FAKULTÄT FÜR !NFORMATIK

M.Sc. Arbeit

Cut Elimination in Inductive Proofs of Weakly Quantified Theorems

ausgeführt am

Institut für Computersprachen
Arbeitsgruppe Theoretische Informatik und Logik
der Technischen Universität Wien

unter der Anleitung von

Univ.Prof. Dr.phil. Alexander Leitsch

durch

Tomer Libal

Wien, 6. Oktober 2008 ....

Tomer Libal Alexander Leitsch

FAKULTÄT FÜR !NFORMATIK

MASTER THESIS

Cut Elimination in Inductive Proofs of Weakly Quantified Theorems

carried out at the

Institute of Computer Languages
Theory and Logic Group
of the Vienna University of Technology

under the instruction of

Univ.Prof. Dr.phil. Alexander Leitsch

by

Tomer Libal

Vienna, October 6, 2008 ....
Tomer Libal Alexander Leitsch

Contents

Kurzfassung

In dieser Arbeit wird erst der Gentzensche Beweis der Konsistenz der Peanoarithmetik vorgestellt, dann wird die Methode auf eine Klasse induktiver Beweise erweitert. Da die Schnittregel in der Peanoarithmetik nicht redundant ist (siehe zB [2]), ist Schnittelimination nur für Teilklassen möglich; hier wird die Teilklasse aller induktiven Beweise von Sätzen ohne starke Quantoren untersucht. Es wird gezeigt, dass bei dieser Klasse alle essentiellen Schnitte eliminiert werden können.

Abstract

After presenting Gentzen's cut elimination theorem and the proof of the consistency of Peano arithmetic, we extend the cut elimination procedure to a certain class of inductive proofs. Cut elimination is possible only on subclasses of all inductive proofs. (see for example [2]). We will investigate the subclass of inductive proofs of theorems without strong quantifiers. We will show that all inductions can be removed following Gentzen's proof of the consistency of Peano arithmetic and therefore, that essential cuts are redundant.

Dedication

I would like to dedicate this master thesis to my parents who, once again, were rewarded for all their love by my departure to a different continent, and to Irina, who has loved me enough to depart together with me.

Acknowledgements

I would like to express my deep and sincere gratitude to my supervisor, Prof. Alexander Leitsch. His guidance and wide knowledge had helped me to obtain the necessary background required in order to complete this thesis. Moreover, I would like to thank him for his patience, despite the many revisions of the thesis.

I would also like to thank Prof. Matthias Baaz for the interesting and inspiring lectures, which have greatly deepened my interest in Proof Theory. I thank also Prof. Alberto Pardo, whose class in Madrid gave me the necessary background to continue.

I would like to thank the participants of the weekly meetings of the CERes Project. The discussions held there have helped me to appreciate and understand more the fascinating field of Proof Theory.

I am very grateful for the opportunity given to me to participate in the Erasmus Mundus program as a scholarship holder. The mobility of the program, which has allowed me to try out different fields in logic before I was committed to proof theory, has been proved vital to the completion of my studies.

Last, but not least, I would like to thank the many people I have meet, both in Spain and in Austria, who have made the last two years more interesting and exciting.

Chapter 1

Introduction

Proof theory is a part of mathematical logic, which deals with proofs as mathematical objects. One of the most prominent proof transformations is cut elimination. A formal proof is built up from the theorem to prove and an inference rule leading to this theorem, which may assume the provability of some other statements. The statements, which are assumed in the cut inference rule, are unique because they might include formulas, which do not appear in the conclusion of the rule. Therefore, when the cut rule is known to be required in a proof, the process of building the proof cannot rely on the formulas appearing in the theorem to be proved only. The cut elimination theorem implies that the cut rule is redundant in a system. Systems, which have the cut elimination theorem, are easily proved to be consistent. Other important applications of cut-free proofs include Herbrand's theorem and decidability, interpolation and definability.

One system, which does not have the cut elimination theorem, is first order Peano arithmetic. Peano arithmetic is the axiomatization of arithmetic, which uses the successor function and inductions. In the second part, we will discuss why Peano arithmetic does not have the cut elimination theorem. Although Peano arithmetic does not have the cut elimination theorem, it can be proved to be consistent. The second part will also contain Getzen's proof of the consistency of Peano arithmetic and a discussion on the meaning of the chosen ordinals used in the proof.

The call for proving Peano arithmetic to be consistent was raised by David Hilbert. Hilbert has proposed, in what is known as Hilbert's program, to formalize all mathematics within a finite consistent set of axioms.

Hilbert has suggested that the consistency of a complicated system can be proved in terms of a simpler system. In 1931 the program was proved to be impossible by the introduction of Gödel's second incompleteness theorem. This has shown that any consistent extension of basic arithmetic cannot prove its own consistency and therefore, it cannot prove the consistency of stronger systems. The question remained is whether it is possible to prove the consistency of Peano arithmetics using only finitistic reasoning. In 1936 Gerhard Gentzen has proved the consistency of PA using elementary mathematics and transfinite induction up to the ordinal ϵ_0. It can be shown that all decreasing sequences of ordinals smaller than ϵ_0 are finite. Therefore, this proof can be considered as elementary, despite involving ordinals bigger than ω.

Besides the important theoretical implications of cut elimination, the constructive nature of some proofs of cut elimination may also give some interesting results. Cuts may be regarded as lemmas, which appear in the proof. As such, they may involve language and axioms other than those of the theorem proved. Fürstenberg's proof of the infinitude of primes, for example, is a proof of a number theory statement obtained by topological means [7]. Constructive methods for cut elimination may produce a proof formalized in the language of the proved theorem. As, for example, the result of applying CERes (a method of cut elimination by resolution) to Fürstenberg's proof in order to obtain Euclid's original proof [8]. The proof is inductive and emphasize the importance of cut elimination in inductive proofs.

The third and last part of the thesis will be devoted to the presentation of a variant of the cut elimination theorem. The subclass investigated here is the set of all inductive proofs. However, the theorems proved by these inductive proofs will be limited to theorems without strong quantifiers. We will give an algorithm, which is capable of eliminating all cuts, despite the presence of inductions. This will mainly be achieved by the elimination of most of the inductions, by a procedure taken from the consistency proof of Gentzen presented in the second part. Our proof of the cut elimination theorem requires the modification of the ordinals assignment as presented in the consistency proof in the second part. A method for the elimination of cuts in proofs of theorems without strong quantifiers is hinted as an exercise in [1], which requires the theorem to be in prenex normal form. We present another method for the elimination of cuts, which does not require the theorem to be in prenex normal form and which tries to give a more direct approach to the problem, which is not based completely on the consistency proof.

In the first part, we will present all the technical terms, which are used in the rest of the thesis. We will conclude the first part with the presentation of the cut elimination theorem for classical logic.

Chapter 2

Classical Logic

Gentzen has proved his results about cut elimination and the consistency of Peano arithmetic by using his first order sequent calculi LK and PA. PA is an extension of LK (but with a limited language), which includes the induction inference rule and some additional axioms. In this section we will formalize LK using sequent notation.

2.1 Sequent Calculus

We will first formalize the language of our logic.

2.1.1 Basic notations

Definition 2.1.1 (The language)**.** Our language will consist of the following symbols:

1. Constants:
   (a) Individual constants: k_i for $i \in \mathbb{N}$.
   (b) Function constants with i argument-places f^i_j for $i, j \in \mathbb{N}$.
   (c) Predicate constants with i argument-places R^i_j for $i, j \in \mathbb{N}$.
2. Variables:
   (a) Free variables: a_i for $i \in \mathbb{N}$.
   (b) Bound variables x_i for $i \in \mathbb{N}$.
3. Logical symbols: $\neg, \vee$ and $\forall$.

4. Auxiliary symbols: '(',')','[',']' and ','.

- Remarks
    1. Although the distinction between bound and free variables is not essential, it simplifies the arguments taken in this thesis.
    2. For convenience we might sometimes omit superscripts and subscripts of functions and predicates, or denote them by a single quote instead of natural numbers.

An expression is any finite sequence of symbols from the language defined above. The next definition is about terms and is given inductively. All inductive definitions will implicitly mean that the objects, which are defined, are only those given by the definition.

Definition 2.1.2 (Terms and semi-terms)**.** Semi-terms are defined inductively as follows:

1. Every individual constant is a semi-term.
2. Bound and free variable are semi-terms.
3. If f^i is a function constant with i argument-places and t_1,..,t_i are semi-terms, then $f^i(t_1, .., t_i)$ is a semi-term.

Semi-terms which do not contain bound variables are called terms.

- Remark -'a' is called fully indicated in P(a) if for some expression 'b', P(b) is obtained by replacing all occurrences of 'a' by 'b'.

Definition 2.1.3 (Formulas and semi-formulas)**.** If R^i is a predicate constant with i argument-places and $t_1, .., t_i$ are terms, then $R^i(t_1, .., t_i)$ is an atomic formula. Formulas and their outermost logical symbols are defined as follows:

1. Every atomic formula is a formula.
2. If A and B are formulas, then $\neg A$ and $A \vee B$ are formulas with $\neg$ and $\vee$ as their outermost logical symbol.
3. If $A(a)$ is a formula with a free variable 'a' being not necessarily fully indicated in A, then $\forall x A(x)$ is a formula with 'x' a bound variable replacing each occurrence of 'a' in A. The outermost logical symbol is $\forall$.

Semi-formulas differ from formulas in containing semi-terms, which are not bound by a quantifier.

- Remarks

  1. A formula or a term without free variables will be called 'closed'. A closed formula is also called a sentence.
  2. $A(x)$ in the above definition is called the scope of the formula $\forall x A(x)$.
  3. For convenience we might sometimes omit parentheses while having $\neg$ and $\forall$ take precedence over $\vee$.

Replacement on positions play a central role in proof transformations. We first introduce the concept of positions for terms.

Definition 2.1.4 (Positions)**.** Positions within semi-terms are defined inductively:

- If t is a variable or a constant symbol then 0 is a position in t and $t.0 = t$.
- Let $t = f(t_1, ..., t_n)$ then 0 is a position in t and $t.0 = t$. Let $\mu : (0, k_1, ..., k_l)$ be a position in a t_j (for $1 \leq j \leq n$) and $t_j.\mu = s$, then $v : (0, j, k_1, ..., k_l)$ is a position in t and $t.v = s$.

A sub-semi-term s of t is a semi-term $t.v = s$ for some position v in t. Positions will be denoted by [,], i.e. $t[r]_v$ denotes the term t after replacing $t.v$ with r.

- Remarks

  1. Sub-formulas are defined in a similar way to sub-terms. However, they are defined up to replacing previously bound variables.
  2. We will use $P(a)$ to represent a term, formula, sequence of formulas or a whole proof where the variable or term a is fully indicated. $P[a]_\lambda$, where λ can be a single position or a set of positions, will represent the case where a is indicated only at position(s) λ.

Example 2.1.5 (Sub-semi-formula)**.** *The following are sub-semi-formulas of the formula* $\forall x A(x) \vee B$*:* $\forall x A(x), A(t), A(x)$*, etc.*

Definition 2.1.6 (Substitutions)**.** A substitution is a mapping σ from the set of free and bound variables to the set of semi-terms such that $\sigma(v) \neq v$ for only a finite number of variables.

Definition 2.1.7 (Logical complexity of formulas)**.** If F is a formula then the complexity comp(F) is the number of logical symbols occurring in F. Later in the thesis we will identify this definition with the definition of grades of formulas.

Let sequences of formulas be represented by the greek letters: Γ, Δ, Π and Λ with possible superscripts and subscripts.

2.1.2 Sequent calculus for classical logic (LK)

Definition 2.1.8 (Sequents)**.** For arbitrary Γ and Δ, $\Gamma \vdash \Delta$ is called a sequent with $\vdash$ called the sequent symbol. Γ and Δ are called the antecedent and the succedent of the sequent. Each formula in Γ and Δ is called a sequent-formula. A sequent will be denoted by the letter 'S' with or without subscripts, i.e. $A \vdash^S B$.

Definition 2.1.9 (Semantics of sequents)**.** Semantically a sequent

$$A_1, .., A_n \vdash^S B_1, ..., B_m$$

stands for formula $F(S)$:

$$\bigwedge_{i=1}^n A_i \to \bigvee_{j=1}^m B_j.$$

In particular, we define M to be the interpretation of S if it is the interpretation of $F(S)$. If $n = 0$ (i.e. the antecedent is empty), we assign $\top$ to $\bigwedge_{i=1}^n A_i$. If $m = 0$ (i.e. the succedent is empty), we assign $\bot$ (falsum) to $\bigvee_{j=1}^m B_j$. The empty sequent $\vdash$ is represented by $\top \to \bot$ which is equivalent to $\bot$. S is true in M if $F(S)$ is true in M and S is valid if $F(S)$ is valid.

Definition 2.1.10 (Atomic sequents)**.** A sequent $A_1, ..., A_n \vdash B_1, ..., B_m$ is called atomic if for all $1 \leq i \leq n$ and $1 \leq j \leq m$, A_i and B_j are atomic.

Definition 2.1.11 (Axiom set)**.** A (possibly infinite) set $\mathcal{A}$ of sequents is called an axiom set if it is closed under substitution. I.e. for every $S \in \mathcal{A}$ and a substitution σ we have $\sigma(S) \in \mathcal{A}$. If $\mathcal{A}$ consists only of atomic sequents it is called an atomic axioms set.

Definition 2.1.12 (Standard axiom set)**.** The standard axiom set is the smallest axiom set containing all sequents of the form $A \vdash A$ for arbitrary formula A.

Definition 2.1.13 (Inference)**.** An inference is an expression of the form:

$$\frac{S_1}{S} \qquad \text{or} \qquad \frac{S_1 \quad S_2}{S}$$

where S_1, S_2 and S are sequents. S_1 and S_2 are called the upper sequents and S is called the lower sequent of this inference.

- Remark - for simplicity we are using Gentzen calculus with the following operators only: $\wedge, \forall, \neg$. The remaining operators $\wedge$, $\rightarrow$ and $\exists$ can be obtained from the other three operators as follows: $A \wedge B \equiv \neg(\neg A \vee \neg B)$, $A \rightarrow B \equiv \neg A \vee B$ and $\exists x A(x) \equiv \neg \forall x \neg A(x)$.

Definition 2.1.14 (Standard LK)**.** The standard (multiplicative) sequent calculus LK contains the standard axioms set and the following rules of inference. Each inference may have auxiliary formulas marked by $+$ and principal formulas marked by $\star$.

1. Structural rules:

   (a) Weakenings:

$$\frac{\Gamma \vdash \Delta}{D^\star, \Gamma \vdash \Delta}\,(\text{w:l}) \qquad \frac{\Gamma \vdash \Delta}{\Gamma \vdash \Delta, D^\star}\,(\text{w:r})$$

   (b) Contractions:

$$\frac{D^+, D^+, \Gamma \vdash \Delta}{D^\star, \Gamma \vdash \Delta}\,(\text{c:l}) \qquad \frac{\Gamma \vdash \Delta, D^+, D^+}{\Gamma \vdash \Delta, D^\star}\,(\text{c:r})$$

   (c) Exchanges:

$$\frac{\Gamma, C^+, D^+, \Pi \vdash \Delta}{\Gamma, D^\star, C^\star, \Pi \vdash \Delta}\,(\text{e:l}) \qquad \frac{\Gamma \vdash \Delta, C^+, D^+, \Lambda}{\Gamma \vdash \Delta, D^\star, C^\star, \Lambda}\,(\text{e:r})$$

   These three rules will be called weak inferences while the others will be called strong inferences.

   (d) Cuts:

$$\frac{\Gamma \vdash \Delta, D^+ \qquad D^+, \Pi \vdash \Lambda}{\Gamma, \Pi \vdash \Delta, \Lambda}\,(\text{cut:D})$$

   D is also called the cut formula of the inference.

2. Logical rules:

   (a) $\neg$-introduction:

$$\frac{\Gamma \vdash \Delta, D^+}{\neg D^\star, \Gamma \vdash \Delta}\ (\neg\text{:l}) \qquad \frac{D^+, \Gamma \vdash \Delta}{\Gamma \vdash \Delta, \neg D^\star}\ (\neg\text{:r})$$

(b) $\vee$-introduction:

$$\frac{C^+, \Gamma \vdash \Delta \qquad D^+, \Pi \vdash \Lambda}{(C \vee D)^\star, \Gamma, \Pi \vdash \Delta, \Lambda}\ (\vee\text{:l})$$

$$\frac{\Gamma \vdash \Delta, C^+}{\Gamma \vdash \Delta, (C \vee D)^\star}\ (\vee : r_1) \qquad \frac{\Gamma \vdash \Delta, D^+}{\Gamma \vdash \Delta, (C \vee D)^\star}\ (\vee : r_2)$$

The $\neg$ and $\vee$ are called propositional inferences.

(c) $\forall$-introduction:

$$\frac{F(t)^+, \Gamma \vdash \Delta}{(\forall x F(x))^\star \Gamma \vdash \Delta}\ (\forall : l) \qquad \frac{\Gamma \vdash \Delta, F(a)^+}{\Gamma \vdash \Delta, (\forall x F(x))^\star}\ (\forall : r)$$

Where 't' is an arbitrary term and 'a' does not occur in the lower sequent. The 'a' in $\forall : r$ is called the eigenvariable of the inference. The condition that 'a' does not occur in the lower sequent is called the eigenvariable condition of the inference. We will also say that the quantifiers in the lower sequents eliminate the eigenvariable or the term in the upper sequents. The $\forall : r$ rule is called a strong quantifier rule while the $\forall : l$ is called a weak quantifier rule.

2.1.3 Derivations and proofs

Definition 2.1.15 (Inference's instances)**.** Suppose

$$\frac{S_1 \qquad S_2}{S}\ (\xi)$$

is a binary inference rule of LK. Let S', S'_1 and S'_2 be instantiations of the schema variables S, S_1 and S_2, then

$$\frac{S'_1 \qquad S'_2}{S'}\ (\xi)$$

is called an instance of ξ, also denoted by $\xi(S'_1, S'_2, S')$. We will refer to both inference rules and their instances as inference rules. The instance of an unary rule is defined analogously.

Definition 2.1.16 (LK-derivations)**.** An LK-derivation is defined as a directed labelled tree where the nodes are labelled by sequents (via the function *seq*) and the edges by inference rules. The label of the root is called the end-sequent. Sequents occurring at the leaves are called initial sequents or axioms. The formal definition is:

- Let v be a node and $seq(v) = S$ for an arbitrary sequent S. Then v is an LK-derivation and v is the root node.

- Let ψ be a derivation tree and v be a leaf in ψ. Let $\xi(S_1, S_2, S)$ be an instance of the binary rule ξ. We extend ψ to ψ' by appending the edges $e_1 : (v, \mu_1)$ and $e_2 : (v, \mu_2)$ to v such that $seq(\mu_1) = S_1$, $seq(\mu_2) = S_2$ and the label of (e_1, e_2) is ξ. ψ' is an LK-derivation with the same root as ψ but with v no longer a leaf. v in ψ is called a ξ-node and μ_1 and μ_2 are leaves.

- The extension by an unary rule is done analogously.

Definition 2.1.17 (LK-sub-derivations)**.** let ψ be an LK-derivation. An LK-sub-derivation of ψ is any sub-tree of ψ.

Definition 2.1.18 (Formal proof)**.** A proof P in LK is an LK-derivation where the leaves are mapped to initial sequents:

The following terminology and conventions will be used all along this thesis:

- If there exists a proof with S being its end-sequent, then S is said to be provable in LK.

- A proof without the cut rule is called cut-free.

- Remark - Due to the difficulty of representing a proof tree graphically. We will use the following notation:

$$\frac{\frac{S_1}{\cdots}}{S_2}$$

  in order to denote a specific branch in the proof tree of S_2. It will also be used to denote a full derivation ending with S_2 where one of the leaves is mapped to S_1.

Definition 2.1.19 (Threads)**.** A sequence of nodes in a proof P is called a thread of P if the following conditions are satisfied:

1. The sequence begins with an initial sequent and ends with an end-sequent.

2. Every sequent in the sequence except the last is an upper sequent of an inference and it is immediately followed by the lower sequent of this inference.

Now we can speak about the order of sequents in the proof:
Let S_1, S_2 and S_3 be sequents in a proof P.

- We say S_1 is above S_2 or S_2 is below S_1 if there is a thread of P containing both S_1 and S_2 in which S_1 appears before S_2. If S_1 is above S_2 and S_2 is above S_3, we say that S_2 is between S_1 and S_3.

- An inference in P is said to be below a sequent S if its lower sequent is below S.

Definition 2.1.20 (Subproofs)**.** Let ψ be a proof. a subproof of ψ is a sub-derivation of ψ which is also a proof.

2.2 $LK_=$ and the Cut Elimination Theorem

We can extend the current system by increasing the number of valid axioms.

Definition 2.2.1 (Axiom systems)**.** For the basic system LK:

1. A finite or infinite set $\mathcal{A}$ of sequents is called an axiom system and each of its sequents is called an axiom of $\mathcal{A}$.

2. A finite (possibly empty) sequence of formulas consisting only of axioms of $\mathcal{A}$ is called an axiom system of $\mathcal{A}$.

3. If there exists an axiom sequence Γ_0 of $\mathcal{A}$ such that $\Gamma_0, \Gamma \vdash \Delta$ is LK-provable, then $\Gamma \vdash \Delta$ is said to be provable from $\mathcal{A}$ in LK.

4. $\mathcal{A}$ is inconsistent if the empty sequent $\vdash$ is provable from $\mathcal{A}$.

5. If $\mathcal{A}$ is not inconsistent, then it is said to be consistent.

6. $LK_{\mathcal{A}}$ is the system obtained from LK by adding $\vdash A$ as initial sequent for all A in $\mathcal{A}$.

7. $LK_{\mathcal{A}}$ is said to be inconsistent if $\vdash$ is $LK_{\mathcal{A}}$-provable, otherwise it is consistent.

Now we can extend LK to include equality. We will first add the equality symbol to the language (as a fixed binary predicate constant, written in infix notation).

Definition 2.2.2 (Equality axioms)**.** The following axiom set $\Gamma_=$ axiomatizes equality:

1. $\vdash s = s$.
2. $s_1 = t_1, ..., s_n = t_n \vdash f(s_1, ..., s_n) = f(t_1, ..., t_n)$ for every n-argument function constant f (for all natural numbers n).
3. $s_1 = t_1, ..., s_n = t_n, R(s_1, ..., s_n) \vdash R(t_1, ..., t_n)$ for every n-argument predicate constant R (for all natural numbers n).

$LK_=$ is obtained from LK by adding the equality axioms as initial sequents.

Proposition 2.2.3. The followings are provable in $LK_=$:

1. Transitivity of $=$.
2. Symmetry of $=$.
3. $s_1 = t_1, ..., s_n = t_n, A(s_1, ..., s_n) \rightarrow A(t_1, ..., t_n)$ for all terms s_i, t_i and formulas $A(a_1, ..., a_n)$.

