FACULTY OF ARTS
Charles University

BACHELOR THESIS

Jiří Rýdl

Aspects of the Cut-Elimination Theorem

Department of Logic

Supervisor of the bachelor thesis: Doc. RNDr. Vítězslav Švejdar, CSc.
Study programme: Logic

Prague 2021

I declare that I carried out this bachelor thesis independently, and only with the cited sources, literature and other professional sources. It has not been used to obtain another or the same degree.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

Prague, July 29, 2021 Jiří Rýdl

I would like to thank Vítězslav Švejdar for gently communicating to me many points of both formal and contentual interest.

Title: Aspects of the Cut-Elimination Theorem

Author: Jiří Rýdl

Department: Department of Logic

Supervisor: Doc. RNDr. Vítězslav Švejdar, CSc.

Abstract: I give a proof of the cut-elimination theorem (Gentzen's *Hauptsatz*) for an intuitionistic multi-succedent calculus. The proof follows the strategy of eliminating topmost maximal-rank cuts that allows for a straightforward way to measure the upper bound of the increase of derivations during the procedure. The elimination of all cut inferences generates a superexponential increase. I follow the structure of the proof for classical logic given in Švejdar's [18], modifying only the critical cases related to two restricted rules. Motivated by the diversity found in the early literature on this topic, I survey selected aspects of various formulations of sequent calculi. These are reflected in the proof of the *Hauptsatz* and its preliminaries. In the end I give one corollary of cut elimination, the Midsequent theorem, which is one of the three applications to be found already in Gentzen's [10].

Keywords: cut rule, sequent calculus, lengths of proofs

Contents

Introduction

In 1935 Gerhard Gentzen published the article [9] where he presented two novel formal systems for classical and intuitionistic first-order logic, natural deduction and sequent calculus. The latter was supposed to provide a convenient formalization of a normalization theorem for derivations in natural deduction style. This central theorem or *Hauptsatz* is usually referred to as the cut-elimination theorem in the setting of a sequent calculus.

Unlike systems consisting of several axiom schemes and one or two inference rules sequent calculi for pure logic are formulated in the opposite way. There is usually one axiom scheme and several inference rules most of which say under which circumstances can a complex formula be introduced into discourse. Besides rules that correspond to the meaning of logical connectives in this sense, there are those which do not make formulas more complex, but rather add or delete their occurrences. The only rule that allows a formula to completely disappear from a derivation is the cut rule. It can be said that this rule allows detours via (possibly) complex formulas that do not occur as components in the final result, and the procedure of cut elimination which Gentzen so famously introduced shows that these detours are unnecessary. This is to say that the rule of cut is redundant and everything that can be proved with it can be proved without it as well.

Derivations containing no instances of the cut rule have a rather determinate structure. If we take the complexity of a formula to be for example the number of its logical symbols (propositional connectives and quantifiers), the formula to be proved controls the complexity and structure of formulas that are contained in its derivation which does not use cut.[1] This is a very important property which for instance directly entails the consistency of first-order logic. More generally, it allows for direct demonstrations of underivability results, such as the law of excluded middle in intuitionistic logic.

On the other hand, there may be a huge expansion in size of proofs.[2] The procedure of cut elimination comprises several transformations which replace certain subderivations by another containing either no cuts or cuts that are simpler than the original. The size of these substitute derivations can be measured in terms of the size of the original subderivations. The iteration of such simplifications yields derivations without cuts altogether. Since at each step we can provide an upper bound on the increase of the subderivation in question and the inductive iterations can be bounded as well, there is a straightforward way to measure the expansion of size during the procedure. This expansion turns out to be superexponential[3] for first-order logic and the height of the stack of exponents is proportionate to the complexity of the most complex cut inferences in the original derivation. It can be said that although the cut rule does not give the calculus in question more inferential power, it can greatly shorten some proofs.

[1] This is more formally treated in the beginning of the first chapter.

[2] I use "derivation" and "proof" interchangeably to refer to the formal objects of logical systems such as well-defined sequences or proof-trees of formulas. On the metalevel, for instance in the phrase "the proof of cut elimination", I do not use the term "derivation".

[3] This means that there is no fixed stack of exponents such that for an arbitrary given derivation the size of the modified derivation containing no cuts fits into this bound.

The central topic of this thesis is to expand on the proof of cut elimination for classical logic given in [18] and modify it for intuitionistic logic, providing the corresponding upper bound on the increase of size of derivations. The particular calculus chosen for this purpose restricts two inference rules but otherwise is kept very close to its classical counterpart. This allows for a very similar treatment of some parts of the proof. And although in intuitionistic logic there are more cases to be distinguished as the two modified rules are not symmetric, the method and basic structure of the proof remains the same as that for classical logic. This is the topic of chapter 2.

Besides exemplifying the procedure on one particular calculus I was also interested in the aspects of different formulations of sequent calculi regarding both the structural features and their influence on the subsequent proofs of elementary properties (such as cut elimination). The analysis of various classes of sequent calculi has been present in logical textbooks already since the early 50's due to great mathematicians like Stephen Cole Kleene ([12]) or Haskell Curry ([6]). In the first chapter I discuss three selected topics that mark some turning points in the development of these systems. Invertible formulations of some rules are a stepping stone for both root-first proof search in propositional fragments and, which is important for us, absorbing the properties of one problematic rule into the others so that it needs not to be assumed. This is the rule of contraction, and systems that have it built into the other rules are called contraction-free calculi. Derivations in these systems behave differently than derivations in those with explicit contraction, such as Gentzen's original formalisms LK and LJ of [9]. Correspondingly, structural properties such as the eliminability of cut may have dissimilar proofs. The last topic of chapter 1 focuses only on intuitionistic systems and it has to do with the problem of staying as close to classical logic as possible. That is to retain, whenever possible, the symmetrical nature of derivations in classical systems.

In the beginning of the last chapter I briefly comment on a different method of proving cut elimination and three notions that play a crucial role in the whole procedure. In the second section of his seminal paper from 1935 Gentzen gave three different applications of his *Hauptsatz*, the decidability of intuitionistic propositional logic, the Midsequent theorem and its consequence in demonstrating the consistency of a particular induction-free arithmetical theory. The proof of the Midsequent theorem has a computational character, the central method being the reordering of rule applications on a given derivation, and it can be regarded as an extension of the procedure of cut elimination. I discuss this in the second part of chapter 3.

1. A survey of sequent calculi

In this chapter I discuss sequent calculi for classical and intuitionistic logic in general. Besides introducing some basic concepts, two sections deal with invertibility and admissibility of inference rules, respectively. These two concepts are related to the particular method of proving cut elimination that is given later. They are also illustrated on a few selected systems that are of historical importance.

First-order formulas are built up from atomic formulas $P(s)$, $Q(y)$, $R(x,t)$ ..., propositional connectives $\neg$, $\wedge$, $\vee$, $\rightarrow$ and quantifiers $\forall$, $\exists$. In some cases the constant $\bot$ is used instead of negation. If the arity of the respective predicate symbols is irrelevant, atomic formulas are sometimes denoted by p, q, r ... Arbitrary formulas are denoted by lowercase Greek letters φ, ψ, χ ..., finite collections of formulas by uppercase Greek letters Γ, Δ, Π ... The only complexity measure of formulas used below is the *depth* $\mathrm{d}(\varphi)$ of a formula φ. This is defined inductively for all formulas by

- $\mathrm{d}(\bot) = \mathrm{d}(p) = 0$ for atomic p,
- $\mathrm{d}(\varphi \wedge \psi) = \mathrm{d}(\varphi \vee \psi) = \mathrm{d}(\varphi \rightarrow \psi) = 1 + \max\{\mathrm{d}(\varphi), \mathrm{d}(\psi)\}$,
- $\mathrm{d}(\neg\psi) = \mathrm{d}(\forall x\psi) = \mathrm{d}(\exists x\psi) = 1 + \mathrm{d}(\psi)$.

Sequent calculi derive objects which are slightly more complex than formulas. A *sequent* $\langle\, \Gamma \Rightarrow \Delta \,\rangle$ is a pair of finite sequences of formulas, Γ is the *antecedent* of a sequent and Δ is its *succedent*. Sometimes I write $S : \langle\, \Gamma \Rightarrow \Delta \,\rangle$, which means that S is the sequent $\langle\, \Gamma \Rightarrow \Delta \,\rangle$. In contrast to Hilbert-style formalisms, sequent calculi usually have few axioms and they mostly consist of a family of inference rules. These are of two sorts, *structural* rules modify sequents by adding or eliminating occurrences of formulas, and *logical* rules construct new formulas from old ones by means of logical symbols. Both structural and logical rules (except for the cut rule) have two parts, the left one for altering antecedents and the right one for altering succedents. Four structural rules, *exchange*, *weakening* (W), *contraction* (C) and *cut*, which are used throughout, are the following:

$$\frac{\langle\, \Pi, \varphi, \psi, \Gamma \Rightarrow \Delta \,\rangle}{\langle\, \Pi, \psi, \varphi, \Gamma \Rightarrow \Delta \,\rangle} \qquad \frac{\langle\, \Gamma \Rightarrow \Delta, \varphi, \psi, \Lambda \,\rangle}{\langle\, \Gamma \Rightarrow \Delta, \psi, \varphi, \Lambda \,\rangle}$$

$$\frac{\langle\, \Gamma \Rightarrow \Delta \,\rangle}{\langle\, \varphi, \Gamma \Rightarrow \Delta \,\rangle} \qquad \frac{\langle\, \Gamma \Rightarrow \Delta \,\rangle}{\langle\, \Gamma \Rightarrow \Delta, \varphi \,\rangle}$$

$$\frac{\langle\, \varphi, \varphi, \Gamma \Rightarrow \Delta \,\rangle}{\langle\, \varphi, \Gamma \Rightarrow \Delta \,\rangle} \qquad \frac{\langle\, \Gamma \Rightarrow \Delta, \varphi, \varphi \,\rangle}{\langle\, \Gamma \Rightarrow \Delta, \varphi \,\rangle}$$

$$\frac{\langle\, \Gamma \Rightarrow \Delta, \varphi \,\rangle \qquad \langle\, \varphi, \Pi \Rightarrow \Lambda \,\rangle}{\langle\, \Pi, \Gamma \Rightarrow \Delta, \Lambda \,\rangle}$$

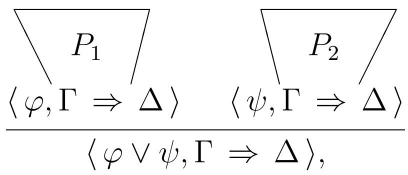
Logical rules introduce complex formulas into antecedents and succedents. They are denoted by L$\circ$ or R$\circ$, depending on which logical symbol $\circ$ is the outermost in the formula introduced into an antecedent (L) or a succedent (R) of a sequent. The following example is the introduction of disjunction to the antecedent, written as L$\vee$.

$$\frac{\langle\, \varphi, \Gamma \Rightarrow \Delta \,\rangle \qquad \langle\, \psi, \Gamma \Rightarrow \Delta \,\rangle}{\langle\, \varphi \vee \psi, \Gamma \Rightarrow \Delta \,\rangle}$$

Besides inference rules with one or two premises there are also sequents which can be asserted independently of any other sequents. These are called *initial sequents* and usually have the structure $\langle\, \varphi, \Gamma \Rightarrow \Delta, \varphi \,\rangle$ for arbitrary formula φ, or $\langle\, \bot, \Gamma \Rightarrow \Delta \,\rangle$.

