alpha_1 \ldots \alpha_n$ where $\alpha_1, .., \alpha_n$ are principal terms and $n \geq 2$. We have:

$$\alpha_1 < (\alpha_1, \beta) < (\alpha_1 \ldots \alpha_n, \beta) = (\alpha, \beta)$$

The first inequality holds according to the induction hypothesis. The second one holds according to the definition of $<$-relation and claim (i). Since α and (α, β) differ in the first principal term, the assertion is established. $\qquad\square$

Definition 22. *The operation $\alpha + \beta$ of ordinal terms $\alpha, \beta \in \mathrm{OT}$ is defined as follows:*

- $\alpha + 0 = 0 + \alpha = \alpha$

- *If $\alpha = (\mu_1, \nu_1) \ldots (\mu_m, \nu_m)$ and $\beta = (\sigma_1, \tau_1) \ldots (\sigma_k, \tau_k)$ where $m, k \geq 1$, then*

$$\alpha + \beta = (\mu_1, \nu_1) \ldots (\mu_j, \nu_j)(\sigma_1, \tau_1) \ldots (\sigma_k, \tau_k)$$

  where j is the largest index less than or equal to m such that $(\mu_j, \nu_j) \geq (\sigma_1, \tau_1)$.

The successor $\alpha + 1$ of ordinal term $\alpha \in \mathrm{OT}$ is defined as follows:

- $\alpha + 1 = \alpha + (0, 0)$

Basic properties of this addition of the ordinal terms are provable in $\mathrm{I}\Sigma_1$ by induction on the length of the ordinal terms ([11], p. 90). Note that the principal terms $\beta = (\beta_1, \beta_2)$ have the property of additive principal numbers. If $\alpha = (\mu_1, \nu_1) \ldots (\mu_m, \nu_m) < (\beta_1, \beta_2)$, then we must have $(\mu_1, \nu_1) < (\beta_1, \beta_2)$ and the definition of addition yields $\alpha + \beta = \beta$.

Definition 23. *We define the term $\Phi\alpha\beta$ where $\alpha, \beta \in \mathrm{OT}$ as follows:*

- *If $\beta = (\beta_1, \beta_2)$ where $\alpha < \beta_1$, then $\Phi\alpha\beta := \beta$.*

- *If $\beta = (\beta_1, \beta_2) + (\gamma + 1)$ where γ denotes a natural number and $\alpha < \beta_1$, then $\Phi\alpha\beta := (\alpha, (\beta_1, \beta_2) + \gamma)$.*

- *We set $\Phi\alpha\beta := (\alpha, \beta)$ in every other case.*

It is clear that $\Phi\alpha\beta$ is a principal term.

Lemma 30. *If $\beta < \gamma$, then $\Phi\alpha\beta < \Phi\alpha\gamma$.*

Proof. (i) Assume that $\Phi\alpha\beta = \beta$. Then, by assumption $\beta < \gamma$, it suffices to show $\gamma \leq \Phi\alpha\gamma$. Let us consider all possible values of $\Phi\alpha\gamma$. If $\Phi\alpha\gamma = \gamma$, we are done. If $\Phi\alpha\gamma$ is equal to (α, γ), Lemma 29 gives us the required result. If $\Phi\alpha\gamma$ is (α, γ_0) where $\gamma = \gamma_0 + 1$, we use Lemma 29 again and we obtain $\gamma = \gamma_0 + 1 \leq (\alpha, \gamma_0) = \Phi\alpha\gamma$.

(ii) Assume now that $\Phi\alpha\beta = (\alpha, \beta_0)$ where $\beta = \beta_0$ or $\beta = \beta_0 + 1$ and $\Phi\alpha\gamma = \gamma = (\gamma_1, \gamma_2)$ where $\alpha < \gamma_1$. The assumption $\beta < \gamma$ gives us $\beta_0 < \gamma$. Altogether, using the definition of the $<$-relation, we obtain:

$$\Phi\alpha\beta = (\alpha, \beta_0) < (\gamma_1, \gamma_2) = \gamma = \Phi\alpha\gamma$$

(iii) Assume that $\Phi\alpha\beta = (\alpha, \beta_0)$ where $\beta = \beta_0$ or $\beta = \beta_0 + 1$ and $\Phi\alpha\gamma = (\alpha, \gamma_0)$ where $\gamma = \gamma_0$ or $\gamma = \gamma_0 + 1$. It suffices to show $\beta_0 < \gamma_0$ to obtain $(\alpha, \beta_0) < (\alpha, \gamma_0)$. The only case when this cannot be trivially seen is $\beta = \beta_0$, $\gamma = \gamma_0 + 1$ and $\beta = \gamma_0$. We show that this case, in fact, cannot occur: By definition, $\Phi\alpha\gamma$ must be $(\alpha, \gamma_0) = (\alpha, (\gamma_1, \gamma_2) + \delta)$ where δ denotes a natural number not equal to 0 since otherwise $\beta = \gamma_0$ would give us $\Phi\alpha\beta = \beta$ and this is not the case. It follows that β must be equal to $\beta_0 + 1$ and we obtain the required result $\beta_0 < \gamma_0$. $\square$

Lemma 31. *We have $\Phi\alpha_1\beta_1 = \Phi\alpha_2\beta_2$ if and only if one of the following conditions is met:*

- $\alpha_1 < \alpha_2$ *and* $\beta_1 = \Phi\alpha_2\beta_2$

- $\alpha_1 = \alpha_2$ *and* $\beta_1 = \beta_2$

- $\alpha_1 > \alpha_2$ *and* $\Phi\alpha_1\beta_1 = \beta_2$

Lemma 32. *We have $\Phi\alpha_1\beta_1 < \Phi\alpha_2\beta_2$ if and only if one of the following conditions is met:*

- $\alpha_1 < \alpha_2$ *and* $\beta_1 < \Phi\alpha_2\beta_2$

- $\alpha_1 = \alpha_2$ *and* $\beta_1 < \beta_2$

- $\alpha_1 > \alpha_2$ *and* $\Phi\alpha_1\beta_1 < \beta_2$

Proof. We prove Lemmas 31 and 32 simultaneously. By definition, $\Phi\alpha_2\beta_2$ is a principal term of the form (γ_1, γ_2) where $\alpha_2 \le \gamma_1$. We know that one of these possibilities must be valid: $\alpha_1 < \alpha_2$, $\alpha_1 = \alpha_2$, $\alpha_1 > \alpha_2$. Assume that $\alpha_1 < \alpha_2$. We want:

$$\Phi\alpha_1\beta_1 = \Phi\alpha_2\beta_2 \quad \equiv \quad \beta_1 = \Phi\alpha_2\beta_2$$

$$\Phi\alpha_1\beta_1 < \Phi\alpha_2\beta_2 \quad \equiv \quad \beta_1 < \Phi\alpha_2\beta_2$$

Since $\Phi\alpha_2\beta_2$ is a principal term (γ_1, γ_2) and transitivity yields $\alpha_1 < \gamma_1$, we obtain $\Phi\alpha_1(\Phi\alpha_2\beta_2) = \Phi\alpha_2\beta_2$ by Definition 23. This observation and Lemma 30 establish the assertions. The other two cases are similar. $\qquad\square$

To finish the demonstration that term $\Phi\alpha\beta$ represents the ordinal denoted by $\Phi_\alpha(\beta)$, we show the following lemma that corresponds to Lemma 21.

Lemma 33. *For every principal term γ there exists unique term α and unique term $\beta < \gamma$ such that $\gamma = \Phi\alpha\beta$.*

Proof. Term γ is a principal term, hence $\gamma = (\alpha, \beta_0)$. Assume first that β_0 is of the form $(\beta_1, \beta_2) + \delta$ where $\alpha < \beta_1$ and δ denotes a natural number. If $\alpha < \beta_1$, then $\beta_1 \ge (0, 0)$ and, thus $\beta_0 > (0, 0)$. We have:

$$\Phi\alpha(\beta_0 + 1) = \Phi\alpha((\beta_1, \beta_2) + (\delta + 1)) = (\alpha, (\beta_1, \beta_2) + \delta) = (\alpha, \beta_0) = \gamma$$

If we show $\beta_0 + 1 < \gamma$, then $\beta_0 + 1$ is our desired term. The definition of $\Phi\alpha\beta_0$ reveals that $\beta_0 \le \Phi\alpha\beta_0 < \Phi\alpha(\beta_0 + 1)$. As $(0, 0) < \beta_0 < \Phi\alpha(\beta_0 + 1)$ and principal terms have the property of the additive principal numbers, we obtain $\beta_0 + (0, 0) = \beta_0 + 1 < \Phi\alpha(\beta_0 + 1)$.

If β_0 has any form except the one above, we have $\Phi\alpha\beta_0 = (\alpha, \beta_0)$ by definition and this equals to γ. It follows from Lemma 29 that $\beta_0 < \gamma$.

The uniqueness of α and β follows from Lemma 31. $\qquad\square$

1.5.2 Formalization of the proof - Preliminaries

We want to show that Gentzen's consistency proof of 1935 can be formalized in III_3 plus transfinite induction up to the supremum of heights of deduction trees that are constructed for sequents that are derivable in PA. The induction formulas of the transfinite induction are at most Δ_3.

Notation: We will use an additional functional symbol $\ulcorner \cdot \urcorner$ that represents a primitive recursive function $\ulcorner \cdot \urcorner : \varphi \to \ulcorner\varphi\urcorner$. This one gives us Gödel numbers for syntactic objects as formulas, sequents, notations for ordinal numbers, deduction trees and similar. We write $\bar\varphi$ instead of $\overline{\ulcorner\varphi\urcorner}$.

The substitution of free variables $x_1, \ldots, x_k$ by numerals $\bar{n}_1, \ldots, \bar{n}_k$ in sequent S or term t is denoted by $S(x_i/\bar{n}_i)$ or $t(x_i/\bar{n}_i)$, respectively. For instance, if S is of the form $D \to D$, then $D \to D(x_i/\bar{n}_i)$ means that we have replaced free variables $x_1, \ldots, x_k$ by $\bar{n}_1, \ldots, \bar{n}_k$ in the antecedent as well as in the succedent occurrence of D. If there is a multiset Δ in the antecedent, then $\Delta \to D(x_i/\bar{n}_i)$ means that we have replaced free variables $x_1, \ldots, x_k$ by $\bar{n}_1, \ldots, \bar{n}_k$ in D and in all formulas contained in Δ.

First, we need some finite representation of deduction trees. If we have a deduction tree T that consists only of one sequent in endform, then the code $\ulcorner T \urcorner$ of T is

$$\langle\, \ulcorner \alpha \urcorner, \ulcorner \Gamma \to 0 = 0 \urcorner, \ulcorner \text{endform} \urcorner, \ulcorner \emptyset \urcorner \,\rangle$$

or

$$\langle\, \ulcorner \alpha \urcorner, \ulcorner \Gamma, 0 = 1 \to 0 = 1 \urcorner, \ulcorner \text{endform} \urcorner, \ulcorner \emptyset \urcorner \,\rangle$$

where $0 = 0$ stands for an arbitrary true atomic sentence, $0 = 1$ stands for an arbitrary false atomic sentence, α is an upper bound on the height of T, sequents $\Gamma \to 0 = 0$ and $\Gamma, 0 = 1 \to 0 = 1$, respectively, denote the endsequent of T and the empty set $\emptyset$ means that no premises were used to derive the endsequent. The angle brackets $\langle\, \ldots \,\rangle$, as expected, code finite sequences.

If we have a deduction tree T with the endsequent $\Gamma \to A \& B$, then the code $\ulcorner T \urcorner$ of T is

$$\langle\, \ulcorner \alpha \urcorner, \ulcorner \Gamma \to A \& B \urcorner, \ulcorner \& R \urcorner, \ulcorner T_1 \urcorner, \ulcorner T_2 \urcorner \,\rangle$$

where α is an upper bound on the height of T, $\& R$ is the derivation rule by which the endsequent $\Gamma \to A \& B$ is derived, variables T_1, T_2 denote deduction trees for the premises of $\& R$. Unary rules are treated similarly.

If the last derivation rule in T is the rule of $\forall R$ or var, let us call them ω-rules, we are not able to enumerate deduction trees for all the premises explicitly. Instead, we adopt a variant of the idea suggested in [12]. Schwichtenberg uses primitive recursive functions in ([12], p. 886) to enumerate codes of deduction trees for premises of the ω-rules. Codes of these primitive recursive functions are then applied to construct a code of the whole deduction tree. We decided to use codes of formulas instead of the codes of functions and since we would not get by with primitive recursive functions here, we will use codes of Σ_1-formulas. Hence, the code of a deduction tree T whose last derivation rule is an ω-rule has the following form:

$$\langle\, \ulcorner \alpha \urcorner, \ulcorner \Gamma \to \forall x F(x) \urcorner, \ulcorner \forall R \urcorner, \ulcorner \varphi_1(x, y) \urcorner \,\rangle$$

or

$$\langle\, \ulcorner \alpha \urcorner, \ulcorner \Delta \to C \urcorner, \ulcorner var \urcorner, \ulcorner \varphi_2(x_1, \ldots, x_k, y) \urcorner \,\rangle$$

where sequent $\Delta \to C$ has k free variables and φ_1, φ_2 are Σ_1-formulas. Roughly speaking, $\varphi_1(x, y)$ represents a function in the way that y is a code of a deduction tree for the premise $\Gamma \to F(\bar{n})$ when $\varphi_1(\bar{n}, y)$ holds. Similarly for $\varphi_2(x_1, \ldots, x_k, y)$: We have a code y of a deduction tree for $\Delta \to C(x_i/\bar{n}_i)$ when $\varphi_2(\bar{n}_1, \ldots, \bar{n}_k, y)$ holds. We will define below a predicate $DT(x)$ that exactly describes which properties x must have to be a code of a proper deduction tree (Definition 24).

There is also a different approach to representation of infinite derivations which was used by Buchholz in his article [2]. He denotes infinite derivations by finite terms. These terms are generated from finite derivations by functional symbols whose implementation performs cut elimination. Finite derivations are viewed as constants, the simplest terms.

With the help of our finite representation, we aim to construct a formula that represents function $\texttt{DedTree}(\ulcorner \vartheta \urcorner)$ whose inputs are codes $\ulcorner \vartheta \urcorner$ of derivations ϑ in PA and output is a code $\ulcorner T \urcorner$ of a deduction tree T for the endsequent of ϑ. Function $\texttt{DedTree}(\ulcorner \vartheta \urcorner)$ knows deduction trees for all mathematical initial sequents. Furthermore, it calls itself recursively and the following auxiliary functions:

1. $\texttt{DedTreeAxiom}(\ulcorner S \urcorner, \ulcorner L \urcorner)$ that accepts codes of a sequent S and of an auxiliary variable L that stores a list of decisions that the function has made during the decomposition of the succedent formula in S. The call

$$\texttt{DedTreeAxiom}(\ulcorner D \to D \urcorner, \ulcorner \emptyset \urcorner)$$

where D is a sentence in $\mathcal{L}$ and $\emptyset$ is an empty list yields a code of a deduction tree for the logical initial sequent $D \to D$.

2. $\texttt{Wk}(\ulcorner T \urcorner, \ulcorner \Delta \urcorner)$ that accepts a code of a deduction tree T and a code of a multiset of formulas Δ and yields a code of tree T where Δ was added to the antecedent of every sequent in T.

3. $\texttt{Ct}(\ulcorner T \urcorner, \ulcorner A \urcorner)$ that accepts a code of a deduction tree T with the endsequent of the form $\Gamma, A, A \to C$ and yields a code of a deduction tree with the endsequent $\Gamma, A \to C$.

4. $\texttt{Elim}(\ulcorner T_1 \urcorner, \ulcorner T_2 \urcorner, \ulcorner D \urcorner)$ that accepts codes of deduction trees T_1 and T_2 whose endsequents are of the form $\Gamma \to D$ and $D, \Delta \to C$, respectively, and yields a code of a deduction tree with the endsequent $\Gamma, \Delta \to C$.

The goal is to find out what assumptions are needed to show that $\texttt{DedTree}$ is total and correct. For this purpose, we will analyse circumstances under

which the auxiliary functions are total and correct. A total function is defined for all possible input values, in our case, for all natural numbers. We are mostly interested in outputs for inputs that code proper derivations of PA or proper deduction trees, depending on the function. If the input is corrupt, we can just tell the function to return 0. A function is correct when it yields a proper deduction tree for a proper input.

Intuitively, we like to think of functions, but, since PA contains no such functions and we do not want to extend the theory, we use, in the end, arithmetic formulas $\mathtt{DedTree}(x, y)$, $\mathtt{DedTreeAxiom}(x, z, y)$, $\mathtt{Wk}(x, z, y)$, $\mathtt{Ct}(x, z, y)$ and $\mathtt{Elim}(x_1, x_2, z, y)$ that represent the functions above. So, we will, in fact, construct these formulas and prove their properties instead of the properties of the functions. We even do not insist on having exactly one y for the given input. The crucial point is that we have some y with the required properties.

Now, we would like to construct an arithmetic formula $DT(x)$ that captures properties of deduction trees and expresses that x is a code of a proper deduction tree. It is natural to define $DT(x)$ in the following way. Rules denoted by P range over derivation rules from Definition 3.

Definition 24.

$$
\begin{aligned}
DT(x) \quad \equiv \quad & x = \langle\ \ulcorner\alpha\urcorner, \ulcorner\Gamma \to 0 = 0\urcorner, \ulcorner\text{endform}\urcorner, \ulcorner\emptyset\urcorner\ \rangle \\
& \qquad\qquad \vee \\
& x = \langle\ \ulcorner\alpha\urcorner, \ulcorner\Gamma, 0 = 1 \to 0 = 1\urcorner, \ulcorner\text{endform}\urcorner, \ulcorner\emptyset\urcorner\ \rangle \\
& \qquad\qquad \vee \\
& x = \langle\ \ulcorner\alpha\urcorner, \ulcorner\Gamma \to C\urcorner, \ulcorner\text{unary rule } P\urcorner,\ y\ \rangle \text{ and} \\
& DT(y),\ y = \langle\ \ulcorner\beta\urcorner, \ulcorner\Delta \to D\urcorner, \ldots\ \rangle,\ \beta < \alpha \text{ and} \\
& \tfrac{\Delta \to D}{\Gamma \to C} \text{ is an instance of the rule of } P \\
& \qquad\qquad \vee \\
& x = \langle\ \ulcorner\alpha\urcorner, \ulcorner\Gamma \to C\urcorner, \ulcorner\text{binary rule } P\urcorner,\ y,\ z\ \rangle \text{ and} \\
& DT(y),\ DT(z),\ \beta < \alpha,\ \gamma < \alpha, \\
& y = \langle\ \ulcorner\beta\urcorner, \ulcorner\Delta \to D\urcorner, \ldots\ \rangle \text{ and} \\
& z = \langle\ \ulcorner\gamma\urcorner, \ulcorner\Sigma \to E\urcorner, \ldots\ \rangle \text{ and} \\
& \tfrac{\Delta \to D \qquad \Sigma \to E}{\Gamma \to C} \text{ is an instance of the rule of } P \\
& \qquad\qquad \vee \\
& x = \langle\ \ulcorner\alpha\urcorner, \ulcorner\Gamma \to \forall w F(w)\urcorner, \ulcorner\forall R\urcorner, \ulcorner\varphi_1(n, y)\urcorner\ \rangle \text{ and} \\
& \forall n\, \exists y\ \varphi_1(n, y) \text{ and} \\
& \forall n\, \forall y\ [\varphi_1(n, y)\ \supset\ DT(y),\ y = \langle\ \ulcorner\alpha_n\urcorner, \ulcorner\Gamma \to F(\bar{n})\urcorner, \ldots\ \rangle,\ \alpha_n < \alpha] \\
& \qquad\qquad \vee \\
& x = \langle\ \ulcorner\alpha\urcorner, \ulcorner\Gamma \to C(x_1, \ldots, x_k)\urcorner, \ulcorner var\urcorner, \ulcorner\varphi_2(n_1, \ldots, n_k, y)\urcorner\ \rangle \text{ and} \\
& \forall n_1 \ldots \forall n_k\, \exists y\ \varphi_2(n_1, \ldots, n_k, y) \text{ and} \\
& \forall n_1 \ldots \forall n_k\, \forall y\ [\varphi_2(n_1, \ldots, n_k, y)\ \supset\ DT(y),\ \beta < \alpha, \\
& \qquad\qquad\qquad\qquad\qquad\qquad y = \langle\ \ulcorner\beta\urcorner, \ulcorner\Gamma \to C(x_i/\bar{n}_i)\urcorner, \ldots\ \rangle]
\end{aligned}
$$

All variables that appear inside the angle brackets $x = \langle \ldots \rangle$, $y = \langle \ldots \rangle$ or $z = \langle \ldots \rangle$ are existentially quantified and bounded by x, y or z, respectively. Hence, more precisely, we should write:

$$\exists x_1 \exists x_2 \exists x_3 \exists x_4 \leq x \quad [\; x = \langle x_1, x_2, x_3, x_4 \rangle \;\&\; OrdNum(x_1) \;\&$$
$$\&\; Sequent(x_2) \;\&\; Rule(x_3) \;\&\; Sigma_1 Fle(x_4) \ldots \;]$$

At the same time, we see that we implicitly use some predicates that tell us that these bound variables have required properties, i.e., they are codes of ordinal numbers, sequents, derivation rules and Σ_1-formulas, respectively. These predicates are Δ_1 in $I\Sigma_1$ and the same holds for the relation that compares (codes of) ordinals. In the sequel, we will use only the intuitive notation of Definition 24 to increase the readability. Note that the only case where free variables are allowed in the endsequent of x is the case when the last derivation rule in x is var.

We see that formula $DT(x)$ defined in this way refers to itself. To avoid this, we apply the Fixed-point theorem ([6], pp. 158–160 or [16], p. 348):

For every arithmetic formula $\psi(v, x_1, .., x_k)$ there is an arithmetic formula $\varphi(x_1, .., x_k)$ such that $\mathbf{Q} \vdash \varphi(x_1, .., x_k) \equiv \psi(\overline{\varphi(x_1, .., x_k)}, x_1, .., x_k)$.

We believe that formula $DT(x)$ will be Π_2 in the end, hence, we choose the following formula $\psi(v, x)$:

$$\psi(v, x) \;=\; x = \langle\, \ulcorner \alpha \urcorner, \ulcorner \Gamma \to 0 = 0 \urcorner, \ulcorner \text{endform} \urcorner, \ulcorner \emptyset \urcorner \,\rangle$$
$$\vee$$
$$x = \langle\, \ulcorner \alpha \urcorner, \ulcorner \Gamma, 0 = 1 \to 0 = 1 \urcorner, \ulcorner \text{endform} \urcorner, \ulcorner \emptyset \urcorner \,\rangle$$
$$\vee$$
$$x = \langle\, \ulcorner \alpha \urcorner, \ulcorner \Gamma \to C \urcorner, \ulcorner \text{unary rule } P \urcorner, \, y \,\rangle, \; Sat_{\Pi,2}(v, y),$$
$$y = \langle\, \ulcorner \beta \urcorner, \ulcorner \Delta \to D \urcorner, \ldots \,\rangle, \; \beta < \alpha \text{ and}$$
$$\tfrac{\Delta \to D}{\Gamma \to C} \text{ is an instance of the rule of } P$$
$$\vee$$
$$x = \langle\, \ulcorner \alpha \urcorner, \ulcorner \Gamma \to C \urcorner, \ulcorner \text{binary rule } P \urcorner, \, y, \, z \,\rangle, \; Sat_{\Pi,2}(v, y), \; Sat_{\Pi,2}(v, z),$$
$$y = \langle\, \ulcorner \beta \urcorner, \ulcorner \Delta \to D \urcorner, \ldots \,\rangle, \; \beta < \alpha,$$
$$z = \langle\, \ulcorner \gamma \urcorner, \ulcorner \Sigma \to E \urcorner, \ldots \,\rangle, \; \gamma < \alpha \text{ and}$$
$$\tfrac{\Delta \to D \qquad \Sigma \to E}{\Gamma \to C} \text{ is an instance of the rule of } P$$
$$\vee$$
$$x = \langle\, \ulcorner \alpha \urcorner, \ulcorner \Gamma \to \forall w F(w) \urcorner, \ulcorner \forall R \urcorner, \ulcorner \varphi_1(n, y) \urcorner \,\rangle \text{ and}$$
$$\forall n \, \exists y \; \varphi_1(n, y) \text{ and}$$
$$\forall n \, \forall y \, [\varphi_1(n, y) \;\supset\; Sat_{\Pi,2}(v, y), \; y = \langle\, \ulcorner \alpha_n \urcorner, \ulcorner \Gamma \to F(\bar{n}) \urcorner, \ldots \,\rangle, \; \alpha_n < \alpha]$$
$$\vee$$
$$x = \langle\, \ulcorner \alpha \urcorner, \ulcorner \Gamma \to C(x_1, \ldots, x_k) \urcorner, \ulcorner var \urcorner, \ulcorner \varphi_2(n_1, \ldots, n_k, y) \urcorner \,\rangle \text{ and}$$
$$\forall n_1 \ldots \forall n_k \, \exists y \; \varphi_2(n_1, \ldots, n_k, y) \text{ and}$$
$$\forall n_1 \ldots \forall n_k \, \forall y \, [\varphi_2(n_1, \ldots, n_k, y) \;\supset\; Sat_{\Pi,2}(v, y), \; \beta < \alpha,$$
$$y = \langle\, \ulcorner \beta \urcorner, \ulcorner \Gamma \to C(x_i / \bar{n}_i) \urcorner, \ldots \,\rangle]$$

The predicate $Sat_{\Pi,2}(v, y)$ is the partial truth predicate for Π_2-formulas and it is Π_2 itself ([6], pp. 50–59 or [16], pp. 337–343). Such a predicate can be formalized in $I\Sigma_1$. The variable v stands for a code of a Π_2-formula with one free variable x. The variable y is the evaluation of x in the sense that the value of x should be the y-th numeral. If we recall that formulas φ_1, φ_2 are Σ_1 and bounded quantifiers, which we implicitly use, do not affect the complexity of formulas, we see that $\psi(v, x)$ is Π_2.

Let us apply the Fixed-point theorem for $\psi(v, x)$. We obtain that there exists an arithmetic formula $\vartheta(x)$ with the following property:

$$\mathsf{Q} \ \vdash \ \vartheta(x) \equiv \psi(\overline{\vartheta(x)}, x)$$

Since we know that $\vartheta(x)$ is Π_2 (in $I\Sigma_1$), we have

$$I\Sigma_1 \ \vdash \ \vartheta(y) \equiv Sat_{\Pi,2}(\overline{\vartheta(x)}, y)$$

Now, when we observe the formula $\psi(\overline{\vartheta(x)}, x)$, we see that it is the same as $\psi(v, x)$ except for the variables v in $Sat_{\Pi,2}(v, y)$ and $Sat_{\Pi,2}(v, z)$ that are replaced by $\overline{\vartheta(x)}$. Hence, $Sat_{\Pi,2}(\overline{\vartheta(x)}, y)$ and $Sat_{\Pi,2}(\overline{\vartheta(x)}, z)$ appear in $\psi(\overline{\vartheta(x)}, x)$ and these are further equivalent to $\vartheta(y)$ and $\vartheta(z)$ respectively. Since $\psi(\overline{\vartheta(x)}, x)$ itself is equivalent to $\vartheta(x)$, we will denote $\vartheta(x)$ by $DT(x)$ and we obtain:

Lemma 34. *The equivalence from Definition 24 that defines formula $DT(x)$, which says that x is a code of a proper deduction tree, is provable in $I\Sigma_1$. Formula $DT(x)$ is Π_2 in $I\Sigma_1$.*

At a metalevel, formula $DT(x)$ contains the information that deduction trees are well-founded. Since every node is assigned an ordinal number that represents the height of the particular subtree and these numbers decrease towards the leaves, we know that x such that $DT(x)$ is well-founded. Nevertheless, this cannot be proved in PA because the proof theoretic ordinal of PA is ε_0 and our upper bound on heights of deduction trees is $\Phi_\omega(0)$. Numbers below $\Phi_\omega(0)$ can be compared in $I\Sigma_1$, see Lemma 28, but $I\Sigma_1$ does not prove that they do not build infinite decreasing sequences.

The main reason for not requiring that formula $DT(x)$ speaks about well-foundedness explicitly is that well-foundedness is a second order property and a consistency proof of PA is trivial when second order properties are allowed. Formula $DT(x)$ appears in the induction formulas that we use below and we want to use only arithmetic induction formulas in the proof. We do not mind applying transfinite induction on the height of deduction trees, but the induction formulas will always be arithmetic and of a bounded complexity.

Eventually, the assertion that gives us the consistency of PA is that for every sequent S derivable in PA we can construct T such that S is the endsequent of T and $DT(\overline{T})$. The point is that sequent $\rightarrow 0 = 1$ has no deduction tree even if we allowed deduction trees with infinite branches. This reasoning is an important improvement of Gentzen's original proof in [4] in which he implicitly uses transfinite induction on the height of deduction trees with induction formulas that explicitly speak about well-foundedness.

In the sequel, we will need the following definition:

Definition 25. *Assume that A is a formula and n is a natural number. Then,* one-step subformula *of A is defined as follows:*

- *If A is of the form $B\&C$, then B, C are one-step subformulas of A.*

- *If A is of the form $\forall x F(x)$, then $F(\bar{n})$ is an one-step subformula of A.*

- *If A is of the form $\neg C$, then C is an one-step subformula of A.*

The subformula *relation is reflexive and transitive closure of the one-step subformula relation.*

We will use relation $One\_Step\_Subfle(x, y)$ that expresses that y is an one-step subformula of x. Furthermore, we use $Subfle(x, y)$ that expresses that y is a subformula of x. They are both Δ_1 in $I\Sigma_1$.

Formula $\texttt{DedTreeAxiom}(s, l, y)$

Notation: We write as $l = l' * \langle \ulcorner E \urcorner \rangle$ to mean that l is a code of a sequence that is built by concatenation of sequences whose codes are l' and $\langle \ulcorner E \urcorner \rangle$, respectively. The second sequence contains only one member, namely, the code of a formula E.