Definition 2.2.4 (Essential cuts)**.** If the cut formula of a cut in $LK_=$ is an atomic formula, then the cut is called inessential. Otherwise, it is called essential.

Proof

- The proof is similar to the proof of cut elimination in LK. The only difference is the addition of the new axioms. In this case the cuts are inessential as they must be over atomic formulas. It should be noted that there is a transformation transforming any proof which contains axioms with logical symbols into a proof with axioms containing only atomic formulas.

Theorem 2.2.5 (Cut Elimination for $LK_=$)**.** If a sequent of $LK_=$ is $LK_=$-provable, then it is $LK_=$-provable without an essential cut.

Definition 2.2.6 (Sub-formula property)**.** A proof is said to have the sub-formula property if all the formulas, which appear in the proof, are sub-formulas of the formulas, which appear in the end-sequent.

- Proofs without cuts have the sub-formula property. Therefore one consequence of the cut elimination theorem is that there is no proof of the empty sequent $\vdash$, implying the consistency of LK.
- Proofs without essential cuts may contain only atomic formulas in addition to the symbols appearing in the end-sequent.

Theorem 2.2.7 (Consistency of LK)**.** LK is consistent.

Chapter 3

Peano Arithmetic

In the previous section we have seen that the cut elimination theorem holds for LK and also implies its consistency. In this section we will extend LK to the system PA (Peano arithmetic). We will continue by discussing why the cut elimination theorem does not hold for PA and we will finish by proving its consistency.

3.1 Peano Arithmetic (PA)

3.1.1 Formalization

Peano arithmetic can be derived from the equality, natural numbers axioms and the axioms of induction. We will limit our language from the previous section (of LK) to the following constants: The individual constant '0', the binary predicate constant '=' and a constant f for every primitive recursive function f, including the binary function constants '+' and '$\cdot$' and the unary function constant '(the successor function).

Definition 3.1.1 (Natural numbers axioms)**.** The natural numbers axiom system is:

1. $\forall x \forall y (x' = y' \rightarrow x = y)$.
2. $\forall x (\neg (x' = 0))$.
3. $\forall x (x + 0 = 0)$.
4. $\forall x \forall y (x + y' = (x + y)')$.
5. $\forall x (x \cdot 0 = 0)$.

6. $\forall x \forall y (x \cdot y' = (x \cdot y) + x)$.

Definition 3.1.2 (Axioms of induction)**.** The axioms of induction are all sentences for all (n+1)-arguments predicate symbols F:

$$\forall z_1 .. \forall z_n \forall x (F(0, z_1, .., z_n) \wedge (\forall y (F(y, z_1, .., z_n) \rightarrow F(y', z_1, .., z_n)) \rightarrow F(x, z_1, .., z_n))$$

We can formulate Peano arithmetic by adding the basic axioms for natural numbers and a new inference rule for induction to the system $LK_=$. However, as we deal with numerals only and we dont need to define equality for arbitrary predicate symbols (only for =), it will be simpler to define PA with an infinite number of axioms. These axioms will then represent the equalities and inequalities between all possible terms.

Definition 3.1.3 (PA)**.** The system PA is obtained from LK (with a limited language) by the addition of extra initial sequents (called mathematical initial sequents) and of a new rule of inference called 'ind'.

- Mathematical initial sequents, which are the defining equations for all primitive recursive functions plus all sequents of the form $\vdash s = t$ where s and t are closed terms of the language denoting the same number and $s = t \vdash$ for terms denoting different numbers.

- ind:

$$\frac{F(a)^+, \Gamma \vdash \Delta, F(a')^+}{F(0)^\star, \Gamma \vdash \Delta, F(s)^\star} \text{(ind)}$$

  where a is not in F(0), Γ or Δ and s is an arbitrary term which may contain a. F(a) is an arbitrary formula of the language. F(a) is called the induction formula and a is called the eigenvariable of the inference. s is called the induction term.

- Remarks
  - In PA there are two kinds of initial sequents: the mathematical sequents defined above and the logical ($A \vdash A$) sequents defined in the previous section.
  - From the definitions above it can be seen that the three groups of axiom systems (for equality, natural numbers and induction) are consistent if and only if $\vdash$ is not provable in PA.

- From now on, whenever we will say provable we will mean provable in PA.

We can now also redefine essential cuts as containing atomic equalities only ($s = t$ for arbitrary terms s and t) and not arbitrary atomic formulas. This is because, in the way we have defined PA, all mathematical axioms are of this form.

Theorem 3.1.4 (Cut Elimination for PA)**.** If a sequent is PA-provable without an induction, then it is PA-provable without an induction and essential cuts.

Proof

- The same proof as for $LK_=$ as the new axioms contain only atomic equalities.

An important lemma, which will be required in order to transform inductions into finite number of cuts, will be given next. It states that for every closed term s of the language of arithmetic, there is a numeral $\overline{n}$ such that $\vdash s = \overline{n}$ is provable in PA.

Lemma 3.1.5. The following holds for arbitrary closed terms s and t:

1. for an arbitrary closed term s there exists a unique numeral $\overline{n}$ such that $s = \overline{n}$ is provable without an essential cut and without ind.

2. Either $\vdash s = t$ or $s = t \vdash$ is provable without an essential cut or ind.

3. If $\vdash s = t$ is provable without an essential cut or ind then for $q[a]_\lambda$ and $r[a]_\theta$, $q[s]_\lambda = r[s]_\theta \vdash q[t]_\lambda = r[t]_\theta$ is provable without an essential cut or ind.

4. For an arbitrary formula F(a), if $\vdash s = t$ is provable without an essential cut or ind, so is $F(s) \vdash F(t)$.

Proof - by induction on the complexity of s.

The following three lemmas, dealing with the replacement of fully indicated free variables in proofs, will be used later in this chapter and in the next one.

Lemma 3.1.6. Let $\Gamma(a) \rightarrow \Delta(a)$ be a provable sequent in which a is fully indicated and let $P(a)$ be a proof ending with $\Gamma(a) \rightarrow \Delta(a)$. Let b be a free variable not occurring in $P(a)$ Then $P(b)$ (the result of replacing all a's in $P(a)$ by b) is a proof whose end-sequent is $\Gamma(b) \rightarrow \Delta(b)$.

Proof - by induction on the length n of $P(a)$.

Lemma 3.1.7. Let t be an arbitrary term. Let $\Gamma(a) \rightarrow \Delta(a)$ be a provable sequent in which a is fully indicated and let $P(a)$ be a proof ending with $\Gamma(a) \rightarrow \Delta(a)$ in which every eigenvariable is different from a and not contained in t. Then $P(t)$ (the result of replacing all a's in $P(a)$ by t) is a proof whose end-sequent is $\Gamma(t) \rightarrow \Delta(t)$.

Proof - by induction on the length n of $P(a)$.

Lemma 3.1.8. Let t be an arbitrary term. Let $\Gamma(a) \rightarrow \Delta(a)$ be a provable sequent in which a is fully indicated and let $P(a)$ be a proof ending with $\Gamma(a) \rightarrow \Delta(a)$. Let $P'(a)$ be a proof obtained from $P(a)$ in which every eigenvariable is different from a and not contained in t. Then $P'(t)$ (the result of replacing all a's in $P'(a)$ by t) is a proof whose end-sequent is $\Gamma(t) \rightarrow \Delta(t)$.

Proof - by induction on the number of eigenvariables occurring in $P(a)$ which are either a or contained in t, using lemmas 3.1.6 and 3.1.7.

3.1.2 Cut elimination in PA

The addition of the new rule of inference (the induction rule) adds a new complexity to our proofs. First, the rule is semantically infinite as it infers the truth value of a formula for all possible terms from the truth value of a certain instance. However, we are already dealing with semantically infinite rules in the form of the quantifier rules. The main difference is that unlike the quantifiers, which range over a domain, the induction ranges over domains of interpretations, which are isomorphic to the standard interpretation (of the natural numbers) only. Therefore, while the quantifier rules remain sound under all possible interpretations, the induction rule will no longer be sound under some interpretations. For example interpretations, which contain more elements in their domain than those that can be build by '0' and the successor function. The addition of the induction rule is also problematic on the syntactic side. Structural rules manipulate only the quantity and order of formulas. Logical rules affect the logical complexity of formulas. The induction rule, although it may change formulas (according to the semantics of Peano arithmetic), does not always affect the logical complexity of formulas (for example, when the induction term is a variable). Therefore, we cannot use the complexity of formulas in order to show the termination of cut elimination in PA.

Our cut elimination proof in the third section will try to solve this by focusing on proofs, where the induction rule infers formulas with finite numerical terms. Such an induction is defined as grounded and can be replaced by a finite number of cuts. We will prove that all inductions in proofs of end-sequents without strong quantifiers have this property. The main aspect of the proof, which is replacing inductions by a finite number of cuts (when it is possible), is taken from Gentzen's proof of the consistency of Peano arithmetic. In the next two sections we will give Gentzen's proof as presented in [1] and discuss the meaning of the ordinals used.

3.2 The Consistency Proof

3.2.1 Ordinal numbers and the length of derivations

In his proof of the consistency of PA, given in the next section, Gentzen uses the fact, that the length of all derivations in PA is bounded by ordinals smaller than ϵ_0. It will be showed that there can be no infinite decreasing sequence of such ordinals. Gentzen uses this in order to show that proofs of the empty sequent do have an infinite decreasing sequences of these ordinals and therefore, such proofs do not exist. This is showed by transforming any proof of the empty sequent into a proof with a smaller ordinal. The first part of the proof deals with assigning these ordinals to proofs, according to their inference rules.

Before we will assign these ordinals to proofs and discuss their relation to the length of the proofs, we will define the ordinals and prove that the set of all ordinals smaller than ϵ_0 does not contain infinitely decreasing sequences.

The ordinals defined here and assigned to proofs later will be used in order to enable us to prove the consistency of PA by a transfinite induction up to ϵ_0, which means an induction over all ordinals up to ϵ_0.

From now on we will use the name ordinal to refer to those ordinals which are less than ϵ_0. When saying decreasing sequences of ordinals we implicitly mean strictly decreasing.

Ordinals are supposed to represent counting into the transfinite. Being countable, there must be some order among them. Frequently the set $\in$-relation is being used. Ordinals are viewed as sets such that the ordinal ρ contains all the ordinals smaller than itself.

Definition 3.2.1 (Types of ordinals)**.** There are three types of ordinals:

1. The ordinal 0.

2. Successor ordinals, i.e. of the form $\rho \cup \{\rho\}$.

3. Limit ordinals, where the least such ordinal is denoted by ω.

One property of ordinals is their normal-form.

Definition 3.2.2 (Cantor normal-form)**.** For every ordinal ρ other than 0 there are uniquely determined ordinals $\rho_1, .., \rho_n$ such that $\rho = \omega^{\rho_1} + .. + \omega^{\rho_n}$ and $\rho_1 \geq .. \geq \rho_n$.

ϵ_0 is defined to be the minimal ordinal which is equal to its limit, i.e. $\epsilon_0 = min\{\rho \mid \omega^\rho = \rho\}$. This, together with the normal-form of ordinals, means that all ordinals smaller than ϵ_0 can be built up from the symbols $0, +$ and ω only.

Definition 3.2.3 (Relations and operations on ordinals)**.** The following is an inductive definition of $=$, $<$, $+$ and $\cdot$.

1. $<$ is a linear ordering with 0 as its least element.

2. $\omega^\mu < \omega^v$ if and only if $\mu < v$.

3. For ordinals $\mu = \omega^{\mu_1} + .. + \omega^{\mu_k}$ and $v = \omega^{v_1} + .. + \omega^{v_l}$, $\mu + v = \omega^{\mu_1} + .. + \omega^{\mu_k} + \omega^{v_1} + .. + \omega^{v_l}$. If μ and v are given in normal form, then their natural sum $\mu \# v = \omega^{\lambda_1} + .. + \omega^{\lambda_{k+l}}$ where $\{\lambda_1, .., \lambda_{k+l}\} = \{\mu_1, .., \mu_k, v_1, .., v_l\}$ and $\lambda_1 \geq .. \geq \lambda_{k+l}$.

4. Let $\mu = \omega^{\mu_1} + .. + \omega^{\mu_j} + \omega^{\mu_{j+1}} + .. + \omega^{\mu_k}$ such that $\mu_j < \mu_{j+1}$ and let μ' be the ordinal obtained from μ by deleting ω^{μ_j} then $\mu = \mu'$.

5. Let $\mu = \omega^{\mu_1} + .. + \omega^{\mu_k}$ and $v = \omega^{v_1} + .. + \omega^{v_l}$ be in normal form. $\mu < v$ if and only if $\omega^{\mu_i} < \omega^{v_i}$ for some i and $\omega^{\mu_j} = \omega^{v_j}$ for all $j < i$. Or, $k < l$ and $\omega^{\mu_i} = \omega^{v_i}$ for all $i \leq k$.

6. Let μ have the normal form $\omega^{\mu_1}+..+\omega^{\mu_k}$ and $v > 0$ then $\mu{\cdot}\omega^{v} = \omega^{\mu_1+v}$.

7. Let $\mu = \omega^{\mu_1}+..+\omega^{\mu_k}$ and $v = \omega^{v_1}+..+\omega^{v_l}$ then $\mu{\cdot}v = \mu{\cdot}\omega^{v_1}+..+\mu{\cdot}\omega^{v_l}$.

8. $(\omega^{\mu})^n$ is defined as $\omega^{\mu} \cdot .. \cdot \omega^{\mu}$ n times where n is a natural number. I.e. $(\omega^{\mu})^n = \omega^{\mu \cdot n}$.

We can carry on the consistency proof of PA by a transfinite induction on the ordinals up to ϵ_0 as ϵ_0 is accessible.

Definition 3.2.4 (Accessibility of ordinals)**.** An ordinal μ is said to be accessible if it has been demonstrated that every (strictly) decreasing sequence of ordinals starting with μ is finite.

The proof that ϵ_0 is accessible will follow closely the proof given in [1].

Lemma 3.2.5. ϵ_0 is accessible

Proof

1. Every decreasing sequence, which starts with a natural number n, is finite. This is because there cannot be a decreasing sequence of size $> n + 1$.

2. A decreasing sequence of ordinals smaller than $\omega + \omega$ is finite because the first term can be either a natural number or of the form $\omega + n$ where n is a natural number. Therefore, either the first term is finite and we can use the previous argument, or there can be at most $n + 1$ ordinals bigger than ω before we reach a natural number and can use the previous argument again.

3. If ordinals μ and v are accessible, then so is $\mu+v$. Let $\mu+v$ be the first element of a decreasing sequence and assuming $\mu \geq v$. Every element can be either smaller than μ or of the form $\omega^{\mu_1} + v_0$ where ω^{μ_1} is the biggest monomial of μ and v_0 is some ordinal smaller or equal to μ. Those there can be at most v_0 elements bigger than μ before we reach an element smaller than μ.

4. If μ is accessible, then so is $\mu \cdot \omega$. Because $\mu \cdot \omega = \omega^{\mu_1+1}$ (where ω^{μ_1} is the largest monomial in μ), any element in a decreasing sequence starting with ω^{μ_1+1} (except ω^{μ_1+1}) must be either smaller than μ or of the form $\omega^{\mu_1} + v_0$ where v_0 is smaller than μ so there can be only a finite number of elements bigger than μ in the sequence.

5. We will define now the notion of n-accessiblity and continue with the proof.

   Definition 3.2.6 (n-accessibility of ordinals)**.** is defined by induction on n:

   - μ is said to be 1-accessible if μ is accessible.
   - μ is said to be (n+1)-accessible if for every v which is n-accessible, $v \cdot \omega^{\mu}$ is n-accessible.

6. If μ is n-accessible and $v < \mu$, then v is n-accessible. The proof is by induction on n. It is clear that if μ is 1-accessible then v is also 1-accessible. Assuming that for every two ordinals μ_1 and μ_2 such that $\mu_2 < \mu_1$, if μ_1 is n-accessible then also μ_2 is n-accessible and assuming μ is (n+1)-accessible, then for every x which is n-accessible, $x \cdot \omega^{\mu}$ is n-accessible. But as $x \cdot \omega^{v} < x \cdot \omega^{\mu}$, by the induction hypothesis $x \cdot \omega^{v}$ is also n-accessible. Therefore v is (n+1)-accessible.

7. Suppose that $\{\mu_m\}$ is an increasing sequence of ordinals with limit μ. If each μ_m is n-accessible then so is μ. This is true because either $\mu \in \{\mu_m\}$ or $\mu = \omega^{v+1}$ and therefore we need to prove that for every (n-1)-accessible x, $x \cdot \omega^{\omega^{v+1}}$ is (n-1)-accessible. As there is $\mu_0 \in \{\mu_m\}$ such that $\mu_0 = \omega^{v} + v_0$ and as it is n-accessible we know that $x \cdot \omega^{\omega^{v}+v_0}$ is (n-1)-accessible. Therefore also $x \cdot \omega^{\omega^{v}+v_0} \cdot \omega^{\omega^{v}+v_0} = x \cdot \omega^{\omega^{v}+\omega^{\omega^{v}+v_0}}$ is (n-1)-accessible, but it is bigger than $x \cdot \omega^{\omega^{v+1}}$ and according to (6) $x \cdot \omega^{\omega^{v+1}}$ is (n-1)-accessible as well.

8. If v is (n+1)-accessible, then so is $v \cdot \omega$. Here we must show that for every n-accessible x, $x \cdot \omega^{v \cdot \omega}$ is n-accessible. Due to (7), it is enough to show that for every m, $x \cdot \omega^{v \cdot m}$ is n-accessible. But $x \cdot \omega^{v \cdot m} = x \cdot (\omega^{v})^{m} = x \cdot \omega^{v} .. \omega^{v}$. As v is (n+1)-accessible, we know that $x \cdot \omega^{v}$ is n-accessible and therefore also $x \cdot \omega^{v} \cdot \omega^{v}$ is n-accessible. Using this step m times we have that $x \cdot (\omega^{v})^{m}$ is n-accessible.

9. 1 is (n+1)-accessible. Because let some μ be n-accessible, then from (8) $\mu \cdot \omega = \mu \cdot \omega^{1}$ is n-accessible and therefore 1 is (n+1)-accessible.

10. Given that $\omega_0 = 1$ and $\omega_{n+1} = \omega^{\omega_n}$ then ω_k is (n-k)-accessible for all $n > k$. This is proved by induction on k. If $k = 0$ then $w_k = 1$ and it is n-accessible for all n according to (9). Now suppose ω_k is (n-k)-accessible, then, as 1 is (n-(k+1))-accessible, we have $1 \cdot \omega^{\omega_k}$ is (n -(k+1))-accessible, i.e. ω_{k+1} is (n - (k+1))-accessible.

11. As a special case of (10) we have that ω_k is accessible for every k.

12. For every decreasing sequence smaller than ϵ_0, there is an ordinal ω_k such that it is bigger than all elements in the sequence. Therefore, we can conclude that ϵ_0 is accessible.

Now we can give Gentzen assignment of ordinals to proofs.

Definition 3.2.7 (Grade and height of formulas and sequents)**.** The grades of formulas and inferences and the height of sequents are defined as follows:

1. The grade of a formula is the number of logical symbols that this formula contains. The grade of a cut is the grade of the cut formula. The grade of an ind is the grade of the induction formula.

2. The height of a sequent S in a proof P denoted by $h(S,P)$ or $h(S)$ is the maximum of the grades of the cuts and inds, which occur in P below S.

Let $\omega_k(x)$ be defined inductively as follows:

- $\omega_0(x) = x$.
- $\omega_{k+1}(x) = \omega^{\omega_k(x)}$.

Definition 3.2.8 (Gentzen's assignment of ordinals)**.** The ordinal assigned to a proof P of a sequent S, denoted by $o(S,P)$ or $o(S)$ is defined as follows:

1. An initial sequent is assigned the ordinal 1.

2. If S is the lower sequent of a weak inference (structural other than a cut), then o(S) is the same as the ordinal of its upper sequent.

3. If S is the lower sequent of $\vee : r$, $\neg : r$, $\neg : l$, $\forall : r$, $\forall : l$ and the upper sequent has ordinal μ, then $o(S) = \mu + 1$.

4. If S is the lower sequent of $\vee : l$ and the upper sequents have ordinals μ and v then $o(S) = \mu \# v$.

5. If S is the lower sequent of a cut and its upper sequents have the ordinals μ and v then $o(S) = \omega_{k-l}(\mu \# v)$, where k and l are the heights of the upper sequents and of S, respectively.

6. If S is the lower sequent of an ind and its upper sequent has the ordinal μ then $o(S) = \omega_{k-l+1}(\mu_1 + 1)$, where k and l are the heights of the upper sequent and of S, respectively and μ has the normal form $\omega^{\mu_1} + .. + \omega^{\mu_n}$.

7. The ordinal of a proof P is the ordinal of its end-sequent.

The ordinals given here can be seen as a bound on the length of proofs of the specific sequents. For example, the use of an unary logical inference increases the length by one, while weak inferences do not increase it at all. The intuition is less clear when it comes to cuts and inductions. In order to estimate the length of proofs, we will refer to the number of rules in a cut and induction free proof. The cut elimination proof for classical logic given by Tait [2] places a bound on the size of the cut-free proofs, according to the most complex cut formula and the initial size of the proof.

Theorem 3.2.9 (Tait's cut elimination for finitary predicate logic)**.** Given a proof P of sequent S with length d and cut complexity ρ, there is a cut-free proof P' of S with length $\leq 2^d_\rho$ where $2^\epsilon_0 = \epsilon$, $2^\epsilon_{k+1} = 2^{2^\epsilon_k}$.

We notice that the induction rule can be replaced by a transfinite derivation:

$$\dfrac{\dfrac{A(0) \vdash A(0) \qquad \dfrac{\ldots}{A(0) \vdash A(1)}}{A(0) \vdash A(1)} \qquad \dfrac{\ldots}{A(1) \vdash A(2)}}{A(0) \vdash A(2)}$$

and so on.

By applying the cut rule with cut formula $A(\overline{n})$, we can replace the induction by a possibly infinite number of cuts $\leq \omega$.

Since all derivations in PA are finite and contains only finite cut formulas and since there are only a finite number of inductions in the proof, we obtain a proof without inductions of length $\leq \omega^2$. Therefore, any proof can be transformed into a cut-free proof of length $\leq 2^{\omega^2}_k < \epsilon_0$ for some k.

Now the bounds given by Gentzen can be better understood:

- Let ρ be the grade of a maximal-grade cut. Let μ and v be the respective lengths of the sub-derivations of its upper sequents. Eliminating this cut can amount, according to Tait, to length $\leq 2^d_\rho < \omega_\rho(\mu \# v)$.

- Let ρ be the grade of an induction formula . Let μ be the length of the sub-derivation of the induction's upper sequent. Eliminating this induction increases the length to length $\leq 2^{\omega^2}_k < \omega_{\rho+1}(\mu_1 + 1)$.