The formula $\varphi \vee \psi$ in the conclusion of L$\vee$ is the *principal* formula of this rule, formulas φ and ψ in its left and right premises are the *active* formulas, and Γ, Δ are *side* formulas (or *contexts*). Every logical rule has one principal formula and one or two active formulas, depending on whether it has one or two premises. Among the structural rules, weakening has no active formula and cut has no principal formula. Its active formula φ is the *cut formula.*

Examples of particular calculi are given in the following sections, for now assume that a calculus C consists of a certain set of inference rules together with the initial sequents. A *derivation* is a finite tree whose nodes and transitions are labeled by sequents and inference rules, respectively. Its leaves are initial sequents and the root is the *endsequent.* A proof (derivation) of a sequent S is a derivation having S as its endsequent. Derivations are denoted by P, P', P_1, P_2 ... The *depth* $\mathrm{d}(P)$ of a derivation P is the maximum number of successive applications of rules in P. This is to say that $\mathrm{d}(P)$ is the length of a maximal branch in P. So initial sequents (viewed as single-noded trees) have depth 0 and if we have for instance the rule L$\vee$ applied on the endsequents of two derivations P_1 and P_2, which can be depicted as



the depth of the resulting derivation P is equal to $1 + \max\{\mathrm{d}(P_1), \mathrm{d}(P_2)\}$.

Although strictly speaking we are dealing exclusively with sequents, they are in a natural correspondence with formulas, and therefore the distinction between derivable sequents and derivable formulas can sometimes be omitted. On one hand, the formula φ being derivable in a calculus C amounts to the sequent $\langle\, \Rightarrow \varphi \,\rangle$ being C-derivable. On the other hand if we have a sequent $\langle\, \Gamma \Rightarrow \Delta \,\rangle$, its standard interpretation is the formula $\bigwedge \Gamma \rightarrow \bigvee \Delta$.

A derivation is called *cut-free* if it contains no instances of the cut rule. The procedure of cut elimination accepts an arbitrary derivation and produces a cut-free derivation of the same endsequent. This is the central topic of the next chapter, for now let me mention one immediate consequence of cut elimination.

The relation of being a *subformula* is defined for arbitrary formulas φ and ψ as follows. Every formula is a subformula of itself, φ is a subformula of $\neg\varphi$, φ and ψ are subformulas of $\varphi \circ \psi$ for any binary connective $\circ$, and $\varphi_x(t)$ is a subformula of $\forall x \varphi$ and $\exists x \varphi$ for any t free for x in φ.

Besides cut all the inference rules that we will encounter, both structural and logical, preserve subformulas of all formulas from their premises onto their conclusions. This means for instance that there are no occurrences of a formula in any derivation that would be more complex than the formulas present in the respective endsequent, which is precisely what the cut rule allows. Although formally this property depends on the set of rules in question, as its demonstration requires checking the rules one by one, let me state it now without proof and without reference to a particular calculus. The following proposition holds for every derivation in the calculi presented in the following sections and chapters.

Proposition 1.0.1 (Subformula property)**.** *Every formula in a cut-free derivation P is a subformula of a formula in the endsequent. If there are no instances of rules for negation and implication in P, every formula in a succedent of a sequent is a subformula of a formula in the succedent of the endsequent, and similarly for antecedents.*

A one-premise (unary) rule R is said to be *admissible* in a calculus C (alternatively, C is *closed* under R), if when the premise of R is C-derivable, so is its conclusion. A two-premise (binary) rule R is admissible in C, if whenever both premises of R are derivable there is also a derivation of its conclusion. A somewhat stronger form of the admissibility of cut for one intuitionistic calculus is established in the next chapter. If cut is admissible for C, every C-derivable sequent has a cut-free proof, and such a proof enjoys the subformula property. One of the immediate consequences of cut admissibility is the consistency of first-order logic, which is usually stated as the underivability of the empty sequent $\langle \Rightarrow \rangle$. Were this sequent derivable, take its cut-free proof P. All formulas in P must be subformulas of formulas in $\langle \Rightarrow \rangle$. Initial sequents may never be empty, hence P is not a derivation, hence $\langle \Rightarrow \rangle$ is not derivable.

1.1 Inversions

In this section and in the following one I discuss four calculi. The first three of them come, respectively, from Gentzen's paper [9] and Kleene's book [12]. As the order of formulas in cedents is usually considered irrelevant I drop the exchange rule in the formulations.

Gentzen's calculus LK for classical logic allows initial sequents of the form $\langle \varphi \Rightarrow \varphi \rangle$ for arbitrary formula φ and consists of the three structural rules of weakening, contraction and cut together with the following logical rules:

$$\frac{\langle \varphi_i, \Gamma \Rightarrow \Delta \rangle}{\langle \varphi_0 \wedge \varphi_1, \Gamma \Rightarrow \Delta \rangle} \qquad \frac{\langle \Gamma \Rightarrow \Delta, \varphi \rangle \quad \langle \Gamma \Rightarrow \Delta, \psi \rangle}{\langle \Gamma \Rightarrow \Delta, \varphi \wedge \psi \rangle}$$

$$\frac{\langle \varphi, \Gamma \Rightarrow \Delta \rangle \quad \langle \psi, \Gamma \Rightarrow \Delta \rangle}{\langle \varphi \vee \psi, \Gamma \Rightarrow \Delta \rangle} \qquad \frac{\langle \Gamma \Rightarrow \Delta, \varphi_i \rangle}{\langle \Gamma \Rightarrow \Delta, \varphi_0 \vee \varphi_1 \rangle}$$

$$\frac{\langle \Gamma \Rightarrow \Delta, \varphi \rangle \quad \langle \psi, \Pi \Rightarrow \Lambda \rangle}{\langle \varphi \to \psi, \Gamma, \Pi \Rightarrow \Delta, \Lambda \rangle} \qquad \frac{\langle \varphi, \Gamma \Rightarrow \Delta, \psi \rangle}{\langle \Gamma \Rightarrow \Delta, \varphi \to \psi \rangle}$$

$$\frac{\langle\, \Gamma \;\Rightarrow\; \Delta, \varphi \,\rangle}{\langle\, \neg\varphi, \Gamma \;\Rightarrow\; \Delta \,\rangle} \qquad \frac{\langle\, \varphi, \Gamma \;\Rightarrow\; \Delta \,\rangle}{\langle\, \Gamma \;\Rightarrow\; \Delta, \neg\varphi \,\rangle}$$

$$\frac{\langle\, \varphi_x(y), \Gamma \;\Rightarrow\; \Delta \,\rangle}{\langle\, \exists x\varphi, \Gamma \;\Rightarrow\; \Delta \,\rangle} \qquad \frac{\langle\, \Gamma \;\Rightarrow\; \Delta, \varphi_x(y) \,\rangle}{\langle\, \Gamma \;\Rightarrow\; \Delta, \forall x\varphi \,\rangle}$$

$$\frac{\langle\, \varphi_x(t), \Gamma \;\Rightarrow\; \Delta \,\rangle}{\langle\, \forall x\varphi, \Gamma \;\Rightarrow\; \Delta \,\rangle} \qquad \frac{\langle\, \Gamma \;\Rightarrow\; \Delta, \varphi_x(t) \,\rangle}{\langle\, \Gamma \;\Rightarrow\; \Delta, \exists x\varphi \,\rangle}$$

The bottom pair of quantifier rules is called *specification* rules, the one above *generalization* rules. In the specification rules the term t is free for x in φ.[1] The variable y in the generalization rules is free for x in φ. It is called the *eigenvariable* of the respective rule instance. This pair of rules is subject to the *eigenvariable condition* which says that y may not occur free in the conclusion of these rules, that is in $\Gamma \cup \Delta \cup \{\forall x\varphi\}$ or $\Gamma \cup \Delta \cup \{\exists x\varphi\}$.

LK and its intuitionistic counterpart LJ are the first two sequent calculi to appear in print, both in the paper [9]. Basically all other formalisms working with sequents derive from and are inspired by these two calculi. The systems which we come across in the following are in each case certain modifications, usually only as regards the choice of a few particular rules. But, as we will see, certain structural properties of sequent calculi can vary with only slightly different formulations of some logical rules.

Let me give two examples of derivations in LK.

Example 1.1.1. The formula $\neg\forall x\neg\varphi \to \exists x\varphi$, which is not provable in intuitionistic logic, has the following derivation in LK:

$$\cfrac{\cfrac{\cfrac{\cfrac{\cfrac{\langle\, \varphi_x(y) \;\Rightarrow\; \varphi_x(y) \,\rangle}{\langle\, \varphi_x(y) \;\Rightarrow\; \exists x\varphi \,\rangle}}{\langle\, \;\Rightarrow\; \exists x\varphi, \neg\varphi_x(y) \,\rangle}}{\langle\, \;\Rightarrow\; \exists x\varphi, \forall x\neg\varphi \,\rangle}}{\langle\, \neg\forall x\neg\varphi \;\Rightarrow\; \exists x\varphi \,\rangle}}{\langle\, \;\Rightarrow\; \neg\forall x\neg\varphi \to \exists x\varphi \,\rangle}$$

It is only important to first apply the specification rule (R∃) and only then the generalization rule (R∀), since the latter would not be applicable if y had free occurrences in φ.