We must introduce an additional relation $Rank(s, r)$ that accepts a code s of a sequent S and holds true if r is a code of the rank of S (Definition 7). The relation is Δ_1 in $I\Sigma_1$.

The definition of $\texttt{DedTreeAxiom}$ formalizes the construction of deduction trees for the logical initial sequents of PA. The construction is described in Section 1.4.2. We add formulas to the left hand side of the list, whose code is l, during the decomposition of the succedent formula. We pick them out from the right hand side of the list when the formula from the antecedent is decomposed. Formula $\texttt{DedTreeAxiom}(s, l, y)$ has the following definition:

$$\texttt{DedTreeAxiom}(s, l, y) \quad \equiv \quad \phi_1(s, l, y) \ \& \ \phi_2(s, l, y)$$

$\phi_1(s, l, y) \quad \equiv$

if s is $\ulcorner A \to B \urcorner$

    if B is a true atomic sentence, then $y = \langle\ \ulcorner 0 \urcorner, s, \ulcorner \text{endform} \urcorner, \ulcorner \emptyset \urcorner\ \rangle$

    &

    if B is a false atomic sentence

        if A is a false atomic sentence, then $y = \langle\ \ulcorner 0 \urcorner, s, \ulcorner \text{endform} \urcorner, \ulcorner \emptyset \urcorner\ \rangle$

        &

        if A is of the form $E \& F$, then

            if $l = l' * \langle \ulcorner E \urcorner \rangle$, then

$$\exists r, z \leq y\ \exists w \leq s\ (\ Rank(s, r)\ \&\ w = \ulcorner E \to 0 = 1 \urcorner\ \&\ \texttt{DedTreeAxiom}(w, l', z)$$
$$\&\ y = \langle\ r, s, \ulcorner \& L_1 \urcorner, z\ \rangle\ )$$

            &

            if $l = l' * \langle \ulcorner F \urcorner \rangle$, then

$$\exists r, z \leq y\ \exists w \leq s\ (\ Rank(s, r)\ \&\ w = \ulcorner F \to 0 = 1 \urcorner\ \&\ \texttt{DedTreeAxiom}(w, l', z)$$
$$\&\ y = \langle\ r, s, \ulcorner \& L_1 \urcorner, z\ \rangle\ )$$

        &

        if A is of the form $\forall x E(x)$

            if $l = l' * \langle \ulcorner E(\bar{n}) \urcorner \rangle$, then

$$\exists r, z \leq y\ \exists w \leq l * \langle \ulcorner 0 = 1 \urcorner \rangle\ (\ Rank(s, r)\ \&\ w = \ulcorner E(\bar{n}) \to 0 = 1 \urcorner\ \&$$
$$\texttt{DedTreeAxiom}(w, l', z)\ \&$$
$$y = \langle\ r, s, \ulcorner \forall L_1 \urcorner, z\ \rangle\ )$$

    &

    if B is a compound formula

        if B is of the form $C \& D$, then

$$\exists r, z_1, z_2 \leq y\ \exists w_1, w_2 \leq s\ (\ Rank(s, r)\ \&\ w_1 = \ulcorner A \to C \urcorner\ \&\ w_2 = \ulcorner A \to D \urcorner\ \&$$
$$\texttt{DedTreeAxiom}(w_1, \langle \ulcorner C \urcorner \rangle * l, z_1)\ \&$$
$$\texttt{DedTreeAxiom}(w_2, \langle \ulcorner D \urcorner \rangle * l, z_2)\ \&$$
$$y = \langle\ r, s, \ulcorner \& R \urcorner, z_1, z_2\ \rangle\ )$$

        &

        if B is of the form $\forall x F(x)$, then

$$\exists r \leq y\ (\ Rank(s, r)\ \&\ y = \langle\ r, s, \ulcorner \forall R \urcorner,$$
$$\ulcorner \texttt{DedTreeAxiom}(\ulcorner A \to F(\bar{n}) \urcorner, \langle \ulcorner F(\bar{n}) \urcorner \rangle * l, y') \urcorner\ \rangle\ )$$

        &

        if B is of the form $\neg C$, then

$$\exists r, z \leq y\ \exists w \leq s * \langle \ulcorner 0 = 1 \urcorner \rangle\ (\ Rank(s, r)\ \&\ w = \ulcorner A, C \to 0 = 1 \urcorner\ \&$$
$$\texttt{DedTreeAxiom}(w, l, z)\ \&$$
$$y = \langle\ r, s, \ulcorner \neg R \urcorner, z\ \rangle\ )$$

Note that we have used the code of the formula $\texttt{DedTreeAxiom}$ in the case where B is of the form $\forall x F(x)$. The formula has two free variables n and y'.

We will discuss below that `DedTreeAxiom` is Δ_1 in $I\Sigma_1$ and, thus, it satisfies our requirements on the complexity. We require Σ_1-formulas to produce codes of deduction trees for the premises of the ω-rules.

$\phi_2(s, l, y) \quad \equiv$

if s is $\ulcorner A, B \to 0 = 1 \urcorner$

  if l is a code of a *non*empty sequence

    if $|B| < |A|$

      if A is of the form $E \& F$

        if $l = l' * \langle \ulcorner E \urcorner \rangle$, then

          $\exists r, z \le y\, \exists w \le s\, (\ Rank(s, r)\, \&\, w = \ulcorner E, B \to 0 = 1 \urcorner\, \&$

                `DedTreeAxiom`$(w, l', z)\, \&$

                $y = \langle\, r, s, \ulcorner \& L_1 \urcorner, z\, \rangle\,)$

        $\&$

        if $l = l' * \langle \ulcorner F \urcorner \rangle$, then

          $\exists r, z \le y\, \exists w \le s\, (\ Rank(s, r)\, \&\, w = \ulcorner F, B \to 0 = 1 \urcorner\, \&$

                `DedTreeAxiom`$(w, l', z)\, \&$

                $y = \langle\, r, s, \ulcorner \& L_1 \urcorner, z\, \rangle\,)$

      $\&$

      if A is of the form $\forall x E(x)$

        if $l = l' * \langle \ulcorner E(\bar{n}) \urcorner \rangle$, then

          $\exists r, z \le y\, \exists w \le l * \langle \ulcorner B \urcorner, \ulcorner 0 = 1 \urcorner \rangle\, (\ Rank(s, r)\, \&\, w = \ulcorner E(\bar{n}), B \to 0 = 1 \urcorner$

                $\&\, $`DedTreeAxiom`$(w, l', z)\, \&$

                $y = \langle\, r, s, \ulcorner \forall L_1 \urcorner, z\, \rangle\,)$

    $\&$

    if $|A| < |B|$

      the same as above, just change A for B

  $\&$

  if l is a code of an empty sequence

    if $A = \neg B$, then

      $\exists r, z \le y\, \exists w \le s\, (\ Rank(s, r)\, \&\, w = \ulcorner B \to B \urcorner\, \&\, $`DedTreeAxiom`$(w, \ulcorner \emptyset \urcorner, z)\, \&$

        $y = \langle\, r, s, \ulcorner \neg L_1 \urcorner, z\, \rangle\,)$

    $\&$

    if $B = \neg A$, then

      $\exists r, z \le y\, \exists w \le s\, (\ Rank(s, r)\, \&\, w = \ulcorner A \to A \urcorner\, \&\, $`DedTreeAxiom`$(w, \ulcorner \emptyset \urcorner, z)\, \&$

        $y = \langle\, r, s, \ulcorner \neg L_1 \urcorner, z\, \rangle\,)$

Formula `DedTreeAxiom`(s, l, y) can be constructed, exactly the same way as formula $DT(x)$ from Definition 24, with the help of the Fixed-point theorem and the partial truth predicates. It does not matter whether we use

$Sat_{\Sigma,1}$ or $Sat_{\Pi,1}$ during the construction. In either case, we obtain a formula that corresponds to $\mathtt{DedTreeAxiom}(s, l, y)$ as defined above. The difference is that the application of $Sat_{\Sigma,1}$ gives us a Σ_1-formula, whereas the application of $Sat_{\Pi,1}$ gives us a Π_1-formula. The conclusion is that $\mathtt{DedTreeAxiom}(s, l, y)$ is Δ_1 in $\mathrm{I}\Sigma_1$.

It can be shown that $\mathtt{DedTreeAxiom}(s, l, y)$ is total by which we mean:

$$\forall s \, \forall l \, \exists y \, \mathtt{DedTreeAxiom}(s, l, y)$$

This is shown by induction on $rk(S)$ where S is a sequent with code s. Induction formula $\forall l \, \exists y \, \mathtt{DedTreeAxiom}(s, l, y)$ is Π_2. If s or l does not satisfy the antecedent of some implication that is a subformula of $\mathtt{DedTreeAxiom}$, the implication returns "true". Hence, if s or l is not of the form that the antecedents require, $\mathtt{DedTreeAxiom}(s, l, y)$ holds for an arbitrary y. If s and l are both of the required form, we can find y such that $\mathtt{DedTreeAxiom}(s, l, y)$ by the induction hypothesis because the rank of the sequent with code w is less than the rank of the sequent with code s.

Let us now move on to the correctness. If we set

- $Sentence(s) \quad \equiv \quad s$ is a code of a sequent that does not contain free variables

- $Sub(s) \quad \equiv \quad s$ is a code of a sequent of the form $A \to B$ where B is a subformula of A such that B is not within the scope of a negation

- $NegSub(s) \quad \equiv \quad s$ is a code of a sequent of the form $A, B \to 0 = 1$ where $\neg B$ is a subformula of A such that $\neg B$ is not within the scope of a negation

- $Choices(l, s) \quad \equiv \quad \exists n \leq l \, (\quad l = \langle \, i_0, \ldots, i_n \, \rangle \, \&$
$(\, s = \ulcorner A \to B \urcorner \vee s = \ulcorner A, B \to 0 = 1 \urcorner) \, \&$
$\forall j \leq n \, (i_j \text{ is a code of a formula}) \, \&$
$\forall j < n \, (i_{j+1} \text{ is not a negated formula}) \, \&$
$\forall j < n \; One\_Step\_Subfle(\, i_{j+1}, i_j \,) \, \&$
$One\_Step\_Subfle(\ulcorner A \urcorner, i_n \,) \, \&$
$[\, (\, Sub(s) \, \& \, i_0 = \ulcorner B \urcorner) \vee (NegSub(s) \, \& \, i_0 = \ulcorner \neg B \urcorner) \,]$
$)$
$\vee \, (\, l = \ulcorner \emptyset \urcorner \, \& \, (s = \ulcorner B \to B \urcorner \vee s = \ulcorner \neg B, B \to 0 = 1 \urcorner) \,)$

then we can prove

$$\forall s \, \forall l \, [\, Sentence(s) \, \& \, (\, Sub(s) \lor NegSub(s) \,) \, \& \, Choices(l, s)$$

$$\supset$$

$$\forall y \, (\, \texttt{DedTreeAxiom}(s, l, y)$$

$$\supset$$

$$DT(y) \, \& \, \text{endsequent of } y \text{ has code } s \, \&$$

$$\text{height of } y \leq \text{rank of sequent with code } s$$

$$)$$

$$]$$

$$(1)$$

by induction on $rk(S)$ where S is a sequent with code s. Since predicates $Sentence(s)$, $Sub(s)$, $NegSub(s)$ and $Choices(l, s)$ are Δ_1 in IΣ_1, the induction formula is Π_2 in IΣ_1. The antecedent of the formula specifies correct inputs. The input $s_1 = \ulcorner D \to D \urcorner$ and $l_1 = \ulcorner \emptyset \urcorner$, where D is a sentence in $\mathcal{L}$, happens to be correct. Since we already know that $\texttt{DedTreeAxiom}$ is total, it follows that we have at least one y such that it is a proper deduction tree for the logical initial sequent $D \to D$.

The following lemma is the conclusion of the analysis of $\texttt{DedTreeAxiom}$:

Lemma 35. *The totality of* $\texttt{DedTreeAxiom}(s, l, y)$ *can be proved in* IΠ_2. *The correctness of* $\texttt{DedTreeAxiom}(s, l, y)$ *represented by formula* (1) *can be proved in* IΠ_2. *The heights of deduction trees for logical initial sequents are always finite.*

Formulas $\texttt{Wk}(x, z, y)$, $\texttt{Ct}(x, z, y)$ **and** $\texttt{MultiCt}(x, z, y)$

Let us move on to formulas $\texttt{Wk}(x, z, y)$ and $\texttt{Ct}(x, z, y)$ that represent structural modifications of deduction trees; namely, y is a code of a deduction tree that was built by modification of every sequent, node, in x with respect to z. Variable z stands for a code of a multiset Δ or a formula A, respectively, and the modification either adds Δ to the antecedent of every sequent in x or it replaces A, A by A in the antecedent of every sequent in x.

First, we state the definitions of the formulas. Rules denoted by P range over derivation rules from Definition 3.

$\mathtt{Wk}(x,z,y) \quad \equiv$

if $x = \langle\, \ulcorner\alpha\urcorner, \ulcorner\Gamma \to C\urcorner, \ulcorner\mathrm{endform}\urcorner, \ulcorner\emptyset\urcorner\,\rangle$
 if $z = \ulcorner\Delta\urcorner$, then
  $y = \langle\, \ulcorner\alpha\urcorner, \ulcorner\Gamma, \Delta \to C\urcorner, \ulcorner\mathrm{endform}\urcorner, \ulcorner\emptyset\urcorner\,\rangle$
&
if $x = \langle\, \ulcorner\alpha\urcorner, \ulcorner\Gamma \to C\urcorner, \ulcorner\mathrm{unary\ rule}\ P\urcorner, x_1\,\rangle$
 if $z = \ulcorner\Delta\urcorner$, then
  $y = \langle\, \ulcorner\alpha\urcorner, \ulcorner\Gamma, \Delta \to C\urcorner, \ulcorner\mathrm{unary\ rule}\ P\urcorner, y_1\,\rangle\ \&\ \mathtt{Wk}(x_1, z, y_1)$
&
if $x = \langle\, \ulcorner\alpha\urcorner, \ulcorner\Gamma \to C\urcorner, \ulcorner\mathrm{binary\ rule}\ P\urcorner, x_1, x_2\,\rangle$
 if $z = \ulcorner\Delta\urcorner$, then
  $y = \langle\ulcorner\alpha\urcorner, \ulcorner\Gamma, \Delta \to C\urcorner, \ulcorner\mathrm{binary\ rule}\ P\urcorner, y_1, y_2\,\rangle\ \&\ \mathtt{Wk}(x_1, z, y_1)\ \&\ \mathtt{Wk}(x_2, z, y_2)$
&
if $x = \langle\, \ulcorner\alpha\urcorner, \ulcorner\Gamma \to \forall w F(w)\urcorner, \ulcorner\forall R\urcorner, \ulcorner\varphi(n,u)\urcorner\,\rangle$, then
 if $z = \ulcorner\Delta\urcorner$, then
  $y = \langle\, \ulcorner\alpha\urcorner, \ulcorner\Gamma, \Delta \to \forall w F(w)\urcorner, \ulcorner\forall R\urcorner, \ulcorner\exists u\,(\,\varphi(n,u)\,\&\,\mathtt{Wk}(u,z,y'))\urcorner\,\rangle$

  Formula $\mathtt{Wk}$ can be constructed the same way as formula $\mathtt{DedTreeAxiom}$ with the help of the Fixed-point theorem and the partial truth predicates. We are able to construct it as Σ_1 or Π_1, hence, it is Δ_1 in $\mathrm{I}\Sigma_1$. Note that we use the code of $\mathtt{Wk}$ in the case when the last derivation rule in x is $\forall R$. There, we always take the Σ_1-variant of $\mathtt{Wk}$ since formulas that give us codes of deduction trees for premises of the ω-rules have to be Σ_1 by definition.

$\mathtt{Ct}(x,z,y) \quad \equiv$

if $z = \ulcorner B\&C\urcorner$, then
 if $x = \langle\, \ulcorner\alpha\urcorner, \ulcorner\Gamma, B\&C, B\&C \to D\urcorner, \ulcorner\mathrm{endform}\urcorner, \ulcorner\emptyset\urcorner\,\rangle$, then
  $y = \langle\, \ulcorner\alpha\urcorner, \ulcorner\Gamma, B\&C \to D\urcorner, \ulcorner\mathrm{endform}\urcorner, \ulcorner\emptyset\urcorner\,\rangle$
 &
 if $x = \langle\, \ulcorner\alpha\urcorner, \ulcorner\Gamma, B\&C, B\&C \to D\urcorner, \ulcorner\mathrm{unary\ rule}\ P\urcorner, x_1\,\rangle$, then
  if $x_1 = \langle\, \ulcorner\beta\urcorner, \ulcorner\Gamma, B\&C, B \to D\urcorner, \dots\,\rangle$, then
   $y = \langle\, \ulcorner\alpha\urcorner, \ulcorner\Gamma, B\&C \to D\urcorner, \ulcorner\&L_2\urcorner, x_1\,\rangle$
  &
  if $x_1 = \langle\, \ulcorner\beta\urcorner, \ulcorner\Gamma', B\&C, B\&C \to D'\urcorner, \dots\,\rangle$, then
   $y = \langle\, \ulcorner\alpha\urcorner, \ulcorner\Gamma, B\&C \to D\urcorner, \ulcorner\mathrm{unary\ rule}\ P\urcorner, y_1\,\rangle\ \&\ \mathtt{Ct}(x_1, z, y_1)$
 &
 if $x = \langle\, \ulcorner\alpha\urcorner, \ulcorner\Gamma, B\&C, B\&C \to E\&F\urcorner, \ulcorner\&R\urcorner, x_1, x_2\,\rangle$, then
  $y = \langle\ulcorner\alpha\urcorner, \ulcorner\Gamma, B\&C \to E\&F\urcorner, \ulcorner\&R\urcorner, y_1, y_2\,\rangle\ \&\ \mathtt{Ct}(x_1, z, y_1)\ \&\ \mathtt{Ct}(x_2, z, y_2)$
 &
 if $x = \langle\, \ulcorner\alpha\urcorner, \ulcorner\Gamma, B\&C, B\&C \to \forall w F(w)\urcorner, \ulcorner\forall R\urcorner, \ulcorner\varphi(n,u)\urcorner\,\rangle$, then
  $y = \langle\, \ulcorner\alpha\urcorner, \ulcorner\Gamma, B\&C \to \forall w F(w)\urcorner, \ulcorner\forall R\urcorner, \ulcorner\exists u\,(\,\varphi(n,u)\,\&\,\mathtt{Ct}(u,z,y'))\urcorner\,\rangle$

&
if $z = \ulcorner \forall w F(w) \urcorner$, then

    $\ldots$

&
if $z = \ulcorner \neg C \urcorner$, then

    $\ldots$

&
if ($z = \ulcorner A \urcorner$ and A is atom), then

    $\ldots$

The cases for $z = \ulcorner \forall w F(w) \urcorner$, $z = \ulcorner \neg C \urcorner$ and $z = \ulcorner A \urcorner$, where A is an atom, are completely analogous to the case for $z = \ulcorner B \& C \urcorner$. The usual construction gives us that formula $\mathtt{Ct}(x, z, y)$ is Δ_1 in $I\Sigma_1$.

Now we want to argue that both formulas $\mathtt{Wk}$ and $\mathtt{Ct}$ are total. When an arbitrary z is fixed, this can be proved by induction on x, the code of the input deduction tree. The induction formulas are Σ_1:

$$\exists y \; \mathtt{Wk}(x, z, y) \qquad\qquad \exists y \; \mathtt{Ct}(x, z, y)$$

Note the case in the definition of both formulas when the endsequent of x is derived by the rule of $\forall R$. The formula that gives us codes of deduction trees for the premises of $\forall R$ in x is $\varphi(n, u)$. To set up some y such that $\mathtt{Wk}(x, z, y)$ or $\mathtt{Ct}(x, z, y)$ in this case, we need to establish the code of $\exists u \, (\, \varphi(n, u) \, \& \, \mathtt{Wk}(u, z, y') \,)$ and $\exists u \, (\, \varphi(n, u) \, \& \, \mathtt{Ct}(u, z, y') \,)$, respectively. We do not even care about the properties of $\varphi(n, u)$; we just need some number that will be included in y. That is why it is sufficient to use the induction on x to prove the totality of both formulas.

The requirements on $\varphi(n, u)$ change essentially when we want to prove the correctness. The correctness of $\mathtt{Wk}$ is expressed in the following way:

$$
\begin{aligned}
\forall x \, \forall z \, [\; & DT(x) \; \& \; x = \langle \ulcorner \alpha \urcorner, \ulcorner \Gamma \to C \urcorner, \ldots \rangle \; \& \; z = \ulcorner \Delta \urcorner \\
& \supset \\
& \forall y \; (\; \mathtt{Wk}(x, z, y) \\
& \qquad \supset \\
& \qquad DT(y) \; \& \; \text{height of } y \text{ is } \alpha \; \& \\
& \qquad \text{endsequent of } y \text{ is } \Gamma, \Delta \to C \\
& \qquad) \\
&]
\end{aligned}
\qquad (2)
$$

The correctness of Ct is analogous:

$$\forall x \, \forall z \; [\; DT(x) \; \& \; x = \langle \ulcorner \alpha \urcorner, \ulcorner \Gamma, D, D \to C \urcorner, \dots \rangle \; \& \; z = \ulcorner D \urcorner$$
$$\supset$$
$$\forall y \; (\; \mathtt{Ct}(x, z, y)$$
$$\supset$$
$$DT(y) \; \& \; \text{height of } y \text{ is } \alpha \; \&$$
$$\text{endsequent of } y \text{ is } \Gamma, D \to C \tag{3}$$
$$)$$
$$]$$

To prove the correctness of both formulas, we choose an arbitrary z, and then, we continue by transfinite induction on the height of the input tree x. The induction formulas are Δ_3.

Why do we have to use transfinite induction on the heights of deduction trees instead of the induction on their codes? Recall the case when the last derivation rule in x is the rule of $\forall R$:

$$x = \langle \; \ulcorner \alpha \urcorner, \ulcorner \Sigma \to \forall w F(w) \urcorner, \ulcorner \forall R \urcorner, \ulcorner \varphi(n, u) \urcorner \; \rangle$$

During the proof of the correctness, we need to assume that the induction hypothesis holds for u_n such that $\varphi(n, u_n)$. This u_n denotes a subtree of x, thus, it has a smaller height than x. On the other hand, since x contains only the code of $\varphi(n, u)$ but no codes u_n of its subtrees, in general, it can occur that $x < u_n$ for some n. Hence, the induction on the codes of input trees does not work for us in this case.

The following lemma is the conclusion of the analysis of Wk and Ct:

Lemma 36. *Assume that the height of x is strictly less than $\Phi_\omega(0)$. The totality of $\mathtt{Wk}(x, z, y)$ and $\mathtt{Ct}(x, z, y)$ can be proved in $\mathrm{I}\Sigma_1$. The correctness of both formulas is represented by (2) and (3), respectively, and it is proved in $\mathrm{I}\Sigma_1 + \mathrm{TI}$ up to $\Phi_\omega(0)$. The induction formulas of the transfinite induction are Δ_3. Structural modification of deduction trees do not change their heights.*

In the sequel, we will mostly use formula $\mathtt{MultiCt}(x, z, y)$ instead of $\mathtt{Ct}(x, z, y)$. Variables x, z, y in $\mathtt{MultiCt}(x, z, y)$ have almost the same meaning as in $\mathtt{Ct}(x, z, y)$ with the only exception that z denotes a multiset instead of a single formula. Hence, formula $\mathtt{MultiCt}(x, z, y)$ represents contraction of several formulas at once.

$\texttt{MultiCt}(x, z, y) \quad \equiv$

if ($z = \ulcorner\Gamma\urcorner$ & $\Gamma = \{A_1, \ldots, A_k\}$), then

    if $k \geq 1$, then

        if $x = \langle\, \ulcorner\alpha\urcorner, \ulcorner\Gamma, \Gamma, \Delta \to C\urcorner, \ulcorner\text{endform}\urcorner, \ulcorner\emptyset\urcorner\,\rangle$, then

            $y = \langle\, \ulcorner\alpha\urcorner, \ulcorner\Gamma, \Delta \to C\urcorner, \ulcorner\text{endform}\urcorner, \ulcorner\emptyset\urcorner\,\rangle$

        &

        if $x = \langle\, \ulcorner\alpha\urcorner, \ulcorner\Gamma, \Gamma, \Delta \to C\urcorner, \ulcorner\text{unary rule } P\urcorner, x_1\,\rangle$, then

            if P is the rule of L_1 applied to A_i from Γ, then

                $y = \langle\, \ulcorner\alpha\urcorner, \ulcorner\Gamma, \Delta \to C\urcorner, \ulcorner\text{suitable rule of } L_2\urcorner, y_1\,\rangle$ &

                $\texttt{MultiCt}(x_1, \ulcorner\{A_1, .., A_{i-1}, A_{i+1}, .., A_k\}\urcorner, y_1)$

            &

            if P is *not* the rule of L_1 applied to A_i from Γ, then

                $y = \langle\, \ulcorner\alpha\urcorner, \ulcorner\Gamma, \Delta \to C\urcorner, \ulcorner\text{unary rule } P\urcorner, y_1\,\rangle$ & $\texttt{MultiCt}(x_1, z, y_1)$

        &

        if $x = \langle\, \ulcorner\alpha\urcorner, \ulcorner\Gamma, \Gamma, \Delta \to E\&F\urcorner, \ulcorner\&R\urcorner, x_1, x_2\,\rangle$, then

            $y = \langle\, \ulcorner\alpha\urcorner, \ulcorner\Gamma, \Delta \to E\&F\urcorner, \ulcorner\&R\urcorner, y_1, y_2\,\rangle$ &

            $\texttt{MultiCt}(x_1, z, y_1)$ & $\texttt{MultiCt}(x_2, z, y_2)$

        &

        if $x = \langle\, \ulcorner\alpha\urcorner, \ulcorner\Gamma, \Gamma, \Delta \to \forall w F(w)\urcorner, \ulcorner\forall R\urcorner, \ulcorner\varphi(n, u)\urcorner\,\rangle$, then

            $y = \langle\, \ulcorner\alpha\urcorner, \ulcorner\Gamma, \Delta \to \forall w F(w)\urcorner, \ulcorner\forall R\urcorner, \ulcorner\exists u\,(\,\varphi(n, u)\, \& \,\texttt{MultiCt}(u, z, y'))\urcorner\,\rangle$

    &

    if $k = 0$, then

        $y = x$

Formula $\texttt{MultiCt}(x, z, y)$ is Δ_1 in $I\Sigma_1$. The totality of $\texttt{MultiCt}$ is expressed as $\forall x \forall z \exists y\ \texttt{MultiCt}(x, z, y)$. The correctness of $\texttt{MultiCt}$ is the following:

$$
\begin{aligned}
\forall x \,\forall z\ [\ & DT(x)\ \&\ x = \langle \ulcorner\alpha\urcorner, \ulcorner\Gamma, \Gamma, \Delta \to C\urcorner, \ldots \rangle\ \&\ z = \ulcorner\Gamma\urcorner \\
& \supset \\
& \forall y\ (\ \texttt{MultiCt}(x, z, y) \\
& \qquad \supset \\
& \qquad DT(y)\ \&\ \text{height of } y \text{ is } \alpha\ \& \\
& \qquad \text{endsequent of } y \text{ is } \Gamma, \Delta \to C \\
& \quad) \\
&\ ]
\end{aligned} \qquad (4)
$$

The totality of $\texttt{MultiCt}(x, z, y)$ is shown by induction on x, the code of the input tree. The induction formula to prove the totality is Π_2. To prove the correctness of $\texttt{MultiCt}(x, z, y)$, we first prove (5) where z is bounded by x:

$$\forall x\, \forall z \leq x\ [\ DT(x)\ \&\ x = \langle \ulcorner \alpha \urcorner, \ulcorner \Gamma \urcorner, \Gamma, \Delta \to C \urcorner, \dots \rangle\ \&\ z = \ulcorner \Gamma \urcorner$$
$$\supset$$
$$\forall y\ (\ \mathtt{MultiCt}(x, z, y)$$
$$\supset$$
$$DT(y)\ \&\ \text{height of } y \text{ is } \alpha\ \&$$
$$\text{endsequent of } y \text{ is } \Gamma, \Delta \to C$$
$$)$$
$$]$$

$$(5)$$

This is proved by transfinite induction on the height of x. The induction formula of the transfinite induction is Δ_3. We also use the fact that $\mathtt{MultiCt}$ is total. If we take some z that is not bounded by x, the antecedent is trivially false. Hence, altogether, we obtain that (4) holds.

A similar "trick" can be used while proving the totality and the correctness of $\mathtt{DedTreeAxiom}$ as well as the totality of $\mathtt{MultiCt}$. Nevertheless, the correctness of $\mathtt{MultiCt}$ is the only case that would spoilt the overall result if we did not use the bounded quantifier: The transfinite induction applied here would require a more complex induction formula than the proof of the correctness of $\mathtt{Elim}$ requires.