The specific bounds themselves were chosen according to the requirements of the proof given next.

3.2.2 The consistency proof of PA

Gentzen's proof of the consistency of PA starts by assuming there is a proof in PA of the empty sequent and then showing that this proof can be reduced to another proof of the empty sequent of smaller ordinal (bound on length). As the ordinals are accessible, we can reduce proofs of the empty sequent only a finite number of times, which leads to a contradiction.Therefore, we conclude that there is no proof in PA of the empty sequent and that PA is consistent.

Before we will carry on with this proof, which will follow closely the proof given in [1] (based on Gentzen's original proof [9]), we will give several more definitions about proofs in PA.

Definition 3.2.10 (Bundles)**.** When referring to the occurrence of a formula in another formula, sequent or proof, we will refer to it as a formula in the other formula, sequent or proof.

1. Successor - For a given formula E in the upper sequent of an inference rule, its successor is defined as follows:
   (a) If E is a cut formula, it has no successor.
   (b) If E is an auxiliary formula of any inference rule other than a cut or exchange, then the principal formula of the rule is the successor of E.
   (c) In the exchange rule given in definition 2.1.14, where the formulas being exchanged are denoted by C and D, the successors for C and D in the upper sequent are C and D (respectively) in the lower sequent.
   (d) Given Γ, Π, Δ and Λ as in definition 2.1.14, if E is the kth formula of Γ, Π, Δ or Λ in the upper sequent, then its successor is the kth formula of Γ, Π, Δ or Λ (respectively) in the lower sequent.
2. A sequent formula is called an initial formula or an end-formula if it occurs, respectively, in an initial sequent or an end-sequent.
3. Bundle - a sequence of formulas in a proof with the following properties is called a bundle:
   (a) The sequence begins with an initial formula or a weakening formula.
   (b) The sequence ends with an end-formula or a cut-formula.

   (c) Every formula in the sequence except the last one is immediately followed by its successor.

4. Ancestor and descendant - let A and B be two formula. A is called ancestor of B and B a descendent of A if A appears above B in the same bundle.

5. Predecessor - If A is the successor of B then B is the predecessor of A.

Definition 3.2.11 (Explicit and implicit bundles)**.** The concept of explicit and implicit bundles:

1. A bundle is called explicit if it ends with an end-formula.

2. It is called implicit if it ends with a cut-formula.

3. An occurrence of a formula in a proof is called implicit if it is contained in an implicit bundle, it is called explicit otherwise.

4. A sequent in a proof is called implicit if it contains a formula from an implicit bundle. Otherwise, it is called explicit.

5. A logical inference in a proof is called explicit or implicit if its principal formula is explicit or implicit (respectively).

Definition 3.2.12 (End-pieces and boundaries)**.** The concept of end-pieces and boundaries:

1. The end-piece of a proof is defined as follows:

   (a) The end-sequent of the proof is contained in the end-piece.
   (b) The upper sequent of an inference other than an implicit logical inference is contained in the end-piece if and only if the lower sequent is contained in the end-piece.
   (c) The upper sequent of an implicit logical inference is not contained in the end-piece.

2. An inference in a proof is said to be in the end-piece if its lower sequent is in the end-piece.

3. Boundary - Let J be an inference in the proof. J is said to belong to the boundary of the proof if its lower sequent is in the end-piece and its upper sequent is not. It should be noted that J must be an implicit logical inference.

4. Suitable cut - a cut in the end-piece is called a suitable cut if each one of the two cut-formulas of this cut has an ancestor, which is the principal formula of a boundary inference.

The proof carries on the reductions by transforming the proof into another proof, such that the ordinal assigned to it (as defined in definition 3.2.8) is decreased. This is done by replacing first all inductions, which appear in the end-piece. Each elimination will create a new bound for the proof, which is strictly smaller than before. We will have to ensure that there are no free variables in the induction term, before we will be able to eliminate an induction. This way the induction is only over a finite number of elements.

Definition 3.2.13 (Grounded induction)**.** An induction is called a grounded induction if the induction term does not contain any free variables. By induction term we refer to s in the following instance of the rule:

$$\frac{\overbrace{F(a), \Gamma \vdash \Delta, F(a')}^{P(a)}}{F(0), \Gamma \vdash \Delta, F(s)} \text{ (induction)}$$

Given a grounded induction, as the term can be evaluated into a numeral, there is a procedure to eliminate it by replacing the induction with a finite number of consecutive cuts.

Definition 3.2.14 (The elimination of grounded inductions)**.** Replacing the derivation

$$\frac{\overbrace{F(a), \Gamma \vdash^{S} \Delta, F(a')}^{P(a)}}{F(0), \Gamma \vdash^{S_0} \Delta, F(s)} \text{ (induction)}$$

by

$$\dfrac{\dfrac{\dfrac{\overbrace{F(0), \Gamma \vdash^{S_1} \Delta, F(\overline{1})}^{P(0)} \qquad \overbrace{F(\overline{1}), \Gamma \vdash \Delta, F(\overline{2})}^{P(0')}}{F(0), \Gamma \vdash^{S_2} \Delta, F(\overline{2})} \text{ (cut)}}{\vdots}}{} \qquad$$

$$\frac{F(0), \Gamma \vdash^{S_n} \Delta, F(\overline{n}) \qquad \overbrace{F(\overline{n}) \vdash F(s)}^{Q}}{F(0), \Gamma \vdash^{S_0} \Delta, F(s)} \text{ (cut)}$$

This procedure will be used in the consistency proof in order to reduce the ordinal of the proof and also in the next section of the thesis, which deals with more general proofs in PA.

The major part of the proof will deal with showing that proofs of the empty sequent are special in a way that they must contain (once inductions and other parts are removed) a suitable cut. This cut is then duplicated in a way that will reduce the ordinal of the proof.

The first property of proofs of the empty sequent that we will show, is that they cannot be "simple".

Definition 3.2.15 (Simple proofs)**.** A proof in PA is simple if no free variable occurs in it and it contains only mathematical initial sequents, weak inferences and inessential cuts.

Lemma 3.2.16. There is no simple proof of $\vdash$.

Proof

- As we have only inessential cuts in the proof and the end-sequent is empty, all formulas in any simple proof of the empty sequent are of the form $s = t$ and are closed. We give a value T to all sequents, in which at least one formula in the antecedent is false or at least one formula in the succedent is true, otherwise it gets the value F. Clearly all mathematical initial sequents get the value T and $\vdash$ is of value F. Clearly exchanges, weakenings and contractions preserve the value as they either change position, add a new formula or delete a duplicated formula. Inessential cuts also preserve the value T as there must be another formula in the context of one of the upper sequents which render the sequent true and appears also in the lower sequent of the cut.

The following lemma will enable us to replace a sub-derivation in a proof by one of lower ordinal and have the whole proof be of lower ordinal.

Lemma 3.2.17. Let P be a proof containing a sequent S_1 and P_1 be a sub-proof of P ending with S_1, such that there is no induction below S_1. Let P_1' be another proof of S_1 and P' be the proof formed by replacing P_1 by P_1'. If $o(S_1, P') < o(S_1, P)$ then $o(P') < o(P)$ (and the same for $\leq$).

Proof

- We will prove by induction on the length of the derivation below S_1, that each sequent S below S_1 is assigned in P' a smaller ordinal than in P.
- This is true for $S = S_1$ by assumption.
- Assuming it is true for derivations of length n, we prove it for derivations of length $n+1$. All inferences except induction are monotonic in with regard to the ordinals assigned to their upper sequents and lower sequent. Therefore, the inequality is preserved.

Before we prove the central lemma of the consistency proof, we remark about the following property of end-pieces of proofs of the empty sequent. All logical inferences must be implicit (as their bundles must end with a cut formula due to the sub-formula property). Therefore, we can easily mark the boundary between the end-piece and the rest of the proof by the first logical inference we encounter when we go up from the end-sequent.

Another assumption is that eigenvariables in proofs are unique and do not appear below the inference rules that have eliminated them.

Definition 3.2.18 (Regular proofs)**.** A proof P is regular if:

1. The eigenvariables of any two distinct $\forall : r$ or induction inference rules are distinct from each other.
2. If a free variable occurs as an eigenvariable of a sequent S of proof P, then it appears in P only in sequents above S.

Lemma 3.2.19. There exists a regular proof for any proof in PA. The regular proof can be obtained by using only a finite process of replacing free variables.

Proof

- By induction on the number l of inference rules of the form $\forall : r$ or ind.
- If $l = 0$ then there is no free variable which is used as an eigenvariable and we are done.
- Otherwise, let us assume it holds for all proofs with at most l such inferences and we prove it for $l + 1$ inferences. We label all lowermost such inferences $J_1, .., J_k$ in P. Looking on all sub-proofs of P, which

end with J_i, we can apply the induction hypothesis in order to obtain a regular sub-proofs, which end with those rules. Making sure none of the regular sub-proofs P_i' contains eigenvariables from another sub-proof by changing them if they do. The last step is to replace all free variables, which are used as eigenvariables in one of the J_i and occur below the J_i by new free variables.

The following lemma will be stated now and a more general version of it will be proved in the next section (in lemma 4.1.22):

Lemma 3.2.20. Given

$$\cfrac{\cfrac{\cdots}{D, \Pi' \vdash \Lambda'}\\\cdots}{R[D]_\lambda, \Pi \vdash \Lambda}$$

where the bundle containing D and $R[D]_\lambda$ does not contain principal formulas of the contraction rule between the two formulas, we can obtain

$$\cfrac{\cfrac{\cdots}{\Pi' \vdash \Lambda'}\\\cdots}{\Pi \vdash \Lambda}$$

with the possible addition of weak inferences only.

The following is the proof of the reduction step. The reduction is made of a sequence of steps. Each step is performed a finite number of times and only after the previous steps were exhausted. None of the steps increases the ordinal and at least one step decreases it. At the end we obtain a proof of a strictly lower ordinal.

Lemma 3.2.21. If P is a proof of $\vdash$, then there is another proof P' of $\vdash$ such that $o(P') < o(P)$.

Proof

1. The first step is to replace all free variables, which are not used as eigenvariables in the end-piece by constants. According to lemma 3.1.7, for each such free variable we can obtain a proof of $\vdash$ where the free variable was replaced by 0. Therefore we can obtain a proof without free variables in the end-piece, which are not being used as eigenvariables.

2. In case we have an induction in the end-piece, we take a lowermost such induction I with upper sequent S and lower sequent S_0:

$$\dfrac{\overset{\displaystyle\underbrace{P(a)}}{F(a), \Gamma \vdash^{S} \Delta, F(a')}}{F(0), \Gamma \vdash^{S_0} \Delta, F(s)}\ (I)$$

As we deal with the end-piece, there are no quantifiers below I and therefore no free variables, which are used as eigenvariables. Because we have eliminated all the remaining free variables as well, the term s is a closed term and by lemma 3.1.5 we know that there exists a numeral $\overline{n}$ such that $\vdash s = \overline{n}$ and $F(\overline{n}) \vdash F(s)$ are provable without an essential cut or ind. Using the transformation that was given in definition 3.2.14, we can replace the induction by a derivation of length $\overline{n}$ containing $\overline{n}$ cuts. Assuming $o(S) = \mu$, the ordinal which was assigned to the lower sequent S_0 of I was $o(S_0) = \omega_{l-k+1}(\mu_1 + 1)$. We notice that all the sequents in the new derivation according to 3.2.14 have the same height l, since all the formulas of the form $F(m)$ for $m = 0, .., \overline{n}$ have the same grade. Therefore, for every $m = 0, .., \overline{n}$ $o(F(m), \Gamma \vdash \Delta, F(m')) = \mu$. Because Q has no essential cut or ind, $o(F(\overline{n}) \vdash F(s), P') = q < \omega$ (where P' is the proof after the transformation given in 3.1.5). Now we can see that each of the lower sequents S_i for $i = 2, .., \overline{n}$ of the new cuts that replaced the ind is of ordinal $o(S_i) = \mu * i$ (where $\mu * i = \mu \# .. \# \mu$ i times). As $\mu * \overline{n} + q < \omega^{\mu_1 + 1}$ we get $o(S_0, P') = \omega_{l-k}(\mu * \overline{n} + q) < \omega_{l-k+1}(\mu_1 + 1) = o(S_0, P)$ and by lemma 3.2.17 $o(P') < o(P)$. So as long there is an ind in the end-piece, we can obtain a proof with a smaller ordinal, otherwise we can assume there is no ind in the end-piece and continue to the next step.

3. In case the end-piece contains an initial logical sequent $D \vdash D$. Both Ds must be cut somewhere above the empty sequent. Assuming the D in the antecedent is cut first (from above), we have:

$$\dfrac{\dfrac{\dfrac{\cdots}{\Gamma \vdash \Delta, D} \qquad \dfrac{\dfrac{D \vdash D}{\cdots}}{D, \Pi \vdash \Lambda}}{\Gamma, \Pi \vdash^{S} \Delta, \Lambda}\ \text{(cut)}}{\dfrac{\cdots}{\vdash}}$$

and as the D in the succedent is cut somewhere below S, we can replace it by:

$$\begin{array}{c} \cdots \\ \hline \Gamma \vdash \Delta, D \\ \hline \text{weakenings and exchanges} \\ \hline \Gamma, \Pi \vdash^{S'} \Delta, \Lambda \\ \hline \cdots \\ \hline \vdash \end{array}$$

We have removed a cut and added weak inferences only, therefore $o(S', P') < o(S, P)$ and by lemma 3.2.17 $o(P') < o(P)$. So we can assume, after applying this step a finite number of times, that there are no logical initial sequents in the end-piece.

4. If there are weakenings in the end-piece. Taking a lowermost weakening W, as the end-sequent is empty, the weakening's formula must be cut below W. I.e. we have:

$$\dfrac{\dfrac{\cdots}{\Gamma \vdash^{S_l} \Delta, R[D]_\lambda} \qquad \dfrac{\dfrac{\dfrac{\dfrac{\cdots}{\Pi' \vdash \Lambda'}}{D, \Pi' \vdash \Lambda'}\,(\mathrm{W})}{\cdots}}{R[D]_\lambda, \Pi \vdash^{S} \Lambda}}{\Gamma, \Pi \vdash^{S_0} \Delta, \Lambda}\,(\mathrm{cut})$$

We have two cases:

(a) There is no contraction in the bundle, which contains D and $R[D]_\lambda$ between W and the cut. Using lemma 3.2.20 we can obtain:

$$\begin{array}{c} \cdots \\ \hline \Pi' \vdash \Lambda' \\ \hline \cdots \\ \hline \Pi \vdash^{S'} \Lambda \\ \hline \text{weakenings and exchanges} \\ \hline \Gamma, \Pi \vdash^{S_0} \Delta, \Lambda \end{array}$$

Given $h(S,P)=k$ and $h(S_0,P)=l$ where $l \leq k$, then $h(S_0,P')=h(S',P')=l$. Hence, the height of all sequents above S_0 in P' is at most the same as the corresponding sequent in P. I.e. for every sequent S_1 in P above S_0 and the corresponding sequent S_1' in P' we have $h(S_1,P) \geq h(S_1',P')$. By induction on the length of the proof down to S_0 (and also because the process of lemma 3.2.20 never adds non-weak inferences), we can show that for every sequent S_1 and its corresponding S_1' in P', we have $o(S_1,P) \geq o(S_1',P')$ and those $o(S,P) \geq o(S',P')$. Therefore, given that $o(S_l,P)=\mu_l$, $o(S,P)=\mu$ and $o(S',P')=\mu'$ then $o(S_0,P)=\omega_{k-l}(\mu_l \# \mu) > \mu \geq \mu' = o(S',P') = o(S_0,P')$ and by lemma 3.2.17 $o(P') < o(P)$.

(b) In case there is a contraction in the thread containing S, let C be the uppermost contraction and we have:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\cdots}{\Gamma' \vdash \Lambda'}}{D,\Gamma' \vdash \Lambda'}\,(W)}{\cdots}}{R[D]_\lambda, R[D]_\lambda, \Gamma'' \vdash \Lambda''}}{R[D]_\lambda, \Gamma'' \vdash^{S} \Lambda''}\,(C)}{\dfrac{\cdots}{P[R[D]_\lambda]_\delta, \Gamma \vdash^{S_0} \Lambda}}$$

Using lemma 3.2.20 on the first part of the proof, we can obtain:

$$\dfrac{\dfrac{\dfrac{\dfrac{\cdots}{\Gamma' \vdash \Lambda'}}{\cdots}}{R[D]_\lambda, \Gamma'' \vdash^{S'} \Lambda''}}{\dfrac{\cdots}{P[R[D]_\lambda]_\delta, \Gamma \vdash^{S_0} \Lambda}}$$

By the same argument of the previous case, we have $o(S',P') \leq o(S,P)$ and by lemma 3.2.17 $o(P') \leq o(P)$.

Therefore, we can conclude that there are no weakenings in the end-piece.

5. It remains to show that there must exist a suitable cut, such that doubling it will reduce the ordinal of the proof. First we notice that P cannot be its own end-piece as otherwise it would be simple and by lemma 3.2.16 there is no simple proof of the empty sequent. The proof would be simple because up to this step it will not contain free

variables, non mathematical initial sequents or weakenings. According to the definition of an end-piece (in proofs of the end-sequent) it cannot contain any logical inference, therefore (and because there are no logical initial sequents) all cuts are inessential. So the proof is not its own end-piece and therefore, as there is at least one implicit logical inference in P (a boundary inference), there is an essential cut in the end-piece, which is suitable. The following lemma states that such a suitable cut exists:

Lemma 3.2.22. Let proof P satisfies the following requirements:

1. P is not its own end-piece.
2. The end-piece of P does not include any logical inference, inductions or weakenings.
3. The end-piece of P does not include any logical axiom.

Then there is a suitable cut in the end-piece of P.

It is apparent that all these requirements hold in our proof of $\vdash$.

Proof

- The proof is by induction on the number of essential cuts in the end-piece.
- If there is only one essential cut C (and there is at least one as discussed above), then it is suitable. The cut formula of this cut contains a predicate symbol other than $=$ and there are no weakenings. This is the only cut in the end-piece which has another cut-formula other than an equation. In case the two cut-formulas are not descendant of boundary inferences and because there is no other essential cut in the end-piece, they must be descendant of logical axioms which contain logical symbols. But this is in contradiction to requirement (3). These inferences must be the lowest such inferences as there is only one essential cut in the end-piece.
- If there are more than one essential cut, take a lowermost one (C). If it is a suitable cut then we are done, otherwise we have:

$$\cfrac{\overbrace{\Gamma \vdash \Delta, D}^{P_1} \qquad \overbrace{D, \Pi \vdash \Lambda}^{P_2}}{\Gamma, \Pi \vdash \Delta, \Lambda}\ (C)$$

As C is not a suitable cut, at least one of the Ds is not a descendant of the principal formula of a boundary inference. We assume that the D in the succedent has this property. Next we prove that: (1) P_1 must contain a boundary inference of P, (2) all boundary inferences of P, which are contained in P_1, are also boundary inferences of P_1 and (3) the end-piece of P_1 is contained in the end-piece of P. Using the induction hypothesis, as P_1 contains one less essential cut, we can conclude that it contains a suitable cut. Because the boundaries in P_1 are also the boundaries in P, it is also a suitable cut in P.

1. P_1 contains a boundary inference of P. This is true because we know that there must be an implicit inference above $\Gamma \vdash \Delta, D$ which is a boundary inference of P. Otherwise, due to requirements (2) and (3), D is a descendant of a logical axiom which occurs in the end-piece of P, which is a contradiction. The inference formula cannot be an ancestor of any of the formulas of $\Gamma \vdash \Delta, D$. This is because we assume it is not of D and if it was of any of the other formulas, there should have been another essential cut below C, which contradicts the fact that it is a lowermost essential cut. Therefore, there must be an essential cut in P_1 and the boundary inference of P is also a boundary inference of P_1.

2. If an inference J in P_1 is a boundary inference of P, then it is a boundary inference of P_1. This is because C is a lowermost essential cut and it cuts D which is not a descendant of a principal formula of a boundary inference. Therefore, all boundary inferences in P_1 must contain a principal formula that ends in an essential cut which is also in P_1.

3. P_1 is not its own end-piece and the end-piece of P_1 is the intersection of P_1 and the end-piece of P. From requirement (1) we know that P is not its own end-piece. Furthermore, from the previous two results we know that as the boundary of P_1 is "contained" in the boundary of P, then the end-piece of P_1 is fully contained in the end-piece of P.

As mentioned before, we can use the induction hypothesis in order to obtain a suitable cut, which is also a suitable cut in P.

Back to our proof. As our current proof of $\vdash$ satisfies all the requirements of the lemma, we have a suitable cut in the end-piece and we define an essential reduction of P as following. By taking a lowermost suitable cut in the end-piece of P (called C), we see that there must be logical rules in both sub-derivations of C with ancestors of the cut-formula of C as principal formulas. The idea of the essential reduction is to replace these logical rules by an additional cut, which is below C and which has a cut-formula of smaller grade than the cut-formula of C. Then it can be shown, that the ordinal of all sequents in the lower part of the proof is strictly smaller by this change. This is true as the height of the original cut did not change and we have made its ordinal smaller. Although there are new cuts, the ordinals of all sequents must be smaller than in the original proof. We will prove by cases according to the outermost logical symbol of the cut -formula of C. We will examine here only $\vee$ and $\forall$:

1. If the outermost symbol is $\vee$, then P is of the form:

$$\dfrac{\dfrac{\dfrac{\dfrac{\cdots}{\Gamma' \vdash \Delta', A}}{\Gamma' \vdash \Delta', A \vee B}\,(\vee : r(I_1))}{\dfrac{\cdots}{\Gamma \vdash^{S_1} \Delta, A \vee B}} \qquad \dfrac{\dfrac{\dfrac{\cdots}{A, \Pi_1' \vdash \Lambda_1'} \quad \dfrac{\cdots}{B, \Pi_2' \vdash \Lambda_2'}}{A \vee B, \Pi' \vdash \Lambda'}\,(\vee : l(I_2))}{\dfrac{\cdots}{A \vee B, \Pi \vdash^{S_2} \Lambda}}}{\dfrac{\Gamma, \Pi \vdash^{S} \Delta, \Lambda}{\dfrac{\cdots}{\Theta \vdash^{S_0} \Xi}}}\,(C)$$

   S_0 is the uppermost sequent below C whose height k is less than the height l of S_1 and S_2. S_0 exists as:

   (a) If all cuts below C are lower than the grade of $A \vee B$, then $S_0 = S$.