[1] Gentzen formulated the specification rules only for variables, the term t that is substituted for x in φ is a variable z.

Example 1.1.2. The formula $(\varphi \to (\varphi \to \psi)) \to (\varphi \to \psi)$ is derivable in LK with the help of weakening and contraction.

$$\dfrac{\langle \varphi \Rightarrow \varphi \rangle \quad \dfrac{\langle \varphi \to \psi \Rightarrow \varphi \to \psi \rangle}{\langle \varphi \to \psi \Rightarrow \varphi \to \psi, \psi \rangle}}{\dfrac{\langle \varphi \to (\varphi \to \psi), \varphi \Rightarrow \psi, \varphi \to \psi \rangle}{\dfrac{\langle \varphi \to (\varphi \to \psi) \Rightarrow \varphi \to \psi, \varphi \to \psi \rangle}{\dfrac{\langle \varphi \to (\varphi \to \psi) \Rightarrow \varphi \to \psi \rangle}{\langle \Rightarrow (\varphi \to (\varphi \to \psi)) \to (\varphi \to \psi) \rangle}}}}$$

Definition 1. *A unary rule R in a sequent calculus is* invertible*, if the provability of its conclusion implies the provability of its premise. A binary rule is* invertible *if both its premises are provable when its conclusion is.*

To rephrase the definition, an invertible rule R does not permit to infer a derivable sequent from underivable premises. Invertibility is a very useful device for the study of decidability of propositional logics (see e.g. [15]) because it is the case that the premises of propositional invertible rules are uniquely determined when their conclusions are written down. The notion of invertibility was introduced by Ketonen in his dissertation thesis from 1944, which in turn became well-known through Bernays' review [1] from 1945.

Notice that the rules L$\wedge$, R$\vee$ and L$\to$ in LK are not invertible. For the first two this is caused by the fact that we do not know which of the two formulas φ or ψ was used to derive the complex formula. The law of excluded middle, $\varphi \vee \neg\varphi$, is a theorem of classical logic. Its derivation in LK may take the following form:

$$\dfrac{\dfrac{\dfrac{\dfrac{\langle \varphi \Rightarrow \varphi \rangle}{\langle \Rightarrow \varphi, \neg\varphi \rangle}}{\langle \Rightarrow \varphi \vee \neg\varphi, \neg\varphi \rangle}}{\langle \Rightarrow \varphi \vee \neg\varphi, \varphi \vee \neg\varphi \rangle}}{\langle \Rightarrow \varphi \vee \neg\varphi \rangle}$$

The rule R$\vee$ was used twice, firstly on φ and secondly on $\neg\varphi$. If we started from the bottom and did not consider that contraction may have been used in the last step, we would have to try to decompose the complex formula using the right rule for disjunction. But neither $\langle \Rightarrow \varphi \rangle$, nor $\langle \Rightarrow \neg\varphi \rangle$ is a theorem of LK. This shows that, as stated, R$\vee$ is not invertible and analogously neither is L$\wedge$. In case of L$\to$ the non-invertibility is caused by the rule having *independent* contexts in the premises, the sequences $\Gamma \cup \Delta$ and $\Pi \cup \Lambda$ may consist of different formulas.

Ketonen considered the following invertible variants of the three rules:

$$\dfrac{\langle \varphi, \psi, \Gamma \Rightarrow \Delta \rangle}{\langle \varphi \wedge \psi, \Gamma \Rightarrow \Delta \rangle} \qquad \dfrac{\langle \Gamma \Rightarrow \Delta, \varphi, \psi \rangle}{\langle \Gamma \Rightarrow \Delta, \varphi \vee \psi \rangle} \qquad \dfrac{\langle \Gamma \Rightarrow \Delta, \varphi \rangle \quad \langle \psi, \Gamma \Rightarrow \Delta \rangle}{\langle \varphi \to \psi, \Gamma \Rightarrow \Delta \rangle}$$

In the presence of weakening and contraction the Ketonen and Gentzen versions of all three rules are inter-derivable, their equivalence is shown in [6]. It is obvious that the premises of the Ketonen rules are determined by their conclusions.

When contrasted with the Gentzen variants, in case of $\wedge$ and $\vee$ this is because the principal formula determines the active formulas, and with $\rightarrow$ because the contexts in the premises are the same as the contexts in the conclusion. In this case we say that L$\rightarrow$ is a *context-sharing* rule.[2]

A proof of invertibility for a few rules is given later, for now I briefly comment on three aspects about invertible rules that will be relevant later. Firstly, it is important that demonstrations of inversions do not require the use of cut. This is because in the proof of cut elimination invertibility is used as a tool in certain transformation steps on a particular derivation, and if the use of inversions introduced new cuts, it might hinder the overall procedure.[3]

A unary rule is called *depth-preserving*[4] (dp-) invertible, if for any given derivation P of its conclusion we can find a derivation P' of its premise such that $\mathrm{d}(P') \leq \mathrm{d}(P)$, and similarly for a binary rule. Depth-preserving inversions are again utilized in the proof of cut elimination. The procedure allows to keep track of the increase of the original derivation in the process of replacing some of its subderivations by new ones, and one of the reasons for this is that some of the rules are dp-invertible.[5]

Lastly, when I speak about invertibility of rules for conjunction, disjunction and implication in what follows, I always mean the formulation using the premises of the respective Ketonen variants. Take for example the original Gentzen L$\rightarrow$. This rule is, as we know, not invertible, but nevertheless it may (and will) be useful to formulate an inversion lemma for its right premise. And since it will be sufficient to know that the derivability of $\langle \varphi \rightarrow \psi, \Pi, \Gamma \Rightarrow \Delta, \Lambda \rangle$ entails the derivability of $\langle \psi, \Pi, \Gamma \Rightarrow \Delta, \Lambda \rangle$, the formulation corresponds to the invertibility of Ketonen's L$\rightarrow$.

1.2 Admissible rules and intuitionistic systems

The admissibility of R with the premise S_1 and the conclusion S in a calculus C can be formally written as

$$C \vdash S_1 \Rightarrow C \vdash S,$$

and similarly if R has two premises. Note that we do not demand that there is a sequence of inferences leading from the premises of R to its conclusion. If this stronger condition obtains we say that R is *derivable* in C. We can appreciate the distinction between these two relations on an example with the cut rule. It is the central topic of the next chapter to show the admissibility of cut for a particular intuitionistic calculus. The proof shows how to transform any derivation into a cut-free one by successively replacing certain subderivations by derivations that only contain simpler cuts (or no cuts at all).[6] It is usually not the case that we could apply a series of inferences to the premises of a cut and obtain its conclusion. In LK this could be done only if the cut formula already appeared

[2] A nice exposition of the reasons that led Gentzen to choose the particular rules for LK is given in [22].

[3] Ketonen used cut to demonstrate invertibility, the first author to prove inversions without reference to the cut rule is Schütte in [17].

[4] The terminology is taken from [20], in [15] the authors use *height-preserving* instead.

[5] Inversion lemmas with explicit reference to the preservation of depth are first given in [6].

[6] A complexity measure for instances of cut is introduced in chapter 2.

in the conclusion (using multiple weakenings and one or two contractions on the premises), as otherwise besides cut no LK-rule permits the elimination of formulas.[7]

A unary rule is *depth-preserving* (dp-) admissible if for a given derivation P of its premise there is a derivation P' of its conclusion such that $\mathrm{d}(P') \leq \mathrm{d}(P)$, and similarly if R is a binary rule.

Example 1.2.1. If the initial sequents of LK are modified so as to allow arbitrary contexts, i.e. if all sequents of the form $\langle \varphi, \Gamma \Rightarrow \Delta, \varphi \rangle$ are considered initial, then weakening becomes dp-admissible in the calculus LK − W. The proof is given in the next chapter for a different calculus, but its structure is the same as in case of LK. We only need to make certain assumptions about free variables in the principal formula φ of weakening explicit, afterwards the demonstration follows by induction on the depth of the derivation of $\langle \Gamma \Rightarrow \Delta \rangle$.

1.2.1 Contraction

Invertible rules have another useful application that concerns itself with the second structural rule, contraction. In the proof of cut elimination for calculi having contraction as an explicit rule (such as LK) there is one case which is rather difficult to handle. I do not go into details, as I have not yet described the procedure of cut elimination, but only make a few remarks.

The central lemma of the proof shows how to eliminate instances of cut whose cut formula has the greatest depth among other cut formulas in the subderivation of this cut.[8] It proceeds by distinguishing cases according to the way this complex cut formula has been introduced into the premises of the cut. One of these is when the right premise S of the cut has been derived by means of contraction on the cut formula φ, i.e. the respective premise S_1 of S contains at least two occurrences of φ in the antecedent. There are two basic methods of eliminating the complex cut that apply to all other cases in Gentzen's proof for LK, but neither is sufficient to remove both occurrences of φ from S_1. In order to be able to deal with this case in a similar manner as the other cases, Gentzen introduced a generalized rule of cut, called mix or multicut, that allows to "cut out" any number of occurrences of a formula in one step. Mix is in the presence of contraction and weakening equivalent to cut and the problematic case is handled by eliminating *all* occurrences of φ from the succedent of the left premise of the cut and the antecedent of S_1 and, if necessary, using some weakenings to restore the original contexts. Due to the overall structure of proving the lemma this step is sufficient to remove the complex cut.

Von Plato showed in [21] that this step actually does not require introducing multicut. He proceeds by analysing the structure of the subderivation above S_1 and demonstrates that the complex cut can be eliminated by the standard methods with the help of appropriate inversion lemmas for L$\wedge$, R$\vee$ and L$\rightarrow$.[9]

However, a rather simpler approach to the proof is available if the calculus C is formulated in such way that contraction is not needed, that is if the logical rules

[7] Contraction only reduces the number of occurrences of a formula.

[8] See Lemma 2.2.1 for a more precise description.