The following lemma is the conclusion of the analysis of $\mathtt{MultiCt}$:

Lemma 37. *Assume that the height of x is strictly less than $\Phi_\omega(0)$. The totality of $\mathtt{MultiCt}(x, z, y)$ can be proved in $\mathrm{I\Pi}_2$. Formula (4) represents the correctness of $\mathtt{MultiCt}(x, z, y)$ and it can be proved in $\mathrm{I\Pi}_2 + \mathrm{TI}$ up to $\Phi_\omega(0)$. The induction formula of the transfinite induction is Δ_3. Structural modification of deduction trees do not change their heights.*

Formula $\mathtt{Elim}(x_1, x_2, z, y)$

We go on to investigate the formula $\mathtt{Elim}(x_1, x_2, z, y)$. Variables x_1 and x_2 are codes of deduction trees whose endsequents $\Gamma \to D$ and $D, \Delta \to C$ are premises of the cut to eliminate. Variable z is a code of the cut formula D and y stands for a code of a deduction tree for $\Gamma, \Delta \to C$ after the cut elimination. First, we define the formula. Then, we take a look at its totality and the correctness. We denote by P a unary derivation rule from Definition 3.

$$\mathtt{Elim}(x_1, x_2, z, y)\ \equiv\ \phi_{\mathrm{atom}}(x_1, x_2, z, y)\ \&\ \phi_{\mathrm{conj}}(x_1, x_2, z, y)\ \&$$
$$\phi_{\mathrm{forall}}(x_1, x_2, z, y)\ \&\ \phi_{\mathrm{neg}}(x_1, x_2, z, y)$$

where

$\phi_{\mathrm{atom}}(x_1, x_2, z, y) \quad \equiv$

if $z = \ulcorner A \urcorner$ & A is an atom

    if $x_1 = \langle \ulcorner \alpha_1 \urcorner, \ulcorner \Gamma \to A \urcorner, \dots \rangle$

        if $x_2 = \langle \ulcorner \alpha_2 \urcorner, \ulcorner A, \Delta \to C \urcorner, \ulcorner \mathrm{endform} \urcorner, \ulcorner \emptyset \urcorner \rangle$

            if $\Gamma, \Delta \to C$ is in endform, then

                $y = \langle \ulcorner 0 \urcorner, \ulcorner \Gamma, \Delta \to C \urcorner, \ulcorner \mathrm{endform} \urcorner, \ulcorner \emptyset \urcorner \rangle$

            &

            if $\Gamma, \Delta \to C$ is *not* in endform, then

                $\mathtt{Wk}(x_1, z_1, y)$ & $z_1 = \ulcorner \Delta \urcorner$

        &

        if $x_2 = \langle \ulcorner \alpha_2 \urcorner, \ulcorner A, \Delta \to C \urcorner, \ulcorner \mathrm{unary\ rule\ } P \urcorner, x_2' \rangle$, then

            $y = \langle \ulcorner \alpha_1 + \alpha_2 \urcorner, \ulcorner \Gamma, \Delta \to C \urcorner, \ulcorner \mathrm{unary\ rule\ } P \urcorner, y' \rangle$ & $\mathtt{Elim}(x_1, x_2', z, y')$

        &

        if $x_2 = \langle \ulcorner \alpha_2 \urcorner, \ulcorner A, \Delta \to C_1 \& C_2 \urcorner, \ulcorner \& R \urcorner, x_2', x_2'' \rangle$, then

            $y = \langle \ulcorner \alpha_1 + \alpha_2 \urcorner, \ulcorner \Gamma, \Delta \to C_1 \& C_2 \urcorner, \ulcorner \& R \urcorner, y', y'' \rangle$ &

            $\mathtt{Elim}(x_1, x_2', z, y')$ & $\mathtt{Elim}(x_1, x_2'', z, y'')$

        &

        if $x_2 = \langle \ulcorner \alpha_2 \urcorner, \ulcorner A, \Delta \to \forall x C(x) \urcorner, \ulcorner \forall R \urcorner, \ulcorner \varphi(n, u) \urcorner \rangle$, then

            $y = \langle \ulcorner \alpha_1 + \alpha_2 \urcorner, \ulcorner \Gamma, \Delta \to \forall x C(x) \urcorner, \ulcorner \forall R \urcorner, \ulcorner \exists u\, (\varphi(n, u)\, \& \, \mathtt{Elim}(x_1, u, z, y'))\urcorner \rangle$

    Since the cut formula is an atom, there are no derivation rules that can modify it. Hence, the elimination of an atomic cut consists in moving the cut upwards along x_2 until a sequent in endform is reached.

$\phi_{\mathrm{conj}}(x_1, x_2, z, y) \quad \equiv$

if $z = \ulcorner E \& F \urcorner$

    if $x_1 = \langle \ulcorner \alpha_1 \urcorner, \ulcorner \Gamma \to E \& F \urcorner, \ulcorner \& R \urcorner, x_1', x_1'' \rangle$

        if $x_2 = \langle \ulcorner \alpha_2 \urcorner, \ulcorner E \& F, \Delta \to C \urcorner, \ulcorner \mathrm{endform} \urcorner, \ulcorner \emptyset \urcorner \rangle$

            $y = \langle \ulcorner 0 \urcorner, \ulcorner \Gamma, \Delta \to C \urcorner, \ulcorner \mathrm{endform} \urcorner, \ulcorner \emptyset \urcorner \rangle$

        &

        if $x_2 = \langle \ulcorner \alpha_2 \urcorner, \ulcorner E \& F, \Delta \to C \urcorner, \ulcorner \mathrm{unary\ rule\ } P \mathrm{\ (not\ applied\ to\ the\ cut\ formula)} \urcorner, x_2' \rangle$

            $y = \langle \ulcorner \Phi_{|E \& F| - 1}(\alpha_1 + \alpha_2) \urcorner, \ulcorner \Gamma, \Delta \to C \urcorner, \ulcorner \mathrm{unary\ rule\ } P \urcorner, y' \rangle$ &

            $\mathtt{Elim}(x_1, x_2', z, y')$

        &

        if $x_2 = \langle \ulcorner \alpha_2 \urcorner, \ulcorner E \& F, \Delta \to C_1 \& C_2 \urcorner, \ulcorner \& R \urcorner, x_2', x_2'' \rangle$

            $y = \langle \ulcorner \Phi_{|E \& F| - 1}(\alpha_1 + \alpha_2) \urcorner, \ulcorner \Gamma, \Delta \to C_1 \& C_2 \urcorner, \ulcorner \& R \urcorner, y', y'' \rangle$ &

              $\mathtt{Elim}(x_1, x_2', z, y')$ & $\mathtt{Elim}(x_1, x_2'', z, y'')$

        &

        if $x_2 = \langle \ulcorner \alpha_2 \urcorner, \ulcorner E \& F, \Delta \to \forall x C(x) \urcorner, \ulcorner \forall R \urcorner, \ulcorner \varphi(n, u) \urcorner \rangle$

            $y = \langle \ulcorner \Phi_{|E \& F| - 1}(\alpha_1 + \alpha_2) \urcorner, \ulcorner \Gamma, \Delta \to \forall x C(x) \urcorner, \ulcorner \forall R \urcorner,$

                                    $\ulcorner \exists u\, (\varphi(n, u)\, \& \, \mathtt{Elim}(x_1, u, z, y'))\urcorner \rangle$

&
if $x_2 = \langle \ulcorner \alpha_2 \urcorner, \ulcorner E \& F, \Delta \to 0 = 1 \urcorner, \ulcorner \& L_1 \urcorner, x_2' \rangle$
    if $x_2' = \langle \ulcorner \gamma \urcorner, \ulcorner E, \Delta \to 0 = 1 \urcorner, \ldots \rangle$
        $\mathtt{Elim}(x_1', x_2', z_1, y) \ \& \ z_1 = \ulcorner E \urcorner$
&
if $x_2 = \langle \ulcorner \alpha_2 \urcorner, \ulcorner E \& F, \Delta \to 0 = 1 \urcorner, \ulcorner \& L_2 \urcorner, x_2' \rangle$
    if $x_2' = \langle \ulcorner \gamma \urcorner, \ulcorner E, E \& F, \Delta \to 0 = 1 \urcorner, \ldots \rangle$
        $\exists y_1, y_2 \, (\, \mathtt{Elim}(x_1, x_2', z, y_1) \ \& \ \mathtt{Elim}(x_1', y_1, z_1, y_2) \ \& \ z_1 = \ulcorner E \urcorner \ \&$
            $\mathtt{MultiCt}(y_2, z_2, y) \ \& \ z_2 = \ulcorner \Gamma \urcorner)$

$\phi_{\text{forall}}(x_1, x_2, z, y) \quad \equiv$

if $z = \ulcorner \forall x F(x) \urcorner$
    if $x_1 = \langle \ulcorner \alpha_1 \urcorner, \ulcorner \Gamma \to \forall x F(x) \urcorner, \ulcorner \forall R \urcorner, \ulcorner \rho(n, u) \urcorner \rangle$
        If the endsequent of x_2 is in endform or the last derivation rule in x_2
        is a unary rule not applied to the cut formula or it is the rule
        of $\& R$ or $\forall R$, then the cases are similar to the analogous cases above.
        &
    if $x_2 = \langle \ulcorner \alpha_2 \urcorner, \ulcorner \forall x F(x), \Delta \to 0 = 1 \urcorner, \ulcorner \forall L_1 \urcorner, x_2' \rangle$
        if $x_2' = \langle \ulcorner \gamma \urcorner, \ulcorner F(\bar{m}), \Delta \to 0 = 1 \urcorner, \ldots \rangle$
            $\exists u \, (\, \rho(m, u) \ \& \ \mathtt{Elim}(u, x_2', z_1, y) \ \& \ z_1 = \ulcorner F(\bar{m}) \urcorner)$
        &
    if $x_2 = \langle \ulcorner \alpha_2 \urcorner, \ulcorner \forall x F(x), \Delta \to 0 = 1 \urcorner, \ulcorner \forall L_2 \urcorner, x_2' \rangle$
        if $x_2' = \langle \ulcorner \gamma \urcorner, \ulcorner F(\bar{m}), \forall x F(x), \Delta \to 0 = 1 \urcorner, \ldots \rangle$
            $\exists y_1, y_2, u \, (\, \mathtt{Elim}(x_1, x_2', z, y_1) \ \& \ \rho(m, u) \ \&$
                $\mathtt{Elim}(u, y_1, z_1, y_2) \ \& \ z_1 = \ulcorner F(\bar{m}) \urcorner \ \&$
                $\mathtt{MultiCt}(y_2, z_2, y) \ \& \ z_2 = \ulcorner \Gamma \urcorner)$

$\phi_{\text{neg}}(x_1, x_2, z, y) \quad \equiv$

if $z = \ulcorner \neg E \urcorner$
    if $x_1 = \langle \ulcorner \alpha_1 \urcorner, \ulcorner \Gamma \to \neg E \urcorner, \ulcorner \neg R \urcorner, x_1' \rangle$
        If the endsequent of x_2 is in endform or the last derivation rule in x_2
        is a unary rule not applied to the cut formula or it is the rule
        of $\& R$ or $\forall R$, then the cases are similar to the analogous cases above.
        &
    if $x_2 = \langle \ulcorner \alpha_2 \urcorner, \ulcorner \neg E, \Delta \to 0 = 1 \urcorner, \ulcorner \neg L_1 \urcorner, x_2' \rangle$
        if $x_2' = \langle \ulcorner \gamma \urcorner, \ulcorner \Delta \to E \urcorner, \ldots \rangle$
            $\mathtt{Elim}(x_2', x_1', z_1, y) \ \& \ z_1 = \ulcorner E \urcorner$
        &

$$\text{if } x_2 = \langle \ulcorner \alpha_2 \urcorner, \ulcorner \neg E, \Delta \to 0 = 1 \urcorner, \ulcorner \neg L_2 \urcorner, x_2' \rangle$$
$$\text{if } x_2' = \langle \ulcorner \gamma \urcorner, \ulcorner \neg E, \Delta \to E \urcorner, \dots \rangle$$
$$\exists y_1, y_2 \, (\, \texttt{Elim}(x_1, x_2', z, y_1) \, \& \, \texttt{Elim}(y_1, x_1', z_1, y_2) \, \& \, z_1 = \ulcorner E \urcorner \, \&$$
$$\texttt{MultiCt}(y_2, z_2, y) \, \& \, z_2 = \ulcorner \Gamma \urcorner)$$

The usual construction with the help of the Fixed-point theorem and the partial truth predicates gives us that formula $\texttt{Elim}(x_1, x_2, z, y)$ is Σ_1 in $\text{I}\Sigma_1$.

Now, we move on to the properties of the formula. It is not possible to prove the totality in the form $\forall x_1 \, \forall x_2 \, \forall z \, \exists y \, \texttt{Elim}(x_1, x_2, z, y)$ because, in the case when the cut formula is $\forall x F(x)$, deduction tree x_1 contains formula $\rho(n, u)$ which may not be total. This causes problems when the last derivation rule in x_2 is $\forall L_1$ or $\forall L_2$ that affects the cut formula $\forall x F(x)$ and transforms it into $F(\bar{m})$ or $F(\bar{m}), \forall x F(x)$, respectively. If there is no u such that $\rho(m, u)$, formula $\texttt{Elim}(x_1, x_2, z, y)$ returns false for an arbitrary y. Hence, we will require y-s such that $\texttt{Elim}(x_1, x_2, z, y)$ only for proper inputs x_1, x_2, z and we will prove the totality and the correctness in the following form:

$$\forall x_1 \forall x_2 \forall z \, (\, DT(x_1) \, \& \, x_1 = \langle \ulcorner \alpha_1 \urcorner, \ulcorner \Gamma \to D \urcorner, \dots \rangle \, \&$$
$$DT(x_2) \, \& \, x_2 = \langle \ulcorner \alpha_2 \urcorner, \ulcorner D, \Delta \to C \urcorner, \dots \rangle \, \& \, z = \ulcorner D \urcorner$$
$$\supset$$
$$\exists y \, \texttt{Elim}(x_1, x_2, z, y) \, \& \, \forall y \, (\, \texttt{Elim}(x_1, x_2, z, y) \hspace{2em} (6)$$
$$\supset$$
$$DT(y) \, \&$$
$$\text{endsequent of } y \text{ is } \Gamma, \Delta \to C \, \&$$
$$\text{height of } y \text{ is } \leq \, \Phi_{|D|-1}(\alpha_1 + \alpha_2)$$
$$)$$
$$)$$

To prove this, we apply three induction arguments of which the second one and the third one are embedded in the first one. We start with induction on the number of the logical operations in the cut formula denoted by z. The induction formula $\theta(z)$ is Π_3:

$$\theta(z) \quad \equiv$$
$$\forall x_1 \, \forall x_2 \, (\, DT(x_1) \, \& \, x_1 = \langle \ulcorner \alpha_1 \urcorner, \ulcorner \Gamma \to D \urcorner, \dots \rangle \, \&$$
$$DT(x_2) \, \& \, x_2 = \langle \ulcorner \alpha_2 \urcorner, \ulcorner D, \Delta \to C \urcorner, \dots \rangle \, \& \, z = \ulcorner D \urcorner$$
$$\supset$$
$$\exists y \, \texttt{Elim}(x_1, x_2, z, y) \, \& \, \forall y \, (\, \texttt{Elim}(x_1, x_2, z, y)$$
$$\supset$$
$$DT(y) \, \&$$
$$\text{endsequent of } y \text{ is } \Gamma, \Delta \to C \, \&$$
$$\text{height of } y \text{ is } \leq \, \Phi_{|D|-1}(\alpha_1 + \alpha_2)$$
$$)$$
$$)$$

Assume that z denotes an atom A. We choose an arbitrary x_1 such that $DT(x_1)$ and $x_1 = \langle \ulcorner \alpha_1 \urcorner, \ulcorner \Gamma \to A \urcorner, \ldots \rangle$. At this point, we apply the transfinite induction on the height of x_2. The induction formula $\xi(x_1, x_2, z)$ is Δ_3:

$$\xi(x_1, x_2, z) \quad \equiv$$

$$DT(x_1) \ \& \ x_1 = \langle \ulcorner \alpha_1 \urcorner, \ulcorner \Gamma \to A \urcorner, \ldots \rangle \ \&$$
$$DT(x_2) \ \& \ x_2 = \langle \ulcorner \alpha_2 \urcorner, \ulcorner A, \Delta \to C \urcorner, \ldots \rangle \ \& \ z = \ulcorner A \urcorner$$
$$\supset$$
$$\exists y \ \mathtt{Elim}(x_1, x_2, z, y) \ \& \ \forall y \ (\ \mathtt{Elim}(x_1, x_2, z, y)$$
$$\supset$$
$$DT(y) \ \&$$
$$\text{endsequent of } y \text{ is } \Gamma, \Delta \to C \ \&$$
$$\text{height of } y \text{ is } \leq \ \alpha_1 + \alpha_2$$
$$)$$

If x_2 is in endform, we consider two cases: (i) The conclusion $\Gamma, \Delta \to C$ of the cut is in endform too or (ii) it is not in endform.

(i) If sequent $\Gamma, \Delta \to C$ is in endform, then there exists y such that $\mathtt{Elim}(x_1, x_2, z, y)$ and every y for which $\mathtt{Elim}(x_1, x_2, z, y)$ holds is of the form $y = \langle \ulcorner 0 \urcorner, \ulcorner \Gamma, \Delta \to C \urcorner, \ulcorner \text{endform} \urcorner, \ulcorner \emptyset \urcorner \rangle$. This is clearly a deduction tree with the required endsequent and the required height.

(ii) Assume that $\Gamma, \Delta \to C$ is not in endform. On the other hand, we know that the endsequent of x_2 is in endform. The only possibility how this can occur is that both formulas A and C are false atomic sentences. Then, the endsequent of x_1 is $\Gamma \to 0 = 1$ and, to obtain a deduction tree for $\Gamma, \Delta \to 0 = 1$, it suffices to apply weakening to x_1: $\mathtt{Wk}(x_1, z_1, y)$ and $z_1 = \ulcorner \Delta \urcorner$. Since $\mathtt{Wk}$ is total, we know that there exists y such that $\mathtt{Wk}(x_1, z_1, y)$ and, for every y of this kind, the correctness of $\mathtt{Wk}$ gives us $DT(y)$ and $y = \langle \ulcorner \alpha_1 \urcorner, \ulcorner \Gamma, \Delta \to 0 = 1 \urcorner, \ldots \rangle$.

Now, we can state the induction hypothesis: For every x_2' such that the height of x_2' is less than the height of x_2 we have $\xi(x_1, x_2', z)$ where x_1 is of the form $\langle \ulcorner \alpha_1 \urcorner, \ulcorner \Gamma \to A \urcorner, \ldots \rangle$, $DT(x_1)$ and z denotes an atom A.

Assume that the last derivation rule in x_2 is a unary rule or the rule of $\&R$. The unary rule cannot be applied to the cut formula since this one is an atom. As an example, we show the case when the last derivation rule in x_2 is the rule of $\neg R$. The other cases are similar.

Assume that x_2 is of the form $\langle\ulcorner\alpha_2\urcorner,\ulcorner A,\Delta\to\neg C\urcorner,\ulcorner\neg R\urcorner,x_2'\rangle$ and $x_2'=$ $\langle\ulcorner\gamma\urcorner,\ulcorner A,C,\Delta\to 0=1\urcorner,\dots\rangle$. Since we have $DT(x_2)$, we also have $DT(x_2')$ and $\gamma<\alpha_2$. Since the height of x_2' is strictly less than the height of x_2, we can apply the induction hypothesis. The induction hypothesis gives us y' such that $\texttt{Elim}(x_1,x_2',z,y')$. Moreover, for every y' of this kind we have $DT(y')$ and $y'=\langle\ulcorner\alpha_1+\gamma\urcorner,\ulcorner\Gamma,C,\Delta\to 0=1\urcorner,\dots\rangle$. The definition of ϕ_{atom} gives us now y such that $\texttt{Elim}(x_1,x_2,z,y)$, namely, it is of the form $y=\langle\ulcorner\alpha_1+\alpha_2\urcorner,\ulcorner\Gamma,\Delta\to\neg C\urcorner,\ulcorner\neg R\urcorner,y'\rangle$. Furthermore, we also have $DT(y)$ because every possible y' that may be contained in y has the properties required by Definition 24.

Assume that the last derivation rule in x_2 is $\forall R$. Then, x_2 is of the form $\langle\ulcorner\alpha_2\urcorner,\ulcorner A,\Delta\to\forall x\,C(x)\urcorner,\ulcorner\forall R\urcorner,\ulcorner\varphi(n,u)\urcorner\rangle$. Obviously, y such that $\texttt{Elim}(x_1,x_2,z,y)$ exists and it has the form

$$\langle\ulcorner\alpha_1+\alpha_2\urcorner,\ulcorner\Gamma,\Delta\to\forall x\,C(x)\urcorner,\ulcorner\forall R\urcorner,\ulcorner\exists u\,(\,\varphi(n,u)\,\&\,\texttt{Elim}(x_1,u,z,y')\,)\urcorner\rangle$$

We want to prove $DT(y)$ for every y of this kind. This requires two things:

(i) $\forall n\,\exists y'\,\exists u\,(\,\varphi(n,u)\,\&\,\texttt{Elim}(x_1,u,z,y')\,)$

   Choose an arbitrary n. We have u such that $\varphi(n,u)$ because $DT(x_2)$ and this also gives us $DT(u)$ and $u=\langle\ulcorner\delta\urcorner,\ulcorner A,\Delta\to C(\bar n)\urcorner,\dots\rangle$ where $\delta<\alpha_2$. Since the height of u is strictly less than the height of x_2, the induction hypothesis yields y' such that $\texttt{Elim}(x_1,u,z,y')$.

(ii) $\forall n\,\forall y'\,(\ \exists u\,(\,\varphi(n,u)\,\&\,\texttt{Elim}(x_1,u,z,y')\,)$
   $$\supset$$
   $DT(y')$ & endsequent of y' is $\Gamma,\Delta\to C(\bar n)$ &
   height of $y'<\alpha_1+\alpha_2$
   $\ )$

   Choose an arbitrary n and y' such that $\exists u\,(\,\varphi(n,u)\,\&\,\texttt{Elim}(x_1,u,z,y')\,)$. As we know $DT(x_2)$, we have $DT(u)$ and $u=\langle\ulcorner\delta\urcorner,\ulcorner A,\Delta\to C(\bar n)\urcorner,\dots\rangle$ where $\delta<\alpha_2$ for any u such that $\varphi(n,u)$. Since u has smaller height than x_2, the induction hypothesis and $\texttt{Elim}(x_1,u,z,y')$ give us $DT(y')$ and $y'=\langle\ulcorner\alpha_1+\delta\urcorner,\ulcorner\Gamma,\Delta\to C(\bar n)\urcorner,\dots\rangle$.

This finishes the first embedded induction whose result is that we have $\theta(z)$ for an atomic cut formula z. Let us state the induction hypothesis of the superior induction: For every cut formula z' that has less logical operations than the cut formula z we have $\theta(z')$. Assume now that z contains at least one

logical operation. To prove $\theta(z)$, we choose an arbitrary x_1 and apply transfinite induction on the height of x_2 again. The induction formula $\chi(x_1, x_2, z)$ is Δ_3:

$$\chi(x_1, x_2, z) \quad \equiv$$

$$DT(x_1) \ \& \ x_1 = \langle \ulcorner \alpha_1 \urcorner, \ulcorner \Gamma \to D \urcorner, \dots \rangle \ \&$$
$$DT(x_2) \ \& \ x_2 = \langle \ulcorner \alpha_2 \urcorner, \ulcorner D, \Delta \to C \urcorner, \dots \rangle \ \& \ z = \ulcorner D \urcorner$$
$$\supset$$
$$\exists y \ \texttt{Elim}(x_1, x_2, z, y) \ \& \ \forall y \ (\ \texttt{Elim}(x_1, x_2, z, y)$$
$$\supset$$
$$DT(y) \ \&$$
$$\text{endsequent of } y \text{ is } \Gamma, \Delta \to C \ \&$$
$$\text{height of } y \text{ is } \leq \ \Phi_{|D|-1}(\alpha_1 + \alpha_2)$$
$$)$$

If the last derivation rule in x_2 does not transform the cut formula D, the argument is the same as for atomic cut formulas. Hence, we will investigate only the six cases when the cut formula is affected.

(i) Assume that the last derivation rule in x_2 is the rule of $\&L_1$. Then, x_1 and x_2 are of the form

$$
\begin{aligned}
x_1 &= \langle \ulcorner \alpha_1 \urcorner, \ulcorner \Gamma \to E \& F \urcorner, \ulcorner \& R \urcorner, x_1^{'}, x_1^{''} \rangle \\
x_1^{'} &= \langle \ulcorner \delta \urcorner, \ulcorner \Gamma \to E \urcorner, \dots \rangle \\
x_2 &= \langle \ulcorner \alpha_2 \urcorner, \ulcorner E \& F, \Delta \to 0 = 1 \urcorner, \ulcorner \& L_1 \urcorner, x_2^{'} \rangle \\
x_2^{'} &= \langle \ulcorner \gamma \urcorner, \ulcorner E, \Delta \to 0 = 1 \urcorner, \dots \rangle \\
z &= \ulcorner E \& F \urcorner
\end{aligned}
$$

Since $DT(x_1)$ and $DT(x_2)$ by assumption, we also have $DT(x_1^{'})$ and $DT(x_2^{'})$. As every y such that $\texttt{Elim}(x_1, x_2, z, y)$ must satisfy $\texttt{Elim}(x_1^{'}, x_2{'}, z_1, y)$ where $z_1 = \ulcorner E \urcorner$, we can apply the induction hypothesis $\theta(z_1)$. We obtain that y-s such that $\texttt{Elim}(x_1^{'}, x_2{'}, z_1, y)$ exist and for each of them we also have $DT(y)$ and $y = \langle \ulcorner \Phi_{|E|-1}(\delta + \gamma) \urcorner, \ulcorner \Gamma, \Delta \to 0 = 1 \urcorner, \dots \rangle$.

(ii) Assume that the last derivation rule in x_2 is the rule of $\&L_2$. Then, x_1 and x_2 are of the form

$$
\begin{aligned}
x_1 &= \langle \ulcorner \alpha_1 \urcorner, \ulcorner \Gamma \to E \& F \urcorner, \ulcorner \& R \urcorner, x_1^{'}, x_1^{''} \rangle \\
x_1^{'} &= \langle \ulcorner \delta \urcorner, \ulcorner \Gamma \to E \urcorner, \dots \rangle \\
x_2 &= \langle \ulcorner \alpha_2 \urcorner, \ulcorner E \& F, \Delta \to 0 = 1 \urcorner, \ulcorner \& L_2 \urcorner, x_2^{'} \rangle \\
x_2^{'} &= \langle \ulcorner \gamma \urcorner, \ulcorner E, E \& F, \Delta \to 0 = 1 \urcorner, \dots \rangle \\
z &= \ulcorner E \& F \urcorner
\end{aligned}
$$

We are looking for y such that $\mathtt{Elim}(x_1, x_2, z, y)$ and for every y of this kind we want $DT(y)$ and $y = \langle \ulcorner \Phi_{|E\&F|-1}(\alpha_1 + \alpha_2) \urcorner, \ulcorner \Gamma, \Delta \to 0 = 1 \urcorner, \ldots \rangle$. To find y such that $\mathtt{Elim}(x_1, x_2, z, y)$, we have to find y_1 and y_2 such that

$$
\left.
\begin{array}{l}
\mathtt{Elim}(x_1, x_2', z, y_1) \quad \& \\
\mathtt{Elim}(x_1', y_1, z_1, y_2) \quad \& \quad z_1 = \ulcorner E \urcorner \quad \& \\
\mathtt{MultiCt}(y_2, z_2, y) \quad \& \quad z_2 = \ulcorner \Gamma \urcorner
\end{array}
\right\} (\star)
$$

The induction hypothesis $\chi(x_1, x_2', z)$ gives us the required y_1 and we also have $DT(y_1)$ and $y_1 = \langle \ulcorner \Phi_{|E\&F|-1}(\alpha_1 + \gamma) \urcorner, \ulcorner \Gamma, E, \Delta \to 0 = 1 \urcorner, \ldots \rangle$. The induction hypothesis $\theta(z_1)$ where $z_1 = \ulcorner E \urcorner$ gives us y_2. Since $\mathtt{MultiCt}$ is total, there exists some y such that $\mathtt{MultiCt}(y_2, z_2, y)$ where $z_2 = \ulcorner \Gamma \urcorner$.

We concentrate now on the properties of y that we have found in the paragraph above. If we have y such that $\mathtt{Elim}(x_1, x_2, z, y)$, it means that we have y_1 and y_2 such that $(\star)$. The induction hypotheses $\chi(x_1, x_2', z)$ and $\theta(z_1)$ give us $DT(y_2)$ and $y_2 = \langle \ulcorner \Phi_{|E|-1}(\delta + \Phi_{|E\&F|-1}(\alpha_1 + \gamma)) \urcorner, \ulcorner \Gamma, \Gamma, \Delta \to 0 = 1 \urcorner, \ldots \rangle$. The correctness of $\mathtt{MultiCt}$ reveals that for every y such that $\mathtt{Elim}(x_1, x_2, z, y)$ we have $y = \langle \ulcorner \Phi_{|E\&F|-1}(\alpha_1 + \alpha_2) \urcorner, \ulcorner \Gamma, \Delta \to 0 = 1 \urcorner, \ldots \rangle$ and $DT(y)$.

(iii), (iv) When the last derivation rule in x_2 is the rule of $\neg L_1$ or $\neg L_2$ that modifies the cut formula $\neg E$, the argument is analogous to (i) and (ii), respectively.

(v) Assume that the last derivation rule in x_2 is the rule of $\forall L_1$. Then, x_1 and x_2 are of the form

$$
\begin{array}{rcl}
x_1 &=& \langle \ulcorner \alpha_1 \urcorner, \ulcorner \Gamma \to \forall x F(x) \urcorner, \ulcorner \forall R \urcorner, \ulcorner \rho(n, u) \urcorner \rangle \\
x_2 &=& \langle \ulcorner \alpha_2 \urcorner, \ulcorner \forall x F(x), \Delta \to 0 = 1 \urcorner, \ulcorner \forall L_1 \urcorner, x_2' \rangle \\
x_2' &=& \langle \ulcorner \gamma \urcorner, \ulcorner F(\bar{m}), \Delta \to 0 = 1 \urcorner, \ldots \rangle \\
z &=& \ulcorner \forall x F(x) \urcorner
\end{array}
$$

We need to find y such that $\exists u\, (\, \rho(m, u) \,\&\, \mathtt{Elim}(u, x_2', z_1, y)\,)$ where $z_1 = \ulcorner F(\bar{m}) \urcorner$. At the same time, we need to show that for every y of this kind we also have $DT(y)$ and $y = \langle \ulcorner \Phi_{|\forall x F(x)|-1}(\alpha_1 + \alpha_2) \urcorner, \ulcorner \Gamma, \Delta \to 0 = 1 \urcorner, \ldots \rangle$. Since $DT(x_1)$, there is some u such that $\rho(m, u)$, $u = \langle \ulcorner \delta \urcorner, \ulcorner \Gamma \to F(\bar{m}) \urcorner, \ldots \rangle$, $\delta < \alpha_1$ and $DT(u)$. When we use the induction hypothesis $\theta(z_1)$, $z_1 = \ulcorner F(\bar{m}) \urcorner$, we obtain that the required y exists and for any such y we have $DT(y)$ and $y = \langle \ulcorner \Phi_{|F(\bar{m})|-1}(\delta + \gamma) \urcorner, \ulcorner \Gamma, \Delta \to 0 = 1 \urcorner, \ldots \rangle$.