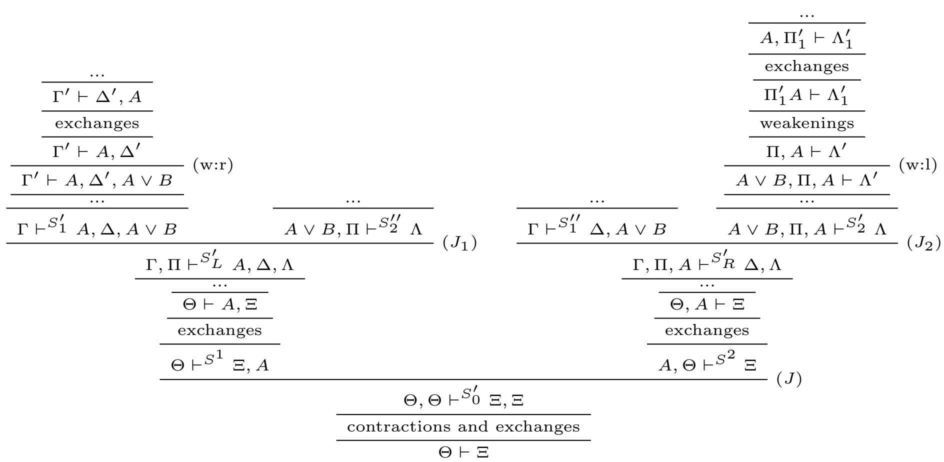
   (b) If not the above, then take the highest cut below S and its lower sequent is S_0.

   (c) We know that either (a) or (b) must be true as the height of the end-sequent is 0 and the height of S is bigger than 0.

   (d) The lower sequent chosen in (a) or (b) may be the end-sequent as well.

S_0 must be a cut as there are no inductions in the end-piece.

We will transform P into P' in such a way that in P' the two $\vee$ inferences will be removed and we will add a new cut below S_0:



The first thing we note is that $h(S_1') = h(S_1'') = h(S_2') = h(S_2'') = h(S_1) = h(S_2) = l$ as we have added only one new cut (J) and its grade is smaller than that of J_1 and J_2. It is also clear than $h(S_0') = h(S_0) = k$. Let $h(S^1) = h(S^2) = m$. It is clear that $m = max(k, grade(A)) < l$.

On the other hand, as we have eliminated one logical inference in each sub-derivation, $o(S_1') < o(S_1)$ and $o(S_2') < o(S_2)$, while $o(S_1'') = o(S_1)$ and $o(S_2'') = o(S_2)$.

Now, given an arbitrary inference

$$\frac{F_1 \qquad F_2}{F}\ (I)$$

in P and its corresponding inference

$$\frac{F_1' \qquad F_2'}{F'}\ (I')$$

in P', such that $h(F_1') = h(F_2') = k_1$ and $h(F') = k_2$. We will show by induction on the length of the derivation, starting with S and ending with S_0, that $o(F', P') < \omega_{l-k_2}(o(F, P))$.

- The idea of this induction is that although we have added a new cut, the accumulated height of the cuts stays the same. This is because the cut we have added is of smaller grade than some other cuts above it. Although we have in the new proof two derivations, they are above a cut whose ordinal is bigger than ω. Therefore, the cut's ordinal cancels the fact that we have two sub derivations instead of one in the original proof.
- We will also use the following two properties of ordinals:

Lemma 3.2.23. Let μ and v be ordinals in normal form different from 0, the following is true:

(a) $\omega_n(\omega_m(\mu)\#\omega_m(v)) \leq \omega_{n+m}(\mu\#v)$.

(b) If $\omega^\mu > v_1$ and $\omega^\mu > v_2$ then $\omega^\mu > v_1\#v_2$.

Proof

- $<$ on ordinals given in normal form is defined such that (1) $\omega^a > \omega^b$ if and only if $a > b$ and (2) By the first monomial that is not equal in the two ordinals. Therefore, in (a) we have $\omega_n(\omega_m(\mu)\#\omega_m(v)) \leq \omega_{n+m}(\mu\#v)$ if and only if $\omega_m(\mu)\#\omega_m(v) \leq \omega_m(\mu\#v)$. They are clearly equal if $m = 0$ and if $m > 0$ then the only monomial in $\omega_m(\mu\#v)$ is clearly bigger than the monomials of $\omega_m(\mu)\#\omega_m(v)$. (b) is true, because we have only one monomial in the bigger ordinal and we know it is strictly bigger than all monomials in the other ordinals. Therefore, it does not matter how many smaller monomials we add to the smaller ordinal.

Back to our inductive proof.

- The base case is clear as $o(S'_L, P') = \omega_{l-k_2}(o(S'_1)\#o(S''_2)) < \omega_{l-k_2}(o(S_1)\#o(S_2)) = \omega_{l-k_2}(\omega_{l-l}(o(S_1)\#o(S_2)) = \omega_{l-k_2}(o(S,P))$.
- We need to prove the step only for cuts, because the ordinals of the rest of the inferences (we dont have inductions) do not include heights and clearly keep the inequality. First we note that for all cuts in P in the specific derivation, except the last one, we have $o(F,P) = o(F_1,P)\#o(F_2,P)$, because all cuts there have the same height. The last cut has $o(S_0) = \omega_{l-k}(o(F_1,P)\#o(F_2,P))$. In P' we have for all cuts $o(F',P') = \omega_{k1-k2}(o(F_1,P)\#o(F_2,P))$. Now assuming we have $o(F',P') < \omega_{l-k_2}(o(F,P))$ for the previous inference. Because the height for the current upper sequents

(in P') is equal to the height of the lower sequent in the previous cut, we have $o(F', P') =$
$\omega_{k_1-k_2}(o(F'_1, P')\#o(F'_2, P')) <$
$\omega_{k_1-k_2}(\omega_{l-k_1}(o(F_1, P))\#\omega_{l-k_1}(o(F_2, P))) \leq_{3.2.23(i)}$
$\omega_{k_1-k_2+l-k_1}(o(F_1, P)\#o(F_2, P)) =$
$\omega_{l-k_2}(o(F_1, P)\#o(F_2, P)) = \omega_{l-k_2}(o(F, P))$ for all cuts except the last.

- For the last cut in the derivation (with lower sequent S_0 in P and the corresponding S^1 and S^2 in P'), we assume the following:
  (a) That the ordinals for the upper sequents of S_0 sums to O.
  (b) That the ordinals for the upper sequents of S^1 sums to O'.
  (c) That the height of the upper sequent of S^1 is m_0.
  (d) Using the fact that above S_0 we have $o(F', P') < \omega_{l-k_2}(o(F, P))$

  We have $o(S_0, P) = \omega_{l-k}(O) = \omega_{l-m_0+m_0-k}(O) = \omega_{m_0-k}(\omega_{l-m_0}(O)) > \omega_{m_0-k}(O') = \omega_{m_0-m+m-k}(O') = \omega_{m-k}(o(S^1, P'))$. Because $o(S_0, P) = \omega_{l-k}(O)$, we also have that $o(S^1, P') < \omega_{l-m}(O)$ and $o(S^2, P') < \omega_{l-m}(O)$.

- The ordinal of S'_0 is therefore $o(S'_0, P') = \omega_{m-k}(o(S^1, P')\#o(S^2, P')) <_{3.2.23(ii)} \omega_{m-k}(\omega_{l-m}(O)) = o(S_0, P)$ as $l > m$. Using lemma 3.2.17 $o(P') < o(P)$.

2. If the outermost symbol is $\forall$ then P is of the form:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\cdots}{\Gamma' \vdash \Delta', F(a)}}{\Gamma' \vdash \Delta, \forall x F(x)}}{\cdots}}{\Gamma \vdash \Delta, \forall x F(x)} \qquad \dfrac{\dfrac{\dfrac{\dfrac{\cdots}{F(s), \Pi' \vdash \Lambda}}{\forall x F(x), \Pi' \vdash \Lambda'}}{\cdots}}{\forall x F(x), \Pi \vdash \Lambda}}{\dfrac{\Gamma, \Pi \vdash \Delta, \Lambda}{\dfrac{\cdots}{\Theta \vdash \Xi}}}$$

We define P' as before, by deleting the $\forall$ inferences and adding a cut below the highest cut below $\Gamma, \Pi \vdash \Delta, \Lambda$:

$$
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\cdots}{\Gamma' \vdash \Delta', F(s)}}{\text{exchanges}}}{\Gamma' \vdash F(s), \Delta'}}{\Gamma' \vdash F(s), \Delta, \forall x F(x)}\,(\text{w:r})}{\cdots}}{\Gamma \vdash F(s), \Delta, \forall x F(x)} \quad \dfrac{\cdots}{\forall x F(x), \Pi \vdash \Lambda}}{\Gamma, \Pi \vdash F(s), \Delta, \Lambda}}{\cdots}}{\Theta \vdash \Xi, F(s)} \quad \dfrac{\dfrac{\dfrac{\dfrac{\cdots}{\Gamma \vdash \Delta, \forall x F(x)} \quad \dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\cdots}{F(s), \Pi' \vdash \Lambda}}{\text{exchanges}}}{\Pi', F(s) \vdash \Lambda}}{\forall x F(x), \Pi', F(s) \vdash \Lambda'}\,(\text{w:l})}{\cdots}}{\forall x F(x), \Pi, F(s) \vdash \Lambda}}{\Gamma, \Pi, F(s) \vdash \Delta, \Lambda}}{\cdots}}{F(s), \Theta \vdash \Xi}}{\Theta, \Theta \vdash \Xi, \Xi}}{\text{contractions and exchanges}}}{\Theta \vdash \Xi}
$$

Because $o(\Gamma' \vdash \Delta', F(s)) = o(\Gamma' \vdash \Delta', F(a))$, we can follow the proof of case (a) and obtain that $o(\Theta \vdash \Xi, P') < o(\Theta \vdash \Xi, P)$ and by lemma 3.2.17 $o(P') < o(P)$.

We have showed that upon reaching the fifth step, there must be a suitable cut in P. The last step showed how to transform this proof into a proof of a smaller ordinal. As none of the steps increases the ordinal and some of them, including the last one, decrease it, we have effectively showed a transformation of P with a smaller ordinal.

By using this lemma and the accessibility of ϵ_0, we prove:

Corollary 3.2.24. PA is consistent

Proof

- In the previous lemma we have showed that for any proof of the empty sequent there is another proof of the empty sequent of a smaller ordinal. However, as the ordinals are all smaller than ϵ_0, there is no infinite decreasing sequence of them. Therefore there is no proof of the empty sequent in PA and it is consistent.

Chapter 4

Proofs of Weakly Quantified Theorems

In this section we will give a cut elimination proof for inductive proofs of theorems without strong quantifiers. First, we will analyze the structure of such proofs and prove some auxiliary lemmas. Then we will prove one of the main lemmas in this thesis, the projection lemma. The following part will deal with a procedure which eliminates certain contractions so we can use the projection lemma. In the last part we will prove the main theorem of this thesis.

4.1 Analysis of the Proofs

4.1.1 Characteristics of inductive proofs of weakly quantified theorems

In the previous section we have seen a procedure for eliminating inductions when the induction term can be evaluated into a numeral. The assumption that induction terms can be evaluated in such a way is not necessarily true. Theorems containing strong quantifiers may have free variables in their end-pieces which cannot be eliminated. Even if the end-sequent is without strong quantifiers, we must look not only on the end-piece, but on the whole proof in order to eliminate all inductions that interfere with cut elimination.
The problem in this case is that although there are no strong quantifiers in the end-sequent, there might be strong quantifiers in implicit bundles, i.e. bundles ending with a cut. These strong quantifiers may introduce eigen-variables, which will appear in an induction term. This makes it impossible

to evaluate the term to a natural number.
The fact that it may happen only in implicit bundles gives us also the solution to the problem. Namely, if this strong quantifier is being cut out, then there must be a weak quantifier on the other branch above the cut. This weak quantifier may introduce a term which might be a suitable replacement for the eigenvariable of the strong quantifier. This is the idea of the "Projection Lemma".
Below we give a definition of a sub-bundle which contains such a weak quantifier.

Definition 4.1.1 (Sub-bundles)**.** - A sub-bundle is a sequence of occurrences of formulas in a proof with the following properties:

1. The sequence may begin with any formula.
2. The sequence may end with any formula.
3. Every formula in the sequence except the last is immediately followed by its successor.

Definition 4.1.2 (Weak Quantifier Formula)**.** - Is a quantifier formula which is contained in a bundle that contains also the introduction of this quantifier by a weak quantifier inference rule.

The notion of a PSB is central to the remaining part of the thesis

Definition 4.1.3 (Problematic quantifiers)**.** A quantifier Q is called problematic for the induction I if Q eliminates a free variable (which is used as an eigenvariable), which appears also in the induction term of I.

Definition 4.1.4 (Problematic cuts)**.** A problematic cut is a cut whose cut-fomula contains a problematic quantifier. A cut-formula may contain quantifiers which are problematic for several different inductions. The two sub-derivations of the upper sequents of the cut are called strong and weak with regard to an induction. The strong side is the side containing the induction itself.

Definition 4.1.5 (Problematic Sub-bundles (PSBs))**.** A PSB is a sub-bundle with the following properties:

1. All the formulas in the sub-bundle contain a weak quantifier Q. Q is called the characteristic quantifier of the PSB.
2. Q is also a problematic quantifier for some induction and is contained in a cut-formula.

3. The sub-bundle starts with Q as the principal formula of a weak quantifier or weakening rule.

4. The sub-bundle ends with a cut formula.

Remark - A PSB is defined for a specific induction that has a problematic quantifier which is the characteristic quantifier of the PSB.

Due to the tree form of the derivation, it is clear that if an induction has more than one problematic quantifier, they must all be contained in the same formula and are cut together. Therefore we can define the characteristic sub-bundle of an induction.

Definition 4.1.6 (Characteristic sub-bundle)**.** The sub-bundle which begins with the right principal induction formula of I and ends with the cut formula of the problematic cut of I is called the characteristic sub-bundle of I. Only inductions with a problematic cut have a characteristic sub-bundle.

Example 4.1.7 (A PSB with a quantified inference rule)**.** *The following example shows a problematic cut. The cut-formula contains the problematic quantifier $\forall xA[B]_\theta(x)$ for the induction. The induction is contained in the strong side of the cut. On the weak side we see a PSB for the induction which begins with a weak quantifier inference rule. The PSB starts with $\forall xA[B]_\theta(x)$ at the lower sequent of the $\forall : l$ inference rule and ends with $R[\forall xA[B]_\theta(x)]_\lambda$ at the cut.*

$$
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\cdots}{B(n),\Gamma''\vdash\Delta'',B(n')}}{B(0),\Gamma''\vdash\Delta'',B(b)}\,(ind)}{\cdots}}{\Gamma'\vdash\Delta',A[B]_\theta(b)}}{\Gamma'\vdash\Delta',\forall xA[B]_\theta(x)}\,(\forall:r)}{\cdots}}{\Gamma\vdash\Delta,R[\forall xA[B]_\theta(x)]_\lambda}
\qquad
\dfrac{\dfrac{\dfrac{\dfrac{\cdots}{A[B]_\theta(t),\Pi'\vdash\Lambda'}}{\forall xA[B]_\theta(x),\Pi'\vdash\Lambda'}\,(\forall:l)}{\cdots}}{R[\forall xA[B]_\theta(x)]_\lambda,\Pi\vdash\Lambda}}{\dfrac{\Gamma,\Pi\vdash\Delta,\Lambda}{\cdots}}\,(cut: R[\forall xA[B]_\theta(x)]_\lambda)
$$

The procedure used in the Projection Lemma and in the elimination of contractions is similar to the procedure taken by Gentzen in his Cut Elimination proof, i.e. propagating the cuts upward in the proof tree. As our proofs contain inductions, we might encounter the same problems we had for cut elimination in PA, as discussed in section 3.1.2. Therefore we must ensure that there are no inductions in the problematic places. Here we take advantage of the fact that the proof is in tree form. We define a well founded order on the inductions, an order that ensures that there will be no other induction in the problematic locations if we process each induction according to it.

The first task is to define when an induction is located in a problematic location in relation to some other induction.

Definition 4.1.8 (Interfering Sub-bundle)**.** When a PSB contains an induction formula, then it is called interfering sub-bundle. - An interfering sub-bundle for some induction over some cut is a PSB for this induction over the same cut which contains an induction formula of another induction.

Example 4.1.9 (Interfering Sub-bundle)**.** *The PSB in example 4.1.7 will be an interfering sub-bundle if it contains an induction. i.e. we would have:*

$$
\dfrac{
\dfrac{\begin{array}{c}
\dfrac{\dfrac{\cdots}{B(n),\Gamma''\vdash\Delta'',B(n')}}{B(0),\Gamma''\vdash\Delta'',B(b)}\,(I_1)\\
\cdots
\end{array}}{\dfrac{\Gamma'\vdash\Delta',A[B]_\theta(b)}{\begin{array}{c}\Gamma'\vdash\Delta',\forall xA[B]_\theta(x)\\ \cdots\end{array}}\,(\forall:r)}
\over \Gamma\vdash\Delta,R[\forall xA[B]_\theta(x)]_\lambda}
\quad
\dfrac{\begin{array}{c}
\dfrac{\dfrac{\cdots}{A[B]_\theta(t),\Pi'\vdash\Lambda'}}{\forall xA[B]_\theta(x),\Pi'\vdash\Lambda'}\,(\forall:l)\\
\cdots\\
\dfrac{C[\forall xA[B]_\theta(x)]_\eta(n),\Pi''\vdash\Lambda'',C[\forall xA[B]_\theta(x)]_\eta(n')}{C[\forall xA[B]_\theta(x)]_\eta(0),\Pi''\vdash\Lambda'',C[\forall xA[B]_\theta(x)]_\eta(c)}\,(I_2)\\
\cdots
\end{array}}{R[\forall xA[B]_\theta(x)]_\lambda,\Pi\vdash\Lambda}
\over
\dfrac{\Gamma,\Pi\vdash\Delta,\Lambda}{\cdots}}\,(cut: R[\forall xA[B]_\theta(x)]_\lambda)
$$

The relationship between the two inductions (I_1 and I_2 in example 4.1.9) is defined as follows:

Definition 4.1.10 (Interfering Induction)**.** An induction I_2 which resides in a PSB of some other induction I_1 is an interfering induction for I_1. We say also in this case that I_2 interferes with I_1.

In example 4.1.9 I_2 interferes I_1.

Inductions may also be problematic to other inductions if they appear below them in a thread in a proof. An induction which does not have problematic inductions of this type is called a suitable induction. Here we give a partial definition and we wait for our definition of the order on inductions for the full definition of a suitable induction.

Definition 4.1.11 (Suitable induction I)**.** An induction I_1 is called suitable with regard to induction I_2 if:

1. Its induction term does not contain the free variable which is used as an eigenvariable by I_2.

2. If it has a characteristic sub-bundle, it does not contain the induction formula of induction I_2.

4.1.2 The order of inductions

Next we would like to define an order between inductions according to these two properties: interferences and suitability. Defining an order between inductions according to induction-interference may not always hold, because we apply transformations on the proof. In general, in case two inductions appear one above the other in some thread, the lower one should be handled first. This is because the lower one may introduce an eigenvariable which is contained in the upper induction term. The lower induction can also occur in a characteristic sub-bundle of the upper induction, This is another reason to handle it first as our algorithm requires an induction free characteristic sub-bundle. On the other hand, when duplicating inductions, an induction which appears above some other induction I may interfere with another induction labeled by I , as happens in the following example:

Example 4.1.12. *Here we investigate the result of Gentzen's transformation on a derivation containing three inductions, one above the other.*

$$\cfrac{\cfrac{\cfrac{\cfrac{\cfrac{\cfrac{C[\forall x A[B](x)](c), \Gamma''' \vdash \Delta''', C[\forall x A[B](x)](c')}{C[\forall x A[B](x)](0), \Gamma''' \vdash \Delta''', C[\forall x A[B](x)](t_3)}\ (I_3)}{\cdots}}{\cfrac{B(b), \Gamma_1'', D[C[\forall x A[B](x)]], \Gamma_2'' \vdash \Delta', B(b')}{B(0), \Gamma_1'', D[C[\forall x A[B](x)]], \Gamma_2'' \vdash \Delta'', B(t_2, a)}\ (I_2)}}{\cdots}}{\cfrac{\cfrac{\Gamma_1', D[C[\forall x A[B](x)]], \Gamma_2' \vdash \Delta', A[B](a)}{\Gamma_1', D[C[\forall x A[B](x)]], \Gamma_2' \vdash \Delta', \forall x A[B](x)}\ (\forall : r)}{\cdots}}$$

$$\cfrac{E[D[C[\forall x A[B](x)]]](d), \Gamma \vdash \Delta, E[D[C[\forall x A[B](x)]]](d')}{E[D[C[\forall x A[B](x)]]](0), \Gamma \vdash \Delta, E[D[C[\forall x A[B](x)]]](t)}\ (I_1)$$

We see a quantifier which is problematic for induction I_2 and appears in the induction formula of I_1. An I_1 induction step will look like:

$$
\dfrac{
\dfrac{
\begin{array}{c}
\dfrac{C[\forall x A[B](x)](c), \Gamma''' \vdash \Delta''', C[\forall x A[B](x)](c')}{C[\forall x A[B](x)](0), \Gamma''' \vdash \Delta''', C[\forall x A[B](x)](t_3)}\,(I_3) \\
\cdots \\
\dfrac{B(b), \Gamma''_1, D[C[\forall x A[B](x)]], \Gamma''_2 \vdash \Delta'', B(b')}{B(0), \Gamma''_1, D[C[\forall x A[B](x)]], \Gamma''_2 \vdash \Delta'', B(t_2, a)}\,(I_2) \\
\cdots \\
\dfrac{\Gamma'_1, D[C[\forall x A[B](x)]], \Gamma'_2 \vdash \Delta', A[B](a)}{\Gamma'_1, D[C[\forall x A[B](x)]], \Gamma'_2 \vdash \Delta', \forall x A[B](x)}\,(\forall : r) \\
\cdots
\end{array}
}{E[D[C[\forall x A[B](x)]]](0), \Gamma \vdash \Delta, E[D[C[\forall x A[B](x)]]](1)}
\quad
\dfrac{
\begin{array}{c}
\dfrac{C[\forall x A[B](x)](c), \Gamma''' \vdash \Delta''', C[\forall x A[B](x)](c')}{C[\forall x A[B](x)](0), \Gamma''' \vdash \Delta''', C[\forall x A[B](x)](t_3)}\,(I_3) \\
\cdots \\
\dfrac{B(b), \Gamma''_1, D[C[\forall x A[B](x)]], \Gamma''_2 \vdash \Delta'', B(b')}{B(0), \Gamma''_1, D[C[\forall x A[B](x)]], \Gamma''_2 \vdash \Delta'', B(t_2, a)}\,(I_2) \\
\cdots \\
\dfrac{\Gamma'_1, D[C[\forall x A[B](x)]], \Gamma'_2 \vdash \Delta', A[B](a)}{\Gamma'_1, D[C[\forall x A[B](x)]], \Gamma'_2 \vdash \Delta', \forall x A[B](x)}\,(\forall : r) \\
\cdots
\end{array}
}{E[D[C[\forall x A[B](x)]]](1), \Gamma \vdash \Delta, E[D[C[\forall x A[B](x)]]](2)}
}{E[D[C[\forall x A[B](x)]]](0), \Gamma, \Gamma \vdash \Delta, \Delta, E[D[C[\forall x A[B](x)]]](2)}\,(cut)
$$

There is a new PSB for induction I_2 which contains an induction formula (of I_3).