[9] These inversions need not preserve depth because in LK weakening is an explicit rule. For details see [21].

are reformulated so as to make contraction *admissible* for C, and invertible rules play a crucial part here. These *contraction-free* systems are discussed already in the early 50's by Curry in [4] (part II, theorem 3) and by Kleene in [12]. Both authors use these calculi to study decidability of propositional fragments, since as contraction is admissible for them (and weakening can be made so easily) and the premises of all propositional rules are determined by their conclusions, it is rather easy to formulate a decision procedure.[10]

For our purpose the advantage of calculi with admissible structural rules lies in that they facilitate the proof of cut elimination to some extent. The first and most obvious point is that some cases need not be considered, secondly one does not have to introduce multicut or produce a deep analysis of one particular case to show that it allows for the standard approach (as von Plato did). But one of the most important points is that the structure of derivations in these systems is determinate in the sense that there are no chains of structural inferences which are not bounded by either of the two measures used in the proof, the depth of derivations and the depth of formulas. As generally these two are the only measures used to give an upper bound of the increase of derivations during cut elimination, the extraction of this bound (based on these particular measures) is not possible for calculi with explicit contraction. There are ways to provide bounds even for such systems, but instead of the depth of derivation one needs to use something else. For instance Buss in [2] uses the number of strong inferences, another possibility is to use what is called logical depth in [20] and in both cases contractions and weakenings are not counted.

Let me briefly discuss an early contraction-free calculus by one of the founding fathers of these systems, Stephen Cole Kleene. This is the intuitionistic system G3 from [12] (§ 80). Since then the class of intuitionistic contraction-free calculi has extended far beyond just one calculus and this class is usually denoted by G3i. I call the particular Kleene's calculus K-G3i, so that the name denotes only one system.

Gentzen gave his proof of cut elimination simultaneously for both classical and intuitionistic logic, using two calculi LK and LJ, respectively. LJ has the same rules as LK with the restriction that *all* succedents may contain at most one formula, in which case the class of provable formulas coincides with the intuitionistically logically valid formulas. Due to this restriction LJ and other systems that satisfy it are called *single-succedent* calculi. K-G3i is a modification of LJ. Kleene uses the Ketonen variants of the three propositional rules and the remaining modification applies only to the left logical rules. It can be stated by requiring that the principal formulas of all left rules must be present in the antecedents of all their premises.
As an example I give the left rule for implication:

$$\frac{\langle\, \varphi \to \psi, \Gamma \;\Rightarrow\; \varphi \,\rangle \qquad \langle\, \psi, \varphi \to \psi, \Gamma \;\Rightarrow\; \Delta \,\rangle}{\langle\, \varphi \to \psi, \Gamma \;\Rightarrow\; \Delta \,\rangle}$$

There is at most one formula in Δ and this is the same in both premises, so these are determined by the conclusion and we can say that it is an example of a context-sharing rule for left implication. K-G3i allows general initial sequents and is therefore closed under weakening.

[10] See [6].

Contraction is also dp-admissible for K-G3i. The requirements on a calculus C to have admissible contraction are stated in a general form in [6], p. 231. Each premise in all rules of C must satisfy one of the following conditions: either the principal formula appears in it as a side formula, or the rule is invertible with respect to this premise. Notice that although the left premise of L$\rightarrow$ is not invertible, it contains $\varphi \rightarrow \psi$ as a side formula, and hence satisfies the requirement. The right premise need not contain $\varphi \rightarrow \psi$ as a side formula because L$\rightarrow$ is dp-invertible with respect to this premise in either case.

A proof of dp-admissibility of contraction proceed by induction on the depth of a derivation. Firstly we distinguish two cases, either the contraction formula φ is principal in the last inference R or it is not. In the latter the induction hypotheses on the premise(s) of R gives derivation(s) with φ contracted, and the subsequent application of R procudes the original conclusion with φ contracted. If φ is principal in R the cases are distinguished relative to its outermost symbol. In each of these cases we either use the invertibility or the second option, namely the presence of φ in the respective premise. For a detailed proof see [15] or [20].

As I remark above, derivations in systems with admissible structural rules have a rather determinate structure. I mentioned three motivations for such systems that are of certain importance. The first is simply that there are good reasons to avoid explicit contraction, one of which was demonstrated on the problematic case in the proof of cut elimination for LK. Secondly, if a propositional system has only logical rules, the number of possibilities in a bottom-up proof search for a given sequent is drastically reduced. The third one is that by avoiding the use of structural rules we can calculate the growth of derivations during certain transformations, particularly cut elimination. Lastly I want to remark that the proof of cut elimination for contraction-free calculi is very similar to calculi whose sequents are pairs of *sets* of formulas, since as contraction becomes implicit in these systems, their derivations behave similarly.

Let me close this part citing a remark of Curry ([4], p. 37) on another feature of contraction-free calculi, one which may be taken to motivate them on the basis of more philosophically oriented considerations. It applies only to calculi such as K-G3i whose all rules contain the principal formula in each premise (whenever this does not go against the single-succedent restriction). Curry notes that the derivations in these calculi are strictly cumulative: "… [in the rules] all constituents present in the conclusion must be present in the premises; and hence, in any derivation, all constituents must be present in the prime statements [= initial sequents]. The essential function of the rules is then to eliminate components."

Example 1.2.2. The following derivation of $(\varphi \rightarrow (\varphi \rightarrow \psi)) \rightarrow (\varphi \rightarrow \psi)$ in K-G3i is an example of this cumulativeness, in single-succedent systems characteristic only for antecedents. Let me denote the premise $\varphi \rightarrow (\varphi \rightarrow \psi)$ by γ and the conclusion $\varphi \rightarrow \psi$ by δ.

$$\dfrac{\dfrac{\dfrac{\langle \gamma, \varphi \Rightarrow \varphi \rangle \qquad \dfrac{\langle \delta, \gamma, \varphi \Rightarrow \varphi \rangle \qquad \langle \psi, \delta, \gamma, \varphi \Rightarrow \psi \rangle}{\langle \delta, \gamma, \varphi \Rightarrow \psi \rangle}}{\langle \gamma, \varphi \Rightarrow \psi \rangle}}{\langle \gamma \Rightarrow \delta \rangle}}{\langle \Rightarrow \gamma \rightarrow \delta \rangle}$$

1.2.2 Multi-succedent calculi

It is natural to ask why restrict all succedents to contain at most one formula when the only rules that permit derivations of formulas which are not intuitionistic laws are the right rules for negation, implication and universal quantifier. Call these the *critical* rules. Gentzen's formulation of LJ was motivated by his proof of cut elimination for both logics. Since the rules of LJ are special cases of rules for LK, the corresponding derivations in LJ are always derivations in LK (but not vice versa). Consequently, the proof of cut elimination for LK yields a proof of cut elimination for LJ as a byproduct, provided the transformations of LJ-derivations only produce LJ-derivations, i.e. that the single-succedent restriction is preserved. This is verified by Gentzen in the last part of [9].

Already in 1937 (published two years later as [3]) Curry noted that the equivalence between LJ and a Hilbert-style axiomatization of intuitionistic logic could be proved without imposing the single-succedent restriction on the non-critical rules. A system for intuitionistic logic which only restricts the critical rules is called *multi-succedent* calculus. Such a system was probably first explicitly formulated in [14] as an auxiliary calculus (*Hilfskalkül*). A systematic study of multi-succedent calculi is first given by Curry in [6] where he develops several formulations of calculi for classical and intuitionistic logic and provides their metalogical analysis. A rather influential calculus called GHPC, whose variants are discussed in [20] and [15], was given by Dragalin in [7].

The main idea behind the study of these systems is that they preserve as much symmetry between the logical rules as possible, and as a result the structure of their derivations is kept quite close to that of classical logic. Accordingly, proofs of their metalogical properties very much resemble the corresponding proofs for classical systems. In the following chapter I give a proof of cut elimination for a particular multi-succedent calculus and this follows basically the same pattern as the proof of cut elimination for classical logic given in [18], deviating only in the critical cases.

An example of an intuitionistic multi-succedent system which is closed under weakening and contraction is the calculus GHPC. Instead of negation Dragalin takes $\bot$ as a primitive constant.[11] GHPC allows two kinds of initial sequents, $\langle\, p, \Gamma \Rightarrow \Delta, p \,\rangle$ where p is an atomic formula and $\langle\, \bot, \Gamma \Rightarrow \Delta \,\rangle$.

$$\frac{\langle\, \varphi, \psi, \Gamma \Rightarrow \Delta \,\rangle}{\langle\, \varphi \wedge \psi, \Gamma \Rightarrow \Delta \,\rangle} \qquad \frac{\langle\, \Gamma \Rightarrow \Delta, \varphi \,\rangle \quad \langle\, \Gamma \Rightarrow \Delta, \psi \,\rangle}{\langle\, \Gamma \Rightarrow \Delta, \varphi \wedge \psi \,\rangle}$$

$$\frac{\langle\, \varphi, \Gamma \Rightarrow \Delta \,\rangle \quad \langle\, \psi, \Gamma \Rightarrow \Delta \,\rangle}{\langle\, \varphi \vee \psi, \Gamma \Rightarrow \Delta \,\rangle} \qquad \frac{\langle\, \Gamma \Rightarrow \Delta, \varphi, \psi \,\rangle}{\langle\, \Gamma \Rightarrow \Delta, \varphi \vee \psi \,\rangle}$$

$$\frac{\langle\, \varphi \rightarrow \psi, \Gamma \Rightarrow \varphi \,\rangle \quad \langle\, \psi, \Gamma \Rightarrow \Delta \,\rangle}{\langle\, \varphi \rightarrow \psi, \Gamma \Rightarrow \Delta \,\rangle} \qquad \frac{\langle\, \varphi, \Gamma \Rightarrow \psi \,\rangle}{\langle\, \Gamma \Rightarrow \Delta, \varphi \rightarrow \psi \,\rangle}$$

$$\frac{\langle\, \psi_x(y), \Gamma \Rightarrow \Delta \,\rangle}{\langle\, \exists x\psi, \Gamma \Rightarrow \Delta \,\rangle} \qquad \frac{\langle\, \Gamma \Rightarrow \psi_x(y) \,\rangle}{\langle\, \Gamma \Rightarrow \Delta, \forall x\psi \,\rangle}$$