(vi) Assume that the last derivation rule in x_2 is the rule of $\forall L_2$. Then, x_1 and x_2 are of the form:

$$
\begin{aligned}
x_1 &= \langle \ulcorner \alpha_1 \urcorner, \ulcorner \Gamma \to \forall x F(x) \urcorner, \ulcorner \forall R \urcorner, \ulcorner \rho(n, u) \urcorner \rangle \\
x_2 &= \langle \ulcorner \alpha_2 \urcorner, \ulcorner \forall x F(x), \Delta \to 0 = 1 \urcorner, \ulcorner \forall L_2 \urcorner, x_2' \rangle \\
x_2' &= \langle \ulcorner \gamma \urcorner, \ulcorner F(\bar{m}), \forall x F(x), \Delta \to 0 = 1 \urcorner, \ldots \rangle \\
z &= \ulcorner \forall x F(x) \urcorner
\end{aligned}
$$

We want to prove $\chi(x_1, x_2, z)$, hence, we are looking for y such that $\texttt{Elim}(x_1, x_2, z, y)$. Further, for any such y we must show $DT(y)$ and $y = \langle \ulcorner \Phi_{|\forall x F(x)| - 1}(\alpha_1 + \alpha_2) \urcorner, \ulcorner \Gamma, \Delta \to 0 = 1 \urcorner, \ldots \rangle$. To find this y, we need y_1, y_2, u such that

$$
\left.
\begin{aligned}
&\texttt{Elim}(x_1, x_2', z, y_1) &&\& &&\rho(m, u) &&\& \\
&\texttt{Elim}(u, y_1, z_1, y_2) &&\& &&z_1 = \ulcorner F(\bar{m}) \urcorner &&\& \\
&\texttt{MultiCt}(y_2, z_2, y) &&\& &&z_2 = \ulcorner \Gamma \urcorner
\end{aligned}
\right\} (\circ)
$$

The assumption $DT(x_1)$ gives us u such that $\rho(m, u)$, $DT(u)$ and $u = \langle \ulcorner \delta \urcorner, \ulcorner \Gamma \to F(\bar{m}) \urcorner, \ldots \rangle$. Induction hypotheses $\chi(x_1, x_2', z)$ and $\theta(z_1)$ yield y_1, y_2 and the totality of $\texttt{MultiCt}$ gives us some y such that $\texttt{Elim}(x_1, x_2, z, y)$.

On the other hand, if we have y such that $\texttt{Elim}(x_1, x_2, z, y)$, we want to show $DT(y)$ and that it has the required endsequent and the required height. If we have y such that $\texttt{Elim}(x_1, x_2, z, y)$, it means that we have u, y_1, y_2 such that $(\circ)$ is valid. The induction hypothesis $\chi(x_1, x_2', z)$ tells us $y_1 = \langle \ulcorner \Phi_{|\forall x F(x)| - 1}(\alpha_1 + \gamma) \urcorner, \ulcorner \Gamma, F(\bar{m}), \Delta \to 0 = 1 \urcorner, \ldots \rangle$ and $DT(y_1)$. The induction hypothesis $\theta(z_1)$ tells us $DT(y_2)$. The endsequent of y_2 is $\Gamma, \Gamma, \Delta \to 0 = 1$ and the height of y_2 is bounded by $\Phi_{|F(\bar{m})| - 1}(\delta + \Phi_{|\forall x F(x)| - 1}(\alpha_1 + \gamma))$ where δ is the height of u. Now, the correctness of $\texttt{MultiCt}$ yields $DT(y)$ and $y = \langle \ulcorner \Phi_{|\forall x F(x)| - 1}(\alpha_1 + \alpha_2) \urcorner, \ulcorner \Gamma, \Delta \to 0 = 1 \urcorner, \ldots \rangle$.

Altogether, we have obtained $\theta(z)$ for any cut formula z. This finishes the proof of the correctness of $\texttt{Elim}(x_1, x_2, z, y)$.

Lemma 38. *Assume that the heights of x_1 and x_2 are strictly less than $\Phi_\omega(0)$. The totality and the correctness of $\texttt{Elim}(x_1, x_2, z, y)$ represented by formula (6) can be proved in $\text{I}\Pi_3 + \text{TI}$ up to $\Phi_\omega(0)$. The induction formulas of the transfinite induction are Δ_3. The height of y, a deduction tree for the conclusion of the cut, is strictly less than $\Phi_\omega(0)$.*

1.5.3 Formula $\texttt{DedTree}(x, y)$

Finally, we are ready to study formula $\texttt{DedTree}(x, y)$ where x denotes a code of a derivation in PA and y denotes a code of a deduction tree for the endsequent of x. The definition of $\texttt{DedTree}(x, y)$ investigates the cases according to the last derivation rule in x:

$$\text{DedTree}(x, y) \quad \equiv \quad \phi_{\text{initial}}(x, y) \ \& \ \phi_{\forall I}(x, y) \ \& \ \phi_{\forall E}(x, y) \ \&$$
$$\phi_{\&I}(x, y) \ \& \ \phi_{\&E}(x, y) \ \& \ \phi_{\neg I}(x, y) \ \& \ \phi_{\neg E}(x, y) \ \&$$
$$\phi_{Ind}(x, y) \ \& \ \phi_{Wk}(x, y) \ \& \ \phi_{Ct}(x, y)$$

where

$$\phi_{\text{initial}}(x, y) \quad \equiv$$

if the endsequent of x is an initial sequent

    if it is a mathematical initial sequent, then

        y is a code of the deduction tree for this mathematical initial
        sequent that exists and is defined unambiguously

    &

    if it is a logical initial sequent of the form $D \to D$

        if $D \to D$ contains no free variables, then

            $\text{DedTreeAxiom}(s, l, y) \ \& \ s = \ulcorner D \to D \urcorner \ \& \ l = \ulcorner \emptyset \urcorner$

        &

        if $D \to D$ contains free variables $x_1, \ldots, x_k$, then

            $y = \langle \ \ulcorner 2 \cdot rk(D) + 1 \urcorner, \ulcorner D \to D \urcorner, \ulcorner var \urcorner, \ulcorner \varphi(n_1, \ldots, n_k, y') \urcorner \rangle$

Formula $\varphi(n_1, \ldots, n_k, y')$, which contains free variables $n_1, \ldots, n_k, y'$, and whose code we use in the definition of $\phi_{\text{initial}}(x, y)$ is an abbreviation for

$$\text{DedTreeAxiom}(\ulcorner D \to D(x_i/\bar{n}_i) \urcorner, \ulcorner \emptyset \urcorner, y')$$

Let us continue:

$$\phi_{\forall I}(x, y) \quad \equiv$$

if the last derivation rule in x is $\forall I$ of the form $\begin{array}{c} \vdots \ x' \\ \Gamma \to F(a) \\ \hline \Gamma \to \forall w F(w) \end{array}$

    if $\Gamma \to \forall w F(w)$ contains no free variables, then

        $\exists z \ (\ \text{DedTree}(x', z) \ \& \ z = \langle \ulcorner \alpha \urcorner, \ulcorner \Gamma \to F(a) \urcorner, \ulcorner var \urcorner, \ulcorner \upsilon(n, u) \urcorner \rangle \ \&$
            $y = \langle \ulcorner \alpha \urcorner, \ulcorner \Gamma \to \forall w F(w) \urcorner, \ulcorner \forall R \urcorner, \ulcorner \upsilon(n, u) \urcorner \rangle \ )$

    &

    if $\Gamma \to \forall w F(w)$ contains free variables $x_1, \ldots, x_k$, then

        $\exists z \ (\ \text{DedTree}(x', z) \ \& \ z = \langle \ulcorner \alpha \urcorner, \ulcorner \Gamma \to F(a) \urcorner, \ulcorner var \urcorner, \ulcorner \rho(a, x_1, .., x_k, z') \urcorner \rangle \ \&$
            $y = \langle \ \ulcorner \alpha + 1 \urcorner, \ulcorner \Gamma \to \forall w F(w) \urcorner, \ulcorner var \urcorner, \ulcorner \varphi(n_1, \ldots, n_k, y') \urcorner \rangle \ )$

Formula $\varphi(n_1, \ldots, n_k, y')$ whose code we use in the definition of $\phi_{\forall I}(x, y)$ is an abbreviation for:

$$y' = \langle \ulcorner \alpha \urcorner, \ulcorner \Gamma \to \forall w F(w)(x_i/\bar{n}_i) \urcorner, \ulcorner \forall R \urcorner, \ulcorner \rho(a, x_1, .., x_k, z')(x_i/\bar{n}_i) \urcorner \rangle$$

In this case, we do not use codes of deduction trees for the premises of the ω-rules; we mean codes u and z' such that $\upsilon(n, u)$ and $\rho(a, x_1, \ldots, x_k, z')$ respectively. We just work with the codes of $\upsilon(n, u)$ and $\rho(a, x_1, \ldots, x_k, z')$, thus, we do not require right now that the formulas yield correct deduction trees u and z'. The next case, however, is different.

$\phi_{\forall E}(x, y) \quad \equiv$

if the last derivation rule in x is $\forall E$ of the form $\quad \dfrac{\substack{\vdots \, x' \\ \Gamma \rightarrow \forall w F(w)}}{\Gamma \rightarrow F(t)}$

    if $\Gamma \rightarrow F(t)$ contains no free variables, then

        $\exists z \, \exists m \, (\, \mathtt{DedTree}(x', z) \,\&\, z = \langle \ulcorner \alpha \urcorner, \ulcorner \Gamma \rightarrow \forall w F(w) \urcorner, \ulcorner \forall R \urcorner, \ulcorner \upsilon(n, u) \urcorner \rangle \,\&$

            value of t is m $\&$ $\upsilon(m, y)$)

    $\&$

    if $\Gamma \rightarrow F(t)$ contains free variables $x_1, \ldots, x_k$, then

        $\exists z \, (\, \mathtt{DedTree}(x', z) \,\&\, z = \langle \ulcorner \alpha \urcorner, \ulcorner \Gamma \rightarrow \forall w F(w) \urcorner, \ulcorner var \urcorner, \ulcorner \rho(x_1, \ldots, x_k, z') \urcorner \rangle \,\&$

            $y = \langle \ulcorner \alpha \urcorner, \ulcorner \Gamma \rightarrow F(t) \urcorner, \ulcorner var \urcorner, \ulcorner \varphi(n_1, \ldots, n_k, y') \urcorner \rangle \,)$

Formula $\varphi(n_1, \ldots, n_k, y')$ whose code we use in the definition of $\phi_{\forall E}(x, y)$ is an abbreviation for:

$$\exists z' \, (\, \rho(n_1, \ldots, n_k, z') \,\&$$
$$z' = \langle \ulcorner \beta \urcorner, \ulcorner \Gamma \rightarrow \forall w F(w)(x_i / \bar{n}_i) \urcorner, \ulcorner \forall R \urcorner, \ulcorner \sigma(n, u) \urcorner \rangle \,\&$$
$$\text{value of } t(x_i / \bar{n}_i) \text{ is } m \,\&$$
$$\sigma(m, y')$$
$$)$$

In general, we do not require that there exist some y, z', y' such that $\upsilon(m, y)$, $\rho(n_1, \ldots, n_k, z')$ and $\sigma(m, y')$, respectively. Hence, we are not able to prove the totality of $\mathtt{DedTree}(x, y)$ without knowing that we deal with proper deduction trees. It follows that, after the definition of the formula, we will prove the totality and the correctness at once in a similar way as we did by $\mathtt{Elim}(x_1, x_2, z, y)$.

$\phi_{\&I}(x, y) \quad \equiv$

if the last derivation rule in x is $\&I$ of the form $\quad \dfrac{\substack{\vdots \, x' \\ \Gamma \rightarrow A} \qquad \substack{\vdots \, x'' \\ \Delta \rightarrow B}}{\Gamma, \Delta \rightarrow A \& B}$

    if $\Gamma, \Delta \rightarrow A \& B$ contains no free variables, then

$$\exists z' \exists z'' \; (\; y = \langle \ulcorner \max\{\alpha,\beta\}+1 \urcorner, \ulcorner \Gamma, \Delta \to A\&B \urcorner, \ulcorner \&R \urcorner, y', y'' \rangle \; \&$$
$$\mathtt{DedTree}(x', z') \; \& \; z' = \langle \ulcorner \alpha \urcorner, \ulcorner \Gamma \to A \urcorner, \dots \rangle \; \&$$
$$\mathtt{DedTree}(x'', z'') \; \& \; z'' = \langle \ulcorner \beta \urcorner, \ulcorner \Delta \to B \urcorner, \dots \rangle \; \&$$
$$\mathtt{Wk}(z', w', y') \; \& \; w' = \ulcorner \Delta \urcorner \; \&$$
$$\mathtt{Wk}(z'', w'', y'') \; \& \; w'' = \ulcorner \Gamma \urcorner$$
$$)$$

&

if $\Gamma, \Delta \to A\&B$ contains free variables $x_1, \dots, x_k$, then
$$\exists z' \exists z'' \; (\; \mathtt{DedTree}(x', z') \; \& \; z' = \langle \ulcorner \alpha \urcorner, \ulcorner \Gamma \to A \urcorner, \ulcorner var \urcorner, \ulcorner \rho'(x_1, \dots, x_k, u) \urcorner \rangle \; \&$$
$$\mathtt{DedTree}(x'', z'') \; \& \; z'' = \langle \ulcorner \beta \urcorner, \ulcorner \Delta \to B \urcorner, \ulcorner var \urcorner, \ulcorner \rho''(x_1, \dots, x_k, v) \urcorner \rangle \; \&$$
$$y = \langle \; \ulcorner \max\{\alpha,\beta\}+1 \urcorner, \ulcorner \Gamma, \Delta \to A\&B \urcorner, \ulcorner var \urcorner, \ulcorner \varphi(n_1, \dots, n_k, y') \urcorner \rangle$$
$$)$$

Formula $\varphi(n_1, \dots, n_k, y')$ whose code we need in the definition of $\phi_{\&I}(x, y)$ is an abbreviation for:

$$\exists u \exists v \; (\; \rho'(n_1, \dots, n_k, u) \; \& \; \rho''(n_1, \dots, n_k, v) \; \&$$
$$\mathtt{Wk}(u, p, u') \; \& \; p = \ulcorner \Delta(x_i/\bar{n}_i) \urcorner \; \&$$
$$\mathtt{Wk}(v, q, v') \; \& \; q = \ulcorner \Gamma(x_i/\bar{n}_i) \urcorner \; \&$$
$$y' = \langle \ulcorner \max\{\alpha,\beta\} \urcorner, \ulcorner \Gamma, \Delta \to A\&B(x_i/\bar{n}_i) \urcorner, \ulcorner \&R \urcorner, u', v' \rangle$$
$$)$$

Let us continue with constructing a deduction tree for the conclusion of the rule of $\&E$.

$\phi_{\&E}(x, y) \quad \equiv$

if the last derivation rule in x is $\&E$ of the form $\begin{array}{c} \vdots\, x' \\ \hline \Gamma \to A\&B \\ \hline \Gamma \to A \end{array}$

if $\Gamma \to A$ contains no free variables, then

if B contains no free variables, then
$$\exists z \; (\; \mathtt{DedTree}(x', z) \; \& \; z = \langle \; \ulcorner \alpha \urcorner, \ulcorner \Gamma \to A\&B \urcorner, \ulcorner \&R \urcorner, y, z' \; \rangle \;)$$

&

if B contains free variables $y_1, \dots, y_l$, then
$$\exists z \exists u \; (\; \mathtt{DedTree}(x', z) \; \&$$
$$z = \langle \; \ulcorner \alpha \urcorner, \ulcorner \Gamma \to A\&B \urcorner, \ulcorner var \urcorner, \ulcorner \upsilon(y_1, \dots, y_l, u) \urcorner \; \rangle \; \&$$
$$\upsilon(0, \dots, 0, u) \; \&$$
$$u = \langle \; \ulcorner \delta \urcorner, \ulcorner \Gamma \to A\&B(y_i/0) \urcorner, \ulcorner \&R \urcorner, y, u' \; \rangle$$
$$)$$

&

if $\Gamma \to A$ contains free variables $x_1, \dots, x_k$, then

if B contains free variables $y_1, \dots, y_l$, then
$$\exists z \; (\; \mathtt{DedTree}(x', z) \; \&$$
$$z = \langle \ulcorner \alpha \urcorner, \ulcorner \Gamma \to A\&B \urcorner, \ulcorner var \urcorner, \ulcorner \rho(x_1, \dots, x_k, y_1, \dots, y_l, u) \urcorner \rangle \; \&$$
$$y = \langle \ulcorner \alpha \urcorner, \ulcorner \Gamma \to A \urcorner, \ulcorner var \urcorner, \ulcorner \varphi(n_1, \dots, n_k, y') \urcorner \rangle$$
$$)$$

Formula $\varphi(n_1, \ldots, n_k, y')$ whose code is needed in the definition of $\phi_{\&E}(x, y)$ is an abbreviation for:

$$\exists u \; (\; \rho(n_1, \ldots, n_k, 0, \ldots, 0, u) \; \& $$
$$u = \langle \ulcorner \beta \urcorner, \ulcorner \Gamma \to A \& B(x_i/\bar{n}_i)(y_i/0) \urcorner, \ulcorner \& R \urcorner, y', y'' \rangle$$
$$)$$

Now, we move on to the cases that make use of $\texttt{Elim}(x_1, x_2, z, y)$. These are the cases when the last derivation rule in x is one of the rules of negation or the induction rule.

$\phi_{\neg I}(x, y) \quad \equiv$

if the last derivation rule in x is $\neg I$ of the form $\quad \dfrac{\begin{matrix} \vdots \; x' \\ A, \Gamma \to B \end{matrix} \qquad \begin{matrix} \vdots \; x'' \\ A, \Delta \to \neg B \end{matrix}}{\Gamma, \Delta \to \neg A}$

    if $\Gamma, \Delta \to \neg A$ contains no free variables

        if B contains no free variables, then

$$\exists z' \, \exists z'' \, \exists u \; (\; y = \langle \ulcorner \Phi_{|B|-1}(\alpha + \beta) \urcorner, \ulcorner \Gamma, \Delta \to \neg A \urcorner, \ulcorner \neg R \urcorner, y' \rangle \; \&$$
$$\texttt{DedTree}(x', z') \; \& \; z' = \langle \ulcorner \alpha \urcorner, \ulcorner A, \Gamma \to B \urcorner, \ldots \rangle \; \&$$
$$\texttt{DedTree}(x'', z'') \; \& \; z'' = \langle \ulcorner \beta \urcorner, \ulcorner A, \Delta \to \neg B \urcorner, \ulcorner \neg R \urcorner, z''' \rangle \; \&$$
$$\texttt{Elim}(z', z''', v, u) \; \& \; v = \ulcorner B \urcorner \; \&$$
$$\texttt{Ct}(u, w, y') \; \& \; w = \ulcorner A \urcorner$$
$$)$$

        &

        if B contains free variables $y_1, \ldots, y_l$, then

$$\exists z' \exists z'' \exists u \, \exists v \, \exists q \; (\; y = \langle \ulcorner \Phi_{|B|-1}(\alpha + \beta) \urcorner, \ulcorner \Gamma, \Delta \to \neg A \urcorner, \ulcorner \neg R \urcorner, y' \rangle \; \&$$
$$\texttt{DedTree}(x', z') \; \&$$
$$z' = \langle \ulcorner \alpha \urcorner, \ulcorner A, \Gamma \to B \urcorner, \ulcorner var \urcorner, \ulcorner \sigma'(y_1, \ldots, y_l, u) \urcorner \rangle \; \&$$
$$\texttt{DedTree}(x'', z'') \; \&$$
$$z'' = \langle \ulcorner \beta \urcorner, \ulcorner A, \Delta \to \neg B \urcorner, \ulcorner var \urcorner, \ulcorner \sigma''(y_1, \ldots, y_l, v) \urcorner \rangle \; \&$$
$$\sigma'(0, \ldots, 0, u) \; \& \; \sigma''(0, \ldots, 0, v) \; \&$$
$$v = \langle \ulcorner \gamma \urcorner, \ulcorner A, \Delta \to \neg B(y_i/0) \urcorner, \ulcorner \neg R \urcorner, v' \rangle \; \&$$
$$\texttt{Elim}(u, v', w', q) \; \& \; w' = \ulcorner B(y_i/0) \urcorner \; \&$$
$$\texttt{Ct}(q, w'', y') \; \& \; w'' = \ulcorner A \urcorner$$
$$)$$

    &

    if $\Gamma, \Delta \to \neg A$ contains free variables $x_1, \ldots, x_k$

        if B contains free variables $y_1, \ldots, y_l$, then

$$\exists z' \exists z'' \; (\; \texttt{DedTree}(x', z') \; \&$$
$$z' = \langle \ulcorner \alpha \urcorner, \ulcorner A, \Gamma \to B \urcorner, \ulcorner var \urcorner, \ulcorner \rho'(x_1, \ldots, x_k, y_1, \ldots, y_l, u) \urcorner \rangle \; \&$$
$$\texttt{DedTree}(x'', z'') \; \&$$
$$z'' = \langle \ulcorner \beta \urcorner, \ulcorner A, \Delta \to \neg B \urcorner, \ulcorner var \urcorner, \ulcorner \rho''(x_1, \ldots, x_k, y_1, \ldots, y_l, v) \urcorner \rangle \; \&$$
$$y = \langle \ulcorner \Phi_{|B|-1}(\alpha + \beta) + 1 \urcorner, \ulcorner \Gamma, \Delta \to \neg A \urcorner, \ulcorner var \urcorner, \ulcorner \varphi(n_1, \ldots, n_k, y') \urcorner \rangle$$
$$)$$

Formula $\varphi(n_1, \ldots, n_k, y')$ whose code is needed in the definition of $\phi_{\neg I}(x, y)$ is an abbreviation for:

$$\exists u \, \exists v \, \exists q \; (\; y' = \langle \ulcorner \Phi_{|B|-1}(\alpha + \beta) \urcorner, \ulcorner \Gamma, \Delta \to \neg A(x_i/\bar{n}_i) \urcorner, \ulcorner \neg R \urcorner, y'' \rangle \; \& $$
$$\rho'(n_1, \ldots, n_k, 0, \ldots, 0, u) \; \& \; \rho''(n_1, \ldots, n_k, 0, \ldots, 0, v) \; \& $$
$$v = \langle \ulcorner \gamma \urcorner, \ulcorner A, \Delta \to \neg B(x_i/\bar{n}_i)(y_i/0) \urcorner, \ulcorner \neg R \urcorner, v' \rangle \; \& $$
$$\mathtt{Elim}(u, v', w', q) \; \& \; w' = \ulcorner B(y_i/0) \urcorner \; \& $$
$$\mathtt{Ct}(q, w'', y'') \; \& \; w'' = \ulcorner A(x_i/\bar{n}_i) \urcorner$$
$$)$$

We will use an auxiliary formula $\mathtt{DedTreeNeg}(s, l, y)$ in the following formula denoted by $\phi_{\neg E}(x, y)$. Formula $\mathtt{DedTreeNeg}(s, l, y)$ holds true when y is a code of a deduction tree for a sequent of the form $\neg\neg D \to D$ whose code is s. Variable l denotes a list with a similar meaning as in $\mathtt{DedTreeAxiom}$ and D is a sentence.

$\phi_{\neg E}(x, y) \quad \equiv$

if the last derivation rule in x is $\neg E$ of the form $\begin{array}{c} \vdots \; x' \\ \hline \Gamma \to \neg\neg A \\ \hline \Gamma \to A \end{array}$

    if $\Gamma \to A$ contains no free variables, then

$$\exists z \, \exists u \; (\; \mathtt{DedTree}(x', z) \; \& $$
$$\mathtt{DedTreeNeg}(s, l, u) \; \& \; s = \ulcorner \neg\neg A \to A \urcorner \; \& \; l = \ulcorner \emptyset \urcorner \; \& $$
$$\mathtt{Elim}(z, u, w, y) \; \& \; w = \ulcorner \neg\neg A \urcorner$$
$$)$$

    &$

    if $\Gamma \to A$ contains free variables $x_1, \ldots, x_k$, then

$$\exists z \; (\; \mathtt{DedTree}(x', z) \; \& $$
$$z = \langle \ulcorner \alpha \urcorner, \ulcorner \Gamma \to \neg\neg A \urcorner, \ulcorner var \urcorner, \ulcorner \rho(x_1, \ldots, x_k, u) \urcorner \rangle \; \& $$
$$y = \langle \ulcorner \Phi_{|\neg\neg A|-1}(\alpha + 2 \cdot rk(A) + 2) \urcorner, \ulcorner \Gamma \to A \urcorner, \ulcorner var \urcorner, \ulcorner \varphi(n_1, \ldots, n_k, y') \urcorner \rangle$$
$$)$$

Formula $\varphi(n_1, \ldots, n_k, y')$ whose code we use in the definition of $\phi_{\neg E}(x, y)$ is an abbreviation for:

$$\exists u \, \exists v \; (\; \rho(n_1, \ldots, n_k, u) \; \& $$
$$\mathtt{DedTreeNeg}(s, l, v) \; \& \; s = \ulcorner \neg\neg A \to A(x_i/\bar{n}_i) \urcorner \; \& \; l = \ulcorner \emptyset \urcorner \; \& $$
$$\mathtt{Elim}(u, v, w, y') \; \& \; w = \ulcorner \neg\neg A(x_i/\bar{n}_i) \urcorner$$
$$)$$

Note that the deduction tree y' for sequent $\Gamma \to A(x_i/\bar{n}_i)$ that formula $\varphi(n_1, \ldots, n_k, y')$ gives us has height at most $\Phi_{|\neg\neg A|-1}(\alpha' + 2 \cdot rk(A) + 2)$

69

where $\alpha' < \alpha$. Since $\alpha' < \alpha$ and $2 \cdot rk(A) + 2$ is a natural number, we have $\alpha' + 2 \cdot rk(A) + 2 < \alpha + 2 \cdot rk(A) + 2$ and the height of y can be bounded by $\Phi_{|\neg\neg A|-1}(\alpha + 2 \cdot rk(A) + 2)$.

Let us state the definition of $\mathtt{DedTreeNeg}(s, l, y)$. The formula formalizes the construction of a deduction tree for a sequent of the form $\neg\neg D \to D$ described in Lemma 18.

$\mathtt{DedTreeNeg}(s, l, y) \quad \equiv$

if s is $\ulcorner \neg\neg D \to B \urcorner$

    if B is a true atomic sentence, then $y = \langle\, \ulcorner 0 \urcorner, s, \ulcorner \mathrm{endform} \urcorner, \ulcorner \emptyset \urcorner \,\rangle$

    &

    if B is a false atomic sentence, then

        $y = \langle \ulcorner rk(\neg\neg D \to 0 = 1) \urcorner, s, \ulcorner \neg L_1 \urcorner, y' \rangle$ &

        $y' = \langle \ulcorner rk(\to \neg D) \urcorner, \ulcorner \to \neg D \urcorner, \ulcorner \neg R \urcorner, y'' \rangle$ &

        $\mathtt{DedTreeAxiom}(s', l, y'')$ & $s' = \ulcorner D \to 0 = 1 \urcorner$

    &

    if B is a compound formula

        if B is of the form $E\&F$, then

            $y = \langle \ulcorner rk(\neg\neg D \to E\&F) \urcorner, s, \ulcorner \& R \urcorner, y', y'' \rangle$ &

            $\mathtt{DedTreeNeg}(w', \langle\, \ulcorner E \urcorner \,\rangle * l, \; y')$ & $w' = \ulcorner \neg\neg D \to E \urcorner$ &

            $\mathtt{DedTreeNeg}(w'', \langle\, \ulcorner F \urcorner \,\rangle * l, \; y'')$ & $w'' = \ulcorner \neg\neg D \to F \urcorner$

        &

        if B is of the form $\forall x F(x)$, then

            $y = \langle\, \ulcorner rk(\neg\neg D \to \forall x F(x)) \urcorner, s, \ulcorner \forall R \urcorner,$

                    $\ulcorner \mathtt{DedTreeNeg}(\ulcorner \neg\neg D \to F(\bar{n}) \urcorner, \langle\, \ulcorner F(\bar{n}) \urcorner \,\rangle * l, y') \urcorner \,\rangle$

        &

        if B is of the form $\neg C$, then

            $y = \langle \ulcorner rk(\neg\neg D \to \neg C) \urcorner, s, \ulcorner \neg R \urcorner, y' \rangle$ &

            $y' = \langle \ulcorner rk(\neg\neg D, C \to 0 = 1) \urcorner, \ulcorner \neg\neg D, C \to 0 = 1 \urcorner, \ulcorner \neg L_1 \urcorner, y'' \rangle$ &

            $y'' = \langle \ulcorner rk(C \to \neg D) \urcorner, \ulcorner C \to \neg D \urcorner, \ulcorner \neg R \urcorner, y''' \rangle$ &

            $\mathtt{DedTreeAxiom}(s', l, y''')$ & $s' = \ulcorner D, C \to 0 = 1 \urcorner$

Formula $\mathtt{DedTreeNeg}(s, l, y)$ is Δ_1 in $\mathrm{I}\Sigma_1$. Its totality, which is denoted by $\forall s\, \forall l\, \exists y\, \mathtt{DedTreeNeg}(s, l, y)$, can be proved in $\mathrm{I}\Pi_2$. If we set

- $Sentence(s) \quad \equiv \quad s$ is a code of a sequent that does not contain free variables

- $SubNeg(s) \quad \equiv \quad s$ is a code of a sequent of the form $\neg\neg D \to B$ where B is a subformula of D such that B is not within the scope of a negation in D

- $ChoicesNeg(l,s) \quad \equiv \quad \exists n \leq l \quad (\quad l = \langle\, i_0, \ldots, i_n\,\rangle \; \& $

$$\forall j \leq n \; (i_j \text{ is a code of a formula}) \; \& $$

$$\forall j < n \; (i_{j+1} \text{ is not a negation}) \; \& $$

$$\forall j < n \; One\_Step\_Subfle(i_{j+1}, i_j) \; \& $$

$$s = \ulcorner \neg\neg D \to B \urcorner \; \& $$

$$i_0 = \ulcorner B \urcorner \text{ and } One\_Step\_Subfle(\ulcorner D \urcorner, i_n) $$

$$) $$

$$\vee \quad (\, l = \ulcorner \emptyset \urcorner \; \& \; s = \ulcorner \neg\neg D \to D \urcorner \,) $$

the correctness of $\texttt{DedTreeNeg}(s,l,y)$ is expressed as

$$\forall s \, \forall l \; [\; Sentence(s) \; \& \; SubNeg(s) \; \& \; ChoicesNeg(l,s) $$
$$\supset$$
$$\forall y \; (\; \texttt{DedTreeNeg}(s,l,y) $$
$$\supset$$
$$DT(y) \; \& \; \text{endsequent of } y \text{ has code } s \; \& $$
$$\text{height of } y \leq \text{rank of sequent with code } s $$
$$) $$
$$] \tag{7}$$

This is shown by induction on $rk(S)$ where S is a sequent whose code is s. Since predicates $Sentence(s)$, $SubNeg(s)$ and $ChoicesNeg(l,s)$ are Δ_1 in $I\Sigma_1$, the induction formula is Π_2. The correctness of $\texttt{DedTreeAxiom}(s,l,y)$ is also used in the proof of the correctness of $\texttt{DedTreeNeg}(s,l,y)$. This does not affect the complexity of the proof since it is proved in $I\Pi_2$ too.