We would like to define an order using a property of inductions which is mostly independent of the form of the proof. This way, transformations will not have much impact on the order and we can easily prove that the order is preserved under the transformations. We notice that a simple property will allow us to define an order between inductions such that: An induction of the biggest label will be a suitable induction for all other inductions. Moreover, It will have no other induction in any of its PSBs and It will prevent the problem stated above, because this property is independent from the form of the derivation. We will order our inductions according to the grade of the induction formula.

Definition 4.1.13 (Order on Inductions)**.** The binary relation $<_I$ over the set of inductions (in a proof) is defined by:

1. $(I_1, I_2) \in <_I$ iff the grade of the induction formula of induction I_2 is bigger than the grade of the induction formula of induction I_1.

2. In case two induction formulas have the same grade and their inductions occur on the same thread, $(I_1, I_2) \in <_I$ iff I_2 is below I_1 in the thread.

3. We extend $<_I$ into a total order in an arbitrary way (for the rest of the pair of inductions of equal formula's grade).

It is clear $<_I$ is well founded.

Definition 4.1.14 (Labels of inductions)**.** Given a proof, we will label all its inductions according to the relation $<_I$.

We will redefine suitable inductions with the use of the new order $<_I$.

Definition 4.1.15 (Suitable induction)**.** An induction I_1 is called suitable if:

1. Its induction term does not contain the free variable which is used as an eigenvariable by an induction of smaller or equal label.

2. If it has a characteristic sub-bundle, it does not contain induction formulas of inductions of smaller or equal labels.

Definition 4.1.16 (Suitable proofs)**.** If all the inductions in a proof:

1. Have no interfering inductions of equal or smaller label.

2. Are suitable.

the proof is called suitable.

Lemma 4.1.17. Given a proof P with all inductions labeled according to $<_I$, we have for all inductions of an arbitrary label l:

1. There is no interfering induction of equal or smaller label.

2. None of the inductions labeled by l contains in its induction term free variables which are used as eigenvariables by inductions of smaller label.

3. None of the inductions labeled by l has an induction formula of another induction of smaller label in its characteristic sub-bundle (if it has one).

Proof

1. Let *ind* be an induction of label l. Assume there is an induction formula F of an induction of label smaller or equal to l in a PSB of *ind*. According to our order, the grade of F is not bigger than the grade of formulas labeled by I. According to the definition of a PSB, the problematic quantified formula A of *ind* must be contained in all formulas in the PSB. A is therefore a sub-formula of F. The induction formula of *ind* must be a strict sub-formula of A as it does not contain the quantifier. Therefore, it must be a strict sub-formula of F, which contradicts the assumption that the grade of F is not bigger than the grade of the induction formula of *ind*.

2. Assume that induction ind_1 of label l contains in its induction term a free variable which is used as an eigenvariable in an induction ind_2 of a smaller label. ind_1 occurs above ind_2 on the same thread and therefore, as ind_1 is of bigger label than ind_2, the grade of ind_2 induction formula F_2 must be smaller than the grade of ind_1 induction formula F_1 (according to the second item in the definition of $<_I$). Both F_1 and F_2 are contained in the same bundle and as F_1 is above F_2 in this bundle, F_1 is a sub-formula of F_2 which contradicts the fact that F_1 is of bigger grade.

3. This proof is similar to the proof of the previous property. The two inductions occur on the same thread. Because they are of different labels, the grades of their induction formulas cannot be equal. As the two induction formulas are in the same bundle, the upper one must be a sub-formula of the lower one. It must be a strict sub-formulas as their grades cannot be equal. We get a contradiction as the upper induction is of bigger label and therefore cannot have a smaller grade on its induction formula.

We will have to make sure that the transformations performed in the rest of this section will not make a suitable proof unsuitable. The next lemma will define the properties that must be kept in a transformation for the suitability of a proof to be preserved.

Lemma 4.1.18. If we obtain a proof P' from a suitable proof P by:

1. Adding only existing inductions and never adding an induction to a thread containing another induction of the same label.

2. Eliminating inductions.

3. Adding or removing any number of other inferences other than induction.

then P' is suitable as well.

Proof

- We can add only existing inductions and therefore the first property of lemma 4.1.17 will always hold. This property is completely independent of the form of the derivation, because it depends only on the grade of formulas.

- The next two properties are not independent of the form of the derivation as there might be inductions of the same label which are not covered by lemma 4.1.17. However, we are restricting our transformations to never add inductions to threads containing other inductions of the same label. Therefore all inductions added, as well as the existing inductions, will remain suitable.

4.1.3 Manipulations on derivations

The next lemmas about bundles show that a derivation remains valid if we remove sub-bundles from it.

Definition 4.1.19 (Bundles removal)**.** Let ψ be a directed labeled tree (as in the definition of an LK-derivation) and β a sub-bundle such that all the occurrences of formulas in β appear in ψ. Then ψ/β is the resulted tree obtained from ψ by removing all the occurrences of formulas from β.

The first lemma is about a sub-bundle which starts and ends with the same formula, for example:

Example 4.1.20. *The derivation ψ containing the sub-bundle β starting and ending with A.*

$$\frac{\dfrac{\Pi_1, A, \Pi_2 \vdash \Lambda}{\cdots}}{\Pi_1', A, \Pi_2' \vdash \Lambda'}$$

Lemma 4.1.21. Given a derivation ψ containing a sub-bundle β that starts and ends with the same formula A and that does not contain the principal formula of a contraction or induction rule, ψ/β is also a derivation.

Proof

- First we examine the sub-bundle β. As it starts and ends with A, A cannot be the principal formula of a logical rule. Otherwise its logical complexity must change. It cannot be the principal formula of a weakening or a cut as each occurrence of A in β (except the first) is the successor of another occurrence of A in β, according to the definition of a sub-bundle. As A cannot be the principal formula of a contraction or induction, we are left only with the possibility that A is the principal formula of an exchange or A is not the principal formula of any rule.

- By induction on the length n of the proof.
- if $n = 0$ then as every sequent is also a derivation, it holds.
- Assuming it holds for all derivations of length n and taking a derivation of length $n + 1$.
- If the lowermost inference rule is J with A not being the principal formula. For example:

$$\dfrac{\dfrac{\Pi_1, A, \Pi_2 \vdash \Lambda}{\cdots}}{\dfrac{\Pi''_1, A, \Pi''_2 \vdash \Lambda''}{\Pi'_1, A, \Pi'_2 \vdash \Lambda'}\ (\mathrm{J})}$$

  we can use the induction hypothesis in order to obtain a derivation of $\Pi''_1, \Pi''_2 \vdash \Lambda''$ from $\Pi_1, \Pi_2 \vdash \Lambda$ and then apply J.

- If the lowermost rule is an exchange with A a principal formula. For example:

$$\dfrac{\dfrac{\Pi_1, A, \Pi_2 \vdash \Lambda}{\cdots}}{\dfrac{\Pi'_1, B, A, \Pi'_2 \vdash \Lambda'}{\Pi'_1, A, B, \Pi'_2 \vdash \Lambda'}\ (\mathrm{exchange : l})}$$

  we can use again the induction hypothesis in order to obtain a derivation of $\Pi'_1, B, \Pi'_2 \vdash \Lambda'$ from $\Pi_1, \Pi_2 \vdash \Lambda$ and we are done.

The second lemma is about sub-bundles in general.

Lemma 4.1.22. Given a derivation ψ containing a sub-bundle β, which does not contain the principal formula of a contraction or induction rule, ψ/β is also a derivation.

Proof

- By induction on the difference n between the logical complexity of the last formula in β and the logical complexity of the first formula in β.
- $n = 0$ then it holds by lemma 4.1.21.

- $n > 0$, then according to the outermost symbol:
  - $\vee$ and the inference rule is on the right, so for example we have:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\Pi_1, A, \Pi_2 \vdash \Lambda}{\cdots}}{\Pi' \vdash \Lambda', A'}}{\Pi' \vdash \Lambda', A' \vee R}\,(\vee : r)}{\cdots}}{\Pi'' \vdash \Lambda''_1, A' \vee R, \Lambda''_2}$$

according to the induction hypothesis we can obtain the derivation:

$$\dfrac{\dfrac{\Pi_1, \Pi_2 \vdash \Lambda}{\cdots}}{\Pi' \vdash \Lambda'}$$

and according to lemma 4.1.21 we can obtain the derivation:

$$\dfrac{\dfrac{\Pi' \vdash \Lambda'}{\cdots}}{\Pi'' \vdash \Lambda''_1, \Lambda''_2}$$

  - $\vee$ and the inference rule is on the left. For example:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\Pi_1, A, \Pi_2 \vdash \Lambda}{\cdots}}{A', \Pi' \vdash \Lambda'} \qquad R, \Gamma \vdash \Delta}{A' \vee R, \Pi', \Gamma \vdash \Lambda', \Delta}\,(\vee : l)}{\cdots}}{\Pi''_1, \Gamma'_1, A' \vee R, \Pi''_2, \Gamma'_2 \vdash \Lambda'', \Delta'}$$

in the same way as before, we can use the induction hypothesis and weakenings in order to obtain the derivation:

$$\dfrac{\dfrac{\dfrac{\Pi_1, \Pi_2 \vdash \Lambda}{\cdots}}{\Pi' \vdash \Lambda'}}{\Pi', \Gamma \vdash \Lambda', \Delta}\,(\text{weakenings})$$

and lemma 4.1.21 in order to obtain the derivation:

$$\dfrac{\dfrac{\Pi', \Gamma \vdash \Lambda', \Delta}{\cdots}}{\Pi''_1, \Gamma'_1, \Pi''_2, \Gamma'_2 \vdash \Lambda'', \Delta'}$$

– ¬, for example

$$\dfrac{\dfrac{\dfrac{\dfrac{\Pi_1, A, \Pi_2 \vdash \Lambda}{\cdots}}{A', \Pi' \vdash \Lambda'}}{\Pi' \vdash \Lambda', \neg A'}\,(\neg : r)}{\dfrac{\cdots}{\Pi'' \vdash \Lambda_1'', \neg A', \Lambda_2''}}$$

again, we can obtain the required subproof by using the induction hypothesis and lemma 4.1.21.

– The rest of the connectives are dealt with in the same way.

The last lemma in this section is a simple weakening lemma.

Lemma 4.1.23. Given $\Pi, \Pi', \Lambda, \Lambda', \Gamma, \Gamma', \Delta, \Delta'$ such that no free variable which is used as an eigenvariable and appears in $\Pi, \Pi', \Lambda, \Lambda'$ also appear in $\Gamma, \Gamma', \Delta, \Delta'$ and given the following derivation:

$$\dfrac{\dfrac{\Pi \vdash \Lambda}{\cdots}}{\Pi' \vdash \Lambda'}$$

we can obtain the derivation:

$$\dfrac{\dfrac{\Pi, \Gamma \vdash \Delta, \Lambda}{\cdots}}{\Pi', \Gamma \vdash \Delta, \Lambda'}$$

Proof

- By induction on the length of the derivation.

4.2 The Projection Lemma

The method applied by Gentzen for the elimination of inductions starts by replacing all free variables which are not being used as eigenvariables by some fixed terms. Free variables which are being used as eigenvariables cannot appear below the inductions, because only the end-piece of the proof is being considered. We are looking at the whole proof and therefore we must deal with those free variables which are being used as eigenvariables as well. We consider only problematic quantifiers and not all strong quantifiers. Because the end-sequent does not contain strong quantifiers, these problematic quantifiers must be cut out. For every problematic quantifier below an induction which is cut out, there must be a dual weak quantifier which is the characteristic quantifier of a PSB of the induction. Because we do not have quantifiers in our axioms, this quantifier must occur as a principal formula either in a weakening rule or a weak quantifier rule. The projection lemma deals with the case, where it occurs as a principal formula of a weak quantifier rule.

Example 4.2.1. *The following example describes a PSB for an induction which begins with an introduction of a weak quantifier with t as the auxiliary formula's term.*

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{A(a),\Gamma\vdash\Delta,A(a')}{A(0),\Gamma\vdash\Delta,A(b)}\,(ind)}{\cdots}}{\Gamma'\vdash\Delta',P[A]_\lambda(b)}}{\Gamma'\vdash\Delta',\forall xP[A]_\lambda(x)}\,(\forall:r)}{\cdots}}{\Gamma''\vdash\Delta'',R[\forall xP[A]_\lambda(x)]_\theta} \quad \dfrac{\dfrac{\dfrac{P[A]_\lambda(t),\Pi\vdash\Lambda}{\forall xP[A]_\lambda(x),\Pi\vdash\Lambda}\,(\forall:l)}{\cdots}}{R[\forall xP[A]_\lambda(x)]_\theta,\Pi'\vdash\Lambda'}}{\Gamma'',\Pi'\vdash\Delta'',\Lambda'}\,(cut:R[\forall xP[A]_\lambda(x)]_\theta)$$

In order to relate the eigenvariable to the weak quantifier, we give the following definition.

Definition 4.2.2 (Free variables and their dual terms)**.** Given a problematic quantifier Q and its set of corresponding characteristic quantifiers (there may be more than one in case of a contraction in the PSB), we examine the strong quantifier rule introducing Q on the strong side of the cut and the weak quantifier rules introducing Q on the weak side. We call all the terms which are eliminated by those weak quantifier rules the dual terms of the eigenvariable eliminated by the strong quantifier rule.

In example 4.2.1 t is the dual term of the free variable b.

Because our proof of the projection lemma is based on propagating cuts up over inferences, we must make sure that there are no inductions there. Our order of eliminating inductions will ensure that. Another issue is that if a formula containing the weak quantifier is being contracted, there might be more than one dual term for the eigenvariable. Therefore there is no way for us to know which instance should be projected. So our second assumption is that there are no contractions in the PSBs of the induction. In order to simplify the proof, contractions appearing in the sub-bundle below the induction will be removed as well. The next section will deal with the elimination of these contractions if they occur.

Definition 4.2.3 (Types of quantifiers in a cut formula)**.** Quantifiers in a cut formula can be either weak or strong quantifiers, based on the type of quantifier rule used to introduce them in the sub-derivation under the induction.

We will also require that if there is more than one strong quantifier in a specific cut formula, we will deal only with the outermost such quantifier.

Definition 4.2.4 (Critical quantifiers)**.** critical formulas over a cut for a problematic quantifier are all sub-formulas of the cut-formula, which contain the problematic quantifier formula as a sub-formula.

In order to deal with all critical quantifiers in an uniform way, we will prove that all of them have some sort of a "PSB".

Lemma 4.2.5. If a strong quantifier formula is a critical formula, then there is a sub-bundle which is contained in the PSB of the problematic quantifier which contains a weak quantifier rule with the critical formula as a principal formula.

Proof

- As the critical formula contains a problematic quantifier formula as a sub-formula, it will also contain the characteristic quantifier of a PSB as a sub-formula. Therefore they both must be on the same sub-bundle and the critical formula must be introduced below the introduction of the characteristic quantifier.

As a consequence of the lemma above, it can be assumed that if all PSBs of this induction contain no contraction then all sub-bundles defined as above for all critical quantifiers of this induction will contain no contraction as well. For the rest of the thesis we will extend the definition of a PSB to include those sub-bundles of critical quantifiers which are contained, as the lemma above proved, in a PSB of a problematic quantifier of the same induction. The dual quantifier formula of the critical quantifier formula will be called the characteristic quantifier of the new PSB. The PSB will begin with the introduction of the quantifier formula.

Definition 4.2.6 (Suitable quantifier)**.** Let P be a suitable proof of sequent S with a PSB ρ for induction I over a problematic cut C. Q is called suitable if:

1. Q is the characteristic quantifier of ρ and it is the outermost critical quantifier in the cut formula of C.
2. ρ contains no contraction or induction.
3. ρ begins with a weak quantifier rule.
4. There are no contractions in the sub-bundle containing the right principal formula of the induction and the cut formula.

Lemma 4.2.7 (The Projection Lemma)**.** Given a suitable proof P of sequent S containing a suitable quantifier Q we can transform P into a suitable proof P' of S where in P':

1. We remove the quantifier rules introducing Q (on both sides of the cut).
2. We replace all occurrences of the eigenvariable of Q by a term t.

As a result, all occurrences of Q are eliminated from the proof.

- Remark - The term t is obtained from the dual term of the eigenvariable in the process of the transformation given in the proof of the lemma.

The following example will demonstrate the application of the lemma.

Example 4.2.8. *Given a derivation such as in example 4.2.1, which satisfies all the requirements of the lemma (and deal with a critical quantifier which is also a problematic quantifier), we can obtain the following derivation according to the Projection Lemma:*

$$\dfrac{\dfrac{\dfrac{\dfrac{A(a),\Gamma \vdash \Delta, A(a')}{A(0),\Gamma \vdash \Delta, A(t')}\,(ind)}{\cdots}}{\Gamma' \vdash \Delta', P[A]_\lambda(t')} \qquad \dfrac{\dfrac{P[A]_\lambda(t'),\Pi \vdash \Lambda}{\cdots}}{P[A]_\lambda(t'),\Pi''' \vdash \Lambda'''}}{\dfrac{\dfrac{\Gamma',\Pi''' \vdash \Delta',\Lambda'''}{\cdot\cdot}}{\Gamma'',\Pi' \vdash \Delta'',\Lambda'}}\,(cut : P[A]_\lambda(t'))$$

Where t' is defined in the proof of the lemma.

Proof

- By induction on the size n of the sub-bundle, which contains the problematic quantifier and is contained in the characteristic sub-bundle of the induction. In example 4.2.1 this is the sub-bundle starting at $\Gamma' \vdash \Delta', \forall x P[A]_\lambda(x)$ and ending at $\Gamma'' \vdash \Delta'', R[\forall x P[A]_\lambda(x)]_\theta$. We call this sub-bundle ISB.

- In the proof we make an extensive use of lemma 4.1.22 which states that we can remove a bundle from a derivation and still have a derivation.

- We also use lemma 4.1.23 which just states we can weaken the context (under some limitations) of a derivation and still have a derivation.

- case $n = 1$ then we have (in case the critical quantifier is a problematic quantifier as well):

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\cdots}{A(a), \Gamma \vdash \Delta, A(a')}}{A(0), \Gamma \vdash \Delta, A(b)}\,(ind)}{\cdots}}{\Gamma' \vdash \Delta', P[A]_\lambda(b)}}{\Gamma' \vdash \Delta', \forall x P[A]_\lambda(x)}\,(\forall : r) \qquad \dfrac{\dfrac{\dfrac{\dfrac{\cdots}{P[A]_\lambda(t), \Pi' \vdash \Lambda'}}{\forall x P[A]_\lambda, (x)\Pi' \vdash \Lambda'}\,(\forall : l)}{\cdots}}{\forall x P[A]_\lambda(x), \Pi'' \vdash \Lambda''}}{\Gamma', \Pi'' \vdash \Delta', \Lambda''}\,(cut : \forall x P[A]_\lambda(x))$$

  and as the PSB does not contain contractions and satisfies all the prerequisites of lemma 4.1.22 we can replace it by:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\cdots}{A(a), \Gamma \vdash \Delta, A(a')}}{A(0), \Gamma \vdash \Delta, A(t)}\,(ind)}{\cdots}}{\Gamma' \vdash \Delta', P[A]_\lambda(t)} \qquad \dfrac{\cdots}{P[A]_\lambda(t), \Pi' \vdash \Lambda'}}{\Gamma', \Pi' \vdash \Delta', \Lambda'}\,(cut : P[A]_\lambda(t))}{\text{(according to lemmas 4.1.22 and 4.1.23 and additional exchanges)}}}{\Gamma', \Pi'' \vdash \Delta', \Lambda''}$$

If the critical quantifier is not a problematic quantifier, then the only change is that it will not affect the induction term.

- We assume that the lemma is true for all derivations with ISB of size $\leq n$ and prove it for size $n+1$. First we notice that according to the lemma the PSB does not contain inductions, contractions and weakenings (Although there might be such inference rules on other formulas in the sequents containing the PSB formulas). We examine the first rule J above the cut and show how to transform the derivation into a derivation with an ISB of size $\leq n$.

  – J is any inference rule, which does not have the cut formula as the principal formula. In this case, J must be either a rule on the left or an exchange. This is because the cut formula is the rightmost formula and according to our calculus it must be the principal formula of any right hand-side rule other than exchange. So we have:

$$
\dfrac{
\dfrac{
\dfrac{
\dfrac{
\dfrac{
\dfrac{\dfrac{\dfrac{\cdots}{A(a), \Gamma \vdash \Delta, A(a')}}{A(0), \Gamma \vdash \Delta, A(b)}\,(ind)}{\cdots}
}{\Gamma' \vdash \Delta', P[A]_\lambda(b)}
}{\Gamma' \vdash \Delta', \forall x P[A]_\lambda(x)}\,(\forall : r)
}{\cdots}
}{\Gamma'' \vdash \Delta'', R[\forall x P[A]_\lambda(x)]_\theta}
}{\Gamma''' \vdash \Delta''', R[\forall x P[A]_\lambda(x)]_\theta}\,(J)
\quad
\dfrac{\dfrac{\dfrac{\dfrac{\cdots}{P[t]_\lambda, \Pi \vdash \Lambda}}{\forall x P[A]_\lambda(x), \Pi \vdash \Lambda}\,(\forall : l)}{\cdots}}{R[\forall x P[A]_\lambda(x)]_\theta, \Pi' \vdash \Lambda'}
}{\Gamma''', \Pi' \vdash \Delta''', \Lambda'}\,(cut : R[\forall x P[A]_\lambda(x)]_\theta)
$$

and we can obtain the derivation:

$$
\dfrac{
\dfrac{
\dfrac{
\dfrac{
\dfrac{\dfrac{\dfrac{\cdots}{A(a), \Gamma \vdash \Delta, A(a')}}{A(0), \Gamma \vdash \Delta, A(b)}\,(ind)}{\cdots}
}{\Gamma' \vdash \Delta', P[A]_\lambda(b)}
}{\Gamma' \vdash \Delta', \forall x P[A]_\lambda(x)}\,(\forall : r)
}{\cdots}
}{\Gamma'' \vdash \Delta'', R[\forall x P[A]_\lambda(x)]_\theta}
\quad
\dfrac{\dfrac{\dfrac{\dfrac{\cdots}{P[t]_\lambda, \Pi \vdash \Lambda}}{\forall x P[A]_\lambda(x), \Pi \vdash \Lambda}\,(\forall : l)}{\cdots}}{R[\forall x P[A]_\lambda(x)]_\theta, \Pi' \vdash \Lambda'}
}{\Gamma'', \Pi' \vdash \Delta'', \Lambda'}\,(cut : R[\forall x P[A]_\lambda(x)]_\theta)
$$

with ISB of size n and we can apply the induction hypothesis in order to obtain:

$$
\dfrac{
\dfrac{
\dfrac{\dfrac{\dfrac{\cdots}{A(a), \Gamma \vdash \Delta, A(a')}}{A(0), \Gamma \vdash \Delta, A(t')}\,(ind)}{\cdots}
}{\Gamma'' \vdash \Delta'', R[P[A]_\lambda(t')]_\theta}
\quad
\dfrac{\dfrac{\dfrac{\cdots}{P[t']_\lambda, \Pi \vdash \Lambda}}{\cdots}}{R[P[A]_\lambda(t')]_\theta, \Pi' \vdash \Lambda'}
}{\Gamma'', \Pi' \vdash \Delta'', \Lambda'}\,(cut : R[P[A]_\lambda(t')]_\theta)
$$

and we can apply J with possibly only additional exchanges in order to obtain $\Gamma''', \Pi' \vdash \Delta''', \Lambda'$. It should be noted that J can also be a binary rule although it is not reflected in the example.