[11] The idea that absurdity is a primitive notion in intuitionistic logic is discussed e.g. in [11], chapter VII.

$$\frac{\langle \forall x\psi, \psi_x(t), \Gamma \Rightarrow \Delta \rangle}{\langle \forall x\psi, \Gamma \Rightarrow \Delta \rangle} \qquad \frac{\langle \Gamma \Rightarrow \Delta, \exists x\psi, \psi_x(t) \rangle}{\langle \Gamma \Rightarrow \Delta, \exists x\psi \rangle}$$

Generalization rules are subject to the eigenvariable condition. There are several differences between the rules of GHPC and LJ (or K-G3i). Firstly, since $\bot$ is primitive, there are additional initial sequents which stand for the rules of negation (below I show that these are derivable in GHPC). Secondly, notice that only the rules R$\rightarrow$ and R$\forall$ have restricted succedents. This is *the* characteristic feature of intuitionistic multi-succedent calculi, to deviate as little as possible from the corresponding classical systems. Preservation of dp-admissibility of weakening is guaranteed by the context Δ in the conclusion of the critical rules, as in the other rules there are arbitrary contexts in both premises and conclusions. Next, since L$\wedge$ and R$\vee$ are formulated in the Ketonen variants these rules need not contain the principal formula in the premises to absorb the properties of contraction. The above given observation by Curry says that their invertibility is sufficient for this purpose. In contrast to K-G3i this observation is made use of also in the case of the right premise of L$\rightarrow$, both premises of L$\vee$ and the left generalization rule.

Example 1.2.3. If we define $\neg\varphi$ as $\varphi \rightarrow \bot$ the following pair of rules for negation from K-G3i are derivable in GHPC:

$$\frac{\langle \neg\varphi, \Gamma \Rightarrow \varphi \rangle}{\langle \neg\varphi, \Gamma \Rightarrow \Delta \rangle} \qquad \frac{\langle \varphi, \Gamma \Rightarrow \ \rangle}{\langle \Gamma \Rightarrow \neg\varphi \rangle}$$

As K-G3i is a single-succedent system, Δ is either empty or consists of an arbitrary formula. In case of R$\neg$ first use dp-admissibility of weakening of GHPC to obtain a derivation of the sequent $\langle \varphi, \Gamma \Rightarrow \bot \rangle$, and then apply R$\rightarrow$ to get the conclusion. The conclusion of L$\neg$ is obtained from the derivation P_1 of its premise as follows:

$$\frac{\begin{array}{c} P_1 \\ \langle \varphi \rightarrow \bot, \Gamma \Rightarrow \varphi \rangle \end{array} \qquad \langle \bot, \Gamma \Rightarrow \Delta \rangle}{\langle \varphi \rightarrow \bot, \Gamma \Rightarrow \Delta \rangle}$$

Notice that we did not use the assumption that Δ contains at most one formula because L$\rightarrow$ is formulated for arbitrary contexts.

A calculus for classical logic can be obtained from GHPC by permitting arbitrary contexts in the premises of the two critical rules R$\rightarrow$ and R$\forall$. In case of L$\rightarrow$ the side formula $\varphi \rightarrow \psi$ in its left premise would become redundant, as the provability of $\langle \Gamma \Rightarrow \Delta, \varphi \rangle$ follows from the provability of $\langle \varphi \rightarrow \psi, \Gamma \Rightarrow \Delta \rangle$ in classical logic.

A detailed proof of cut elimination for GHPC is given in [8]. Compared to the corresponding proof for classical systems with admissible contraction two subcases in one of the four main cases need to be further divided according to the outermost symbol of the cut formula. These concern the critical rules R$\rightarrow$ and R$\forall$. Otherwise no modifications are needed and the structure of the proof remains the same. Cut eliminations for both classical and intuitionistic multi-succedent calculi for calculi based on sequences with admissible structural rules is proved in [20] (systems G3c and m-G3i) and [15] (systems G3c and G3im).

2. The cut-elimination theorem

This chapter is devoted to a detailed proof of cut elimination for one particular multi-succedent calculus for intuitionistic logic that I call smG3i. It begins with the formulation of smG3i together with a few observations regarding its structural properties. In the central second section I give the proof of the *Hauptsatz*, together with an upper bound on the increase of depths of derivations after the cuts have been eliminated.

2.1 The calculus smG3i

The intuitionistic calculus smG3i[1] has the following logical rules:

$$\frac{\langle \varphi_i, \Gamma \Rightarrow \Delta \rangle}{\langle \varphi_0 \wedge \varphi_1, \Gamma \Rightarrow \Delta \rangle} \qquad \frac{\langle \Gamma \Rightarrow \Delta, \varphi \rangle \quad \langle \Gamma \Rightarrow \Delta, \psi \rangle}{\langle \Gamma \Rightarrow \Delta, \varphi \wedge \psi \rangle}$$

$$\frac{\langle \varphi, \Gamma \Rightarrow \Delta \rangle \quad \langle \psi, \Gamma \Rightarrow \Delta \rangle}{\langle \varphi \vee \psi, \Gamma \Rightarrow \Delta \rangle} \qquad \frac{\langle \Gamma \Rightarrow \Delta, \varphi_i \rangle}{\langle \Gamma \Rightarrow \Delta, \varphi_0 \vee \varphi_1 \rangle}$$

$$\frac{\langle \Gamma \Rightarrow \Delta, \varphi \rangle \quad \langle \psi, \Pi \Rightarrow \Lambda \rangle}{\langle \varphi \to \psi, \Pi, \Gamma \Rightarrow \Delta, \Lambda \rangle} \qquad \frac{\langle \varphi, \Gamma \Rightarrow \psi \rangle}{\langle \Gamma \Rightarrow \Delta, \varphi \to \psi \rangle}$$

$$\frac{\langle \varphi_x(y), \Gamma \Rightarrow \Delta \rangle}{\langle \exists x \varphi, \Gamma \Rightarrow \Delta \rangle} \qquad \frac{\langle \Gamma \Rightarrow \varphi_x(y) \rangle}{\langle \Gamma \Rightarrow \Delta, \forall x \varphi \rangle}$$

$$\frac{\langle \varphi_x(t), \Gamma \Rightarrow \Delta \rangle}{\langle \forall x \varphi, \Gamma \Rightarrow \Delta \rangle} \qquad \frac{\langle \Gamma \Rightarrow \Delta, \varphi_x(t) \rangle}{\langle \Gamma \Rightarrow \Delta, \exists x \varphi \rangle}$$

Sequents are built from sets of formulas. The system allows two types of initial sequents, Ax : $\langle p, \Gamma \Rightarrow \Delta, p \rangle$ for an atomic formula p and L$\bot$: $\langle \bot, \Gamma \Rightarrow \Delta \rangle$, in both cases for arbitrary sets of formulas Γ and Δ. Generalization rules are subject to the eigenvariable condition, and in the specification rules t is free for x in φ. The calculus with cut is denoted by smG3i + Cut.

Definition 2. *The* cut rank *of a derivation P,* $\mathrm{r}(P)$*, is equal to* 0 *in case P is cut-free. Otherwise, it is computed as:*

$$\mathrm{r}(P) = 1 + \max\{\mathrm{d}(\varphi); \varphi \text{ is a cut formula in } P\}.$$

Cut rank is the most important complexity measure of derivations in the proof of cut elimination. The exact formulation above is quite useful in the sense that whenever a derivation has zero cut rank, it contains no cuts.

[1] mG3i denotes the class of intuitionistic multi-succedent systems (e.g. in [20]), "s" stands for "set". For instance in contrast to GHPC the rule L$\to$ in smG3i does not need to contain the principal formula in the left premise because contraction is implicit (see chapter 1). Similarly the principal formula does not need to occur in the premises of the specification rules.

Example 2.1.1. The left rule for implication, L→, is not invertible with respect to its left premise in smG3i, and the sequent $S : \langle\, p \to q \;\Rightarrow\; p \to q \,\rangle$ demonstrates this. As I mentioned in the previous chapter we can formulate inversion lemmas even for non-invertible rules, but the statements can only relate to their respective Ketonen variants. For L→ the context-sharing version is

$$\frac{\langle\, \Gamma \;\Rightarrow\; \Delta, \varphi \,\rangle \qquad \langle\, \psi, \Gamma \;\Rightarrow\; \Delta \,\rangle}{\langle\, \varphi \to \psi, \Gamma \;\Rightarrow\; \Delta \,\rangle},$$

so the lemma about the invertibility of S with respect to its left premise would claim the derivability of $S_1 : \langle \;\Rightarrow\; p \to q, p \,\rangle$. However, there is no cut-free proof of S_1 in intuitionistic logic due to the restricted R→.

Example 2.1.2. The following rule is admissible in smG3i:

$$\frac{\langle\, \Gamma \;\Rightarrow\; \Delta, \bot \,\rangle}{\langle\, \Gamma \;\Rightarrow\; \Delta \,\rangle}$$

Let P be a proof of $S : \langle\, \Gamma \;\Rightarrow\; \Delta, \bot \,\rangle$. The proof proceeds by induction on $\mathrm{d}(P)$. If S is an initial sequent, so is $\langle\, \Gamma \;\Rightarrow\; \Delta \,\rangle$. Otherwise S has been obtained through a logical inference R, for instance L→, in which case P ends as follows:

$$\frac{\begin{array}{c} P_1 \\ \langle\, \Gamma_1 \;\Rightarrow\; \Delta_1, \varphi, \bot \,\rangle \end{array} \qquad \begin{array}{c} P_2 \\ \langle\, \psi, \Gamma_2 \;\Rightarrow\; \Delta_2, \bot \,\rangle \end{array}}{\langle\, \varphi \to \psi, \Gamma \;\Rightarrow\; \Delta, \bot \,\rangle},$$

where $\Gamma_1 \cup \Gamma_2 = \Gamma$ and $\Delta_1 \cup \Delta_2 = \Delta$. Induction hypotheses allows us to remove $\bot$ from both premises of S, and the required endsequent is obtained again via an application of L→. The same pattern works also for every other rule of smG3i, first use IH on the premise(s) of R, then conclude with R. If R is one of the two restricted rules, it suffices to derive the conclusion without $\bot$ in the succedent. Also the new derivations do not have greater depth.