We go on to define the case when the last derivation rule in x is the induction rule.

$$\phi_{Ind}(x,y) \quad \equiv$$

if the last derivation rule in x is the induction rule of the form $\dfrac{\begin{array}{c}\vdots\; x' \\ \Gamma \to F(0)\end{array} \qquad \begin{array}{c}\vdots\; x'' \\ F(a), \Delta \to F(a+1)\end{array}}{\Gamma, \Delta \to F(t)}$

if $\Gamma, \Delta \to F(t)$ contains no free variables, then

$$\exists z' \exists z'' \exists m \; (\; \texttt{DedTree}(x', z') \; \& \; \texttt{DedTree}(x'', z'') \; \& $$
$$z'' = \langle \ulcorner \beta \urcorner, \ulcorner F(a), \Delta \to F(a+1) \urcorner, \ulcorner var \urcorner, \ulcorner \rho(n,u) \urcorner \rangle \; \& $$
$$\text{value of } t \text{ is } m \; \& \; \psi_{Ind}(z', \ulcorner \rho(n,u) \urcorner, m, y) $$
$$)$$

$$\&$$

if $\Gamma, \Delta \to F(t)$ contains free variables $x_1, \ldots, x_k$, then

$$\exists z' \exists z'' \; (\; \texttt{DedTree}(x', z') \; \& $$
$$z' = \langle \ulcorner \alpha \urcorner, \ulcorner \Gamma \to F(0) \urcorner, \ulcorner var \urcorner, \ulcorner \rho'(x_1, \ldots, x_k, u) \urcorner \rangle \; \& $$
$$\texttt{DedTree}(x'', z'') \; \& $$
$$z'' = \langle \ulcorner \beta \urcorner, \ulcorner F(a), \Delta \to F(a+1) \urcorner, \ulcorner var \urcorner, \ulcorner \rho''(a, x_1, \ldots, x_k, v) \urcorner \rangle \; \& $$
$$y = \langle\; \ulcorner \Phi_{|F(x)|}(\max\{\alpha, \beta\} + \omega) \urcorner, \ulcorner \Gamma, \Delta \to F(t) \urcorner, \ulcorner var \urcorner, $$
$$\ulcorner \exists m \, (\, t(x_i/\bar{n}_i) = m \; \& \; \vartheta_{Ind}(\ulcorner \rho' \urcorner, \ulcorner \rho'' \urcorner, m, n_1, \ldots, n_k, y')) \urcorner \; \rangle $$
$$)$$

Now, the task is to define formulas

$$\psi_{Ind}(z', \ulcorner \rho(n, u) \urcorner, m, y)$$

$$\vartheta_{Ind}(\ulcorner \rho' \urcorner, \ulcorner \rho'' \urcorner, m, n_1, \ldots, n_k, y')$$

Let us begin with the first one. Formula $\psi_{Ind}(z', \ulcorner \rho(n, u) \urcorner, m, y)$ is defined as follows:

$$\psi_{Ind}(z', \ulcorner \rho(n, u) \urcorner, m, y) \quad \equiv$$

$$\exists u \, \exists v \, \exists w \, [\; m = 0 \supset (\; \mathtt{Wk}(z', q, y) \; \& \; q = \ulcorner \Delta \urcorner \;)$$
$$\&$$
$$m \neq 0 \supset (\; \rho(m-1, u) \; \& \; \psi_{Ind}(z', \ulcorner \rho(n, u) \urcorner, m-1, v) \; \&$$
$$\mathtt{Elim}(v, u, p, w) \; \& \; p = \ulcorner F(\overline{m-1}) \urcorner \; \&$$
$$\mathtt{MultiCt}(w, r, y) \; \& \; r = \ulcorner \Delta \urcorner$$
$$)$$
$$]$$

The second formula $\vartheta_{Ind}(\ulcorner \rho' \urcorner, \ulcorner \rho'' \urcorner, m, n_1, \ldots, n_k, y')$ is of the form:

$$\vartheta_{Ind}(\ulcorner \rho' \urcorner, \ulcorner \rho'' \urcorner, m, n_1, \ldots, n_k, y') \quad \equiv$$

$$\exists u \, \exists v \, \exists w \, \exists y'' \, [\; m = 0 \supset (\; \rho'(n_1, \ldots, n_k, u) \; \&$$
$$\mathtt{Wk}(u, q, y') \; \& \; q = \ulcorner \Delta(x_i / \bar{n}_i) \urcorner$$
$$)$$
$$\&$$
$$m \neq 0 \supset (\; \vartheta_{Ind}(\ulcorner \rho' \urcorner, \ulcorner \rho'' \urcorner, m-1, n_1, \ldots, n_k, y'') \; \&$$
$$\rho''(m-1, n_1, \ldots, n_k, v) \; \&$$
$$\mathtt{Elim}(y'', v, p, w) \; \& \; p = \ulcorner F(\overline{m-1})(x_i / \bar{n}_i) \urcorner \; \&$$
$$\mathtt{MultiCt}(w, r, y') \; \& \; r = \ulcorner \Delta(x_i / \bar{n}_i) \urcorner$$
$$)$$
$$]$$

To complete the definition of $\mathtt{DedTree}(x, y)$, we state the cases for the rules of contraction and weakening.

$\phi_{Wk}(x, y) \quad \equiv$

if the last derivation rule in x is Wk of the form $\begin{array}{c} \vdots \ x^{'} \\ \hline \Gamma \to C \\ \hline A, \Gamma \to C \end{array}$

    if $A, \Gamma \to C$ contains no free variables, then

        $\exists z \,(\, \mathtt{DedTree}(x^{'}, z) \,\&\, \mathtt{Wk}(z, w, y) \,\&\, w = \ulcorner A \urcorner \,)$

    &

    if $A, \Gamma \to C$ contains free variables $x_1, \ldots, x_k$, then

        $\exists z \,(\, \mathtt{DedTree}(x^{'}, z) \,\&\, z = \langle \ulcorner \alpha \urcorner, \ulcorner \Gamma \to C \urcorner, \ulcorner var \urcorner, \ulcorner \rho(x_1, \ldots, x_k, u) \urcorner \rangle \,\&$
           $y = \langle \ulcorner \alpha \urcorner, \ulcorner A, \Gamma \to C \urcorner, \ulcorner var \urcorner, \ulcorner \varphi(n_1, \ldots, n_k, y^{'}) \urcorner \rangle \,)$

Formula $\varphi(n_1, \ldots, n_k, y^{'})$ that we use in the definition of $\phi_{Wk}(x, y)$ is an abbreviation for:

$$\exists u \,(\, \rho(n_1, \ldots, n_k, u) \,\&\, \mathtt{Wk}(u, w, y^{'}) \,\&\, w = \ulcorner A(x_i / \bar{n}_i) \urcorner \,)$$

$\phi_{Ct}(x, y) \quad \equiv$

if the last derivation rule in x is Ct of the form $\begin{array}{c} \vdots \ x^{'} \\ \hline A, A, \Gamma \to C \\ \hline A, \Gamma \to C \end{array}$

    if $A, \Gamma \to C$ contains no free variables, then

        $\exists z \,(\, \mathtt{DedTree}(x^{'}, z) \,\&\, \mathtt{Ct}(z, w, y) \,\&\, w = \ulcorner A \urcorner \,)$

    &

    if $A, \Gamma \to C$ contains free variables $x_1, \ldots, x_k$, then

        $\exists z \,(\, \mathtt{DedTree}(x^{'}, z) \,\&\, z = \langle \ulcorner \alpha \urcorner, \ulcorner A, A, \Gamma \to C \urcorner, \ulcorner var \urcorner, \ulcorner \rho(x_1, \ldots, x_k, u) \urcorner \rangle \,\&$
           $y = \langle \ulcorner \alpha \urcorner, \ulcorner A, \Gamma \to C \urcorner, \ulcorner var \urcorner, \ulcorner \varphi(n_1, \ldots, n_k, y^{'}) \urcorner \rangle \,)$

Formula $\varphi(n_1, \ldots, n_k, y^{'})$ that we use in the definition of $\phi_{Ct}(x, y)$ is an abbreviation for:

$$\exists u \,(\, \rho(n_1, \ldots, n_k, u) \,\&\, \mathtt{Ct}(u, w, y^{'}) \,\&\, w = \ulcorner A(x_i / \bar{n}_i) \urcorner \,)$$

The construction of $\mathtt{DedTree}(x, y)$ requires two applications of the Fixed-point theorem. First, we have to build formulas $\psi_{Ind}(z^{'}, \ulcorner \rho(n, u) \urcorner, m, y)$ and $\vartheta_{Ind}(\ulcorner \rho^{'} \urcorner, \ulcorner \rho^{''} \urcorner, m, n_1, \ldots, n_k, y^{'})$ that are part of the definition of ϕ_{Ind}. We see that the definition of $\psi_{Ind}(z^{'}, \ulcorner \rho(n, u) \urcorner, m, y)$ refers to ψ_{Ind} itself and the definition of $\vartheta_{Ind}(\ulcorner \rho^{'} \urcorner, \ulcorner \rho^{''} \urcorner, m, n_1, \ldots, n_k, y^{'})$ refers to ϑ_{Ind}. These two formulas are built in the usual way by applying the Fixed-point theorem and the partial truth predicates for Σ_1-formulas. The second step is to repeat an analogous process to obtain $\mathtt{DedTree}(x, y)$ that is Σ_1 too.

73

We proceed to prove the correctness of $\mathtt{DedTree}(x,y)$ that is expressed by the following formula:

$$\forall x \, [\; x \text{ is a derivation of PA}$$
$$\supset$$
$$\exists y \, \mathtt{DedTree}(x,y) \; \& \; \forall y \; (\; \mathtt{DedTree}(x,y)$$
$$\supset$$
$$DT(y) \; \& \; \text{height of } y < \; \Phi_\omega(0) \; \& \quad (8)$$
$$\text{the endsequent of } y \text{ is the same}$$
$$\text{as the endsequent of } x$$
$$)$$
$$]$$

This is proved by induction on the height of x, which is a natural number, and the complexity of the induction formula is Π_2. This proof makes use of the totality and the correctness of all formulas that have been analysed so far: $\mathtt{DedTreeAxiom}(s,l,y)$, $\mathtt{Wk}(x,z,y)$, $\mathtt{Ct}(x,z,y)$, $\mathtt{MultiCt}(x,z,y)$, $\mathtt{Elim}(x_1,x_2,z,y)$. Since the proof of the correctness of $\mathtt{Elim}$ has the highest complexity, we obtain eventually that the whole consistency proof of 1935 can be formalized in the theory where the correctness of $\mathtt{Elim}$ can be proved.

Let us present the correctness proof for the case when the last derivation rule in x is the induction rule of the form

$$\frac{\begin{array}{cc} \vdots \, x' & \vdots \, x'' \\ \Gamma \to F(0) & F(a), \Delta \to F(a+1) \end{array}}{\Gamma, \Delta \to F(t)}$$

Assume that the endsequent $\Gamma, \Delta \to F(t)$ contains free variables $x_1, \ldots, x_k$. Our aim is to prove that (1) there exists some y such that $\mathtt{DedTree}(x,y)$ and (2) for any such y we have $DT(y)$, the height of y is strictly less than $\Phi_\omega(0)$ and its endsequent is $\Gamma, \Delta \to F(t)$.

We will work with the definition of ϕ_{Ind} on page 71. To find y such that $\mathtt{DedTree}(x,y)$, we need to find z' and z'' of the required form such that $\mathtt{DedTree}(x',z')$ and $\mathtt{DedTree}(x'',z'')$. These deduction trees z' and z'' are obtained by the induction hypothesis applied to x' and x'' respectively. Then, the y that we are looking for exists. Namely, it is built of the heights of z' and z'', of the endsequent of x, of a code of the rule of var and of a code of the formula

$$\exists m \, (\, t(x_i/\bar{n}_i) = m \; \& \; \vartheta_{Ind}(\ulcorner \rho' \urcorner, \ulcorner \rho'' \urcorner, m, n_1, \ldots, n_k, y'))$$

that can be constructed as above.

Assume now that we have an arbitrary y such that $\mathtt{DedTree}(x, y)$. This means that we also have z' and z'' such that $\mathtt{DedTree}(x', z')$, $\mathtt{DedTree}(x'', z'')$ and

$$
\begin{aligned}
z' &= \langle \ulcorner \alpha \urcorner, \ulcorner \Gamma \to F(0) \urcorner, \ulcorner var \urcorner, \ulcorner \rho'(x_1, \ldots, x_k, u) \urcorner \rangle \\
z'' &= \langle \ulcorner \beta \urcorner, \ulcorner F(a), \Delta \to F(a+1) \urcorner, \ulcorner var \urcorner, \ulcorner \rho''(a, x_1, \ldots, x_k, v) \urcorner \rangle
\end{aligned}
$$

When we use the induction hypothesis for x' and x'', we further obtain $DT(z')$ and $DT(z'')$ which imply that formulas ρ' and ρ'' are total and yield proper deduction trees. The induction hypothesis also gives us $\alpha, \beta < \Phi_\omega(0)$.

It is easy to see that the height of y is less than $\Phi_\omega(0)$ and the endsequent of y is $\Gamma, \Delta \to F(t)$. The important part to show is $DT(y)$. This entails two items:

(2a) $\forall n_1 \ldots n_k \exists y' \exists m \, (\, t(x_i/\bar{n}_i) = m \,\, \& \,\, \vartheta_{Ind}(\ulcorner \rho' \urcorner, \ulcorner \rho'' \urcorner, m, n_1, \ldots, n_k, y') \,)$

(2b) $\forall n_1 \ldots n_k \forall y' \, [\, \exists m \, (\, t(x_i/\bar{n}_i) = m \,\, \& \,\, \vartheta_{Ind}(\ulcorner \rho' \urcorner, \ulcorner \rho'' \urcorner, m, n_1, \ldots, n_k, y') \,)$
$\qquad\qquad \supset$
$\qquad\qquad DT(y') \,\, \& \,\, \text{endsequent of } y' \text{ is } \Gamma, \Delta \to F(t)(x_i/\bar{n}_i) \,\, \& \,\,$
$\qquad\qquad \text{height of } y' < \Phi_{|F(x)|}(\max\{\alpha, \beta\} + \omega)$
$\qquad \,]$

The definition of ϑ_{Ind} is on page 72. Both items are easily obtained from the auxiliary statement below. Recall that deduction trees are allowed to use the rule of *term* (p. 24) implicitly that replaces closed terms by their corresponding values. The statement to prove is the following:

$$
\begin{aligned}
\forall m \, \forall n_1 \ldots n_k \, [\; &\exists y' \; \vartheta_{Ind}(\ulcorner \rho' \urcorner, \ulcorner \rho'' \urcorner, m, n_1 \ldots n_k, y') \,\, \& \\
&\forall y' \, (\; \vartheta_{Ind}(\ulcorner \rho' \urcorner, \ulcorner \rho'' \urcorner, m, n_1, \ldots, n_k, y') \\
&\qquad \supset \\
&\qquad DT(y') \,\, \& \\
&\qquad \text{endsequent of } y' \text{ is } \Gamma, \Delta \to F(\bar{m})(x_i/\bar{n}_i) \,\, \& \\
&\qquad \text{height of } y' \text{ is } < \; \Phi_{|F(x)|}(\max\{\alpha, \beta\} + m) \\
&\quad) \\
\quad] &
\end{aligned}
\qquad (9)
$$

To prove (9), we choose arbitrary natural numbers $n_1, \ldots, n_k$ and then we proceed by induction on m. The induction formula is Π_2, but the complexity of the whole proof is affected by the fact that we use the correctness of $\mathtt{Elim}$.

Assume that $m = 0$. The totality of ρ', which follows from $DT(z')$, gives us u such that $\rho'(n_1, \ldots, n_k, u)$. Since $\mathtt{Wk}$ is total, we obtain y' such that

75

$\vartheta_{Ind}(\ulcorner\rho'\urcorner,\ulcorner\rho''\urcorner,0,n_1,\ldots,n_k,y')$. On the other hand, assume now that we have y' with exactly this property. Since $DT(z')$, we also have $DT(u)$ for any u such that $\rho'(n_1,\ldots,n_k,u)$. The endsequent of u is $\Gamma \to F(0)(x_i/\bar{n}_i)$ and the height of u is strictly less than α. The correctness of $\mathtt{Wk}$ yields the following properties of y': $DT(y')$, the endsequent of y' is $\Gamma,\Delta \to F(0)(x_i/\bar{n}_i)$ and the height of y' is the same as the height of u.

Assume that $m \neq 0$. The induction hypothesis for $m-1$ gives us y'' such that $DT(y'')$. Its endsequent is $\Gamma,\Delta \to F(\overline{m-1})(x_i/\bar{n}_i)$ and the height is less than $\Phi_{|F(x)|}(\max\{\alpha,\beta\}+m-1)$. Formula ρ'' is total and this gives us v such that $\rho''(m-1,n_1,\ldots,n_k,v)$. $DT(z'')$ yields $DT(v)$. The endsequent of v is $F(\overline{m-1}),\Delta \to F(\bar{m})(x_i/\bar{n}_i)$ and its height is $\delta < \beta$. Furthermore, the correctness of $\mathtt{Elim}$ yields w such that $DT(w)$. The endsequent of w is $\Gamma,\Delta,\Delta \to F(\bar{m})(x_i/\bar{n}_i)$ and the height is at most $\Phi_{|F(x)|-1}(\Phi_{|F(x)|}(\max\{\alpha,\beta\}+m-1)+\delta)$ that is strictly less than $\Phi_{|F(x)|}(\max\{\alpha,\beta\}+m)$. Since $\mathtt{MultiCt}$ is total, we obtain y' such that $\vartheta_{Ind}(\ulcorner\rho'\urcorner,\ulcorner\rho''\urcorner,m,n_1,\ldots,n_k,y')$. On the other hand, if we take y' such that $\vartheta_{Ind}(\ulcorner\rho'\urcorner,\ulcorner\rho''\urcorner,m,n_1,\ldots,n_k,y')$, the correctness of $\mathtt{MultiCt}$ yields $DT(y')$. The endsequent of y' is $\Gamma,\Delta \to F(\bar{m})(x_i/\bar{n}_i)$ and its height is the same as the height of w, we mean, strictly less than $\Phi_{|F(x)|}(\max\{\alpha,\beta\}+m)$.

At the end, we want to state the following theorem that summarizes the result of this section.

Theorem 3. *Gentzen's consistency proof of 1935 can be formalized in $\mathrm{I\Pi}_3+\mathrm{TI}$ up to $\Phi_\omega(0)$. The induction formulas of the transfinite induction are Δ_3.*

We do not rule out that the analysis of Gentzen's cut elimination strategy may be done better, but this seems to be quite difficult.

Part 2

Comparison between Tait's and Gentzen's cut elimination strategy in classical propositional logic

Abstract The most problematic part of Gentzen's consistency proof of 1935 is Hilfssatz, the cut elimination theorem, that eliminates uppermost cuts regardless of the complexity. The analysis of the cut elimination strategy of Hilfssatz, which is described in the previous part, showed that Gentzen implicitly applied transfinite induction up to α, $\varepsilon_0 \leq \alpha \leq \Phi_\omega(0)$, in the consistency proof. It must be stressed that Gentzen himself does not speak about any transfinite induction in connection with this proof. We know that if he had applied Tait's cut elimination strategy, the one that decreases the cut-rank of the derivation, he would obtain transfinite induction up to ε_0.

In this part, we will deal with the question to what extent cut-free derivations differ when they are produced by distinct cut elimination strategies, particularly we are interested in Gentzen's strategy and Tait's strategy. We show that both strategies yield the same cut-free derivations in classical propositional logic. Hence, not only are the heights of cut-free derivations the same but also their structures.

Our proof applies an elimination algorithm of a single cut inspired by the method of Buss that makes global changes to the derivations. This algorithm is deterministic. A cut elimination strategy is a list of properties that a cut must have to be eliminated in a particular state. We will use only strategies that are nondeterministic in the sense that any cut with suitable properties can be chosen for elimination. We define a strategy, which we will call *general cut elimination strategy*, that includes both investigated strategies, the one of Gentzen and the one of Tait. We prove that general cut elimination strategy has the weak Church-Rosser property in classical propositional logic. It can be seen that it also has the strong normalization property. Weak Church-Rosser property and strong normalization yield the Church-Rosser property that ensures that normal forms, in our case cut-free propositional derivations, are given unambiguously.

2.1 Introduction

We have analysed Gentzen's consistency proof of 1935 in the previous text. The most problematic part of the proof is Hilfssatz that represents cut elimination in an infinitary calculus. Hilfssatz is remarkable because of the cut elimination strategy that Gentzen had applied there. He always eliminates an uppermost cut, i.e., a cut whose premises have cut-free derivations. The usual cut elimination strategy, known as Tait's strategy, is the one that eliminates one of the most complex cuts such that there are only simpler cuts above it. We say that Tait's strategy decreases the cut-rank, i.e., the complexity of cut formulas in the derivation. Our attempts to analyse Gentzen's cut elimination strategy in the infinitary calculus reveal that there may be differences between both strategies. Whereas the application of Gentzen's strategy in Hilfssatz gave us $\Phi_\omega(0)$ as an upper bound on heights of deduction trees for sequents derivable in PA, Tait's strategy yields ε_0. It is not ruled out that some more sophisticated analysis of Gentzen's strategy would give us ε_0 too. The question is whether and how both strategies differ, especially whether it is possible that Gentzen's strategy is really less efficient in the sense that it yields cut-free derivations whose heights cannot be bounded by iterated exponentiation of ω.

Since the infinitary calculus is more abstract than finite calculi in which the performance of both strategies was unknown to us either, we decided to study the cut elimination strategies in classical propositional and classical predicate logic. In this part, we switch to classical propositional logic and show that both strategies yield the same cut-free derivations. Not only is the height of the cut-free derivations the same but they also have the same structure. Our proof uses the fact that we are able to organize the cut elimination procedure in classical propositional logic in the way that there arise one or two simpler cuts instead of the eliminated cut. Moreover, they are situated exactly at the position of the original cut that was eliminated. The crucial point is that they are not distributed throughout the whole derivation and we know what is above and below them.

However, one cannot use the same approach in predicate logic because of eigenvariables that may occur in the derivation. We will explain this in detail after the elimination of a single cut is defined. The algorithm for elimination of a single cut, which we will use, is inspired by the method of Buss in ([3], pp. 37-40). It plays an important role because, for the most part, the proof is based on it.

Baaz and Leitsch ([1], pp. 93-104) deal with the comparison of Tait's

and Gentzen's cut elimination strategies in finite calculi too. In contrast to us, they investigate the number of elimination steps that are necessary to construct a cut-free derivation in predicate logic. They found out that none of the strategies is faster than the other. The strategies are incomparable in the sense that there exist derivations on which they differ significantly: one can find a sequence of derivations where Tait's strategy eliminates all cuts by taking a number of steps that is elementary in terms of the original derivation and, on the other hand, Gentzen's strategy needs a number of steps that is nonelementary. There exists an example too where this works the other way around.

It must be mentioned that Baaz and Leitsch use nondeterministic elimination of a single cut, i.e., there are more different treatments of one cut and they are allowed to select from them arbitrarily. This fact itself has influence on the length of the cut elimination procedure. We are of the opinion that when we want to compare the influence of cut elimination strategies on the cut elimination process, we have to suppress all other aspects that can also affect the result. That is why we insist on a deterministic elimination of a single cut. Besides, it is easy to see that a nondeterministic elimination of a single cut can lead to different cut-free derivations regardless of the cut elimination strategy.

Let us summarize main aspects that effect the cut elimination procedure. These aspects are contained in answers to the three following questions: (1) According to which property do we choose a cut that will be eliminated in the particular state? There may be, of course, more cuts with the suitable property in our derivation. This leads to the second question. (2) Which particular cut of those with the suitable property do we choose? Finally, after we have chosen a cut, we can proceed to the last question. (3) What is the algorithm for elimination of the chosen cut? The elimination of the chosen cut removes the cut and we possibly obtain new cuts with simpler cut formulas.

The first question is answered by the cut elimination strategy. We observe two strategies: Gentzen's strategy that chooses one of the uppermost cuts and Tait's strategy that chooses one of the most complex cuts such that there are only simpler cuts above them. As far as the second question is concerned, we postulate that both our strategies are nondeterministic, i.e., any random cut from cuts with the suitable property may be picked. This nondeterminism is a part of the strategy. The third question, which asks for an algorithm for elimination of a single cut, permits two solutions. We can either define a deterministic algorithm that allows only one way of elimina-

tion, or we can define a nondeterministic one that suggests more different treatments of the chosen cut. As stated above, we are interested only in differences caused by the application of distinct strategies. This convinced us to define a deterministic algorithm for elimination of a single cut and use it independently of the strategy.

2.2 Preliminaries

For the purpose of the following investigation, we allow sequents to have a multiset in the antecedent as well as in the succedent:

Definition 26. *A sequent is an expression of the form $\Gamma \to \Delta$ where $\Gamma = \{A_1, \ldots, A_n\}$ are antecedent formulas and $\Delta = \{B_1, \ldots, B_k\}$ are succedent formulas. We view Γ and Δ as multisets and they can be empty.*

Since we work in classical propositional logic, we do not need all logical operations. We will use only disjunction and negation. Hence, our language is $\{\vee, \neg\}$ and we define the following calculus:

Definition 27. *A derivation of $\Gamma \to \Delta$ in classical propositional logic is a tree that consists of sequents from Definition 26. Each sequent is either an initial sequent of the form $A \to A$ where A is a propositional variable or it is derived from previous sequents using one of the following inference rules:*

Structural rules: Logical rules:

$$\frac{\Sigma, \Sigma, \Gamma \to \Delta, \Theta, \Theta}{\Sigma, \Gamma \to \Delta, \Theta} \; Ct \qquad \frac{\Gamma \to \Delta, A}{\neg A, \Gamma \to \Delta} \; \neg L \qquad \frac{A, \Gamma \to \Delta}{\Gamma \to \Delta, \neg A} \; \neg R$$

$$\frac{\Gamma \to \Delta}{\Sigma, \Gamma \to \Delta, \Theta} \; Wk \qquad \frac{A, \Gamma \to \Delta \quad B, \Gamma \to \Delta}{A \vee B, \Gamma \to \Delta} \; \vee L \qquad \frac{\Gamma \to \Delta, A, B}{\Gamma \to \Delta, A \vee B} \; \vee R$$

$$\frac{\Gamma \to \Delta, A \quad A, \Gamma \to \Delta}{\Gamma \to \Delta} \; cut$$

The rules of Ct and Wk are called weak inference rules*. The rule of cut and logical inference rules are collectively called* strong inference rules*. Sequent $\Gamma \to \Delta$ is called the* endsequent *and it is situated in the root of the derivation.*

We do not need the rule of Exchange since we use multisets of formulas.

Definition 28. *Formula $\neg A$ or $A \vee B$ in the lower sequent of a logical inference rule is called* principal formula *of that rule. Formulas in Σ, Θ in the lower sequent of the weak inference rules are also called principal formulas. The rule of cut has no principal formula.*

Formulas A, B in the upper sequent of a logical inference rule are called auxiliary formulas *of that rule. Formulas in Σ, Θ and formula A in the upper sequent of the structural rules are also called auxiliary formulas. The rule of Wk has no auxiliary formulas.*

Formulas in multisets Γ, Δ are called side formulas.

Now, we will define the height of a derivation from Definition 27. In principle, the definition is the same as for an infinitary derivation. The only difference is that we used no weak inference rules in infinitary derivations and here, we have some. We postulate that weak inference rules do not increase the height of the derivation.

Definition 29. *The* height *of a derivation from Definition 27 is defined inductively as follows: Each initial sequent is assigned the value 0. If the premise of a unary strong inference rule has n, then the conclusion of this rule has $n+1$. If the premises of a binary strong inference rule have n and m, respectively, then the conclusion is assigned $\max\{n+1, m+1\}$. If the premise of a weak inference rule has n, then the conclusion of this rule has n again. The height of a derivation is the natural number of its endsequent.*

We write as $\vdash^n \Gamma \to \Delta$ to mean that we have a derivation with endsequent $\Gamma \to \Delta$ whose height is at most n.

Definition 30. *If an occurrence of C in the lower sequent of an inference rule is a side formula of this rule, then the corresponding occurrences of C in the upper sequent of this rule are called its* immediate ancestors. *If C belongs to principal formulas of Ct, then the corresponding occurrences of C among auxiliary formulas are called its immediate ancestors. If C is the principal formula of a logical inference rule, then the auxiliary formulas of this rule are called its immediate ancestors.*

The ancestor relation *is reflexive and transitive closure of the immediate ancestor relation.*

Definition 31. *Assume that we have a derivation P in the calculus from Definition 27. We choose an occurrence of formula B in P. A* thread *for this occurrence of formula B are all occurrences of B in P that are ancestors of the chosen occurrence of B. The chosen occurrence of B is called the* root *of the thread.*

Every thread has the form of a tree. It can branch either in the rules of contraction or in the rules with two premises. If B is a propositional variable, the leaves of every thread for B must be either in initial sequents or they

must be in conclusions of weakenings. Otherwise, if B is not a propositional variable, the leaves are in conclusions of weakenings or in conclusions of logical inference rules.