– J is an inference rule with the cut formula as a principal formula. If J is an exchange (on the right) then we have:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dots}{A(a), \Gamma \vdash \Delta, A(a')}}{A(0), \Gamma \vdash \Delta, A(b)}\,(ind)}{\dots}}{\Gamma' \vdash \Delta', P[A]_\lambda(b)}}{\Gamma' \vdash^{S_1} \Delta', \forall x P[A]_\lambda(x)}\,(\forall : r)}{\dots}}{\Gamma'' \vdash^{S_0} \Delta'', R[\forall x P[A]_\lambda(x)]_\theta}}{\dots}}{\Gamma''' \vdash \Delta''', R[\forall x P[A]_\lambda(x)]_\theta, C}}{\Gamma''' \vdash \Delta''', C, R[\forall x P[A]_\lambda(x)]_\theta}\,(J) \quad \dfrac{\dfrac{\dfrac{\dfrac{\dots}{P[A]_\lambda(t), \Pi \vdash \Lambda}}{\forall x P[A]_\lambda(x), \Pi \vdash \Lambda}\,(\forall : l)}{\dots}}{R[\forall x P[A]_\lambda(x)]_\theta, \Pi' \vdash \Lambda'}}{\Gamma''', \Pi' \vdash \Delta''', C, \Lambda'}\,(cut : R[\forall x P[A]_\lambda(x)]_\theta)$$

The sequent S_0 is the lowermost sequent above J which has $R[\forall x P[A]_\lambda(x)]_\theta$ as its right principal formula. S_0 can also be S_1. S_0 exists as the bundle containing $R[\forall x P[A]_\lambda(x)]_\theta$ contains also $\forall x P[A]_\lambda(x)$ and therefore $R[\forall x P[A]_\lambda(x)]_\theta$ must be the right principal formula of a logical rule or equal to $\forall x P[A]_\lambda(x)$. As we have chosen S_0 to be the lowermost sequent which contains $R[\forall x P[A]_\lambda(x)]_\theta$ at its rightmost position, the sub-bundle containing $R[\forall x P[A]_\lambda(x)]_\theta$, beginning with S_0 and ending with J, contains no contractions.
We replace it by the following derivation with ISB of size $\leq n$:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dots}{A(a), \Gamma \vdash \Delta, A(a')}}{A(0), \Gamma \vdash \Delta, A(b)}\,(ind)}{\dots}}{\Gamma' \vdash \Delta', P[A]_\lambda(b)}}{\Gamma' \vdash \Delta', \forall x P[A]_\lambda(x)}\,(\forall : r)}{\dots}}{\Gamma'' \vdash^{S_0} \Delta'', R[\forall x P[A]_\lambda(x)]_\theta} \quad \dfrac{\dfrac{\dfrac{\dfrac{\dots}{P[A]_\lambda(t), \Pi \vdash \Lambda}}{\forall x P[A]_\lambda(x), \Pi \vdash \Lambda}\,(\forall : l)}{\dots}}{R[\forall x P[A]_\lambda(x)]_\theta, \Pi' \vdash \Lambda'}}{\Gamma'', \Pi' \vdash \Delta'', \Lambda'}\,(cut : R[\forall x P[A]_\lambda(x)]_\theta)$$

now the induction hypothesis can be applied in order to obtain the following derivation (According also to lemmas 4.1.22 and 4.1.23):

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dots}{A(a), \Gamma \vdash \Delta, A(a')}}{A(0), \Gamma \vdash \Delta, A(t')}\,(ind)}{\dots}}{\Gamma'' \vdash^{S_0} \Delta'', R[P[A]_\lambda(t')]_\theta} \quad \dfrac{\dfrac{\dfrac{\dots}{P[A]_\lambda(t'), \Pi \vdash \Lambda}}{\dots}}{R[P[A]_\lambda(t')]_\theta, \Pi' \vdash \Lambda'}}{\Gamma'', \Pi' \vdash \Delta'', \Lambda'}\,(cut : R[P[A]_\lambda(t')]_\theta)}{\dfrac{\text{(According to lemmas 4.1.22 and 4.1.23 and additonal exchanges)}}{\Gamma''', \Pi' \vdash \Delta''', C, \Lambda'}}$$

– J cannot be $w : r$ or $c : r$.
– If the rule is $\neg : r$, we have the derivation:

$$
\dfrac{
\begin{array}{c}
\cdots \\ \hline
A(a), \Gamma \vdash \Delta, A(a') \\ \hline
A(0), \Gamma \vdash \Delta, A(b) \quad (ind) \\ \hline
\cdots \\ \hline
\Gamma' \vdash \Delta', P[A]_\lambda(b) \\ \hline
\Gamma' \vdash \Delta', \forall x P[A]_\lambda(x) \quad (\forall : r) \\ \hline
\cdots \\ \hline
R[\forall x P[A]_\lambda(x)]_\theta, \Gamma'' \vdash \Delta'' \\ \hline
\Gamma'' \vdash \Delta'', \neg R[\forall x P[A]_\lambda(x)]_\theta \quad (\neg : r)
\end{array}
\qquad
\begin{array}{c}
\cdots \\ \hline
P[A]_\lambda(t), \Pi \vdash \Lambda \\ \hline
\forall x P[A]_\lambda(x), \Pi \vdash \Lambda \quad (\forall : l) \\ \hline
\cdots \\ \hline
\Pi' \vdash \Lambda', R[\forall x P[A]_\lambda(x)]_\theta \\ \hline
\neg R[\forall x P[A]_\lambda(x)]_\theta, \Pi' \vdash \Lambda' \quad (\neg : l) \\ \hline
\cdots \\ \hline
\neg R[\forall x P[A]_\lambda(x)]_\theta, \Pi'' \vdash \Lambda''
\end{array}
}{
\Gamma'', \Pi'' \vdash \Delta'', \Lambda''
} \; (cut : \neg R[\forall x P[A]_\lambda(x)]_\theta)
$$

we will obtain the following derivation with ISB of size n:

$$
\dfrac{
\begin{array}{c}
\cdots \\ \hline
P[A]_\lambda(t), \Pi \vdash \Lambda \\ \hline
\forall x P[A]_\lambda(x), \Pi \vdash \Lambda \quad (\forall : l) \\ \hline
\cdots \\ \hline
\Pi' \vdash \Lambda', R[\forall x P[A]_\lambda(x)]_\theta
\end{array}
\qquad
\begin{array}{c}
\cdots \\ \hline
A(a), \Gamma \vdash \Delta, A(a') \\ \hline
A(0), \Gamma \vdash \Delta, A(b) \quad (ind) \\ \hline
\cdots \\ \hline
\Gamma' \vdash \Delta', P[A]_\lambda(b) \\ \hline
\Gamma' \vdash \Delta', \forall x P[A]_\lambda(x) \quad (\forall : r) \\ \hline
\cdots \\ \hline
R[\forall x P[A]_\lambda(x)]_\theta, \Gamma'' \vdash \Delta''
\end{array}
}{
\Gamma'', \Pi' \vdash \Delta'', \Lambda'
} \; (cut : \neg R[\forall x P[A]_\lambda(x)]_\theta)
$$

and apply the induction hypothesis, noting also that the derivation between the two $\neg R[\forall x P[A]_\lambda(x)]_\theta$ on the right in the previous derivation satisfies all of lemma 4.1.22 prerequisites, to obtain the following derivation:

$$
\dfrac{
\dfrac{
\begin{array}{c}
\cdots \\ \hline
P[A]_\lambda(t'), \Pi \vdash \Lambda \\ \hline
\cdots \\ \hline
\Pi' \vdash \Lambda', R[P[A]_\lambda(t')]_\theta
\end{array}
\qquad
\begin{array}{c}
\cdots \\ \hline
A(a), \Gamma \vdash \Delta, A(a') \\ \hline
A(0), \Gamma \vdash \Delta, A(t') \quad (ind) \\ \hline
\cdots \\ \hline
R[P[A]_\lambda(t')]_\theta, \Gamma'' \vdash \Delta''
\end{array}
}{
\Gamma'', \Pi' \vdash \Delta'', \Lambda'
} \; (cut : \neg R[P[A]_\lambda(t')_\theta)
}{
\dfrac{\text{(According to lemmas 4.1.22 and 4.1.23 and additonal exchanges)}}{\Gamma'', \Pi'' \vdash \Delta'', \Lambda''}
}
$$

– The rule cannot be a $\forall : r$ as we deal with the outermost critical quantifier.
– The rule is $\forall : l$ so we have (notice we have changed the places of the two branches):

$$
\dfrac{
\begin{array}{c}
\cdots \\ \hline
P[A]_\lambda(t(c), c), \Pi(c) \vdash \Lambda(c) \\ \hline
\forall x P[A]_\lambda(x, c), \Pi(c) \vdash \Lambda(c) \quad (\forall : l) \\ \hline
\cdots \\ \hline
\Pi' \vdash \Lambda', R[\forall x P[A]_\lambda(x)]_\theta(c) \\ \hline
\Pi' \vdash \Lambda', \forall y R[\forall x P[A]_\lambda(x)]_\theta(y) \quad (\forall : r) \\ \hline
\cdots \\ \hline
\Pi'' \vdash \Lambda'', \forall y R[\forall x P[A]_\lambda(x)]_\theta(y)
\end{array}
\qquad
\begin{array}{c}
\cdots \\ \hline
A(a), \Gamma \vdash \Delta, A(a') \\ \hline
A(0), \Gamma \vdash \Delta, A(b) \quad (ind) \\ \hline
\cdots \\ \hline
\Gamma' \vdash \Delta', P[A]_\lambda(b) \\ \hline
\Gamma' \vdash \Delta', \forall x P[A]_\lambda(x) \quad (\forall : r) \\ \hline
\cdots \\ \hline
R[\forall x P[A]_\lambda(x)]_\theta(s), \Gamma'' \vdash \Delta'' \\ \hline
\forall y R[\forall x P[A]_\lambda(x)]_\theta(y), \Gamma'' \vdash \Delta'' \quad (\forall : l)
\end{array}
}{
\Pi'', \Gamma'', \vdash \Lambda'', \Delta''
} \; (cut :)
$$

The above transformation is the transformation where the term t may change as it may contain a free variable which is replaced in this step. We know from lemma 3.1.8 that if there is a proof of $\Pi' \vdash \Lambda', R[\forall x P[A]_\lambda(x)]_\theta(c)$ there is a proof of $\Pi' \vdash \Lambda', R[\forall x P[A]_\lambda(x)]_\theta(s)$, as c was a fully indicated new free variable. So we can replace the derivation by the following derivation with ISB of size n:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dots}{P[A]_\lambda(t(s),s),\Pi(s)\vdash\Lambda(s)}}{\forall x P[A]_\lambda(x,s),\Pi(s)\vdash\Lambda(s)}\ (\forall:l)}{\dots}}{\Pi'\vdash\Lambda', R[\forall x P[A]_\lambda(x)]_\theta(s)} \qquad \dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dots}{A(a),\Gamma\vdash\Delta,A(a')}}{A(0),\Gamma\vdash\Delta,A(b)}\ (ind)}{\dots}}{\Gamma'\vdash\Delta',P[A]_\lambda(b)}}{\Gamma'\vdash\Delta',\forall x P[A]_\lambda(x)}\ (\forall:r)}{\dots}}{R[\forall x P[A]_\lambda(x)]_\theta(s),\Gamma''\vdash\Delta''}}{\Pi',\Gamma'',\vdash\Lambda',\Delta''}\ (cut: R[\forall x P[A]_\lambda(x)]_\theta(s))$$

now the induction hypothesis and lemmas 4.1.22 and 4.1.23 can be used in order to obtain the derivation:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dots}{P[A]_\lambda(t'(s),s),\Pi(s)\vdash\Lambda(s)}}{\dots}}{\Pi'\vdash\Lambda', R[P[A]_\lambda(t'(s))]_\theta(s)} \qquad \dfrac{\dfrac{\dfrac{\dfrac{\dots}{A(a),\Gamma\vdash\Delta,A(a')}}{A(0),\Gamma\vdash\Delta,A(t'(s))}\ (ind)}{\dots}}{R[P[A]_\lambda(t'(s))]_\theta(s),\Gamma''\vdash\Delta''}}{\dfrac{\dfrac{\Pi',\Gamma'',\vdash\Lambda',\Delta''}{\text{(According to lemmas 4.1.22 and 4.1.23 and additonal exchanges)}}}{\Pi'',\Gamma'',\vdash\Lambda'',\Delta''}}\ (cut: R[P[A]_\lambda(t'(s))]_\theta(s))$$

As was mentioned before, the term $t(c)$ with c being fully indicated in t is changed here into $t(s)$.

– The rule is $\vee : r_1$:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dots}{A(a),\Gamma\vdash\Delta,A(a')}}{A(0),\Gamma\vdash\Delta,A(b)}\ (ind)}{\dots}}{\Gamma'\vdash\Delta',P[A]_\lambda(b)}}{\Gamma'\vdash\Delta',\forall x P[A]_\lambda(x)}\ (\forall:r)}{\dots}}{\Gamma''\vdash\Delta'',R[\forall x P[A]_\lambda(x)]_\theta}}{\Gamma''\vdash\Delta'',R[\forall x P[A]_\lambda(x)]_\theta\vee W}\ (\vee:r_1) \qquad \dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dots}{P[A]_\lambda(t),\Pi\vdash\Lambda}}{\forall x P[A]_\lambda(x),\Pi\vdash\Lambda}\ (\forall:l)}{\dots}}{R[\forall x P[A]_\lambda(x)]_\theta,\Pi_1'\vdash\Lambda_1'} \quad \dfrac{\dots}{W,\Pi_2'\vdash\Lambda_2'}}{R[\forall x P[A]_\lambda(x)]_\theta\vee W,\Pi_1',\Pi_2'\vdash\Lambda_1',\Lambda_2'}\ (\vee:l)}{\dfrac{\dots}{R[\forall x P[A]_\lambda(x)]_\theta\vee W,\Pi''\vdash\Lambda''}}}{\Gamma'',\Pi''\vdash\Delta'',\Lambda''}\ (cut)$$

and we obtain the following derivation with ISB of size $= n$:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dots}{A(a),\Gamma\vdash\Delta,A(a')}}{A(0),\Gamma\vdash\Delta,A(b)}\ (ind)}{\dots}}{\Gamma'\vdash\Delta',P[A]_\lambda(b)}}{\Gamma'\vdash\Delta',\forall x P[A]_\lambda(x)}\ (\forall:r)}{\dots}}{\Gamma''\vdash\Delta'',R[\forall x P[A]_\lambda(x)]_\theta} \qquad \dfrac{\dfrac{\dfrac{\dfrac{\dots}{P[A]_\lambda(x),\Pi\vdash\Lambda}}{\forall x P[A]_\lambda(x),\Pi\vdash\Lambda}\ (\forall:l)}{\dots}}{R[\forall x P[A]_\lambda(x)]_\theta,\Pi_1'\vdash\Lambda_1'}}{\Gamma'',\Pi_1'\vdash\Delta'',\Lambda_1'}\ (cut: R[\forall x P[A]_\lambda(x)]_\theta)$$

again, according to the induction hypothesis and lemmas 4.1.22 and 4.1.23 and additonal exchanges and weakenings we can obtain the derivation:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dots}{A(a), \Gamma \vdash \Delta, A(a')}}{A(0), \Gamma \vdash \Delta, A(t')}\,(ind)}{\dots}}{\Gamma' \vdash \Delta', P[A]_\lambda(t')}}{\Gamma' \vdash \Delta', P[A]_\lambda(t')}\,(\forall : r)}{\dots}}{\Gamma'' \vdash \Delta'', R[P[A]_\lambda(t')]_\theta} \quad \dfrac{\dfrac{\dfrac{\dots}{P[A]_\lambda(t'), \Pi \vdash \Lambda}}{\dots}}{R[P[A]_\lambda(t')]_\theta, \Pi'_1 \vdash \Lambda'_1}}{\dfrac{\dfrac{\dfrac{\Gamma'', \Pi'_1 \vdash \Delta'', \Lambda'_1}{\text{(weakenings)}}}{\Gamma'', \Pi'_1, \Pi'_2 \vdash \Delta'', \Lambda'_1, \Lambda'_2}}{\dfrac{\text{(lemmas 4.1.22 and 4.1.23 and additonal exchanges)}}{\Gamma'', \Pi'' \vdash \Delta'', \Lambda''}}}\,(cut : R[P[A]_\lambda(t')]_\theta)$$

- The rest of the inferences are dealt with in a similar way.
- P' is suitable as we have satisfied all requirements of lemma 4.1.18. We have not added or changed inductions.

4.3 Elimination of Contractions

4.3.1 Analyzing the method of the elimination of contractions

In the projection lemma we have assumed, that none of the sub-bundles containing both the cut-formula and the quantifier has any contraction. We will achieve this result by replacing contractions by cuts. The idea of the transformation is that a sub-bundle which contains contractions branches into two sub-bundles. We will show we can transform such a derivation into a derivation where each sub-bundle is distinct from the other including the cut-formula. Then we can double the cut in order to have the same end-sequent.

Example 4.3.1 (Contraction over a dual term)**.** *The following example shows the problem of having two different PSBs for the same quantifier.*

$$
\dfrac{
\dfrac{
\dfrac{
\dfrac{
\dfrac{A(a),\Gamma \vdash \Delta, A(a')}{A(0),\Gamma \vdash \Delta, A(b)}\,(ind)
}{\cdots}
}{\Gamma' \vdash \Delta', P[A]_\lambda(b)}
}{\Gamma' \vdash \Delta', \forall x P[A]_\lambda(x)}\,(\forall : r)
}{\dfrac{\cdots}{\Gamma'' \vdash \Delta'', R[\forall x P[A]_\lambda(x)]_\theta}}
\quad
\dfrac{
\dfrac{
\dfrac{
\dfrac{P[A]_\lambda(s), P[A]_\lambda(t), \Pi'''' \vdash \Lambda''''}{\forall x P[A]_\lambda(x), P[A]_\lambda(t), \Pi'''' \vdash \Lambda''''}\,(\forall : l)
}{\dfrac{\cdots}{P[A]_\lambda(t), \forall x P[A]_\lambda(x), \Pi''' \vdash \Lambda'''}}
}{\forall x P[A]_\lambda(x), \forall x P[A]_\lambda(x), \Pi''' \vdash \Lambda'''}\,(\forall : l)
}{
\dfrac{
\dfrac{\cdots}{\dfrac{\forall x P[A]_\lambda(x), \forall x P[A]_\lambda(x), \Pi'' \vdash \Lambda''}{\forall x P[A]_\lambda(x), \Pi'' \vdash \Lambda''}\,(contraction)}
}{\dfrac{\cdots}{R[\forall x P[A]_\lambda(x)]_\theta, \Pi' \vdash \Lambda'}}
}
$$

$$
\dfrac{\Gamma'' \vdash \Delta'', R[\forall x P[A]_\lambda(x)]_\theta \qquad R[\forall x P[A]_\lambda(x)]_\theta, \Pi' \vdash \Lambda'}{\Gamma'', \Pi' \vdash \Delta'', \Lambda'}\,(cut : R[\forall x P[A]_\lambda(x)]_\theta)
$$

The transformation we are going to show is straightforward but imposes several difficulties. First, we must ensure that the new proofs will be suitable, as this property is assumed in many lemmas in this thesis. Second, we must show that the process terminates as we are adding many contractions in each step. We will deal with a contracted formula by pushing the contraction downward the derivation until we can cut each of its instances separately. The following example shows the general idea of the transformation:

Example 4.3.2 (Elimination of contractions). *Here we have a sub-bundle that branches into two different sub-bundles because of the contraction:*

$$
\dfrac{\dfrac{\dfrac{\vdash A, A}{\vdash A}\,(contraction)}{\vdash A \lor B}\,(\lor : r_1)}{\neg(A \lor B) \vdash}\,(\neg : l)
$$

We wish to remove the contraction by duplicating the end formula of this sub-bundle:

$$
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\vdash A, A}{\vdash A, A \lor B}\,(\lor : r_1)}{\vdash A \lor B, A}\,(exchange)}{\vdash A \lor B, A \lor B}\,(\lor : r_1)}{\neg(A \lor B) \vdash A \lor B}\,(\neg : l)}{\neg(A \lor B), \neg(A \lor B) \vdash}\,(\neg : l)
$$

Given that the cut formula was $\neg A \lor B$, we can do the same cut twice and have no contraction.

4.3.2 Handling contractions over eigenvariables

Another difficulty we may encounter in eliminating contractions is that a contraction may also eliminate a free variable that is used as an eigenvariable and not only eliminate a formula. Consider for example the following situation:

Example 4.3.3 (Contraction over an eigenvariable)**.** *Here one instance of the free variable b is being eliminated by the contraction.*

$$\dfrac{\dfrac{\vdash A(b), A(b)}{\vdash A(b)}\ \textit{(contraction)}}{\vdash \forall x A(x)}\ (\forall : r)$$

If we try to follow the same procedure as in example 4.3.2, we will violate the eigenvariable condition. We will show that free variables which are used as eigenvariables, can be replaced by other variables, even if they are not fully indicated.

In order to give the definition of polarity-balanced formulas, we will add to PA the logical symbols $\wedge$ and $\exists$. After we have proved lemma 4.3.8, we will revert to our original calculus.

Definition 4.3.4 (Negation normal form)**.** The set of formulas in negation normal form (nnf) is being defined inductively:

1. For an atomic formula A, A and $\neg A$ are in nnf.
2. If A and B are in nnf, then so are $A \vee B$ and $A \wedge B$.
3. If A is in nnf, then so are $\forall x A$ and $\exists x A$.

Formulas which do not contain any logical symbol other than $\neg$ will be called negative atomic formulas.