A sequent S is *regular* if no variable has both free and bound occurrences in S. A proof P of a sequent S is *regular* if no variable has both free and bound occurrences in P, instances of generalization rules in P have different eigenvariables and these variables occur only in the respective subderivations of the generalization rules.

Instances of cut are eliminated from derivations of regular sequents. This restriction is not limiting, since by renaming bound variables in non-regular sequents we can always arrive at regular ones. However, it is a necessary restriction as is demonstrated by the sequent $S : \langle\, \forall x \forall y (Q(y) \wedge R(x)) \;\Rightarrow\; R(y) \,\rangle$, where Q and R are unary predicates.[2] Assume P is a cut-free proof of S. Since cut-free proofs enjoy the subformula property and $\to$ does not occur in S, every formula in an antecedent of a sequent of P is a subformula of a formula in the endsequent of P, and similarly for the succedents. The only subformula of $R(y)$ is $R(y)$. But $R(y)$ is not a subformula of $\forall x \forall y (Q(y) \wedge R(x))$ because y is not free for x in $\forall y (Q(y) \wedge R(x))$. This argument shows that S has no cut-free proof.

[2] This is an example from Kleene's [12].

Regular derivations are useful in the proof of cut elimination, as well as in some of its preliminaries. The following lemma shows that given a proof P of a regular sequent S we can assume that P is a regular proof.

Lemma 2.1.3. *Every proof of a regular sequent S can be modified into a regular proof P' of S whose depth and cut rank remain unchanged.*

Proof. The idea is to first rename the bound variables that also have free occurrences in P by fresh ones, so that the sets of free and bound variables in P are disjoint. Then, if necessary, rename the eigenvariables in the subderivations of those rules that generalize over the same variable, or those whose eigenvariables also occur in another part of P. If always the topmost instances of such rules are processed at a time, and if the renaming is made by fresh variables which do not occur anywhere in P, the eigenvariable condition guarantees that the modified subtrees remain derivations of the *same* sequents. Therefore P' is a regular proof of S. A detailed proof can be found in [18] (Lemma 3.3.7). □

In the proof of cut elimination it is sometimes necessary to substitute a term occurring in one part of a derivation into another part of the same derivation. For this reason it is convenient to make explicit the conditions under which we can substitute terms throughout an entire proof. These are spelled out in the following lemma.

Lemma 2.1.4. *If a regular sequent S has a proof P, a variable x is not generalized in P, and no variable in t is either quantified or generalized in P, then $P[x/t]$ is a proof of $S[x/t]$. The substitution does not change the depth or cut rank of the original derivation.*

Proof. Again, I only give the basic idea, a detailed proof can be found in [15]. We can assume by Lemma 1.1 that P is a regular proof. Since no variable in t is quantified in P, t can be substituted for x in every formula. The most important thing to verify is the preservation of correct applications of generalization rules. Firstly, since by assumption x is not generalized in P, each R∀ and L∃ generalizes over the same variable before and after the substitution. And the condition that no eigenvariable occurs free in the conclusion of any application of R∀ or L∃ is also satisfied, because no variable in t is generalized in P. □

The initial sequents of smG3i denoted by Ax may seem somewhat restrictive. It is natural to require from any calculus of intuitionistic logic to derive all sequents whose antecedents and succedents have an *arbitrary* formula in common. Still the restriction on atomic formulas has its reasons, as it will become clear in the proof of inversion lemma, and it does not deprive smG3i of any inferential power.

Example 2.1.5. By induction on $\mathrm{d}(\varphi)$ we can show that all sequents of the form $S : \langle\, \varphi \;\Rightarrow\; \varphi \,\rangle$ are derivable in smG3i. If φ is an atomic formula, S is an instance of Ax. Among the binary connectives let me consider the case $\varphi \equiv \psi \to \chi$. By induction hypothesis there are derivations P_1 and P_2 of the sequents $\langle\, \psi \;\Rightarrow\; \psi \,\rangle$

and $\langle\, \chi \;\Rightarrow\; \chi \,\rangle$. The sequent $\langle\, \psi \to \chi \;\Rightarrow\; \psi \to \chi \,\rangle$ is constructed as follows:

$$\dfrac{\dfrac{\begin{array}{c} P_1 \\ \langle\, \psi \;\Rightarrow\; \psi \,\rangle \end{array} \qquad \begin{array}{c} P_2 \\ \langle\, \chi \;\Rightarrow\; \chi \,\rangle \end{array}}{\langle\, \psi, \psi \to \chi \;\Rightarrow\; \chi \,\rangle}}{\langle\, \psi \to \chi \;\Rightarrow\; \psi \to \chi \,\rangle}$$

Let $\varphi \equiv \forall x \psi$ and assume y is not quantified in ψ. Induction hypothesis gives a derivation P_1 of the sequent $S_1 : \langle\, \psi_x(y) \;\Rightarrow\; \psi_x(y) \,\rangle$. The result is obtained as follows:

$$\dfrac{\dfrac{\begin{array}{c} P_1 \\ \langle\, \psi_x(y) \;\Rightarrow\; \psi_x(y) \,\rangle \end{array}}{\langle\, \forall x \psi \;\Rightarrow\; \psi_x(y) \,\rangle}}{\langle\, \forall x \psi \;\Rightarrow\; \forall x \psi \,\rangle}$$

The same holds for the symbols $\wedge, \vee$ and $\exists$ as well.

The following two lemmas concern some basic properties of smG3i. The calculus is formulated so as to have admissible all structural rules, working with sets means that contraction and exchange are implicit, and dp-admissibility of weakening is guaranteed by the fact that initial sequents allow arbitrary contexts. This property of smG3i proves quite useful in the proof of cut elimination.

Lemma 2.1.6. *If a sequent $S : \langle\, \Gamma \;\Rightarrow\; \Delta \,\rangle$ has a regular proof P in* smG3i + Cut *and* Π, Λ *are arbitrary sets of formulas such that they contain no free variable that is generalized in P, there is also a proof P' of the sequent $S' : \langle\, \Pi, \Gamma \;\Rightarrow\; \Delta, \Lambda \,\rangle$ such that* $\mathrm{d}(P') \leq \mathrm{d}(P)$ *and* $\mathrm{r}(P') \leq \mathrm{r}(P)$.

Proof. By induction on the depth P. If S is an initial sequent, so is S'. Let the final inference of P be L$\wedge$:

$$\dfrac{\begin{array}{c} P_1 \\ \langle\, \varphi, \Gamma_1 \;\Rightarrow\; \Delta \,\rangle \end{array}}{\langle\, \varphi \wedge \psi, \Gamma_1 \;\Rightarrow\; \Delta \,\rangle},$$

with $\{\varphi \wedge \psi\} \cup \Gamma_1 = \Gamma$. If $\varphi \in \Pi$, it suffices to use the induction hypotheses (IH) on the premise, adding the set $\Pi \cup \{\varphi \wedge \psi\}$ to its antecedent and Λ to its succedent. Otherwise IH applied on the endsequent of P_1 gives the sequent $\langle\, \varphi, \Pi, \Gamma_1 \;\Rightarrow\; \Delta, \Lambda \,\rangle$, and the ensuing application of L$\wedge$ produces the required sequent $S' : \langle\, \varphi \wedge \psi, \Pi, \Gamma_1 \;\Rightarrow\; \Delta, \Lambda \,\rangle$. It holds for the obtained proof P' that $\mathrm{d}(P') \leq \mathrm{d}(P)$ and $\mathrm{r}(P') = \mathrm{r}(P)$.
Let P end with R$\forall$:

$$\dfrac{\begin{array}{c} P_1 \\ \langle\, \Gamma \;\Rightarrow\; \varphi_x(y) \,\rangle \end{array}}{\langle\, \Gamma \;\Rightarrow\; \Delta_1, \forall x \varphi \,\rangle}$$

By assumption $y \notin \mathrm{Var}(\Pi \cup \Lambda)$. Apply IH on the antecedent of the premise to obtain the sequent $\langle\, \Pi, \Gamma \Rightarrow \varphi_x(y) \,\rangle$, and conclude with R$\forall$ to derive the required $\langle\, \Pi, \Gamma \Rightarrow \Delta_1, \Lambda, \forall x\varphi \,\rangle$.

Similarly we can show that the rest of the rules preserve adding formulas to both cedents as well. □

Whenever there is an appeal to dp-admissibility of weakening in the rest of this chapter, e.g. "S_1 is obtained from S_2 by dp-admissibility of weakening", I shall simply write "S_1 is obtained from S_2 by (through) W".

The last ingredience is the inversion lemma for six rules. Again, although some of them are, in fact, not invertible, for the transformations needed in cut elimination the inversions formulated for their Ketonen variants suffice.

Lemma 2.1.7. *Whenever in cases (i)-(v) a sequent in the left column has a proof P with depth n and cut rank k, each sequent in the same row in the right column has a proof P' with depth and cut rank bounded by n and k, respectively. The same holds for (vi) provided P is regular, t is free for x in φ and no variable in t is generalized or quantified in P.*

(i) $\langle\, \varphi \wedge \psi, \Gamma \Rightarrow \Delta \,\rangle$ $\qquad$ $\langle\, \varphi, \psi, \Gamma \Rightarrow \Delta \,\rangle$

(ii) $\langle\, \varphi \vee \psi, \Gamma \Rightarrow \Delta \,\rangle$ $\qquad$ $\langle\, \varphi, \Gamma \Rightarrow \Delta \,\rangle$, $\langle\, \psi, \Gamma \Rightarrow \Delta \,\rangle$

(iii) $\langle\, \Gamma \Rightarrow \Delta, \varphi \wedge \psi \,\rangle$ $\qquad$ $\langle\, \Gamma \Rightarrow \Delta, \varphi \,\rangle$, $\langle\, \Gamma \Rightarrow \Delta, \psi \,\rangle$

(iv) $\langle\, \Gamma \Rightarrow \Delta, \varphi \vee \psi \,\rangle$ $\qquad$ $\langle\, \Gamma \Rightarrow \Delta, \varphi, \psi \,\rangle$

(v) $\langle\, \varphi \to \psi, \Gamma \Rightarrow \Delta \,\rangle$ $\qquad$ $\langle\, \psi, \Gamma \Rightarrow \Delta \,\rangle$

(vi) $\langle\, \exists x\varphi, \Gamma \Rightarrow \Delta \,\rangle$ $\qquad$ $\langle\, \varphi_x(t), \Gamma \Rightarrow \Delta \,\rangle$

Proof. In each case we proceed by induction on n. Also in each case we are given a derivation of the respective sequent and for this derivation I always use the same name P, though it is important to remember that P differs from case to case.