Since the notion of thread is crucial and there are different versions of its definition in the literature, we want to provide an example:

$$
Wk \, \frac{A \to \boxed{A}}{A \to \boxed{A}, C} \qquad \frac{C \to C}{C \to C, \boxed{A}} \, Wk
$$
$$
\frac{\boxed{A \vee C} \to \boxed{A}, C}{\boxed{A \vee C}, \neg A \to C} \, {\neg L}
$$
$$
\vee L
$$

Here, we have marked two different threads. The first one starts in the endsequent with formula $A \vee C$. Then, we go one inference rule up and we reach the leaf because $A \vee C$ is decomposed by the rule of $\vee L$.

The second example is a thread for the propositional variable A. It starts in the premise of the rule of $\neg L$ and it has two branches. The right hand side branch ends when A is inferred by the rule of weakening. The left hand side branch ends in the initial sequent $A \to A$.

Definition 32. *The* cut-rank *of a derivation P is defined as*

$$
sup \left\{ \, |C| + 1, \; C \text{ is a cut formula in } P \, \right\}
$$

where $|C|$ is the number of logical operations in formula C. The complexity of a cut *with cut formula C is $|C| + 1$.*

Definition 33. *(i) Assume that B is a formula and $k \geq 0$. We denote by B^k exactly k occurrences of B.*

(ii) Assume that Θ is a multiset of formulas, B is a formula and $k \geq 0$. Assume that Θ contains at least k occurrences of B. We denote by $\Theta^{-(B)^k}$ the multiset Θ from which k occurrences of B are deleted.

2.3 Elimination of a single cut

Now we introduce an algorithm to eliminate a single cut. The algorithm is inspired by the cut elimination procedure described in [3]. It makes global changes to the derivations: The thread for the cut formula is replaced by threads for some of its subformulas and these subformulas become then cut formulas of new cuts.

The elimination of a single cut will be defined by cases according to the outermost logical operation in the cut formula.

Definition 34. *Assume that k, h are natural numbers and we have a cut of the following form:*

$$\frac{\vdots \, Q \qquad \vdots \, R}{\Gamma \to \Delta, \neg B \qquad \neg B, \Gamma \to \Delta} \ cut$$
$$\frac{}{\Gamma \to \Delta}$$

Then, this cut is eliminated as follows:

Every sequent $\Theta \to \Lambda$ in Q where Λ contains exactly k occurrences of formula $\neg B$ that belong to the thread for the cut formula $\neg B$ is changed to $B^k, \Theta \to \Lambda^{-(\neg B)^k}$ where all k thread members $\neg B$ are removed from Λ.

The rules of $\neg R$ whose principal formulas $\neg B$ belong to the thread for the cut formula $\neg B$ become redundant:

$$\frac{B, \Theta \to \Lambda}{\Theta \to \Lambda, \neg B} \ \neg R \quad \rightsquigarrow \quad \frac{B^k, B, \Theta \to \Lambda^{-(\neg B)^k}}{B^{k+1}, \Theta \to \Lambda^{-(\neg B)^k}} \quad \rightsquigarrow \quad B^{k+1}, \Theta \to \Lambda^{-(\neg B)^k}$$

Every sequent $\Theta \to \Lambda$ in R where Θ contains exactly h occurrences of formula $\neg B$ that belong to the thread for the cut formula $\neg B$ is changed to $\Theta^{-(\neg B)^h} \to \Lambda, B^h$ where all h thread members $\neg B$ are removed from Θ.

The rules of $\neg L$ whose principal formulas $\neg B$ belong to the thread for the cut formula $\neg B$ become redundant:

$$\frac{\Theta \to \Lambda, B}{\neg B, \Theta \to \Lambda} \ \neg L \quad \rightsquigarrow \quad \frac{\Theta^{-(\neg B)^h} \to \Lambda, B, B^h}{\Theta^{-(\neg B)^h} \to \Lambda, B^{h+1}} \quad \rightsquigarrow \quad \Theta^{-(\neg B)^h} \to \Lambda, B^{h+1}$$

The rules of contraction and weakening in both derivations Q, R applied to formulas $\neg B$ that belong to the thread for the cut formula $\neg B$ change to contractions and weakenings applied to B. Initial sequents are not violated since they contain only propositional variables.

In this way, we obtain a derivation Q' of $B, \Gamma \to \Delta$ and a derivation R' of $\Gamma \to \Delta, B$. The original cut on $\neg B$ is replaced by the following cut with smaller complexity:

$$\frac{\vdots \, R' \qquad \vdots \, Q'}{\Gamma \to \Delta, B \qquad B, \Gamma \to \Delta} \ cut$$
$$\frac{}{\Gamma \to \Delta}$$

Definition 35. *Assume that k, h are natural numbers and we have a cut of the following form:*

$$\frac{\begin{array}{cc} \vdots Q & \vdots R \\ \Gamma \to \Delta, B \vee C & B \vee C, \Gamma \to \Delta \end{array}}{\Gamma \to \Delta} \ cut$$

Then, this cut is eliminated as follows:

Every sequent $\Theta \to \Lambda$ in Q where Λ contains exactly k occurrences of formula $B \vee C$ that belong to the thread for the cut formula $B \vee C$ is changed to $\Theta \to \Lambda^{-(B \vee C)^k}, B^k, C^k$ where all k thread members $B \vee C$ are removed from Λ.

The rules of $\vee R$ whose principal formulas $B \vee C$ belong to the thread for the cut formula $B \vee C$ become redundant:

$$\frac{\Theta \to \Lambda, B, C}{\Theta \to \Lambda, B \vee C} \vee R \quad \rightsquigarrow \quad \frac{\Theta \to \Lambda^{-(B \vee C)^k} B^k, B, C^k, C}{\Theta \to \Lambda^{-(B \vee C)^k}, B^{k+1}, C^{k+1}} \quad \rightsquigarrow \quad \Theta \to \Lambda^{-(B \vee C)^k}, B^{k+1}, C^{k+1}$$

We modify R in two ways. Every sequent $\Theta \to \Lambda$ in R where Θ contains exactly h occurrences of formula $B \vee C$ that belong to the thread for the cut formula $B \vee C$ is changed to $B^h, \Theta^{-(B \vee C)^h} \to \Lambda$ and $C^h, \Theta^{-(B \vee C)^h} \to \Lambda$, respectively, where all h thread members $B \vee C$ are removed from Θ.

The rules of $\vee L$ whose principal formulas $B \vee C$ belong to the thread for the cut formula $B \vee C$ become redundant together with the derivation of one of their premises:

(1) The right hand side premise of $\vee L$ whose principal formula $B \vee C$ belongs to the thread for the cut formula and the whole derivation of this premise become redundant during the construction of R', the derivation of $B, \Gamma \to \Delta$:

$$\frac{B, \Theta \to \Lambda \quad C, \Theta \to \Lambda}{B \vee C, \Theta \to \Lambda} \vee L \quad \rightsquigarrow \quad \frac{B, B^h, \Theta^{-(B \vee C)^h} \to \Lambda \quad C, B^h, \Theta^{-(B \vee C)^h} \to \Lambda}{B^{h+1}, \Theta^{-(B \vee C)^h} \to \Lambda} \quad \rightsquigarrow \quad B^{h+1}, \Theta^{-(B \vee C)^h} \to \Lambda$$

(2) The left hand side premise of $\vee L$ whose principal formula $B \vee C$ belongs to the thread for the cut formula and the whole derivation of this premise become redundant during the construction of R'', the derivation of $C, \Gamma \to \Delta$:

$$\frac{B, \Theta \to \Lambda \quad C, \Theta \to \Lambda}{B \vee C, \Theta \to \Lambda} \vee L \quad \rightsquigarrow \quad \frac{B, C^h, \Theta^{-(B \vee C)^h} \to \Lambda \quad C, C^h, \Theta^{-(B \vee C)^h} \to \Lambda}{C^{h+1}, \Theta^{-(B \vee C)^h} \to \Lambda} \quad \rightsquigarrow \quad C^{h+1}, \Theta^{-(B \vee C)^h} \to \Lambda$$

The rules of contraction and weakening in both derivations Q, R applied to formulas $B \vee C$ that belong to the thread for the cut formula $B \vee C$ change

to contractions and weakenings applied to B and C. Initial sequents are not violated since they contain only propositional variables.

In this way, we obtain three derivations: Q' of $\Gamma \to \Delta, B, C$, derivation R' of $B, \Gamma \to \Delta$ and derivation R'' of $C, \Gamma \to \Delta$. The original cut on $B \vee C$ is replaced by two simpler cuts:

$$
\cfrac{\cfrac{\raise2pt{\vdots}\,Q' \qquad \cfrac{\raise2pt{\vdots}\,R' \\ B, \Gamma \to \Delta}{B, \Gamma \to \Delta, C}\,Wk}{\Gamma \to \Delta, C} \qquad \raise2pt{\vdots}\,R'' \\ C, \Gamma \to \Delta}{\Gamma \to \Delta}
$$

Definition 36. *Assume that k is a natural number and we have a cut of the following form:*

$$
\cfrac{\raise2pt{\vdots}\,Q \qquad \raise2pt{\vdots}\,R \\ \Gamma \to \Delta, A \qquad A, \Gamma \to \Delta}{\Gamma \to \Delta}\,cut
$$

where A is a propositional variable. Then, this cut is eliminated as follows:

Every sequent $\Theta \to \Lambda$ in R where Θ contains exactly k occurrences of formula A that belong to the thread for the cut formula A is changed to $\Gamma, \Theta^{-(A)^k} \to \Lambda, \Delta$ where all k thread members A are removed from Θ. Multisets Γ, Δ are side formulas of the cut.

This modification violates the initial sequents, the endsequent of R and possibly weak inference rules that contain ancestors of the cut formula A among their principal formulas. This is fixed by the following adjustments:

Weak inference rules that contain ancestors of the cut formula A among their principal formulas either remain weak inference rules or, if multisets Σ, Π are empty, they become redundant:

$$
\cfrac{A^l, A^l, \Sigma, \Sigma, \Theta \to \Lambda, \Pi, \Pi}{A^l, \Sigma, \Theta \to \Lambda, \Pi}\,Ct \quad \rightsquigarrow \quad \cfrac{\Gamma, \Sigma, \Sigma, \Theta^{-(A)^k} \to \Lambda, \Pi, \Pi, \Delta}{\Gamma, \Sigma, \Theta^{-(A)^k} \to \Lambda, \Pi, \Delta}\,Ct
$$

$$
\cfrac{\Theta \to \Lambda}{A^l, \Sigma, \Theta \to \Lambda, \Pi}\,Wk \quad \rightsquigarrow \quad \cfrac{\Gamma, \Theta^{-(A)^k} \to \Lambda, \Delta}{\Gamma, \Sigma, \Theta^{-(A)^k} \to \Lambda, \Pi, \Delta}\,Wk
$$

The endsequent $A, \Gamma \to \Delta$ of R is changed to $\Gamma, \Gamma \to \Delta, \Delta$ and the required endsequent $\Gamma \to \Delta$ is obtained by contraction.

Initial sequents of the form $B \to B$ where B is different from A are changed to $\Gamma, B \to B, \Delta$ and this is derived from $B \to B$ by weakening.

Initial sequents of the form $A \to A$ where the antecedent formula A is an ancestor of the cut formula A are changed to $\Gamma \to \Delta, A$. This is the endsequent of Q.

The result of the elimination described above can be schematically represented as follows. The antecedent formula A in the displayed initial sequent $A \to A$ is an ancestor of the cut formula A:

$$
\begin{array}{c}
\begin{array}{c}
\dfrac{A \to A \qquad B \to B}{
\begin{array}{cc}
\vdots & \vdots
\end{array}
}
\\[2pt]
\vdots R \\
\dfrac{}{A, \Gamma \to \Delta}
\end{array}
\\
\dfrac{\Gamma \to \Delta, A \qquad \qquad A, \Gamma \to \Delta}{\Gamma \to \Delta} \; cut
\end{array}
\quad \rightsquigarrow \quad
\begin{array}{c}
\dfrac{\vdots Q}{\Gamma \to \Delta, A} \qquad \dfrac{B \to B}{\Gamma, B \to B, \Delta} \; Wk
\\
\vdots \\
\dfrac{\vdots}{\Gamma, \Gamma \to \Delta, \Delta}
\\
\dfrac{}{\Gamma \to \Delta} \; Ct
\end{array}
$$

This definition completes the sequence of definitions that deal with the elimination of a single cut.

We shall now explain why a similar elimination of a single cut is not possible in predicate logic. Assume that we have a cut of the form

$$
\dfrac{\Gamma \to \Delta, \exists x B(x) \qquad \exists x B(x), \Gamma \to \Delta}{\Gamma \to \Delta} \; cut
$$

and the rules of $\exists R$ and $\exists L$ have the following definition:

$$
\dfrac{\Gamma \to \Delta, B(t)}{\Gamma \to \Delta, \exists x B(x)} \; \exists R
\qquad \qquad
\dfrac{B(a), \Gamma \to \Delta}{\exists x B(x), \Gamma \to \Delta} \; \exists L
$$

Variable a is an eigenvariable and, therefore, it must not occur in the conclusion of the rule of $\exists L$. Formula $\exists x B(x)$ can be viewed as an abbreviation for a disjunction, possibly infinite, hence, the idea could be to try to eliminate the cut as if the cut formula was a disjunction. The first task is to find all formulas $B(t_1), \ldots, B(t_k)$ in Q that are auxiliary formulas of the rules of $\exists R$ whose principal formulas $\exists x B(x)$ belong to the thread for the cut formula $\exists x B(x)$. We wish to build derivation Q' of $\Gamma \to \Delta, B(t_1), \ldots, B(t_k)$ by transforming every sequent $\Theta \to \Lambda$ in Q into $\Theta \to \Lambda^{-(\exists x B(x))^n}, B(t_1)^n, \ldots, B(t_k)^n$ where Λ contains exactly n ancestors $\exists x B(x)$ of the cut formula $\exists x B(x)$. Now, the problem is that terms $t_1, \ldots, t_k$ may contain eigenvariables of some rules of $\exists L$ in Q. Assume that we have the following rule in Q:

$$
\dfrac{C(a), \Sigma \to \Pi}{\exists x\, C(x), \Sigma \to \Pi} \; \exists L
$$

Further assume that term t_1 contains a, the eigenvariable of this rule. We denote it by $t_1(a)$. Our intended transformation would clearly break the eigenvariable condition, according to which a must not occur in the conclusion of $\exists L$. Hence, we would *not* obtain a correct derivation:

$$\frac{C(a), \Sigma \to \Pi}{\exists x\, C(x), \Sigma \to \Pi}\, \exists L \quad \leadsto \quad \frac{C(a), \Sigma \to \Pi^{-(\exists x B(x))^n}, B(t_1(a))^n, \ldots, B(t_k)^n}{\exists x\, C(x), \Sigma \to \Pi^{-(\exists x B(x))^n}, B(t_1(a))^n, \ldots, B(t_k)^n}$$

Note that if we maintain the derivation in the *eigenvariable normal form*, i.e., all eigenvariables are distinct and occur only above the rules in which they serve as eigenvariables, there is no problem to transform R into derivations with endsequents $B(t_i), \Gamma \to \Delta$, $1 \leq i \leq k$, which would be the right hand side premises of cuts on formulas $B(t_1), \ldots, B(t_k)$.

The above example shows intuitively why the upper bound on heights of cut-free derivations is much bigger in predicate logic than in propositional logic. Cut formulas can be inferred at different positions in the derivation. In predicate logic, we cannot contract auxiliary formulas of the rules that infer different occurrences of the cut formula to a single occurrence and make this occurrence the cut formula of a new simpler cut. New simpler cuts must be introduced exactly at the positions where different occurrences of the cut formula are inferred. Thus, we may obtain any finite number of new cuts. This is in contrast to propositional logic where we are able to arrange the cut elimination in the way that we always obtain at most two simpler cuts.

Let us get back to the propositional calculus.

Lemma 39. *Assume that we have a derivation P whose last inference rule is a cut ϑ of the form:*

$$\frac{\vdots Q \qquad\qquad \vdots R}{\vdash^{n_1} \Gamma \to \Delta, B \qquad \vdash^{n_2} B, \Gamma \to \Delta}{\vdash^{n_i+1} \Gamma \to \Delta}\, \vartheta$$

where $\max\{n_1, n_2\} = n_i$. The elimination of ϑ according to Definitions 34,35,36 does not increase the cut-rank of P. Moreover:

- *If $|B| > 0$, then we obtain $\vdash^{n_i+2} \Gamma \to \Delta$ after the elimination of ϑ and ϑ is replaced by at most two simpler cuts above each other.*

- *If $|B| = 0$, then we obtain $\vdash^{2 \cdot n_i} \Gamma \to \Delta$ after the elimination of ϑ and ϑ disappears.*

Proof. (i) Assume that the cut formula is $\neg B$:

$$\frac{\vdots Q \qquad\qquad \vdots R}{\vdash^{n_1} \Gamma \to \Delta, \neg B \qquad \vdash^{n_2} \neg B, \Gamma \to \Delta}{\vdash^{n_i+1} \Gamma \to \Delta} \; \vartheta$$

The elimination replaces the thread for the cut formula $\neg B$ by a thread for a simpler formula B that is placed on the other side of the sequent arrow. In this way, we obtain derivation Q' of $B, \Gamma \to \Delta$ and derivation R' of $\Gamma \to \Delta, B$. The construction of Q' and R' reveals that their heights are bounded by the heights of Q and R, respectively. They also contain no new cuts. Hence we have:

$$\frac{\vdots R' \qquad\qquad \vdots Q'}{\vdash^{n_2} \Gamma \to \Delta, B \qquad \vdash^{n_1} B, \Gamma \to \Delta}{\vdash^{n_i+1} \Gamma \to \Delta} \; \text{cut}$$

(ii) Assume that the cut formula is $B \vee C$:

$$\frac{\vdots Q \qquad\qquad \vdots R}{\vdash^{n_1} \Gamma \to \Delta, B \vee C \qquad \vdash^{n_2} B \vee C, \Gamma \to \Delta}{\vdash^{n_i+1} \Gamma \to \Delta} \; \vartheta$$

Every formula $B \vee C$ in Q that is an ancestor of the cut formula $B \vee C$ is replaced by B, C during the elimination. Derivation R is modified in two different ways: The thread for the cut formula $B \vee C$ is first replaced by a thread for B and then by a thread for C. This procedure gives us derivation Q' of $\Gamma \to \Delta, B, C$, derivation R' of $B, \Gamma \to \Delta$ and derivation R'' of $C, \Gamma \to \Delta$. These three derivations are applied to build two simpler cuts:

$$\frac{\dfrac{\vdots Q' \qquad \dfrac{\vdots R'}{\dfrac{\vdash^{n_2} B, \Gamma \to \Delta}{\vdash^{n_2} B, \Gamma \to \Delta, C} \; Wk}}{\vdash^{n_1} \Gamma \to \Delta, B, C \qquad \vdash^{\max\{n_1+1, n_2+1\}=n_i+1} \Gamma \to \Delta, C} \; \text{cut} \qquad \dfrac{\vdots R''}{\vdash^{n_2} C, \Gamma \to \Delta}}{\vdash^{\max\{n_i+2, n_2+1\}=n_i+2} \Gamma \to \Delta} \; \text{cut}$$

No new cuts except for the displayed ones are built during the elimination. The height of the derivation after the elimination is at most n_i+2 where n_i+1 is the bound on the height of the original derivation before the elimination step.

(iii) Assume that the cut formula is a propositional variable A:

$$\frac{\begin{array}{cc} \dfrac{A \to A}{\vdots} & \dfrac{B \to B}{\vdots} \\ \end{array}}{\dfrac{\vdots Q \qquad\qquad \vdots R}{\dfrac{\vdash^{n_1} \Gamma \to \Delta, A \qquad \vdash^{n_2} A, \Gamma \to \Delta}{\vdash^{n_i+1} \Gamma \to \Delta}} \; \vartheta}$$

where the whole derivation of the right hand side premise is denoted by R and the antecedent formula A in the displayed initial sequent is an ancestor of the cut formula A. The elimination of ϑ consists in deleting the thread for the cut formula A in R and adjusting the initial sequents and the endsequent of R with the help of weak inference rules.

The result of the elimination is:

$$
\cfrac{
  \cfrac{
    \cfrac{\vdots \, Q}{\vdash^{n_1} \; \Gamma \to \Delta, A} \qquad \cfrac{B \to B}{\Gamma, B \to B, \Delta} \, Wk
  }{
    \vdots \qquad\qquad\qquad \vdots
  } \\[2pt]
  \cfrac{\vdots}{\cfrac{\Gamma, \Gamma \to \Delta, \Delta}{\Gamma \to \Delta} \, Ct}
}{}
$$

Now, we estimate the height of this derivation. The derivation is made of R where some initial sequents are replaced by a derivation whose height is at most n_1. Hence, the initial sequents in R obtain n_1 instead of 0 and it is easy to prove by induction on the height of R that the height of the resulted derivation is at most $n_1 + n_2 \leq 2 \cdot n_i$. $\qquad\square$

Lemma 40. *Assume that we have a derivation $\vdash^n \Gamma \to \Delta$ whose cut-rank is $d > 1$. Then, using Tait's cut elimination strategy and Definitions 34, 35, 36, we can build a derivation $\vdash^{2n} \Gamma \to \Delta$ whose cut-rank is $d-1$.*

Proof. Proceed by induction on the structure of $\vdash^n \Gamma \to \Delta$ and apply Lemma 39. $\qquad\square$

Lemma 41. *Assume that we have a derivation $\vdash^n \Gamma \to \Delta$ whose cut-rank is 1, i.e., it contains only cuts on propositional variables. Then, using Tait's cut elimination strategy and Definitions 34, 35, 36, we can build a derivation $\vdash^{2n} \Gamma \to \Delta$ whose cut-rank is 0, i.e., it contains no cuts.*

Proof. Proceed by induction on the structure of $\vdash^n \Gamma \to \Delta$ and apply Lemma 39. $\qquad\square$

Previous lemmas give us that we have an elementary upper bound for the height of cut-free derivations in classical propositional logic that are constructed with the help of Tait's strategy and the cut elimination method from Definitions 34, 35, 36. Namely, the bound is $2^{2^{d-1}n}$ where n is the height and d is the cut-rank of the original derivation with cuts.

The rest of this section deals with the circumstances under which the elimination of a single cut makes some sequents or subderivations of the original derivation redundant. We have four kinds of this redundancy:

(1) The premise and the conclusion of the rule whose principal formula belongs to the thread for the cut formula acquire the same form during the elimination. Since we do not use the rule of repetition, one of these sequents is redundant and we keep only one copy of it.

The second and the third case are relevant only to the elimination of a cut

$$
\begin{array}{c}
\vdots \\
\dfrac{\Sigma \to \Pi}{} \\[2pt]
\end{array}
$$

$$
\dfrac{\dfrac{\vdots\,\beta \qquad \vdots\,\gamma}{B,\Theta \to \Lambda \qquad C,\Theta \to \Lambda}}{B \vee C, \Theta \to \Lambda}\ \vee L
$$

$$
\dfrac{\Gamma \to \Delta, B\vee C \qquad\qquad B\vee C, \Gamma \to \Delta}{\Gamma \to \Delta}\ cut
$$

where the cut formula is a disjunction. The whole derivation of $B \vee C, \Gamma \to \Delta$ is denoted by R. The derivations of $B, \Theta \to \Lambda$ and $C, \Theta \to \Lambda$ are denoted by β and γ respectively. Assume that the principal formula of the displayed rule of $\vee L$ in R is an ancestor of the cut formula $B \vee C$ and there is no rule of $\vee L$ in R below this one whose principal formula $B \vee C$ is an ancestor of the cut formula too. Furthermore, we use the same notation as in Definition 35.

For the purposes of this explanation, we use sequent $\Sigma \to \Pi$ from β. The sequent is certainly missing in derivation R'' of $C, \Gamma \to \Delta$; we say that $\Sigma \to \Pi$ is not included in R''. As far as the derivation R' of $B, \Gamma \to \Delta$ is concerned, two possibilities may occur:

(2) Sequent $\Sigma \to \Pi$, to be more precise, its modified form $B^k, \Sigma^{-(B\vee C)^k} \to \Pi$, is included in R'. Since the derivation after the elimination step is built of R', among others, sequent $\Sigma \to \Pi$ does not get lost. By means of R', its modified form is part of the derivation after the elimination step. The redundancy in this particular case consists in being excluded from exactly one of the derivations R', R''.

(3) Sequent $\Sigma \to \Pi$ or any of its modified forms is not in R'. Hence, the elimination step makes this occurrence of $\Sigma \to \Pi$ disappear. It is completely excluded from the derivation after the elimination. Lemma 42 tells us when this is the case.

The fourth item is relevant only to the elimination of cuts on propositional variables. This case is rather unimportant for us, but we state it to make the enumeration complete. We use the notation of Definition 36:

(4) If the cut formula is a propositional variable A and there is no initial sequent $A \to A$ in R whose antecedent formula A is an ancestor of the cut formula A, then the whole derivation Q is *not* part of the derivation after the elimination step.

Lemma 42. *Assume that we have a derivation P whose last inference rule is a cut of the form*

$$
\dfrac{\overset{\vdots Q}{\Gamma \to \Delta, B_1 \vee B_2} \qquad \overset{\vdots R}{B_1 \vee B_2, \Gamma \to \Delta}}{\Gamma \to \Delta} \; \vartheta
$$

where the cut formula is a disjunction. Then, an occurrence of an arbitrary sequent $\Sigma \to \Pi$ from R disappears after the elimination step, i.e., this particular occurrence of the sequent in the original or modified form is not present in the derivation after the elimination step, if and only if R contains one of the following patterns:

or

where the principal formulas $B_1 \vee B_2$ of the displayed rules of $\vee L$ belong to the thread for the cut formula and there are no other inference rules of $\vee L$ that derive an ancestor $B_1 \vee B_2$ of the cut formula between the displayed rules of $\vee L$.

Proof. Assume that there is in R one of the above patterns, let us say the left one, and we want to eliminate cut ϑ. We need to transform R into derivations R' of $B_1, \Gamma \to \Delta$ and R'' of $B_2, \Gamma \to \Delta$. The modification of R to R' and R'' is achieved by replacing the thread for the cut formula $B_1 \vee B_2$ by threads for B_1 and B_2, respectively. Assume that we encounter sequent S_1 during this process. If we change $B_1 \vee B_2$ to B_2, the derivation of the left hand side premise is redundant and we do not include it into R''. If we change $B_1 \vee B_2$ to B_1, we continue doing this along the derivation of the left hand side premise until we encounter S_2. Now, we must choose the derivation of the left hand side premise again. Since $\Sigma \to \Pi$ does not get into R' nor R'',

91

it is not in P after the elimination of ϑ. The case with sequents S_3 and S_4 is similar.

On the other hand, assume that P does not contain an occurrence of $\Sigma \to \Pi$ after the elimination of ϑ. We want to show that this particular occurrence is a part of the structure as displayed above.

Since the transformation of Q does not delete any subderivation of P, sequent $\Sigma \to \Pi$ must belong to R. Moreover, it must be in the derivation of a premise of the rule of $\vee L$ that introduces an ancestor $B_1 \vee B_2$ of the cut formula. If there was only one rule of $\vee L$ in R that introduces an ancestor $B_1 \vee B_2$ of the cut formula, no sequent would disappear because the derivation of the left hand side premise of this rule ends up as a part of R' and, similarly, the derivation of the right hand side premise ends up in R''. Thus, there must be at least two such rules in R above each other.

When we transform R into derivations of $B_i, \Gamma \to \Delta$, $i = 1$ or $i = 2$, we replace the whole thread for the cut formula $B_1 \vee B_2$ by B_i and we always follow the path that contains B_i when we run across the rule of $\vee L$ that derives an ancestor $B_1 \vee B_2$ of the cut formula. This means that there must be at least two applications of the rule of $\vee L$ that derive an ancestor $B_1 \vee B_2$ of the cut formula under the occurrence of $\Sigma \to \Pi$. When we check the auxiliary formulas of these rules, we mean the auxiliary formulas that lie directly on the path leading to $\Sigma \to \Pi$, at least one of them must differ from the others. Hence, we are able to find the required pattern in R. $\qquad\square$

2.4 Propositional logic: Comparison of Tait's and Gentzen's cut elimination strategy

We have already introduced the cut elimination strategies of Tait and Gentzen, respectively:

Definition 37. Tait's elimination strategy *selects one of the most complex cuts such that there are only cuts of smaller complexity above it if any and this one is then eliminated according to Definitions 34, 35, 36.*

Definition 38. Gentzen's elimination strategy *selects an uppermost cut such that there are no other cuts above it and this one is then eliminated according to Definitions 34, 35, 36.*

The only difference between them is in the decision which cut to eliminate in a particular state. Each strategy is nondeterministic in the sense that it

selects an arbitrary cut from those that satisfy the required property for elimination.

In fact, we will examine a more general cut elimination strategy below that covers both strategies mentioned above. Baaz and Leitsch mention this strategy in ([1], p. 104); the author has also invented it independently of them. This strategy is called *general cut elimination*. It allows to choose an arbitrary cut such that there are only simpler cuts above it if any. Hence, the chosen cut does not have to be one of the most complex cuts in the derivation, but it may be, and it does not have to be an uppermost cut, but it may be. This strategy is also considered to be nondeterministic. When the algorithm has chosen the cut which is going to be eliminated, Definitions 34, 35, 36 are applied again.

Definition 39. General cut elimination strategy *selects an arbitrary cut such that there are only cuts of smaller complexity above it if any and this one is then eliminated according to Definitions 34, 35, 36.*

Since only cuts of smaller complexity than the eliminated one are created and reproduced during the elimination based on the general cut elimination strategy, general cut elimination strategy always terminates. This means that general cut elimination has the strong normalization property.