Lemma 4.3.5. Every formula in LK extended by the symbols $\wedge$ and $\exists$ is equivalent to another formula in nnf.

Proof - by decomposing formulas according to De-Morgan rules and the logical equivalence between $\neg\forall x A$ and $\exists x \neg A$.

Definition 4.3.6 (Polarity-balanced)**.** Let A and B be two formulas and A_{nnf} and B_{nnf} be their equivalent formulas in nnf. Let α and β be the sets containing all the negative atomic sub-formulas of A_{nnf} and B_{nnf}. Then:

- A and B are polarity-balanced if the intersection of α and β is empty.
- Let S be a sequent containing A and B. Then A and B are polarity-balanced with respect to S if:
  - A and B are polarity-balanced and are each in the succedent and antecedent of S.
  - Both A and $\neg B$ and $\neg A$ and B are polarity-balanced and A and B are both in the succedent or antecedent of S.

We will revert back to our original definition of PA.

Lemma 4.3.7. Let β be an arbitrary sub-bundle and A and B two formulas occurring in all sequents of this sub-bundle. If A and B are polarity-balanced with regard to a sequent S in β, then any pair of a sub-formula of A and a sub-formula of B will be polarity-balanced with regard to any ancestor of S in β in which they are both contained.

Proof

- The proof is by induction on the size n of β.
- If $n = 0$ then it is clear.
- If $n > 0$ then we assume that it holds for sub-bundles of size up to n and prove it for size $n + 1$ by cases according to the inference which has the last sequent S of β as its lower sequent.
- It is clear that none of the inference rules, except $\neg$, can change the polarity-balance of formulas with respect to the sequent they occur in. $\neg$ also does not change the polarity-balance as one formula must change sides in the sequent. Therefore, its negative atomic formulas cannot be identical if they have not been identical before.
- Therefore we can apply the induction hypothesis and prove the lemma. Because the upper sequent contains the sub-formulas of A and B and they are polarity-balanced with regard to the upper sequent, the induction hypothesis can be applied in order to prove that all sub-formulas of A and B are polarity-balanced in all sequents of β in which they occur.

Lemma 4.3.8. Let P be a proof of a sequent S containing two formulas $A(a)$ and $B(a)$ which are polarity-balanced with regard to S and in which a is fully indicated. We can obtain a proof P' of S' where in S', $B(a)$ is replaced by $B(b)$ for a free variable b, which does not occur in P.

Proof

- Let β be the set of all threads in P containing sub-formulas of either $A(a)$ or $B(a)$ which also contain the free variable a. Let α be the set containing the sequence of sequents which is the union of all threads in β.

- According to lemma 4.3.7 if $A(a)$ and $B(a)$ are polarity-balanced with regard to S, then their sub-formulas, which appear both in a sequent S' which is an ancestor of S, will be polarity-balanced with regard to S'.

- We prove it by induction on the length n of α.

- If $n = 1$ and let S_0 be the only sequent contained in α. Let $A'(a)$ and $B'(a)$ be the two sub-formulas of $A(a)$ and $B(a)$ contained in S_0, then we may have one of the following:

  – $A'(a)$ (or $B'(a)$) is the principal formula of a weakening, we replace it by $A'(b)$ where b is a new free variable not appearing in P.

  – S_0 cannot be a logical axiom, because in this case $B'(a)$ is equal to $A'(a)$ and being polarity-balanced with regard to S_0, they cannot appear each in the antecedent and succedent of S_0.

  – S_0 cannot be a mathematical axiom as none of the mathematical axioms contains more than one formula.

  – S_0 is the lower sequent of an induction with $A'(a)$ (or $B(a)$) as the right principal formula. We replace $A'(a)$ with $A'(b)$ with b being a new free variable not appearing in P.

- If $n > 0$, we assume the lemma is true for α of length n and prove it for length $n+1$. Let S_0 be the last sequent contained in α. Let $A'(a)$ and $B'(a)$ be the two sub-formulas of $A(a)$ and $B(a)$ contained in S_0.

- We examine the inference J which has S_0 as its lower sequent. According to lemma 4.3.7 all inference rules preserve the polarity-balance property.

- Let Let $A''(a)$ and $B''(a)$ be the two sub-formulas of $A'(a)$ and $B'(a)$ contained in upper sequent of S_0.

- By the induction hypothesis, we can replace $A''(a)$ (or $B''(a)$) by $A''(b)$ in the upper sequent of J. Where b is a new free variable not appearing in P.
- We can then apply J in order to obtain a proof P' of S'.

We should remark that an elimination of a contraction in the procedure (given in the previous examples) will double the number of contractions on other bundles, which contain the formula at the end of the bundles. For example:

Example 4.3.9 (Doubling contractions)**.** *The process shown in example 4.3.2 works in a straightforward way until we duplicate a binary inference rule. In this case a whole proof and all its context must be duplicated. The following derivation shows the result of eliminating a contraction, transforming:*

$$\dfrac{P,\Gamma \vdash \Delta \qquad \dfrac{Q,Q,\Pi \vdash \Lambda}{Q,\Pi \vdash \Lambda}\ \textit{(contraction)}}{P \vee Q,\Gamma,\Pi \vdash \Delta,\Lambda}\ (\vee : l)$$

into:

$$\dfrac{P,\Gamma \vdash \Delta \qquad \dfrac{\dfrac{\dfrac{P,\Gamma \vdash \Delta \qquad Q,Q,\Pi \vdash \Lambda}{P \vee Q,\Gamma,Q,\Pi \vdash \Delta,\Lambda}\ (\vee : l)}{\textit{(exchanges)}}}{Q,P \vee Q,\Gamma,\Pi \vdash \Delta,\Lambda}}{P \vee Q,\Gamma,P \vee Q,\Gamma,\Pi \vdash \Delta,\Delta,\Lambda}\ (\vee : l)$$

4.3.3 The elimination of contractions

If we assume that there were some contractions in sub-bundles containing P, then all of them will be doubled in the process. But although we have duplicated a whole derivation, it has been done for the newly created sub-bundle only. The original sub-bundle stays the same (still branches only once). The process can be seen as taking a stick with many leaves and side branches, which branches also at the end into two branches. Then splitting it in a way such that each part will contain all the branches and leaves of the original stick. Therefore both individual sticks contain one less branching although the number of sticks was doubled.

We can see that the only formulas we get duplicated in the context formulas in example 4.3.9 are formulas not belonging to any sub-bundle containing $P \vee Q$. Therefore, contracting these formulas will not add a new contraction in the sub-bundles containing $P \vee Q$.

Definition 4.3.10 (Formula trees)**.** A basic formula tree for a derivation ψ is the tree obtained from ψ when considering the occurrences of formulas as nodes and their successor relation as branches. I.e.

1. Any formula contained in a sequent of ψ can be the root of the tree.
2. The basic formula tree contains an occurrence of a formula A_2 if it contains an occurrence of a formula A_1 and A_1 is the successor of A_2 in ψ.
3. Let A be an occurrence of a formula in the basic formula tree which was a principal formula in ψ. We label the edges ending with A in the formula tree by the inferences in ψ having this occurrence of A as principal formula.

A formula tree is obtained from a basic formula tree by removing all edges which are labeled by an exchange or not labeled at all.
The set containing all formula occurrences in ψ, which appear in the formula tree is called the projection of the formula tree in ψ.

Example 4.3.11 (A formula tree)**.** *The following tree is a formula tree for the second derivation given in example 4.3.9.*

$$\begin{array}{c} P \vee Q \\ \overbrace{P \quad Q} \end{array}$$

The two edges are labeled by $\vee : l$*.*

Definition 4.3.12 (Identity of formula trees)**.** Two formula trees T_1 and T_2 for a derivation ψ are identical if they are syntactically identical.

The two formula trees rooted at the two $P \vee Q$ in example 4.3.9 are identical according to this definition.

Definition 4.3.13 (Disjoint projections of formula trees)**.** As the projections in ψ of two formula trees are sets, they are disjoint if the sets are disjoint.

Lemma 4.3.14. Let ψ:

$$\dfrac{\dfrac{Q, Q, \Pi \vdash \Lambda}{Q, \Pi \vdash \Lambda}\ \text{(contraction)}}{\dfrac{\cdots}{R, \Pi' \vdash \Lambda'}}$$

be a derivation where Q and R are on the far ends of the same sub-bundle, such that this sub-bundle does not contain a contraction (except the one shown) or an induction. Let T be the formula tree for ψ rooted at R and ending below the contraction.

We can obtain the following derivation without this contraction and without any new contraction on each of the sub-bundles starting with Q and ending with R:

$$\dfrac{\dfrac{Q, Q, \Pi \vdash \Lambda}{\cdots}}{R, R, \Pi' \vdash \Lambda'}$$

The new derivation will have a newly created formula tree which is identical to T and its projection is disjoint to the projection of T.

- Remarks

  1. The lemma shows the case where the sub-bundle starts and ends with formulas in the antecedent of the sequents. However, both can be in the succedent or each in the antecedent and succedent.
  2. The lemma takes one sub-bundle that branches at the end into two sub-bundles via the contraction rule and creates two different sub-bundles without this branching.

Proof

- If there are two identical formulas with the same free variable, which is used as eigenvariable, we can use lemma 4.3.8 in order to obtain a derivation with the two formulas containing two different eigenvariables. The reason we can use the lemma is that two identical formulas on the same side of a sequent are polarity-balanced with regard to the sequent.

- By induction on the number n of inferences in the derivation below the contraction.

- If $n = 0$ then we delete the contraction to obtain $Q, Q, \Pi \vdash \Lambda$. As both formula trees are of one element only they are clearly identical and their projections are disjoint.

- if $n > 0$ then assume the lemma is correct for all derivations of size $< n$ and prove if for size n according to the last inference J:
  - R is not the principal formula of J, then we have:

$$\dfrac{\dfrac{\dfrac{\dfrac{Q, Q, \Pi \vdash \Lambda}{Q, \Pi \vdash \Lambda}\text{(contraction)}}{\cdots}}{R, \Pi'' \vdash \Lambda''}}{R, \Pi' \vdash \Lambda'}\text{(J)}$$

    By using the induction hypothesis we can obtain $R, R, \Pi'' \vdash \Lambda''$ without the contraction and with a new identical formula tree with a disjoint projection. We need to apply J in order to obtain $R, R, \Pi' \vdash \Lambda'$ (and maybe exchanges as well). As J does not have a formula from one of the formula trees as principal formula (except maybe exchanges), the two trees are identical and their projections are disjoint.
  - R is the principal formula of J:
  - J is exchange then we have:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{Q, Q, \Pi \vdash \Lambda}{Q, \Pi \vdash \Lambda}\text{(contraction)}}{\cdots}}{R, \Pi'' \vdash^{S_1} \Lambda''}}{\cdots}}{A, R, \Pi' \vdash^{S_0} \Lambda'}}{R, A, \Pi' \vdash \Lambda'}\text{(J)}$$

    S_1 is the lowermost sequent above S_0 with R on the leftmost position. S_1 must exist as either $R \equiv Q$ or it must be the principal formula of a logical rule and must be on the leftmost position. We know from the induction hypothesis that we can obtain a derivation of $R, R, \Pi'' \vdash \Lambda''$. We know also (from lemma 4.1.22) that we can obtain a derivation of $R, R, A, \Pi' \vdash \Lambda'$. This step may add to the formula trees only exchanges so they are identical and their projections are disjoint.
  - J cannot be a contraction.
  - J cannot be weakening as R is on the same bundle as Q.
  - J is a logical inference, for example:
  - J is $\forall : r$. We have the following derivation:

$$\dfrac{\dfrac{\dfrac{\dfrac{\Gamma \vdash \Delta, A(a), A(a)}{\Gamma \vdash \Delta, A(a)}\,\text{(c:r)}}{\cdots}}{\Gamma' \vdash \Delta', A(a)}}{\Gamma' \vdash \Delta', \forall x A(x)}\,(\forall : r)$$

As specified, we can use lemma 4.3.8 in order to obtain a derivation with end-seuqent $\Gamma \vdash \Delta, A(a), A(b)$ and therefore we have:

$$\dfrac{\dfrac{\dfrac{\dfrac{\Gamma \vdash \Delta, A(a), A(b)}{\cdots}}{\Gamma' \vdash \Delta', A(a), A(b)}}{\Gamma' \vdash \Delta', A(a), \forall x A(x)}\,(\forall : r)}{\Gamma' \vdash \Delta', \forall x A(x), \forall x A(x)}\,(\text{exchange and } \forall : r)$$

– J is $\vee : l$ i.e.

$$\dfrac{P, \Gamma \vdash \Delta \qquad \dfrac{\dfrac{\dfrac{Q, Q, \Pi \vdash \Lambda}{Q, \Pi \vdash \Lambda}\,\text{(contraction)}}{\cdots}}{R, \Pi' \vdash \Lambda'}}{P \vee R, \Gamma, \Pi' \vdash \Delta, \Lambda'}\,(\vee : l)$$

By the induction hypothesis we can obtain a derivation of $R, R, \Pi' \vdash \Lambda'$ without the contraction and with a new identical formula tree rooted at R. We apply $\vee : l$ with $P, \Gamma \vdash \Delta$ twice to obtain $P \vee R, P \vee R, \Gamma, \Gamma, \Pi' \vdash \Delta, \Delta, \Lambda'$. Now, as the sequences of formulas Γ and Δ are context formulas for P and the two $P \vee R$, they are not included in any sub-bundle containing any of the $P \vee R$ and they can be contracted:

$$\dfrac{\dfrac{\dfrac{P, \Gamma \vdash \Delta \qquad \dfrac{\dfrac{\dfrac{P, \Gamma \vdash \Delta \qquad \dfrac{\dfrac{Q, Q, \Pi \vdash \Lambda}{\cdots}}{R, R, \Pi' \vdash \Lambda'}}{P \vee R, \Gamma, R, \Pi' \vdash \Delta, \Lambda'}\,(\vee : l)}{\text{(exchanges)}}}{R, P \vee R, \Gamma, \Pi' \vdash \Delta, \Lambda'}}{P \vee R, \Gamma, P \vee R, \Gamma, \Pi' \vdash \Delta, \Delta, \Lambda'}\,(\vee : l)}{\text{(contractions and exchanges)}}}{P \vee R, P \vee R, \Gamma, \Pi' \vdash \Delta, \Lambda'}$$

This step is the only step we contract formulas. However, none of the formulas contracted occurs in a bundle containing any of the $P \vee R$. We also duplicate the proof of $P, \Gamma \vdash \Delta$. Therefore, the bundles containing P may contain contractions as well, which will be doubled. The two formula trees rooted at $P \vee R$ are identical and their projections are disjoint as we have added one $\vee : l$ to each (in addition some exchanges which are ignored). The added contractions are, as we have mentioned, on formulas not contained in these formula trees.

– J is $\vee : r_2$ i.e.

$$\dfrac{\dfrac{\dfrac{\dfrac{\Pi \vdash \Lambda, Q, Q}{\Pi \vdash \Lambda, Q}\ (\text{contraction})}{\cdots}}{\Pi' \vdash \Lambda', R}}{\Pi' \vdash \Lambda', P \vee R}\ (\vee : r_2)$$

By the induction hypothesis we obtain $\Pi' \vdash \Lambda', R, R$ and by applying $\vee : r$ twice and some exchanges we obtain $\Pi' \vdash \Lambda', P \vee R, P \vee R$. The two formula trees are identical and their projections are disjoint as we have added one $\vee : r$ to each sub-bundle and exchanges.

– J is $\neg : l$

$$\dfrac{\dfrac{\dfrac{\dfrac{\Pi \vdash \Lambda, Q, Q}{\Pi \vdash \Lambda, Q}\ (\text{contraction})}{\cdots}}{\Pi' \vdash \Lambda', R}}{\neg R, \Pi' \vdash \Lambda'}\ (\neg : l)$$

As before, we use the induction hypothesis in order to obtain $\Pi' \vdash \Lambda', R, R$ and apply $\neg : l$ twice with an exchange. Because we have added one $\neg : l$ to each sub-bundle and exchanges, the formula trees are identical and their projections are disjoint.

The next lemmas prove that we can also eliminate all contractions in all sub-bundles containing an occurrence of formula R (which is doubled after the elimination of each contraction). As we know that each elimination of a contraction can double the number of the other contractions, we will have to label all contractions. The labeling will ensure that in each elimination of a contraction we may double only contractions of a different label. This is because each bundle contains only one contraction of each label. The following corollary follows from lemma 4.3.14.

Corollary 4.3.15. By assuming some labeling on all the contractions that are on sub-bundles which contain the formula R, the process given in lemma 4.3.14 will result in two identical formula trees. Each formula tree will contain exactly the original contractions except c.

Using this corollary, we now prove that we can eliminate all contractions in all sub-bundles of a derivation in a similar manner. The proof will be given by induction on the labels. Elimination of one labeled contraction may increase the number of contractions with a different label only. Therefore, if we process the contractions in an orderly manner, we are ensured this process terminates.

Lemma 4.3.16. Suppose we have a derivation of $R^k, \Pi \vdash \Lambda$ where the number of contraction in the formula tree of each occurrence of R is exactly m. Suppose also that all formula trees of occurrences of Rs are identical and their projections are disjoint. We can obtain a proof of $R^{(2^m * k)}, \Pi \vdash \Lambda$ with these contractions removed and with no new contraction in a formula tree of some occurrence of R.

Proof

- Our first step will be to label the these contractions (in the formula trees of the R^k). Because all the formula trees are identical, we can execute such an uniform labeling. Given one formula tree, we label each of its m contractions by an unique natural number such that contractions which occur below other contractions are being labeled first. The proof will eliminate in parallel one contraction in all formula trees of occurrences of R.

- The proof is by induction on m.

- if $m = 0$, we have a contraction free derivation of $R^k, \Pi \vdash \Lambda$, because we do not have any contraction in the derivation.

- If $m > 0$ then we assume that the lemma is true for all proofs with m contractions in all the Rs sub-bundles sets and we prove it for $m+1$ contractions. Let n be the biggest contraction's label.

- Each occurrence of R is contained in a sub-bundle which contains a contraction labeled by n. As the projections of the formula trees are disjoint we apply lemma 4.3.14 k times in order to obtain a proof of $R^{(2*k)}, \Pi \vdash \Lambda$. According to corollary 4.3.15, all the $2k$ formula trees are identical and contain $m-1$ contractions.Therefore, we can apply the induction hypothesis in order to obtain a proof of $R^{(2^{m-1}*2*k)}, \Pi \vdash \Lambda \equiv R^{(2^m*k)}, \Pi \vdash \Lambda$.

- We notice that although some occurrences of R are considered as context formulas when applying lemma 4.3.14, there are no new contractions on them. This is so because we have added contractions only on the context of formulas which are auxiliary in a sub-bundle of some occurrence of R. All the Rs are on disjoint sub-bundles because their projections of formula trees are disjoint.

Given this procedure of eliminating contractions, we can now prove the contraction elimination lemma. The idea is that using lemma 4.3.16 we can replace a proof of $R, \Pi \vdash \Lambda$ with m contractions in sub-bundles containing R by a proof of $R^{2^m}, \Pi \vdash \Lambda$ without these contractions. We duplicate the cut with R as cut formula 2^m times and obtain a proof of the same end-sequent. The process eliminates certain contractions but add many others (both in lemma 4.3.14 and in the following lemma). However, these contractions do no occur in sub-bundles containing the cut-formula. The process also duplicates existing derivations but nonetheless , we show that the algorithm terminates.

In the projection lemma we assume that not only all PSBs are contraction free but the sub-bundle between the induction and the cut-formula is contraction free as well. Given a proof P containing n inductions labeled by l and m problematic cuts labeled by l, we examine the total number of problematic contractions divided into two groups. All the contractions which are on some sub-bundle below an induction and a problematic cut (both labeled by l) are in one group. All contractions which are on some PSB of an induction labeled by l are in the second group. We will eliminate all these contractions by working on a specific label of inductions. We execute two algorithms, both iterating on all problematic cuts which are critical (as will be defined next). The first algorithm will eliminate all contractions of the first group, while increasing the number of contractions of the second group. The second algorithm will eliminate all contractions of the second group, while increasing the number of inductions and problematic cuts. At the end of the process, we will have a proof which does not have any contraction in the problematic places for all inductions of label l.

In the process of the contraction elimination lemma we are contracting many formulas and duplicating many subproofs. We must ensure that all these duplications and contractions will not render the proof unsuitable. We must ensure aso that when executing each algorithm, the total number of critical problematic cuts will decrease.

We will give one proof for both algorithms:

Definition 4.3.17 (Critical cuts, sub-bundles and contractions)**.** For a label l and all problematic cuts labeled by l, we define critical sub-bundles as one of the following:

1. As strong critical sub-bundles we define all sub-bundles ending with a problematic cut labeled by l and beginning with an induction labeled by l.

2. All PSBs of inductions of label l are defined as weak critical sub-bundles.

We define as critical all contractions appearing in a critical sub-bundle. We define as strongly (weakly) critical all problematic cuts labeled by l, whose cut-formula ends a strong (weak) critical sub-bundle, which contains critical contractions.

- Remarks

  1. Please note that we might have critical sub-bundles without critical contractions. The problematic cuts below these sub-bundles will not be defined as critical in this case.
  2. A cut can be both strongly and weakly critical.

Lemma 4.3.18 (Contraction elimination lemma)**.** Let P be a suitable proof containing m critical cuts labeled by the biggest induction label l. The cuts are critical either according to the strong or to the weak definition, but not both. We can obtain a suitable proof P' of the same end-sequent as P and without any critical cut (strong or weak) labeled by l and with the number of (weak or strong) critical cuts labeled by l having the following property $Prop$: If there were no (weak or strong) critical cuts in P then there will be no (weak or strong) critical cuts in P'.

Proof

- The proof is by induction on m.

- $m = 0$. There are no critical cuts and we have obtained the required proof with no further transformation.

- Assuming the lemma is true for m critical cuts, we prove it for $m + 1$ critical cuts.

- There can be a dual case where the contracted formula is on the other side of the sequents and this is dealt in a similar manner.