Let me consider (i) in detail. The base step of the induction is very similar in each case, so I discuss it only here. $S : \langle\, \varphi \wedge \psi, \Gamma \Rightarrow \Delta \,\rangle$ is an initial sequent, so either $\bot \in \Gamma$ or there is an atom p such that $p \in \Gamma \cap \Delta$. In both cases it holds that $S' : \langle\, \varphi, \psi, \Gamma \Rightarrow \Delta \,\rangle$ is also an initial sequent. Note that it may very well happen that $\varphi \in \Gamma$ or $\psi \in \Gamma$, but this is inconsequential to the proof. The base step of the induction is actually the reason why initial sequents are formulated with an *atomic* formula occurring in both cedents. If we allowed the cedents to share an arbitrary formula, which is e.g. what Gentzen did in his LK, then the proof of the inversion lemma would not go through.

In the induction step we distinguish whether $\varphi \wedge \psi$ is a principal formula in the last inference R of P or not. If the former holds, then R is L$\wedge$ and it suffices to take its premise $\langle\, \varphi, \Gamma \Rightarrow \Delta \,\rangle$ (alternatively with ψ in place of φ) and the rest follows by W. So assume that $\varphi \wedge \psi$ is not the principal formula of R.

Let R be L$\rightarrow$, then P has the following structure:

$$\dfrac{\dfrac{P_1}{\langle\, \varphi \wedge \psi, \Gamma_1 \Rightarrow \Delta_1, \gamma \,\rangle} \qquad \dfrac{P_2}{\langle\, \delta, \varphi \wedge \psi, \Gamma_2 \Rightarrow \Delta_2 \,\rangle}}{\langle\, \varphi \wedge \psi, \gamma \rightarrow \delta, \Gamma \Rightarrow \Delta \,\rangle},$$

Note that $\varphi \wedge \psi$ may not occur as a context in one of the premises, but this is not relevant to the following approach. Induction hypotheses (IH) applied on the premises gives derivations P_1' and P_2' of the sequents $S_1' : \langle\, \varphi, \psi, \Gamma_1 \Rightarrow \Delta_1, \gamma \,\rangle$ and $S_2' : \langle\, \varphi, \psi, \delta, \Gamma_2 \Rightarrow \Delta_2 \,\rangle$ with depths bounded by $\mathrm{d}(P_1)$ and $\mathrm{d}(P_2)$, respectively. It may happen that φ or ψ occurs in the contexts Γ_1 or Γ_2, or that one them is the formula δ. In any case the rule L$\rightarrow$ remains applicable on S_1' and S_2' and its conclusion is the sequent $\langle\, \varphi, \psi, \gamma \rightarrow \delta, \Gamma \Rightarrow \Delta \,\rangle$. The same pattern works also for any other rule R, For instance if R is unary, IH on its premise gives a derivation of the respective sequent without the complex formula $\varphi \wedge \psi$ as a context, and the subsequent application of R finishes the proof.

Let us move to case (ii). If $\varphi \vee \psi$ is principal in the last inference R of P, simply take its premises. Otherwise let R be for instance L$\exists$, then P ends as follows:

$$\dfrac{\dfrac{P_1}{\langle\, \varphi \vee \psi, \chi_x(y), \Gamma \Rightarrow \Delta \,\rangle}}{\langle\, \varphi \vee \psi, \forall x \chi, \Gamma \Rightarrow \Delta \,\rangle}$$

IH on the endsequent of P_1 produces derivations P_2 and P_3 of the sequents $S_2 : \langle\, \varphi, \chi_x(y), \Gamma \Rightarrow \Delta \,\rangle$ and $S_3 : \langle\, \psi, \chi_x(y), \Gamma \Rightarrow \Delta \,\rangle$, respectively. Potential implicit contractions of either φ or ψ do hinder the procedure. The eigenvariable condition guarantees that the rule L$\exists$ is applicable on both S_2 and S_3. Other rules can be dealt with in a similar manner. Also case (iii) proceeds the same way.

I show an example case of (iv), others are dealt with as in (i). The last inference R of P is the restricted rule R$\rightarrow$:

$$\dfrac{\dfrac{P_1}{\langle\, \gamma, \Gamma \Rightarrow \delta \,\rangle}}{\langle\, \Gamma \Rightarrow \Delta_1, \varphi \vee \psi, \gamma \rightarrow \delta \,\rangle}$$

In this case there is naturally no appeal to IH, as $\varphi \vee \psi$ does not occur in the premise of R, but we can derive $\langle\, \Gamma \Rightarrow \Delta_1, \varphi, \psi, \gamma \rightarrow \delta \,\rangle$ directly from the endsequent of P_1.

Case (v) also follows the standard pattern, if $\varphi \rightarrow \psi$ occurs in a premise of the last inference rule R of P, use IH on this premise to obtain the simplified sequent, and conclude with the application of R.

In the final case we make use of substitution lemma. Let $\exists x \varphi$ be principal in the last rule R of P, i.e. R is L$\exists$. P ends as follows:

$$\dfrac{\dfrac{P_1}{\langle\, \varphi_x(y), \Gamma \Rightarrow \Delta \,\rangle}}{\langle\, \exists x \varphi, \Gamma \Rightarrow \Delta \,\rangle}$$

Let t be a term such that no variable in t is generalized or quantified in P. Since P is regular, the variable y is not generalized in P_1 and we can use substitution lemma on P_1 and t to obtain a derivation $P_1[y/t]$ of the sequent $\langle \varphi_x(t), \Gamma \Rightarrow \Delta \rangle$. Assume now that R is the rule R∀:

$$\dfrac{\begin{array}{c} P_1 \\ \langle \exists x\varphi, \Gamma \Rightarrow \psi_v(y) \rangle \end{array}}{\langle \exists x\varphi, \Gamma \Rightarrow \Delta, \forall v\psi \rangle}$$

Let t be a term satisfying the assumption of the lemma. In particular this means that $y \notin \mathrm{Var}(t)$. Use IH on the endsequent of P_1, obtain a derivation of the sequent $\langle \varphi_x(t), \Gamma \Rightarrow \psi_v(y) \rangle$, and conclude with R∀. Similarly for all other cases where $\exists x\varphi$ is not principal in R. □

2.2 Cut elimination for smG3i with an upper bound

In this section I give a proof of cut elimination for the calculus smG3i + Cut together with an upper bound on the increase of depths of derivations during the procedure. Most of the lemmas discussed in the previous section are used throughout, although I always try to explicitly state when a lemma is employed. First let me briefly mention the overall method which is followed in the course of the proof.

We are given an arbitrary derivation of smG3i + Cut. In the first step we mark all the cut inferences whose cut formulas have maximal depth d such that $1 + d$ is the cut rank of the entire derivation. Among these maximal-rank cuts we select a topmost one, i.e. one such that its respective subderivation does not contain another maximal-rank cut. The principal lemma of the entire proof establishes that this subderivation P can be transformed into a derivation P' (of the same endsequent) with $\mathrm{r}(P') < \mathrm{r}(P)$. Two basic methods are employed, the permutation of the cut upwards over another inference rule, and the replacement of a complex cut for cuts on formulas of lower complexity. Once the cut rank of P has been decreased by at least 1, the procedure is iterated for the rest of the maximal-rank cuts, always processing a topmost one at a time. This way the cut rank of the entire derivation is decreased, and the outer induction on the cut rank finishes the proof.

If we take the superexponential function with base 2 defined as

$$2^n_0 = n, \quad 2^n_{m+1} = 2^{2^n_m},$$

it is shown that any derivation P of a regular sequent S can be transformed into a cut-free P' of S with $\mathrm{d}(P) \leq 2^{\mathrm{d}(P)}_{\mathrm{r}(P)}$. The central lemma reduces a topmost maximal-rank cut, which causes at most doubles the depth of the respective subderivation. The reduction of all maximal-rank cuts in P by at least 1 causes an exponential increase of P, and its iteration until a cut-free P' is obtained generates a superexponential increase.

The organization follows the proof of cut elimination for classical logic given in [18].[3]

[3] This was first given in Ivo Kylar's master's thesis [13].

Lemma 2.2.1 (Cut reduction)**.** *Let P be a regular proof of $S : \langle \Pi, \Gamma \Rightarrow \Delta, \Lambda \rangle$ whose last inference is a cut on a formula φ such that for every other cut formula ψ in P it holds that* $\mathrm{d}(\psi) < \mathrm{d}(\varphi)$*:*

$$\dfrac{\begin{matrix} P_1 \\ \langle \Gamma \Rightarrow \Delta, \varphi \rangle \end{matrix} \qquad \begin{matrix} P_2 \\ \langle \varphi, \Pi \Rightarrow \Lambda \rangle \end{matrix}}{\langle \Pi, \Gamma \Rightarrow \Delta, \Lambda \rangle}$$

Denote $k = \mathrm{d}(P_1) + \mathrm{d}(P_2)$. Then there is a proof P' of S with $\mathrm{r}(P') < \mathrm{r}(P)$ and $\mathrm{d}(P') \leq k$.

Proof. We proceed by induction on k, which is sometimes called the *cut level* of the respective cut inference. There are four main cases, each of which nests into several subcases.

1. Either premise is an initial sequent.
2. The cut formula φ is not principal in the left premise.
3. φ is principal in the left premise and not principal in the right premise.
4. φ is principal in both premises.[4]

Case 1 concerns arbitrary arrangements in the base step of the induction and it is also the easiest to deal with.