We will focus on the weak Church-Rosser property of general cut elimination. Objects that are going to be rewritten are the whole derivations and the only rewriting rules are the elimination steps described in Definitions 34, 35, 36. We aim to prove that general cut elimination has the weak Church-Rosser property that says that if we have a derivation and we apply two different elimination steps, we reach two different states that can be both further rewritten so that, in a finite number of steps, the derivations will be the same again. Strong normalization and the weak Church-Rosser property give us that the normal forms (cut-free derivations) are given unambiguously. This means that Tait's and Gentzen's cut elimination strategy, respectively, yield cut-free derivations not only of the same height, but also of the same form.

Theorem 4. *General cut elimination in the classical propositional logic calculus from Definition 27 has the weak Church-Rosser property when the cut elimination algorithm from Definitions 34, 35, 36 is applied.*

Proof. Assume that we have a derivation P. We choose two different cuts ϑ_1 and ϑ_2 from P such that there are only simpler cuts above each of them if any. If ϑ_1 and ϑ_2 are not above each other, we obtain the same derivation regardless of which one is eliminated first.

If they are above each other, cut ϑ_1 will always be above ϑ_2, we go on to prove the assertion by cases, based on the outermost logical operations in cut formulas of ϑ_1 and ϑ_2. Note that the cut formula of ϑ_1 must contain less logical operations than the cut formula of ϑ_2.

Furthermore, we distinguish two parts of P, the black one and the blue one, in every investigated case. Sequents in the black part are modified only when ϑ_2 is eliminated and the elimination of ϑ_1 does not modify them. On the other hand, sequents in the blue part are modified whenever ϑ_1 or ϑ_2 is eliminated.

Note that the modification of P during the elimination of ϑ_1 and ϑ_2 has two aspects: (i) First, the form of sequents is changed and, (ii) second, subderivations of P, possibly built of modified sequents, may be moved and interconnected again or they may become redundant. We are interested in the transformation of sequents in the blue part since these are the only sequents in P whose form is modified twice. The first time by the elimination of ϑ_1 and the second time by the elimination of ϑ_2 or the other way around. We will also pay attention to the transformation of the overall structure of P.

The notation for transformation of sequents is in Definition 33. The following scheme

$$S_i$$
$$\wr \vartheta$$
$$S_j$$

means that sequent S_i is transformed into sequent S_j after the elimination of cut ϑ. Similarly for derivations.

(1) Assume that the cut formula of ϑ_1 is a propositional variable A and the cut formula of ϑ_2 is $\neg B$:

Propositional variables A in the antecedents of the displayed initial sequents $A \rightarrow A$ are ancestors of the cut formula A. There can be any finite number of these initial sequents in the derivation. The derivation of the first premise of ϑ_2 is denoted by Q and the derivation of the second premise of ϑ_2 is denoted by R.

94

Let us focus on the case where ϑ_1 is in Q. The other case is treated in an analogous way. Assume that $\Theta \to \Lambda$ is an arbitrary sequent from the blue part. When we first eliminate ϑ_2, sequent $\Theta \to \Lambda$ is changed to $B^i, \Theta \to \Lambda^{-(\neg B)^i}$ where all i formulas $\neg B \in \Lambda$ that belong to the thread for the cut formula $\neg B$ are deleted. This means that the first premise of ϑ_1 is changed to $B^j, \Sigma \to A, \Pi^{-(\neg B)^j}$ and the second premise is changed to $B^j, A, \Sigma \to \Pi^{-(\neg B)^j}$ where Π contains exactly j occurrences of $\neg B$ that belong to the thread for the cut formula $\neg B$. Cut ϑ_1 modified in this way is denoted by ϑ_1'. Since the elimination of ϑ_2 makes Q', derivation Q after the elimination, the derivation of the right hand side premise of the simpler cut and, further, it makes R', derivation R after the elimination, the derivation of the left hand side premise of the simpler cut, cut ϑ_1, in the form of ϑ_1', is in the derivation of the right hand side premise of the simpler cut. The cut formula A of ϑ_1' is the same as the cut formula of ϑ_1. The elimination of ϑ_2 did not change the thread for A. Derivation P after the elimination of ϑ_2 has the form as displayed on the left hand side below:

The elimination of ϑ_2 is followed by the elimination of ϑ_1'. The whole thread for the cut formula A in the blue part is deleted and this violates all initial sequents $A \to A$ whose antecedent formulas are ancestors of the cut formula A. Since the side formulas of ϑ_1' are added to all sequents in the blue part, the violated initial sequents are transformed into the left hand side premise of ϑ_1' for which we have a derivation by assumption. If there are no initial sequents $A \to A$ in the blue part whose antecedent formulas are ancestors of the cut formula A, the left hand side premise of ϑ_1' and its whole derivation are redundant and excluded from the derivation after the elimination.

An arbitrary sequent $\Theta \to \Lambda$ from the blue part is modified in the following way:

$$\Theta \to \Lambda$$
$$\wr \vartheta_2$$
$$B^i, \Theta \to \Lambda^{-(\neg B)^i}$$
$$\wr \vartheta_1'$$
$$\Sigma, B^j, B^i, \Theta^{-(A)^n} \to \Lambda^{-(\neg B)^i}, \Pi^{-(\neg B)^j}$$

where $\Sigma, B^j, \Pi^{-(\neg B)^j}$ are side formulas of ϑ'_1 and multiset Θ contains exactly n occurrences of A that are ancestors of the cut formula A. Although i occurrences of B in the antecedent can be in fact i occurrences of A, these occurrences of B and the ancestors of the cut formula A belong to different threads, and thus, we are sure that formulas B^i are not removed. The thread for B's starts in the cut on B that we obtain after the elimination of ϑ_2, whereas the thread for the ancestors of the cut formula A starts in ϑ'_1.

The second possibility is to begin with the elimination of ϑ_1. The elimination of ϑ_1 deletes the same thread for the cut formula A as the elimination of ϑ'_1 and the same initial sequents $A \to A$ are changed. Only the side formulas of ϑ_1, which are add to all sequents in the blue part, differ from the side formulas of ϑ'_1. The modified initial sequents $A \to A$ are changed to the left hand side premise of ϑ_1 for which we have a derivation that is changed only once, namely, when ϑ_2 is eliminated. This derivation is attached to all modified initial sequents. Derivation P after the elimination of ϑ_1 has the form as displayed on the left hand side below:

$$
\begin{array}{cc}
\vdots & \cfrac{C \to C}{\Sigma, C \to C, \Pi}\,Wk \\
\cfrac{}{\Sigma \to \Pi, A} & \\
\vdots \qquad\qquad \vdots & \\
\vdots & \\
\cfrac{\Sigma, \Sigma \to \Pi, \Pi}{\Sigma \to \Pi}\,Ct & \\
\vdots \qquad\qquad \vdots\,R & \\
\cfrac{\Gamma \to \Delta, \neg B \qquad \neg B, \Gamma \to \Delta}{\Gamma \to \Delta}\,\vartheta_2 &
\end{array}
\qquad \overset{\vartheta_2}{\rightsquigarrow} \qquad
\begin{array}{cc}
\vdots & \cfrac{C \to C}{B^j, \Sigma, C \to C, \Pi^{-(\neg B)^j}}\,Wk \\
\cfrac{}{B^j, \Sigma \to A, \Pi^{-(\neg B)^j}} & \\
\vdots \qquad\qquad \vdots & \\
\vdots & \\
\cfrac{B^j, B^j, \Sigma, \Sigma \to \Pi^{-(\neg B)^j}, \Pi^{-(\neg B)^j}}{B^j, \Sigma \to \Pi^{-(\neg B)^j}}\,Ct & \\
\vdots\,R' \qquad\qquad \vdots & \\
\cfrac{\Gamma \to \Delta, B \qquad B, \Gamma \to \Delta}{\Gamma \to \Delta} &
\end{array}
$$

The elimination of ϑ_1 is followed by the elimination of ϑ_2. The elimination of ϑ_2 changes now the side formulas of ϑ_1 that we have added to all sequents in the blue part during the elimination of ϑ_1 from Σ, Π to $B^j, \Sigma, \Pi^{-(\neg B)^j}$. Furthermore, the derivation of the left hand side premise of ϑ_2 becomes the derivation of the right hand side premise of the new simpler cut and vice versa because the threads for the cut formulas $\neg B$ are replaced by threads for simper formulas B that are placed on the other side of the sequent arrow. Hence, the blue part is in the derivation of the right hand side premise of the cut on B again.

An arbitrary sequent $\Theta \to \Lambda$ form the blue part is modified in the following way which is exactly the same result as above:

$$\Theta \to \Lambda$$
$$\wr\, \vartheta_1$$
$$\Sigma, \Theta^{-(A)^n} \to \Lambda, \Pi$$
$$\wr\, \vartheta_2$$
$$\Sigma, B^j, B^i, \Theta^{-(A)^n} \to \Lambda^{-(\neg B)^i}, \Pi^{-(\neg B)^j}$$

(2) The case when the cut formula of ϑ_1 is a propositional variable A, the cut formula of ϑ_2 is $B \vee C$ and ϑ_1 is in the derivation of the left hand side premise of ϑ_2 is analogous to (1):

$$
\cfrac{
\cfrac{
\cfrac{
\cfrac{\cfrac{\overline{A \to A}}{\vdots} \quad \cfrac{\overline{D \to D}}{\vdots}}{\Sigma \to \Pi, A} \quad \cfrac{\vdots}{A, \Sigma \to \Pi}
}{\Sigma \to \Pi} \, \vartheta_1
}{
\cfrac{\vdots \, Q}{\Gamma \to \Delta, B \vee C} \qquad\qquad \cfrac{\vdots \, R}{B \vee C, \Gamma \to \Delta}
}
}{\Gamma \to \Delta} \, \vartheta_2
$$

Variable A in the antecedent of the displayed initial sequent $A \to A$ is an ancestor of the cut formula A.

(3) Assume that the cut formula of ϑ_1 is a propositional variable A and the cut formula of ϑ_2 is $B \vee C$. This time, cut ϑ_1 is in R:

$$
\cfrac{
\cfrac{\vdots \, Q}{\Gamma \to \Delta, B \vee C} \qquad
\cfrac{
\cfrac{
\cfrac{\cfrac{\overline{A \to A}}{\vdots} \quad \cfrac{\overline{D \to D}}{\vdots}}{\Sigma \to \Pi, A} \quad \cfrac{\vdots}{A, \Sigma \to \Pi}
}{\Sigma \to \Pi} \, \vartheta_1
}{
\cfrac{\vdots \, R}{B \vee C, \Gamma \to \Delta}
}
}{\Gamma \to \Delta} \, \vartheta_2
$$

Variable A in the antecedent of the displayed initial sequent $A \to A$ is an ancestor of the cut formula A. There can be any finite number of these initial sequents in the derivation. The derivation of the first premise of ϑ_2 is denoted by Q and the derivation of the second premise of ϑ_2 is denoted by R.

We already know that the elimination of ϑ_2 modifies R in two ways: First, it replaces formulas of the form $B \vee C$ that belong to the thread for the cut formula $B \vee C$ by B. Second, the same formulas are replaced by C. In this way, we obtain a derivation R' of $B, \Gamma \to \Delta$ and a derivation R'' of $C, \Gamma \to \Delta$. Some parts of R, including cut ϑ_1 itself, may be missing from R' or R'' or both.

We focus only on the modification of R that replaces $B \vee C$ by B, i.e., the construction of R'. The investigation of R'' is completely analogous and independent of R'. We want to show that if we first eliminate ϑ_2 and then possibly two cuts of the form of ϑ_1, one in R' and the second one in R'', we

obtain the same derivation as if we started with the elimination of ϑ_1 and then continued with the elimination of ϑ_2.

There may be the rules of $\vee L$ in R of the following form:

$$\frac{B, \Phi \to \Psi \qquad C, \Phi \to \Psi}{B \vee C, \Phi \to \Psi} \vee L \qquad\qquad (*)$$

whose principal formulas $B \vee C$ are ancestors of the cut formula $B \vee C$. When we say that a rule is of the form as described in $(*)$, we mean that it is the rule of $\vee L$ whose principal formula belongs to the thread for the cut formula $B \vee C$. All sequents in R that are in the derivation of the right hand side premise $C, \Phi \to \Psi$ of these rules are, by definition, not included in R'. We consider three cases:

(3a) There is a rule of $\vee L$ in R as described in $(*)$ such that cut ϑ_1 is in the derivation of its right hand side premise $C, \Phi \to \Psi$.

(3b) There is *no* rule of $\vee L$ in R as described in $(*)$ such that cut ϑ_1 is in the derivation of its right hand side premise $C, \Phi \to \Psi$. Assume that there are rules of $\vee L$ as described in $(*)$ in the blue part. We consider an arbitrary initial sequent $A \to A$ from the blue part whose antecedent formula A is an ancestor of the cut formula A such that it is in the derivation of $C, \Phi \to \Psi$, the right hand side premise of $\vee L$ as described in $(*)$.

(3c) There is *no* rule of $\vee L$ in R as described in $(*)$ such that cut ϑ_1 is in the derivation of its right hand side premise $C, \Phi \to \Psi$. We consider an arbitrary initial sequent $A \to A$ from the blue part whose antecedent formula A is an ancestor of the cut formula A such that it is *not* in the derivation of $C, \Phi \to \Psi$, the right hand side premise of $\vee L$ as described in $(*)$. The derivation of the left hand side premise $\Sigma \to \Pi, A$ of ϑ_1, however, may contain the rules of $\vee L$ as described in $(*)$. This case is analogous to (1) and (2).

(3a) Assume that cut ϑ_1 is in the derivation of $C, \Phi \to \Psi$, the right hand side premise of $\vee L$ as described in $(*)$. The situation is schematically represented as follows:

$$\frac{\Sigma \to \Pi, A \qquad A, \Sigma \to \Pi}{\Sigma \to \Pi} \vartheta_1$$

$$\frac{B, \Phi \to \Psi \qquad\qquad C, \Phi \to \Psi}{B \vee C, \Phi \to \Psi} \vee L$$

$$\frac{\Gamma \to \Delta, B \vee C \qquad B \vee C, \Gamma \to \Delta}{\Gamma \to \Delta} \vartheta_2$$

The elimination of ϑ_2 does not include the derivations of premises $C, \Phi \to \Psi$ as described in $(*)$ to R', hence, we obtain:

$$
\cfrac{
  \cfrac{
    \cfrac{
      \begin{array}{c} \vdots \\ B^{a+1}, \Phi^{-(B \vee C)^a} \to \Psi \end{array}
    }{}
  }{}
}{}
$$

$$
\begin{array}{ccc}
 & \vdots\, R' & \\
\begin{array}{c} \vdots\, Q' \\ \Gamma \to \Delta, B, C \end{array} & \cfrac{B, \Gamma \to \Delta}{B, \Gamma \to \Delta, C}\, Wk & \vdots\, R'' \\
\end{array}
$$

Multiset Φ contains exactly a occurrences of $B \vee C$ that are ancestors of the cut formula $B \vee C$. Since $C, B^a, \Phi^{-(B \vee C)^a} \to \Psi$, the modified right hand side premise of $\vee L$, and its whole derivation are redundant, cut ϑ_1 is excluded from R'.

Now, when we begin with the elimination of ϑ_1, we know that only the part above the conclusion $\Sigma \to \Pi$ of ϑ_1 is modified. Nevertheless, the modified part is still in the derivation of $C, \Phi \to \Psi$ as described in $(*)$. Since the subsequent elimination of ϑ_2 makes the derivation of $C, \Phi \to \Psi$ redundant in R', the modified part does not get into R' either. This is our desired result.

(3b) Assume that there is *no* rule of $\vee L$ in R as described in $(*)$ such that ϑ_1 is in the derivation of its right hand side premise $C, \Phi \to \Psi$. Assume that we have an initial sequent $A \to A$ in the blue part whose antecedent formula is an ancestor of the cut formula A. This initial sequent is in the derivation of $C, \Phi \to \Psi$ as described in $(*)$:

$$
\begin{array}{c}
\begin{array}{cc}
E \to E & A \to A \\
\vdots & \vdots \\
B, \Phi \to \Psi & C, \Phi \to \Psi \\
\end{array} \\
\cline{1-1}
B \vee C, \Phi \to \Psi \quad \vee L
\end{array}
$$

$$
\begin{array}{cc}
\vdots & \vdots \\
\Sigma \to \Pi, A & A, \Sigma \to \Pi \\
\cline{1-2}
\Sigma \to \Pi \quad \vartheta_1
\end{array}
$$

$$
\begin{array}{cc}
\vdots\, Q & \vdots\, R \\
\Gamma \to \Delta, B \vee C & B \vee C, \Gamma \to \Delta \\
\cline{1-2}
\Gamma \to \Delta \quad \vartheta_2
\end{array}
$$

We start with the elimination of ϑ_2. The elimination of ϑ_2 and the construction of R' change the left hand side premise of ϑ_1 from $\Sigma \to \Pi, A$ to $B^k, \Sigma^{-(B \vee C)^k} \to \Pi, A$ and the right hand side premise $A, \Sigma \to \Pi$ is changed to $A, B^k, \Sigma^{-(B \vee C)^k} \to \Pi$ where multiset Σ contains exactly k occurrences of $B \vee C$ that are ancestors of the cut formula $B \vee C$. Cut ϑ_1 after the elimination of ϑ_2 is denoted by ϑ_1'. We know that the derivations of sequents of

the form $C, \Phi \to \Pi$ as described in $(*)$ do not get into R', hence, the chosen initial sequent $A \to A$ does not get into R' either. The only modification of the thread for the cut formula A that the elimination of ϑ_2 has made is that some branches of the thread are shortened. No new members were added and all members that were not deleted during the elimination of ϑ_2 are still in the thread.

The subsequent elimination of ϑ'_1 attaches now the derivation of the left hand side premise of ϑ'_1 to all initial sequents $A \to A$ that were not excluded from R' and that are changed to $B^k, \Sigma^{-(B\vee C)^k} \to \Pi, A$ by the elimination of ϑ'_1. There is on the left hand side the derivation after the elimination of ϑ_2. The elimination of ϑ_2 is followed by the elimination of ϑ'_1:

$$
\begin{array}{ccc}
\left.
\begin{array}{c}
E \to E \\
\vdots \\
\dfrac{\quad}{B^{a+1}, \Phi^{-(B\vee C)^a} \to \Psi} \\
\end{array}
\right.
&
&
\left.
\begin{array}{c}
\dfrac{E \to E}{B^k, \Sigma^{-(B\vee C)^k}, E \to E, \Pi}\; Wk \\
\vdots \\
\dfrac{\quad}{B^k, \Sigma^{-(B\vee C)^k}, B^{a+1}, \Phi^{-(B\vee C)^a}, -(A)^h \to \Psi, \Pi}
\end{array}
\right.
\end{array}
$$

$$
\dfrac{B^k, \Sigma^{-(B\vee C)^k} \to \Pi, A \qquad A, B^k, \Sigma^{-(B\vee C)^k} \to \Pi}{B^k, \Sigma^{-(B\vee C)^k} \to \Pi}\; \vartheta'_1
\qquad \overset{\vartheta'_1}{\leadsto} \qquad
\dfrac{B^k, B^k, \Sigma^{-(B\vee C)^k}, \Sigma^{-(B\vee C)^k} \to \Pi, \Pi}{B^k, \Sigma^{-(B\vee C)^k} \to \Pi}\; Ct
$$

$$
\dfrac{\Gamma \to \Delta, B, C \quad \dfrac{B, \Gamma \to \Delta}{B, \Gamma \to \Delta, C}\,Wk}{\dfrac{\Gamma \to \Delta, C \qquad C, \Gamma \to \Delta}{\Gamma \to \Delta}}
$$

with the derivations Q', R', R''.

Multiset Φ contains exactly a occurrences of $B \vee C$ that belong to the thread for the cut formula $B \vee C$ and, furthermore, it contains exactly h occurrences of A that belong to the thread for the cut formula A. Note that cut formula A and cut formula B belong to different threads.

The second possibility is to start with the elimination of ϑ_1. The elimination of ϑ_1 attaches the derivation of the left hand side premise $\Sigma \to \Pi, A$ of ϑ_1 to all initial sequents $A \to A$ that are transformed into $\Sigma \to \Pi, A$ during the elimination of ϑ_1. This includes the investigated initial sequent $A \to A$ too. At the same time, the elimination of ϑ_1 does not affect the thread for the cut formula $B \vee C$ and no rules of $\vee L$ whose principal formulas belong to the thread for the cut formula $B \vee C$ are violated. The elimination of ϑ_1 modifies only the side formulas of inference rules except for the rules whose principal formula is an ancestor of the cut formula A. All rules of $\vee L$ from the blue

part are preserved at the positions where they were before the elimination of ϑ_1.

Now, we can continue with the elimination of ϑ_2. The derivation of $\Sigma, C, \Phi^{-(A)^h} \to \Psi, \Pi$, the right hand side premise $C, \Phi \to \Psi$ as described in $(*)$ modified after the elimination of ϑ_1, does not get into R'. This means that the modified sequent $A \to A$ and its derivation do not get into R' either. There is on the left hand side the derivation after the elimination of ϑ_1. The elimination of ϑ_1 is followed by the elimination of ϑ_2:

$$
\begin{array}{c}
\dfrac{\dfrac{E \to E}{\Sigma, E \to E, \Pi}\ Wk}{\vdots} \qquad \dfrac{\vdots}{\Sigma \to \Pi, A} \\[2ex]
\dfrac{\Sigma, B, \Phi^{-(A)^h} \to \Psi, \Pi \qquad \Sigma, C, \Phi^{-(A)^h} \to \Psi, \Pi}{\Sigma, B \vee C, \Phi^{-(A)^h} \to \Psi, \Pi}\ \vee L \\[2ex]
\vdots \\
\dfrac{\dfrac{\Sigma, \Sigma \to \Pi, \Pi}{\Sigma \to \Pi}\ Ct}{} \\
\vdots Q \qquad \vdots \\
\dfrac{\Gamma \to \Delta, B \vee C \qquad B \vee C, \Gamma \to \Delta}{\Gamma \to \Delta}\ \vartheta_2
\end{array}
\qquad \overset{\vartheta_2}{\rightsquigarrow} \qquad
\begin{array}{c}
\dfrac{E \to E}{B^k, \Sigma^{-(B \vee C)^k}, E \to E, \Pi}\ Wk \\[2ex]
\vdots \\
B^k, \Sigma^{-(B \vee C)^k}, B^{a+1}, \Phi^{-(A)^h}, -(B \vee C)^a \to \Psi, \Pi \\[2ex]
\vdots \\
\dfrac{B^k, B^k, \Sigma^{-(B \vee C)^k}, \Sigma^{-(B \vee C)^k} \to \Pi, \Pi}{B^k, \Sigma^{-(B \vee C)^k} \to \Pi}\ Ct \\[2ex]
\vdots \\
\vdots Q' \qquad \dfrac{B, \Gamma \to \Delta}{B, \Gamma \to \Delta, C}\ Wk \qquad \vdots \\
\dfrac{\Gamma \to \Delta, B, C \qquad \dfrac{\Gamma \to \Delta, C}{}\qquad C, \Gamma \to \Delta}{\Gamma \to \Delta}
\end{array}
$$

Let us now look at the form of sequents that are not made redundant during the elimination of ϑ_2. An arbitrary sequent $\Theta \to \Lambda$ from the blue part that is included in R' is transformed in the following way:

$$
\Theta \to \Lambda
$$
$$
\downdownarrows \vartheta_2
$$
$$
B^i, \Theta^{-(B \vee C)^i} \to \Lambda
$$
$$
\downdownarrows \vartheta_1'
$$
$$
B^i, \Theta^{-(B \vee C)^i, -(A)^j}, B^k, \Sigma^{-(B \vee C)^k} \to \Lambda, \Pi
$$

Multisets $\Sigma^{-(B \vee C)^k}$, B^k, Π are the side formulas of ϑ_1'. Multiset Θ contains exactly j occurrences of A that are ancestors of the cut formula A and i occurrences of $B \vee C$ that are ancestors of the cut formula $B \vee C$. The antecedent formulas B^i and the ancestors of the cut formula A belong to different threads. This is why the elimination of ϑ_1' cannot remove B^i from the antecedent, even if formulas A and B had the same form.

101

When we begin with the elimination of ϑ_1, the chosen arbitrary sequent $\Theta \to \Lambda$ is transformed in the following way:

$$\Theta \to \Lambda$$

$$\wr \vartheta_1$$

$$\Sigma, \Theta^{-(A)^j} \to \Lambda, \Pi$$

$$\wr \vartheta_2$$

$$B^i, \Theta^{-(A)^j, -(B\vee C)^i}, B^k, \Sigma^{-(B\vee C)^k} \to \Lambda, \Pi$$

Multisets Σ, Π are side formulas of ϑ_1. Since there is no difference whether we first remove j occurrences of A and then i occurrences of $B \vee C$ from Θ or the other way around, we obtain the required result.

(4) Assume that the cut formula of ϑ_1 is $D \vee E$ and the cut formula of ϑ_2 is $B \vee C$:

$$\cfrac{\cfrac{\begin{matrix}\vdots\ \alpha\\ \Sigma \to \Pi, D\vee E\end{matrix} \qquad \begin{matrix}\vdots\ \beta\\ D\vee E, \Sigma \to \Pi\end{matrix}}{\Sigma \to \Pi}\vartheta_1}{}$$

$$\cfrac{\begin{matrix}\vdots\ Q\\ \Gamma \to \Delta, B\vee C\end{matrix} \qquad \begin{matrix}\vdots\ R\\ B\vee C, \Gamma \to \Delta\end{matrix}}{\Gamma \to \Delta}\vartheta_2$$

We have $|D \vee E| < |B \vee C|$. The derivation of the first premise of ϑ_1 is denoted by α and the derivation of the second premise of ϑ_1 is denoted by β. Similarly, the derivation of the first premise of ϑ_2 is denoted by Q and the derivation of the second premise of ϑ_2 is denoted by R.

Neither the elimination of ϑ_2 nor the elimination of ϑ_1 can exclude any sequent S that belongs to α from the derivation after the elimination. There will always be in the derivation after the elimination some modified form of S. On the other hand, the elimination of ϑ_1 may make some sequents from β redundant because there may be the rules of $\vee L$ of the form

$$\cfrac{D, \Phi \to \Psi \qquad E, \Phi \to \Psi}{D \vee E, \Phi \to \Psi}\vee L \tag{$\diamond$}$$

in β whose principal formulas $D \vee E$ are ancestors of the cut formula $D \vee E$. When we say that a rule is of the form as described in ($\diamond$), we mean that it is the rule of $\vee L$ whose principal formula belongs to the thread for the cut formula $D \vee E$. The derivation can be schematically represented as follows:

$$\cfrac{\begin{matrix}\vdots\\ \Theta_1 \to \Lambda_1\end{matrix} \qquad \begin{matrix}\vdots\\ \Theta_2 \to \Lambda_2\end{matrix}}{}$$

$$\cfrac{\cfrac{\begin{matrix}\vdots\\ D, \Phi \to \Psi\end{matrix} \qquad \begin{matrix}\vdots\\ E, \Phi \to \Psi\end{matrix}}{D\vee E, \Phi \to \Psi}\vee L}{}$$

$$\cfrac{\cfrac{\begin{matrix}\vdots\ \alpha\\ \Sigma \to \Pi, D\vee E\end{matrix} \qquad \begin{matrix}\vdots\ \beta\\ D\vee E, \Sigma \to \Pi\end{matrix}}{\Sigma \to \Pi}\vartheta_1}{}$$

$$\cfrac{\begin{matrix}\vdots\ Q\\ \Gamma \to \Delta, B\vee C\end{matrix} \qquad \begin{matrix}\vdots\ R\\ B\vee C, \Gamma \to \Delta\end{matrix}}{\Gamma \to \Delta}\vartheta_2$$

Let us start with the elimination of ϑ_1. It builds three new derivations: the first one is α' of $\Sigma \to \Pi, D, E$, the second one is β' of $D, \Sigma \to \Pi$ and the third one is β'' of $E, \Sigma \to \Pi$. All sequents that are in the derivations of premises of the form of $E, \Phi \to \Psi$ as described in $(\diamond)$ are redundant in β' and, on the other hand, all sequents that are in the derivations of premises of the form of $D, \Phi \to \Psi$ as described in $(\diamond)$ are redundant in β''. The subsequent elimination of ϑ_2 does not exclude anything else from the blue part. It only modifies sequents that are still available after the elimination of ϑ_1. The derivation after the elimination of ϑ_1 is shown in the figure below. It is the part above the vertical arrow. The elimination of ϑ_1 is followed by the elimination of ϑ_2:

$$
\begin{array}{c}
\vdots \\
\dfrac{D^a, \Theta_1^{-(D\vee E)^a} \to \Lambda_1}{\vdots} \\
\dfrac{}{D^{j+1}, \Phi^{-(D\vee E)^j} \to \Psi} \quad \dfrac{E^b, \Theta_2^{-(D\vee E)^b} \to \Lambda_2}{\vdots} \\
\end{array}
$$

$$
\cfrac{
\cfrac{\Sigma \to \Pi, D, E}{\vdots\ \alpha'} \quad
\cfrac{\cfrac{D, \Sigma \to \Pi}{D, \Sigma \to \Pi, E}\ Wk}{\vdots\ \beta'} \quad
\cfrac{E^{j+1}, \Phi^{-(D\vee E)^j} \to \Psi}{\cfrac{E, \Sigma \to \Pi}{\vdots\ \beta''}}
}{
\cfrac{\Sigma \to \Pi, E}{\cfrac{\Sigma \to \Pi}{
\cfrac{\vdots}{\cfrac{\Gamma \to \Delta, B\vee C \quad \cfrac{B\vee C, \Gamma \to \Delta}{\vdots\ R}}{\Gamma \to \Delta}}\ \vartheta_2
}}
}
$$

$$
\left\downarrow \vartheta_2\right.
$$

$$
\begin{array}{c}
\vdots \\
\dfrac{D^a, \Theta_1^{-(D\vee E)^a} \to \Lambda_1^{-(B\vee C)^c}, B^c, C^c}{\vdots} \\
D^{j+1}, \Phi^{-(D\vee E)^j} \to \Psi^{-(B\vee C)^i}, B^i, C^i
\end{array}
$$

$$
\cfrac{
\cfrac{\Sigma \to \Pi^{-(B\vee C)^h}, B^h, C^h, D, E}{\vdots} \quad
\cfrac{\cfrac{D, \Sigma \to \Pi^{-(B\vee C)^h}, B^h, C^h}{D, \Sigma \to \Pi^{-(B\vee C)^h}, B^h, C^h, E}\ Wk}{\vdots} \quad
\cfrac{E^{j+1}, \Phi^{-(D\vee E)^j} \to \Psi^{-(B\vee C)^i}, B^i, C^i}{\cfrac{\vdots}{E, \Sigma \to \Pi^{-(B\vee C)^h}, B^h, C^h}}
}{
\cfrac{\Sigma \to \Pi^{-(B\vee C)^h}, B^h, C^h, E}{
\cfrac{\Sigma \to \Pi^{-(B\vee C)^h}, B^h, C^h}{
\cfrac{\cfrac{\vdots}{\Gamma \to \Delta, B, C} \quad \cfrac{\cfrac{B, \Gamma \to \Delta}{B, \Gamma \to \Delta, C}\ Wk \quad \cfrac{\vdots\ R''}{C, \Gamma \to \Delta}}{\cfrac{\Gamma \to \Delta, C}{\Gamma \to \Delta}}}{\Gamma \to \Delta}
}}
}
$$

All ancestors of the cut formulas in the multisets in the figure above are processed according to Definition 35.