- Taking a lowermost critical cut below critical sub-bundles containing m critical contractions. We know from lemma 4.3.16 that if there is a proof of $R, \Pi \vdash \Lambda$ containing m contractions in sub-bundles containing R, we can obtain a proof of $R^{2^m}, \Pi \vdash \Lambda$ without these m contractions and with new contractions only on formulas which do not occur in any sub-bundle containing R. The other requirement of the lemma (that all formula trees are identical and their projections are disjoint) is also true, because we deal here only with one formula and so with only one formula tree. So given proof P:

$$\frac{\dfrac{\ldots}{\Gamma \vdash \Delta, R} \qquad \dfrac{\text{(Proof containing some } m \text{ contractions on sub-bundles containing } R\text{)}}{R, \Pi \vdash \Lambda}}{\Gamma, \Pi \vdash \Delta, \Lambda} \ (cut : R)$$

We can duplicate the cut and the subproof of $\Gamma \vdash \Delta, R$ 2^m times in order to obtain the following proof $P'(t)$, $t < 2^m$, given inductively:

  - $P'(0) =$

$$\frac{\dfrac{\ldots}{\Gamma \vdash \Delta, R} \qquad \dfrac{\text{(Proof with no problematic contractions)}}{R^{2^m}, \Pi \vdash \Lambda}}{R^{2^m-1}, \Gamma, \Pi \vdash \Delta, \Lambda} \ (cut : R)$$

  - $P'(t) =$

$$\dfrac{\dfrac{\dfrac{\dfrac{\ldots}{\Gamma \vdash \Delta, R} \qquad \dfrac{\text{P'(t-1)}}{R^{2^m-t}, \Gamma, \Pi \vdash \Delta, \Lambda}}{R^{2^m-(t+1)}, \Gamma, \Gamma, \Pi \vdash \Delta, \Delta, \Lambda} \ (cut : R)}{\text{(contractions and exchanges)}}}{R^{2^m-(t+1)}, \Gamma, \Pi \vdash \Delta, \Lambda}$$

- We have duplicated the subproof ending with $\Gamma \vdash \Delta, R$ many times. However, we can apply the induction hypothesis in order to obtain a proof of $\Gamma \vdash \Delta, R$ with no (strong or weak) critical cuts and with *Prop*. This is because this subproof contains only m critical cuts. *Prop* is clearly maintained.

- We have added many new cuts but none of them is critical (strong or weak), because the transformation in lemma 4.3.16 does not add contractions on sub-bundles ending with an R. If the original cut was not a (weak or strong) critical cut, then none of the new cuts will be (weak or strong), therefore $Prop$ is maintained so far as well.

- The original cut is no longer critical (strong or weak). Therefore, we can use the induction hypothesis in order to obtain a proof P' with no critical cuts at all, as we have one less critical cut. If the original cut was also not (weak or strong) critical, it will also not be a (weak or strong) critical after the transformation and the property $Prop$ is preserved in P'.

- P' is suitable as we have added inductions only on threads which did not contain already inductions of the added induction label, as required by lemma 4.1.18.

Corollary 4.3.19. Let P be a suitable proof and let l be the inductions' biggest label. We can obtain a suitable proof P' of the same end-sequent as P which is without any critical cut, weak or strong, labeled by l .

Proof

- Applying lemma 4.3.18 on strong critical cuts labeled by l and then applying it again on weak critical cuts labeled by l. The second application is done where there are no longer strong critical cuts and at the end we have no critical cut of any type labeled by l.

4.4 Cut Elimination

4.4.1 The elimination of inductions in proofs of weakly quantified theorems

In this section we prove a cut elimination theorem for proofs of theorems without strong quantifiers. We will give first two auxiliary lemmas. The first deals with the elimination of one induction and the second deals with the elimination of all inductions of the biggest induction's label. We will require that the inductions will not have critical contractions or other inductions interfering with them. We can achieve these two requirements by processing induction's labels one at a time, eliminating all contractions in all inductions of this label (using corollary 4.3.19) and then eliminating all inductions of this label.

Remark - In our cut elimination proof we are interested only in eliminating inductions which prevent us from eliminating the cuts in Gentzen's procedure. I.e. the inductions which are above some cut in the proof. All inductions which interfere with an induction, which is above a cut, will be above a cut as well according to their definition.

Lemma 4.4.1. Let P be a labeled (inductions) suitable proof of an end-sequent S such that S is without strong quantifiers and without free variables. If P contains an induction that is not interfered by any other induction and does not have critical contractions of any kind, then we can obtain a suitable proof of S, with this induction eliminated and new inductions of smaller labels only.

- Remarks - the requirement, that the induction is not interfered by any other induction, is equal to the requirement that this induction is of the biggest label and that the proof is suitable.

Proof

- We know that as S does not contain strong quantifiers, any strong quantifier in the proof must be in an implicit bundle and be cut out.
- The proof is by induction on the number of critical quantifiers of the induction in the proof.
- If there are no critical quantifiers for the induction, its term must be a ground term. By using the procedure of Gentzen in his proof of the consistency of Peano Arithmetic (lemma 3.2.14), we replace the induction by a finite number of cuts.
- The proof obtained from the application of Gentzen transformation is a suitable proof. This is because the transformation never adds inductions to threads which already contain inductions of these labels and we can use lemma 4.1.18.
- According to the procedure, the derivation of the top sequent of the induction is duplicated. According to our order, only smaller inductions are added. Because we deal with the biggest label and two inductions on the same thread must be of different labels.

- By assuming that the lemma holds for all inductions with n critical quantifiers, we consider inductions with $n+1$ critical quantifiers.
- We take one quantifier Q, which appears in the topmost cut below the induction and which is the outermost such quantifier. We consider the relevant PSB of the induction that contains this quantifier. We know that this PSB does not contain inductions or contractions. As the quantifier is the outermost critical quantifier, this PSB begins either with a weakening that introduces the formula that contains (or is) Q or with a quantifier inference rule that introduces Q:
  - The PSB begins with a weakening that introduce Q so we have:

$$\dfrac{\dfrac{\dfrac{\vdots}{\vdots}\,(\text{ind})}{\Gamma \vdash \Delta, R[P[Q]_\eta]_\lambda} \qquad \dfrac{\dfrac{\dfrac{\dfrac{\cdots}{\Pi \vdash \Lambda}}{P[Q]_\eta, \Pi \vdash \Lambda}\,(\text{Weakening})}{\cdots}}{R[P[Q]_\eta]_\lambda, \Pi' \vdash \Lambda'}}{\Gamma, \Pi' \vdash \Delta, \Lambda'}\,(cut : R[P[Q]_\eta]_\lambda)$$

    as the derivation on the right does not contain contractions or inductions in the PSB, according to lemma 4.1.22 we can replace it by the following derivation:

$$\dfrac{\dfrac{\dfrac{\dfrac{\cdots}{\Pi \vdash \Lambda}}{\text{(according to lemma 4.1.22)}}}{\Pi' \vdash \Lambda'}}{\Gamma, \Pi' \vdash \Delta, \Lambda'}\,(\text{weakenings and exchanges})$$

    We have eliminated this induction and no induction or cut was added or moved, so we have obtained a suitable proof of S with this induction eliminated.
  - The PSB begins with a quantifier rule and contains no induction or contraction. We apply the projection lemma (4.2.7) on the outermost critical quantifier. We can use this lemma in order to obtain a suitable proof with this quantifier eliminated. We can now use the induction hypothesis, in order to obtain a suitable proof of S with this induction eliminated.

The next lemma, which is based on the previous one, states that we can eliminate all inductions of the same label.

Lemma 4.4.2. Let P be a labeled (inductions) suitable proof of an end-sequent S. We assume that S is without strong quantifiers and without free variables. We assume further that all the inductions of the biggest label l do not contain critical contractions of any kind. Then we can obtain a suitable proof of S such that all its inductions are smaller than l.

Proof

- The proof is by induction on the number of inductions labeled by l.
- If there are no inductions of this label then we leave the proof as it is.
- By assuming the lemma is true for proofs where there are n inductions labeled by the biggest label, we prove it for $n+1$ inductions.
- We take one induction labeled by l. Because the proof is suitable we know that it is not being interfered by any other induction. As it has no critical contractions, we can use lemma 4.4.1 in order to obtain a suitable proof P'. P' does not contain any induction of a label greater than l and contains only n inductions of label l. Therefore we can apply the induction hypothesis and obtain a suitable proof, which contains only inductions of labels smaller than l.

4.4.2 The cut elimination theorem

In the last section we prove a cut elimination theorem for inductive proofs of theorems without strong quantifiers. This proof is by induction on the ordinal assigned to the proof. We will redefine the ordinal assignment of Gentzen, which was given in definition 3.2.8:

Definition 4.4.3 (Assignment of ordinals)**.** We will assign to proofs the same ordinals as were assigned in Gentzen's assignment (definition 3.2.8), but with one exception regarding the ordinal assigned to the lower sequent of an induction rule. Let h_P be the maximal height in P. We define a function t_P inductively as following: $t_P(1) = h_P + 2$; $t_P(n) = \Sigma_{i=1}^{n-1} t_P(i) + h_P + 2$. Let I be an induction labeled by l and let S_1 and S be the upper and lower sequents of I, respectively. Then, $o(S) = \omega_{t_P(l)}(o(S_1))$.

As will be shown, this assignment has the property that proofs will always be assigned a smaller ordinal than proofs containing inductions of bigger labels. The assignment is dependent on a specific proof because of the usage of the function t_P, but the next lemma shows that in the case of the transformations we have defined in this chapters, the function's range is never increased.

Lemma 4.4.4. Let P be a suitable proof and T be one of the transformations, which were presented in this chapter. Then , for every natural number n, we have $t_{T(P)}(n) \leq t_P(n)$.

Proof

- $T(P)$ is a suitable proof and all the new inductions in $T(P)$ can only be duplications of inductions in P. Therefore, the labeling is preserved. It should be noted that transformations may eliminate a label.
- None of the transformations given in this chapter increases the height of the proof.
- Therefore, given that the labeling is identical, then we can prove by induction on n that $t_{T(P)}(n) = \Sigma_{i=1}^{n-1} t_{T(P)}(i) + h_{T(P)} + 2 \leq \Sigma_{i=1}^{n-1} t_P(i) + h_P + 2 = t_P(n)$.

The following definition and lemmas will explain the choice of the function t_P.

Definition 4.4.5 (Degrees of proofs)**.** We first assign degrees to sequents in the proof. The degree of S, $d(S)$:

1. If S is an initial sequent then $d(S) = 1$.
2. If S is the lower sequent of a binary rule other than a cut and the upper sequents are S_1 and S_2, then $d(S) = max(d(S_1), d(S_2))$.
3. If S is the lower sequent of an unary rule other than an induction and the upper sequent is S_1, then $d(S) = d(S_1)$.
4. If S is the lower sequent of a cut, the upper sequents are S_1 and S_2 and the heights of the upper and lower sequents are k and l, respectively, then $d(S) = \omega_{k-l}(max(d(S_1), d(S_2)))$.
5. If S is the lower sequent of an induction of label l and the upper sequent is S_1, then $d(S) = \omega_{t_P(l)}(d(S_1))$.

The degree of P is the degree of its end-sequent.

Using the notion of a degree, we want to show that a proof containing an induction of the biggest label l, must be of a certain degree, which is much bigger than the degree of proofs not containing inductions of this label. The next three lemmas will define this relation.

Lemma 4.4.6. Let P be a suitable proof, then $o(P) < \omega^{d(P)}$

Proof

- The proof is by induction on the length of P. Let S be the end-sequent of P.
- If S is an initial sequent, then $o(P) = 1 < \omega^1 = \omega^{d(P)}$.
- We assume the lemma is true for all suitable proofs of length $< n$ and we prove it for suitable proofs of length n.
- The proof is by cases according to the last inference J in P.
- For rules other than cuts and inductions, we will show only for $\vee : l$. Let S_1, S_2 and S be the upper and lower sequents of J and we assume that $o(S_1) \leq o(S_2) < \omega^{d(S_2)}$. Then, $o(S) = o(S_1)\#o(S_2) <_{3.2.23(b)} \omega^{d(S_2)} \leq \omega^{max(d(S_1),d(S_2))} = \omega^{d(S)}$.
- J is a cut and let S_1 and S_2 be the upper sequents and k and l be the heights of the upper and lower sequents, respectively. We assume that $o(S_1) \leq o(S_2) < \omega^{d(S_2)}$. Then $o(S) = \omega_{k-l}(o(S_1)\#o(S_2)) <_{3.2.23(b)} \omega_{k-l}(\omega^{d(S_2)}) \leq \omega_{k-l}(max(\omega^{d(S_1)}, \omega^{d(S_2)}) = \omega^{d(S)}$.
- J is an induction of label l and S_1 is the upper sequent. We assume that $o(S_1) < \omega^{d(S_1)}$. Then $o(S) = \omega_{t_P(l)}(o(S_1)) < \omega_{t_P(l)}(\omega^{d(S_1)}) = \omega^{d(S)}$.

Lemma 4.4.7. Let P be a suitable proof, h_P be the maximal height of sequents in P and L be the set containing all labels of inductions used in P. Then $d(P) \leq \omega_{h_P+\Sigma_{l\in L}(t_P(l))}(1)$.

Proof

- Let τ be a maximal thread in P, which is constructed as follows: For a binary rule, take the upper sequent which is of the maximum degree.
- Let S be a sequent in τ. m_S is defined as the difference between the height of S in P and h_P. L_S is defined as the set of labels of inductions above S in τ.
- Using an induction over the number of sequents above a sequent S in the thread, we prove that $d(S) \leq \omega_{m_S+\Sigma_{l\in L_S}(t_P(l))}(1)$.
- If S is an axiom, then $d(S) = 1 \leq \omega_0(1)$.

- None of the rules, except inductions and cuts, affect the degree of the proof, the height or the number of inductions above the rule. Therefore, the inequality does not change.

- The last inference is a cut and let k and l be the heights of its upper and lower sequents. The upper sequent S_1 of the cut is in τ only if $d(S) = \omega_{k-l}(d(S_1))$, because $d(S) = \omega_{k-l}(max(d(S_1), d(S_2)))$. Therefore, we can use the induction hypothesis and obtain $d(S) = \omega_{k-l}(d(S_1)) \leq \omega_{(k-l)+m_{S_1}+\Sigma_{l\in L_{S_1}}(t_P(l))}(1) = \omega_{m_S+\Sigma_{l\in L_S}(t_P(l))}(1)$. Because $\Sigma_{l\in L_{S_1}}(t_P(l)) = \Sigma_{l\in L_S}(t_P(l))$ and $m_S = m_{S_1} + (k-l)$.

- The last inference is an induction of label l and S_1 is the upper sequent. As P is suitable, we know that there is no other induction of label l above S in τ. Therefore, using the induction hypothesis, we get that $d(S) = \omega_{t_P(l)}(d(S_1)) \leq \omega_{t_P(l)+\Sigma_{i\in L_{S_1}}(t_P(i))+m_{S_1}}(1) \leq \omega_{\Sigma_{i\in L_S}(t_P(i))+m_S}(1)$. Because $m_{S_1} \leq m_S$ and $\Sigma_{i\in L_S}(t_P(i)) = \Sigma_{i\in L_{S_1}}(t_P(i)) + t_P(l)$.

- Therefore, $d(P) \leq \omega_{h_P+\Sigma_{l\in\tau}(t_P(l))}(1) \leq \omega_{h_P+\Sigma_{l\in P}(t_P(l))}(1)$

Lemma 4.4.8. Let P be a suitable proof containing an induction of the biggest label l, then $o(P) \geq \omega_{t_P(l)}(1)$.

Proof

- The proof is by induction on the number of sequents below an induction of label l in P.

- If the induction is the last inference and the upper sequent of the induction is S_1, then $o(P) = \omega_{t_P(l)}(o(S_1)) \geq \omega_{t_P(l)}(1)$.

- Assuming it holds for n sequents, it will also hold for $n+1$ as all the inferences are monotonic in both arguments.

Corollary 4.4.9. Let P be a suitable proof and P' be a proof with inductions labeled by the same labels except of the biggest label l. Then, $\omega^{o(P')} < o(P)$.

Proof

- Let L be the set containing all labels in P. From lemma 4.4.6 we know that $\omega^{o(P')} < \omega_2(d(P'))$. According to lemma 4.4.7, we get that $\omega_2(d(P')) \leq \omega_{\Sigma_{i\in L\setminus\{l\}} t_{P'}(i)+h_{P'}+2}(1)$. Lemma 4.4.4 tells us that $\omega_{\Sigma_{i\in L\setminus\{l\}} t_{P'}(i)+h_{P'}+2}(1) \leq \omega_{\Sigma_{i\in L\setminus\{l\}} t_P(i)+h_P+2}(1)$. According to the definition of the function t_P, we have $\omega_{\Sigma_{i\in L\setminus\{l\}} t_P(i)+h_P+2}(1) = \omega_{t_P(l)}(1)$ and the result is obtained by using lemma 4.4.8 as $\omega_{t_P(l)}(1) \leq o(P)$.

We have shown that the elimination of all inductions of the biggest label significantly decreases the ordinal of the proof. We will now claim that the transformation of the elimination of contractions, although increasing the ordinal of the proof, does so by a lesser degree. We will first prove the following simple result and then prove the claim.

Lemma 4.4.10. Let P be a suitable proof. Then, $o(P) \geq d(P)$.

Proof

- Proof by induction on the derivation P of S.
- If S is an initial sequent, then $o(S) = d(S) = 1$.
- S is the lower sequent of an unary rule, other than an induction and S_1 is the upper sequent. We assume that $o(S_1) \geq d(S_1)$. Then $o(S) = o(S_1) + b \geq d(S_1) = d(S)$ ($b \in \{0, 1\}$).
- S is the lower sequent of a binary rule. We can use the same argument as the natural sum of ordinals is always bigger than their maximum.
- The same also holds for inductions and cuts as it involves the same argument as above, but as the argument of the function ω_x (ω_{k-l} and $\omega_{t_P(l)}$).

Lemma 4.4.11. Let P be a suitable proof. The resulted proof P', obtained by eliminating all critical contractions as shown in corollary 4.3.19, satisfies $o(P') < \omega^{o(P)}$.

Proof sketch

- Our aim is to show that $d(P) = d(P')$. Lemma 4.4.6 tells us that $o(P') < \omega^{d(P')} = \omega^{d(P)}$. Using lemma 4.4.10 we get that $o(P') < \omega^{d(P)} \leq \omega^{o(P)}$.
- It can be shown, for each transformation involved in the proof of corollary 4.3.19, that the degree of the proof does not change. This is true because the transformations involve only:

1. The addition of an arbitrary number of different rules, other than cuts and inductions.
2. The duplication of existing inductions, but never on the same thread. I.e. new threads are created with the duplicated inductions.
3. The duplication of existing cuts.

- Therefore, the degrees are equal and as was claimed above, $o(P') < \omega^{o(P)}$.

Theorem 4.4.12 (Cut Elimination in Inductive Proofs)**.** Let S be a provable sequent in PA without strong quantifiers or free variables. Then S is provable without an essential cut.

Proof

- The first step is to label all the original inductions by labels according to their order. We notice that as there is only one induction in each label-group and the inductions are being labeled according to $<_I$, there can be no two inductions of the same label on the same thread and the proof is suitable. Inductions which are below all cuts are excluded from the labeling, because there is no need to eliminate them in order to use Gentzen's cut elimination.
- We assign an ordinal to the proof using the new ordinal assignment (definition 4.4.5).
- The proof is by a transfinite induction on $o(P)$.
- The first step is to eliminate all free variables which are not being used as eigenvariables as is done in Gentzen proof of the consistency of PA (using lemma 3.1.7).
- If $o(P) < \omega$, then there are no inductions or essential cuts.
- Otherwise, we assume that the lemma holds for all suitable proofs P', such that $o(P') < o(P)$ and we prove that for P.
- If there is no induction above any of the cuts in the proof, then we apply cut elimination (lemma 3.1.4) in order to get a proof P' of S without essential cuts.
- Otherwise we consider all inductions labeled by l.

- As the proof is suitable and l is the biggest label, we know that the PSBs of all inductions labeled by l do not contain inductions. Therefore we use corollary 4.3.19 in order to obtain a suitable proof P_1 without critical contractions. According to lemma 4.4.11, $o(P_1) < \omega^{o(P)}$.
- Now we can use lemma 4.4.2 and obtain a suitable proof P_2, which has only inductions of smaller labels. According to corrolary 4.4.9, $\omega^{o(P_2)} < o(P_1)$.
- Therefore, $o(P_2) < o(P)$ and by the induction hypothesis, P_2 can be transformed into a proof without essential cuts.

Chapter 5

Possible Extensions and Improvements

In this thesis we have established an algorithm for the elimination of cuts in a specific subclass of proofs in PA. Of course, it is not very interesting to consider only inductive proofs of weakly quantified theorems. However, it might be possible to study and describe general inductive proofs by using the algorithm for cut elimination of weakly quantified theorems. The following extensions can be divided into computerizable ones and theoretical ones. The computerizable deals with programming the transformation. The theoretical ones will attempt to use the method in order to study and describe general inductive proofs.

- Among the computerizable improvements we can list:

    1. Improve the efficiency of the contraction elimination procedure. Although the procedure has an exponential complexity, it never adds new information so the proof can be represented in a much more compact and efficient way.
    2. To program the algorithm.
    3. In the proof we have assumed that the end-sequent contains no strong quantifier. Therefore, all strong quantifiers must be cut out. A more refined requirement should be given instead, because it is possible that although there are strong quantifiers there will be no free variables in the induction terms.

- Among the theoretical improvements we can list:

1. It is possible to eliminate all cuts when using PA_∞. I.e. LK extended by the ω-rule and some additional axioms, in order to replace the induction and quantifier inference rules as is done in [2]. It is clear that any proof in PA can be transformed into a proof in PA_∞. But as the resulted proofs in PA_∞ may be infinite derivations, it may not be possible to transform them into a proof in PA. The infinite derivations themselves are sometimes of little use. It might be possible, by using cut elimination for weakly quantified theorems, to obtain cut free proofs for specific instances of strongly quantified theorems. Moreover, we can try to use them in order to study, describe or limit the infinite proofs of strongly quantified theorems.
2. The proof was given for Peano arithmetic only but it should be possible to extend it to Heyting arithmetic as well.

Bibliography

[1] Gaisi Takeuti, *Proof Theory (chapters 1 and 2)*. North Holland, 2nd Edition, 1987.

[2] W. W. Tait, *Normal Derivability in Classical Logic. The Syntax and Semantics of Infinitary Languages, pages 204-236*. Springer-Verlag, 1989.

[3] H. Schwichtenberg, *Cut Elimination. Handbook of Mathematical Logic, pages 871 - 895*. North Holland, 1982.

[4] A. S. Troelstra and H. Schwichtenberg, *Basic Proof Theory (chapter 10)*. Cambridge University Press, 2nd Edition, 2000.

[5] Jean-Yves Girard, *Proof Theory and Logical Complexity, Volume I (chapters 6 and 7)*. Bibliopolis, 1987.

[6] W. Pohlers, *A Short Course in Ordinal Analysis, Proof Theory, Editors: Aczel, Simmons and Wainer, pages 27 - 79)*. Cambridge University Press, 1992.

[7] H. Fürstenberg *On the Infinitude of Primes*. American Mathematical Monthly 62 353, 1955.

[8] Matthias Baaz, Stefan Hetzl, Alexander Leitsch, Clemens Richter, and Hendrik Spohr. *Ceres: An Analysis of Fürstenbergs Proof of the infinity of Primes, pages 160-175*. Theoretical Computer Science 403 (2-3), 2008.

[9] Gerhard Gentzen. *Die Widerspruchfreiheit der reinen Zahlentheorie, pages 493-565*. Mathematische Annalen 112, 1936.