The second possibility is to begin with the elimination of ϑ_2. The elimination of ϑ_2 changes derivation Q of $\Gamma \rightarrow \Delta, B \vee C$ to derivation Q' of $\Gamma \rightarrow \Delta, B, C$. All modified versions of sequents from the blue part are included in Q'. Cut ϑ_1 is transformed into ϑ_1' whose left hand side premise is $\Sigma \rightarrow \Pi^{-(B \vee C)^h}, B^h, C^h, D \vee E$ and the right hand side premise is $D \vee E, \Sigma \rightarrow \Pi^{-(B \vee C)^h}, B^h, C^h$. Multiset Π contains exactly h occurrences of $B \vee C$ that belong to the thread for the cut formula $B \vee C$. Only the thread for the cut formula $B \vee C$ is replaced by the elimination of ϑ_2. This means that principal formulas of only such inference rules are changed whose principal formulas belonged to the thread for the cut formula $B \vee C$ before the elimination. The other rules remain essentially the same because only their side formulas are modified - some formulas are added to their side formulas and members of the thread for the cut formula $B \vee C$ are deleted. The derivation of the left hand side premise of ϑ_1' is denoted by α_1 and the derivation of the right hand side premise of ϑ_1' is denoted by β_1. The derivation after the elimination of ϑ_2 has the following form:

$$
\dfrac{
  \dfrac{
    \dfrac{
      \begin{array}{cc}
        \vdots & \vdots \\
        \Theta_1 \rightarrow \Lambda_1^{-(B \vee C)^c}, B^c, C^c & \Theta_2 \rightarrow \Lambda_2^{-(B \vee C)^d}, B^d, C^d \\
        \vdots & \vdots \\
        D, \Phi \rightarrow \Psi^{-(B \vee C)^i}, B^i, C^i \quad & \quad E, \Phi \rightarrow \Psi^{-(B \vee C)^i}, B^i, C^i
      \end{array}
    }{D \vee E, \Phi \rightarrow \Psi^{-(B \vee C)^i}, B^i, C^i} \vee L
  }{}
}{}
$$

The subsequent elimination of ϑ_1' makes derivations of the premises of the form of $E, \Phi \rightarrow \Psi$ as described in ($\diamond$) redundant in β_1', the derivation of $D, \Sigma \rightarrow \Pi^{-(B \vee C)^h}, B^h, C^h$. Similarly for β_1'', the derivation of $E, \Sigma \rightarrow \Pi^{-(B \vee C)^h}, B^h, C^h$, where derivations of the premises of the form of $D, \Phi \rightarrow \Psi$ as described in ($\diamond$) are redundant. Since the elimination of ϑ_2 did not affect the thread for the cut formula $D \vee E$, the modified versions of the rules of $\vee L$ that were in P before any elimination and whose principal formulas $D \vee E$ are ancestors of the cut formula $D \vee E$ fulfill this condition. The subsequent elimination of ϑ_1' yields:

$$
D^a, \Theta_1^{-(D\vee E)^a} \to \Lambda_1^{-(B\vee C)^c}, B^c, C^c
$$
$$
\vdots
$$
$$
D^{j+1}, \Phi^{-(D\vee E)^j} \to \Psi^{-(B\vee C)^i}, B^i, C^i
$$
$$
\vdots\, \beta'_1
$$
$$
\dfrac{D, \Sigma \to \Pi^{-(B\vee C)^h}, B^h, C^h}{D, \Sigma \to \Pi^{-(B\vee C)^h}, B^h, C^h, E}\ Wk
$$

$$
\vdots\, \alpha'_1
$$
$$
\Sigma \to \Pi^{-(B\vee C)^h}, B^h, C^h, D, E
$$
$$
\Sigma \to \Pi^{-(B\vee C)^h}, B^h, C^h, E
$$
$$
\Sigma \to \Pi^{-(B\vee C)^h}, B^h, C^h
$$

$$
E^b, \Theta_2^{-(D\vee E)^b} \to \Lambda_2^{-(B\vee C)^d}, B^d, C^d
$$
$$
\vdots
$$
$$
E^{j+1}, \Phi^{-(D\vee E)^j} \to \Psi^{-(B\vee C)^i}, B^i, C^i
$$
$$
\vdots\, \beta''_1
$$
$$
E, \Sigma \to \Pi^{-(B\vee C)^h}, B^h, C^h
$$

$$
\dfrac{\Gamma \to \Delta, B, C}{\Gamma \to \Delta, C} \qquad
\vdots\, R' \quad
\dfrac{B, \Gamma \to \Delta}{B, \Gamma \to \Delta, C}\ Wk \qquad
\vdots\, R'' \quad
C, \Gamma \to \Delta
$$
$$
\Gamma \to \Delta
$$

Let us take an arbitrary sequent $\Theta \to \Lambda$ from α. The elimination of ϑ_1 and ϑ_2 changes the form of the sequent in the following way:

$$
\Theta \to \Lambda
$$
$$
\wr\, \vartheta_2
$$
$$
\Theta \to \Lambda^{-(B\vee C)^k}, B^k, C^k
$$
$$
\wr\, \vartheta'_1
$$
$$
\Theta \to \Lambda^{-(B\vee C)^k, -(D\vee E)^l}, B^k, C^k, D^l, E^l
$$

$$
\Theta \to \Lambda
$$
$$
\wr\, \vartheta_1
$$
$$
\Theta \to \Lambda^{-(D\vee E)^l}, D^l, E^l
$$
$$
\wr\, \vartheta_2
$$
$$
\Theta \to \Lambda^{-(D\vee E)^l, -(B\vee C)^k}, D^l, E^l, B^k, C^k
$$

Multiset Λ contains exactly k occurrences of $B \vee C$ that belong to the thread for the cut formula $B \vee C$ and it also contains exactly l occurrences of $D \vee E$ that belong to the thread for the cut formula $D \vee E$. Succedent occurrences of B, C are not removed during the elimination of ϑ'_1 because they belong to threads different from the one for $D \vee E$. Since $|D|, |E| < |D \vee E| < |B \vee C|$, the succedent occurrences of D, E are not removed during the elimination of ϑ_2. Hence, the sequents in α acquire the same form regardless of the order of elimination.

Let us now deal with the form of an arbitrary sequent $\Theta \to \Lambda$ in β that is not excluded from β', the derivation of $D, \Sigma \to \Pi$:

$$\Theta \to \Lambda$$

$$\wr\, \vartheta_1$$

$$D^u, \Theta^{-(D\vee E)^u} \to \Lambda$$

$$\wr\, \vartheta_2$$

$$D^u, \Theta^{-(D\vee E)^u} \to \Lambda^{-(B\vee C)^v}, B^v, C^v$$

$$\Theta \to \Lambda$$

$$\wr\, \vartheta_2$$

$$\Theta \to \Lambda^{-(B\vee C)^v}, B^v, C^v$$

$$\wr\, \vartheta_1'$$

$$D^u, \Theta^{-(D\vee E)^u} \to \Lambda^{-(B\vee C)^v}, B^v, C^v$$

Multiset Θ contains exactly u occurrences of $D\vee E$ that belong to the thread for the cut formula $D \vee E$ and multiset Λ contains exactly v occurrences of $B \vee C$ that belong to the thread for the cut formula $B \vee C$. It is sufficient to investigate only sequents in β' since β'' is completely analogous.

(5) Assume that the cut formula of ϑ_1 is $D \vee E$ and the cut formula of ϑ_2 is $B \vee C$:

$$
\cfrac{
  \cfrac{\vdots\, \gamma \qquad \vdots\, \delta}{}
}{}
$$

$$
\begin{array}{c}
\dfrac{\vdots\, Q}{\Gamma \to \Delta, B\vee C} \qquad
\dfrac{\begin{array}{c}\dfrac{\Sigma \to \Pi, D\vee E \qquad D\vee E, \Sigma \to \Pi}{\Sigma \to \Pi}\,\vartheta_1 \\ \vdots\, R \\ B\vee C, \Gamma \to \Delta\end{array}}{} \\[2pt]
\dfrac{}{\Gamma \to \Delta}\,\vartheta_2
\end{array}
$$

We have $|D \vee E| < |B \vee C|$. The derivation of the left hand side premise of ϑ_2 is denoted by Q and the derivation of the right hand side premise of ϑ_2 is denoted by R. The derivations of the premises of ϑ_1 are γ, δ respectively. Cut ϑ_1 is in R.

In comparison to the other cases, the elimination of each cut ϑ_1, ϑ_2 is now able to make some sequents redundant. The elimination of ϑ_1 makes such sequents redundant in δ', the derivation of $D, \Sigma \to \Pi$, that are in the derivations of the right hand side premises $E, \Omega \to \Upsilon$ of the rules of $\vee L$ in δ. The rules of $\vee L$ must have the form

$$\frac{D, \Omega \to \Upsilon \qquad E, \Omega \to \Upsilon}{D \vee E, \Omega \to \Upsilon}\,\vee L \qquad\qquad (\dagger)$$

and their principal formulas $D \vee E$ are ancestors of the cut formula $D \vee E$. This is analogous for δ'', the derivation of $E, \Sigma \to \Pi$. The elimination of ϑ_1 makes such sequents redundant in δ'' that are in the derivations of the left hand side premises $D, \Omega \to \Upsilon$ of the rules of $\vee L$ in δ whose principal formulas $D \vee E$ are ancestors of the cut formula $D \vee E$. When we say that a rule is of the form as described in ($\dagger$), we mean that it is the rule of $\vee L$ whose principal formula belongs to the thread for the cut formula $D \vee E$.

106

Furthermore, there may be in the black part as well as in the blue part of R the rules of $\vee L$ of the form

$$\frac{B, \Phi \to \Psi \qquad C, \Phi \to \Psi}{B \vee C, \Phi \to \Psi} \vee L \qquad\qquad (\triangleleft)$$

whose principal formulas $B \vee C$ are ancestors of the cut formula $B \vee C$. Again, when we say that a rule is of the form as described in ($\triangleleft$), we mean that it is the rule of $\vee L$ whose principal formula belongs to the thread for the cut formula $B \vee C$.

The elimination of ϑ_2 makes such sequents redundant in R', the derivation of $B, \Gamma \to \Delta$, that are in the derivations of the right hand side premises $C, \Phi \to \Psi$ of the rules of $\vee L$ as described in ($\triangleleft$). On the other hand, the sequents from the derivations of the left hand side premises $B, \Phi \to \Psi$ of the rules of $\vee L$ as described in ($\triangleleft$) are redundant in R'', the derivation of $C, \Gamma \to \Delta$.

Now, the rules of $\vee L$ of both kinds (†) and ($\triangleleft$) may be above each other. We want to argue that, regardless of the order of elimination, the same sequents are made redundant. Let us focus only on the constructions of δ' and R'. The constructions of δ'' and R'' are treated in an analogous way.

(5a) Assume that cut ϑ_1 is in the derivation of $C, \Phi \to \Psi$, the right hand side premise of $\vee L$ as described in ($\triangleleft$). The situation is schematically represented as follows:

$$
\cfrac{
  \cfrac{
    \overset{\vdots\ \gamma}{\Sigma \to \Pi, D \vee E} \qquad \overset{\vdots\ \delta}{D \vee E, \Sigma \to \Pi}
  }{\Sigma \to \Pi} \vartheta_1
}{}
$$

$$
\cfrac{
  \cfrac{
    \overset{\vdots}{B, \Phi \to \Psi} \qquad \overset{\vdots}{C, \Phi \to \Psi}
  }{B \vee C, \Phi \to \Psi} \vee L
}{}
$$

$$
\cfrac{
  \overset{\vdots\ Q}{\Gamma \to \Delta, B \vee C} \qquad \overset{\vdots\ R}{B \vee C, \Gamma \to \Delta}
}{\Gamma \to \Delta} \vartheta_2
$$

The elimination of ϑ_2 does not include the derivations of the premises of the form of $C, \Phi \to \Psi$ as described in ($\triangleleft$) in R' and, hence, cut ϑ_1 or any of its modified forms do not get into R'.

When we begin with the elimination of ϑ_1, only derivations γ, δ are modified. Since the black part is not affected by the elimination of ϑ_1, the simpler cuts that we obtain after the elimination of ϑ_1 are still in the derivation of $C, \Phi \to \Psi$ as described in ($\triangleleft$). The subsequent elimination of ϑ_2 makes the derivations of the premises of the form of $C, \Phi \to \Psi$ as described in ($\triangleleft$), including the blue part, redundant in R' again.

(5b) Assume that there is *no* rule of $\vee L$ in R as described in ($\triangleleft$) such that ϑ_1 is in the derivation of its right hand side premise $C, \Phi \rightarrow \Psi$. This means that a modified form of ϑ_1 is included in $R^{'}$ after the elimination of ϑ_2. However, there may be some rules of $\vee L$ that derive an ancestor $B \vee C$ of the cut formula $B \vee C$ in γ, δ. We can schematically represent it as follows:

$$
\cfrac{
\cfrac{
\cfrac{\vdots \qquad\quad \vdots}{\cfrac{B, \Phi \rightarrow \Psi \qquad C, \Phi \rightarrow \Psi}{B \vee C, \Phi \rightarrow \Psi} \, \vee L}
\quad \vdots \gamma
}{\Sigma \rightarrow \Pi, D \vee E}
\quad
\cfrac{
\cfrac{\vdots \qquad\quad \vdots}{\cfrac{B, \Omega \rightarrow \Upsilon \qquad C, \Omega \rightarrow \Upsilon}{B \vee C, \Omega \rightarrow \Upsilon} \, \vee L}
\quad \vdots \delta
}{D \vee E, \Sigma \rightarrow \Pi} \, \vartheta_1
}{
\cfrac{\cfrac{\vdots Q}{\Gamma \rightarrow \Delta, B \vee C} \qquad \cfrac{\vdots R}{B \vee C, \Gamma \rightarrow \Delta} \, \vartheta_2}{\Gamma \rightarrow \Delta}
}
$$

The most interesting case is when we take a branch in R such that it contains $D \vee E, \Sigma \rightarrow \Pi$, the right hand side premise of ϑ_1, and some rules of $\vee L$ of the form of (†) and ($\triangleleft$) at the same time. Let us choose such a branch and let us fix the rule of the form of ($\triangleleft$) such that it is at the lowest position among the rules of the form of ($\triangleleft$) in the chosen branch. We will denote the rule by $\vee L_{\vartheta_2}$.

Furthermore, assume that there is a rule of the form of (†) in the chosen branch that is below $\vee L_{\vartheta_2}$. Let us denote it by $\vee L_{\vartheta_1}$. Assume that $\vee L_{\vartheta_2}$ is in the derivation of the right hand side premise of $\vee L_{\vartheta_1}$. Other layouts, however, are treated in an analogous way. We can schematically represent it as follows:

$$
\cfrac{
\cfrac{
\cfrac{\vdots \qquad\quad \vdots}{\cfrac{B, \Phi \rightarrow \Psi \qquad C, \Phi \rightarrow \Psi}{B \vee C, \Phi \rightarrow \Psi} \, \vee L}
\quad \vdots \gamma
}{\Sigma \rightarrow \Pi, D \vee E}
\quad
\cfrac{
\cfrac{D, \Phi_1 \rightarrow \Psi_1 \qquad \cfrac{\cfrac{\vdots \quad \vdots}{\cfrac{B, \Phi_2 \rightarrow \Psi_2 \quad C, \Phi_2 \rightarrow \Psi_2}{B \vee C, \Phi_2 \rightarrow \Psi_2} \vee L_{\vartheta_2}} \quad E, \Phi_1 \rightarrow \Psi_1}{D \vee E, \Phi_1 \rightarrow \Psi_1} \vee L_{\vartheta_1}
\quad \vdots \delta
}{D \vee E, \Sigma \rightarrow \Pi} \, \vartheta_1
}{
\cfrac{\cfrac{\vdots Q}{\Gamma \rightarrow \Delta, B \vee C} \qquad \cfrac{\vdots R}{B \vee C, \Gamma \rightarrow \Delta} \, \vartheta_2}{\Gamma \rightarrow \Delta}
}
$$

Let us begin with the elimination of ϑ_1. The construction of $\delta^{'}$ replaces the thread for the cut formula $D \vee E$ by the thread for D. This process always follows the premise of the rule of $\vee L$ of the form as described in (†) that

108

contains the auxiliary formula D. The other premise and its derivation are excluded from δ'. This means that $\vee L_{\vartheta_2}$ does not get into δ' either. There may be in δ some other rules of $\vee L$ as described in ($\triangleleft$) whose modified variants, i.e., side formulas have been changed but the principal formulas remained untouched, are included in δ'. The subsequent elimination of ϑ_2 makes derivations of their premises of the form of $C, \Phi \rightarrow \Psi$ redundant in the derivation of $B, \Gamma \rightarrow \Delta$.

The derivation after the elimination of ϑ_1 is shown in the figure below. It is the part above the vertical arrow. The elimination of ϑ_1 is followed by the elimination of ϑ_2:

$$
\cfrac{
\cfrac{
\cfrac{
\begin{array}{cc}
\vdots & \vdots \\
B, \Phi \rightarrow \Psi^{-(D \vee E)^i}, D^i, E^i & C, \Phi \rightarrow \Psi^{-(D \vee E)^i}, D^i, E^i
\end{array}
}{B \vee C, \Phi \rightarrow \Psi^{-(D \vee E)^i}, D^i, E^i} \; \vee L
}{
\cfrac{\vdots\; \gamma'}{\Sigma \rightarrow \Pi, D, E}
}
\quad
\cfrac{
\cfrac{\vdots\; \delta'}{\Sigma \rightarrow \Pi, E}
}{ }
}{ }
$$

Multisets Ψ and Φ_1 contain exactly i, a occurrences of $D \vee E$ that belong to the thread for the cut formula $D \vee E$, respectively. Multisets Σ, Φ, Φ_1 contain exactly k, j, b occurrences of $B \vee C$ that belong to the thread for the cut formula $B \vee C$, respectively.

Let us now begin with the elimination of ϑ_2. The construction of R' transforms cut ϑ_1 into ϑ_1' whose left hand side premise is $B^k, \Sigma^{-(B \vee C)^k} \to \Pi, D \vee E$ and the right hand side premise is $D \vee E, B^k, \Sigma^{-(B \vee C)^k} \to \Pi$. The construction of R' also changes the side formulas of all inference rules whose principal formulas do *not* belong to the thread for the cut formula $B \vee C$. The thread for the cut formula $D \vee E$ is not affected by the construction of R' and the rule of $\vee L_{\vartheta_1'}$, rule $\vee L_{\vartheta_1}$ after the elimination of ϑ_2, has a principal formula $D \vee E$ that is still an ancestor of the cut formula $D \vee E$. The derivation after the elimination of ϑ_2 has the following form:

$$
\cfrac{
\cfrac{\vdots}{B^{j+1}, \Phi^{-(B \vee C)^j} \to \Psi}
\qquad
\cfrac{
\cfrac{
\cfrac{\vdots}{B^k, \Sigma^{-(B \vee C)^k} \to \Pi, D \vee E}
\qquad
\cfrac{
\cfrac{
\cfrac{\vdots}{D, B^b, \Phi_1^{-(B \vee C)^b} \to \Psi_1} \qquad \cfrac{\cfrac{\vdots}{B^{h+1}, \Phi_2^{-(B \vee C)^h} \to \Psi_2} \quad \cfrac{\vdots}{E, B^b, \Phi_1^{-(B \vee C)^b} \to \Psi_1}}{\quad}}{D \vee E, B^b, \Phi_1^{-(B \vee C)^b} \to \Psi_1} \vee L_{\vartheta_1'}
}{\vdots}}{D \vee E, B^k, \Sigma^{-(B \vee C)^k} \to \Pi}
}{B^k, \Sigma^{-(B \vee C)^k} \to \Pi} \vartheta_1'
}{}
$$

Multiset Φ_2 contains exactly h occurrences of $B \vee C$ that belong to the thread for the cut formula $B \vee C$.

The subsequent elimination of ϑ_1' makes now derivations of all premises $E, \Omega \to \Upsilon$ of the rules of $\vee L$ as displayed in (†) redundant during the construction of the derivation for $D, B^k, \Sigma^{-(B \vee C)^k} \to \Pi$. Since the elimination of ϑ_2 did not violate the thread for the cut formula $D \vee E$, the modified forms of the same derivations are redundant as in the case when we started with the elimination of ϑ_1. The derivation after the elimination of ϑ_1' is shown in the figure below:

Assume now that $\Theta \to \Lambda$ is an arbitrary sequent from γ that is not excluded from R'. The modification of this sequent during the elimination of both cuts proceeds as follows:

$$\Theta \to \Lambda$$
$$\wr \vartheta_1$$
$$\Theta \to \Lambda^{-(D \vee E)^p}, D^p, E^p$$
$$\wr \vartheta_2$$
$$B^q, \Theta^{-(B \vee C)^q} \to \Lambda^{-(D \vee E)^p}, D^p, E^p$$

$$\Theta \to \Lambda$$
$$\wr \vartheta_2$$
$$B^q, \Theta^{-(B \vee C)^q} \to \Lambda$$
$$\wr \vartheta_1'$$
$$B^q, \Theta^{-(B \vee C)^q} \to \Lambda^{-(D \vee E)^p}, D^p, E^p$$

Multiset Θ contains exactly q occurrences of $B \vee C$ that belong to the thread for the cut formula $B \vee C$. Multiset Λ contains exactly p occurrences of $D \vee E$ that belong to the thread for the cut formula $D \vee E$.

Furthermore, assume that $\Phi \to \Psi$ is an arbitrary sequent from δ that is made redundant neither during the construction of R' nor during the construction of δ'. The modification of this sequent proceeds as follows:

$$\Phi \to \Psi$$
$$\wr \vartheta_1$$
$$D^c, \Phi^{-(D \vee E)^c} \to \Psi$$
$$\wr \vartheta_2$$
$$B^d, D^c, \Phi^{-(D \vee E)^c, -(B \vee C)^d} \to \Psi$$

$$\Phi \to \Psi$$
$$\wr \vartheta_2$$
$$B^d, \Phi^{-(B \vee C)^d} \to \Psi$$
$$\wr \vartheta_1'$$
$$B^d, D^c, \Phi^{-(B \vee C)^d, -(D \vee E)^c} \to \Psi$$

Multiset Φ contains exactly c occurrences of $D \vee E$ that belong to the thread for the cut formula $D \vee E$ and it also contains d occurrences of $B \vee C$ that belong to the thread for the cut formula $B \vee C$. On the left hand side, the elimination of ϑ_2 does not delete the antecedent occurrences of D because $|D| < |B \vee C|$. There is a similar situation on the right hand side, too. We are not able to compare the number of logical operations in B and $D \vee E$. However, we know that the antecedent occurrences of B belong to the thread different from the thread for the cut formula $D \vee E$.

(6) The case when the cut formula of ϑ_1 is $\neg C$ and the cut formula of ϑ_2 is $\neg B$ is easy since no part of P can be made redundant. We have $|\neg C| < |\neg B|$:

$$
\cfrac{
  \cfrac{
    \vdots \alpha \qquad \vdots \beta \\
    \Sigma \to \Pi, \neg C \qquad \neg C, \Sigma \to \Pi
  }{\Sigma \to \Pi} \vartheta_1
}{}
\qquad
\cfrac{\vdots Q}{\Gamma \to \Delta, \neg B} \qquad \cfrac{\vdots R}{\neg B, \Gamma \to \Delta}
$$

$$\Gamma \to \Delta \quad \vartheta_2$$

$$\text{or}$$

$$
\cfrac{
  \cfrac{
    \vdots \gamma \qquad \vdots \delta \\
    \Sigma \to \Pi, \neg C \qquad \neg C, \Sigma \to \Pi
  }{\Sigma \to \Pi} \vartheta_1
}{}
$$

$$\cfrac{\vdots Q}{\Gamma \to \Delta, \neg B} \qquad \cfrac{\vdots R}{\neg B, \Gamma \to \Delta}$$

$$\Gamma \to \Delta \quad \vartheta_2$$

(7) The case when the cut formula of ϑ_1 is $C \vee D$ and the cut formula of ϑ_2 is $\neg B$ is analogous to (4). We have $|C \vee D| < |\neg B|$:

$$
\cfrac{
  \vdots \alpha \qquad \vdots \beta \\
  \Sigma \to \Pi, C \vee D \qquad C \vee D, \Sigma \to \Pi
}{\Sigma \to \Pi} \vartheta_1
$$

$$\cfrac{\vdots Q}{\Gamma \to \Delta, \neg B} \qquad \cfrac{\vdots R}{\neg B, \Gamma \to \Delta}$$

$$\Gamma \to \Delta \quad \vartheta_2$$

$$\text{or}$$

$$
\cfrac{
  \vdots \gamma \qquad \vdots \delta \\
  \Sigma \to \Pi, C \vee D \qquad C \vee D, \Sigma \to \Pi
}{\Sigma \to \Pi} \vartheta_1
$$

$$\cfrac{\vdots Q}{\Gamma \to \Delta, \neg B} \qquad \cfrac{\vdots R}{\neg B, \Gamma \to \Delta}$$

$$\Gamma \to \Delta \quad \vartheta_2$$

(8) The case when the cut formula of ϑ_1 is $\neg B$ and the cut formula of ϑ_2 is $C \vee D$ is easy since no blue sequents can be made redundant. We have $|\neg B| < |C \vee D|$:

$$
\cfrac{
  \vdots \alpha \qquad \vdots \beta \\
  \Sigma \to \Pi, \neg B \qquad \neg B, \Sigma \to \Pi
}{\Sigma \to \Pi} \vartheta_1
$$

$$\cfrac{\vdots Q}{\Gamma \to \Delta, C \vee D} \qquad \cfrac{\vdots R}{C \vee D, \Gamma \to \Delta}$$

$$\Gamma \to \Delta \quad \vartheta_2$$

(9) The case when the cut formula of ϑ_1 is $\neg B$, the cut formula of ϑ_2 is $C \vee D$ and ϑ_1 is in the derivation of the right hand side premise of ϑ_2 is analogous to (5). We have $|\neg B| < |C \vee D|$:

$$
\cfrac{
  \vdots \alpha \qquad \vdots \beta \\
  \Sigma \to \Pi, \neg B \qquad \neg B, \Sigma \to \Pi
}{\Sigma \to \Pi} \vartheta_1
$$

$$\cfrac{\vdots Q}{\Gamma \to \Delta, C \vee D} \qquad \cfrac{\vdots R}{C \vee D, \Gamma \to \Delta}$$

$$\Gamma \to \Delta \quad \vartheta_2$$

$\square$

Bibliography

[1] Baaz, M., Leitsch, A.: Methods of Cut-Elimination. Springer (2011)

[2] Buchholz, W.: Notation systems for infinitary derivations. Archive for Mathematical Logic **30**, 277–296 (1991)

[3] Buss, S.: An Introduction to Proof Theory. In: Buss, S., Abramsky, S., Artemov, S. (eds.) Handbook of Proof Theory. Elsevier, Amsterdam (1998)

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

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

[6] Hájek, P., Pudlák, P.: Metamathematics of First-Order Arithmetic. Springer (1993)

[7] Horská, A.: Where is the Gödel-Point hiding: Gentzen's consistency proof of 1936 and his representation of constructive ordinals. Springer (2014)

[8] Menzler-Trott, E.: Gentzen's Problem. Mathematische Logik im nationalsozialistischen Deutschland. Birkhäuser Verlag (2001)

[9] von Plato, J.: From Hauptsatz to Hilfssatz. In: Kahle, R., Rathjen, M. (eds.) Gentzen's Centenary - The Quest for consistency. Springer (2015)

[10] Schütte, K.: Beweistheoretische Erfassung der unendlichen Induktion in der Zahlentheorie. Mathematische Annalen **122**, 369–389 (1951)

[11] Schütte, K.: Proof Theory. Springer (1977)

[12] Schwichtenberg, H.: Some applications of cut-elimination. In: Barwise, J. (eds.) Handbook of Mathematical Logic. North-Holland, Amsterdam (1977)

[13] Schwichtenberg, H.: Normalization. In: Bauer, F.L. (ed.) Logic, Algebra and Computation. Springer (1991)

[14] Siders Kanckos, A.: Gentzen's Consistency Proofs for Arithmetic. In: Peliš, M. (eds.) The Logica Yearbook 2009, pp. 109–119. College Publications, London (2010)

[15] Sørensen, M.H., Urzyczyn, P.: Lectures on the Curry-Howard Isomorphism. Elsevier, Amsterdam (2006)

[16] Švejdar, V.: Logika, neúplnost, složitost a nutnost. Academia, Prague (2002)

[17] Tait, W.: Gentzen's original consistency proof and the Bar Theorem. In: Kahle, R., Rathjen, M. (eds.) Gentzen's Centenary - The Quest for consistency. Springer (2015)

[18] Veblen, O.: Continuous Increasing Functions of Finite and Transfinite Ordinals. Transactions of the American Mathematical Society **9**, 280–292 (1908)