In case 2 the subcases are distinguished according to the last inference R of P_1. The general pattern which works for all cases is the following: since φ is not principal in P_1, it has to appear in the succedent(s) of the premise(s) of R. We are proving the reduction lemma by induction on the cut level, hence the induction hypotheses can be applied on any derivation ending with a topmost maximal-rank cut inference whose cut level is strictly less than k. Therefore taking a premise of $\langle \Gamma \Rightarrow \Delta, \varphi \rangle$ together with $\langle \varphi, \Pi \Rightarrow \Lambda \rangle$ and cutting out φ produces a derivation on which IH can be applied. This yields a derivation of a sequent dependent on R whose depth is bounded by $k - 1$, and the subsequent application of R gives the endsequent $\langle \Pi, \Gamma \Rightarrow \Delta, \Lambda \rangle$ of the original cut satisfying both conditions required by the lemma.

Case 3 proceeds as the previous one, most subcases are symmetric but R$\rightarrow$ and R$\forall$ have to be further analyzed. This is also the only part of the proof of cut elimination which is characteristic of intuitionistic logic. The reason for the failure of the straightforward approach of permuting cut over R$\rightarrow$ or R$\forall$ without distinguishing the last inference of P_1 is the following: if S_1 is a premise of $\langle \varphi, \Pi \Rightarrow \Lambda \rangle$ and S_2 is the result of applying cut on $(\langle \Gamma \Rightarrow \Delta, \varphi \rangle, S_1, \varphi)$, we can apply IH on this derivation of S_2. However, in the case when R is R$\rightarrow$ or R$\forall$ these rules may not be applicable on the endsequent of the obtained derivation because of the possibly nonempty context Δ in its succedent.

In the last case 4 the cut formula φ has just been introduced into the premises S_1 and S_2 of the cut inference. Here the idea is to analyze the premises of S_1 and S_2. If φ appears in their contexts, which is the most involved case, the idea

[4] In classical logic cases 2 and 3 are symmetric.

is to use inversions to obtain simpler sequents with no occurrence of φ and which possibly contain only the immediate subformulas of φ. These get eliminated by the application of one or two cut inferences and that way we obtain the original endsequent (possibly through W). Note that such an introduction of new cuts does not hinder the overall procedure (and is in fact necessary), because the cut formulas of these inferences are strictly simpler than φ.

Case 1. Either premise is an initial sequent.

1.1 The left premise S_1 is an initial sequent. S_1 contains either an atom p in both cedents, or $\bot$ in the antecedent. If p is the cut formula φ, the endsequent is obtained through W. Otherwise the endsequent is an initial sequent, and the same is true if $\bot \in \Gamma$.

1.2 The right premise S_2 is an initial sequent. The cases $p \in \Pi \cap \Lambda$ for an atom p and $\bot \in \Pi$ are dealt with exactly as above. What remains is the case when the cut formula φ is the constant $\bot$. This means that S_1 is $\langle \Gamma \Rightarrow \Delta, \bot \rangle$, by example 2.1.2 there is a derivation P_1' of $S' : \langle \Gamma \Rightarrow \Delta \rangle$ with $\mathrm{d}(P_1') \leq \mathrm{d}(P_1)$, and the endsequent is obtained through W.

Case 2. The cut formula φ is not principal in the left premise. In this case it is irrelevant whether the principal formula ψ of R also appears in its premise(s), because the general approach is to apply first cut and then R, at which point ψ gets implicitly contracted. So I formulate the cases without ψ recurring in the premises.

2.1 L$\wedge$ (the principal formula of R is $\psi \wedge \chi$)

$$\dfrac{\dfrac{\begin{matrix} P_3 \\ \langle \psi, \Gamma_1 \Rightarrow \Delta, \varphi \rangle \end{matrix}}{\langle \psi \wedge \chi, \Gamma_1 \Rightarrow \Delta, \varphi \rangle} \qquad \begin{matrix} P_2 \\ \langle \varphi, \Pi \Rightarrow \Lambda \rangle \end{matrix}}{\langle \psi \wedge \chi, \Pi, \Gamma_1 \Rightarrow \Delta, \Lambda \rangle},$$

with $\{\psi \wedge \chi\} \cup \Gamma_1 = \Gamma$. Denote $S_3 : \langle \psi, \Gamma_1 \Rightarrow \Delta, \varphi \rangle$, $S_2 : \langle \varphi, \Pi \Rightarrow \Lambda \rangle$. Because $\mathrm{d}(P_3) + \mathrm{d}(P_2) < k = \mathrm{d}(P_1) + \mathrm{d}(P_2)$, we can apply the induction hypotheses on the triple (S_3, S_2, φ), which produces a proof P_4 of $S' : \langle \psi, \Pi, \Gamma_1 \Rightarrow \Delta, \Lambda \rangle$ with $\mathrm{r}(P_4) < \mathrm{r}(P)$. If $\psi \in \Pi$, the result is obtained through W. Otherwise applying L$\wedge$ on S' gives the proof P' of original endsequent. Also it holds that $\mathrm{d}(P_4) \leq k - 1$, hence $\mathrm{d}(P') \leq k$. The transformation of the original proof P can be depicted as follows:

$$\dfrac{\dfrac{\begin{matrix} P_3 \\ \langle \psi, \Gamma_1 \Rightarrow \Delta, \varphi \rangle \end{matrix} \qquad \begin{matrix} P_2 \\ \langle \varphi, \Pi \Rightarrow \Lambda \rangle \end{matrix}}{\langle \psi, \Pi, \Gamma_1 \Rightarrow \Delta, \Lambda \rangle}}{\langle \psi \wedge \chi, \Pi, \Gamma_1 \Rightarrow \Delta, \Lambda \rangle}$$

Cut is permuted over L∧. For the sake of completeness I give the final proof P':[5]

$$\dfrac{\begin{matrix} P_4 \\ \langle\, \psi, \Pi, \Gamma_1 \Rightarrow \Delta, \Lambda \,\rangle \end{matrix}}{\langle\, \psi \wedge \chi, \Pi, \Gamma_1 \Rightarrow \Delta, \Lambda \,\rangle}$$

2.2 R∧
P has the following structure:

$$\dfrac{\dfrac{\begin{matrix} P_3 \\ \langle\, \Gamma \Rightarrow \Delta_1, \psi, \varphi \,\rangle \end{matrix} \quad \begin{matrix} P_4 \\ \langle\, \Gamma \Rightarrow \Delta_1, \chi, \varphi \,\rangle \end{matrix}}{\langle\, \Gamma \Rightarrow \Delta_1, \psi \wedge \chi, \varphi \,\rangle} \qquad \begin{matrix} P_2 \\ \langle\, \varphi, \Pi \Rightarrow \Lambda \,\rangle \end{matrix}}{\langle\, \Pi, \Gamma \Rightarrow \Delta_1, \Lambda, \psi \wedge \chi \,\rangle},$$

with $\{\psi \wedge \chi\} \cup \Delta_1 = \Delta$. In the case of binary rules, such as R∧, cut has to permuted above both premises of these rules. This means that it gets duplicated, instead of one occurrence of cut in the original proof there are two in the transformed proof. Again, this does not hinder the elimination of these cuts, since both have strictly lower cut level and IH guarantees that they can be replaced by derivations of the same endsequents with strictly lower cut ranks. This transformation produces the following derivation:

$$\dfrac{\dfrac{\begin{matrix} P_3 \\ \langle\, \Gamma \Rightarrow \Delta_1, \psi, \varphi \,\rangle \end{matrix} \quad \begin{matrix} P_2 \\ \langle\, \varphi, \Pi \Rightarrow \Lambda \,\rangle \end{matrix}}{\langle\, \Pi, \Gamma \Rightarrow \Delta_1, \Lambda, \psi \,\rangle} \qquad \dfrac{\begin{matrix} P_4 \\ \langle\, \Gamma \Rightarrow \Delta_1, \chi, \varphi \,\rangle \end{matrix} \quad \begin{matrix} P_2 \\ \langle\, \varphi, \Pi \Rightarrow \Lambda \,\rangle \end{matrix}}{\langle\, \Pi, \Gamma \Rightarrow \Delta_1, \Lambda, \chi \,\rangle}}{\langle\, \Pi, \Gamma \Rightarrow \Delta_1, \Lambda, \psi \wedge \chi \,\rangle}$$

If $\psi \in \Lambda$ or $\chi \in \Delta_1$, the result follows by W.[6] Otherwise IH on both immediate subderivations separately gives derivations of the respective endsequents, each of which has strictly lower cut rank and has depth bounded by $k - 1$. Then P' is obtained by R∧, and so $\mathrm{d}(P') \leq k$.

2.3 L∨ – similar to 2.2

2.4 R∨ – similar to 2.1

2.5 L→

$$\dfrac{\dfrac{\begin{matrix} P_3 \\ \langle\, \Gamma_1 \Rightarrow \Delta_1, \psi, \varphi \,\rangle \end{matrix} \quad \begin{matrix} P_4 \\ \langle\, \chi, \Gamma_2 \Rightarrow \Delta_2, \varphi \,\rangle \end{matrix}}{\langle\, \psi \to \chi, \Gamma_0 \Rightarrow \Delta, \varphi \,\rangle} \qquad \begin{matrix} P_2 \\ \langle\, \varphi, \Pi \Rightarrow \Lambda \,\rangle \end{matrix}}{\langle\, \psi \to \chi, \Pi, \Gamma_0 \Rightarrow \Delta, \Lambda \,\rangle},$$

[5] In the following I usually only give the original proof P and the transformation. I think it is illustrative to see the transformation step, it indicates the ideas behind the general pattern of the proof and the structure of the few final inferences of P' can be inferred from it.

[6] This may happen in every one of the discussed cases, so I will not consider it again.

with $\Gamma_1 \cup \Gamma_2 = \Gamma_0$, $\Delta_1 \cup \Delta_2 = \Delta$ and $\{\psi \to \chi\} \cup \Gamma_0 = \Gamma$. Again, first rearrange the proof so that cuts on φ precede L$\to$.

$$\dfrac{\dfrac{\overset{P_3}{\langle \Gamma_1 \Rightarrow \Delta_1, \psi, \varphi \rangle} \quad \overset{P_2}{\langle \varphi, \Pi \Rightarrow \Lambda \rangle}}{\langle \Pi, \Gamma_1 \Rightarrow \Delta, \Lambda, \psi \rangle} \qquad \dfrac{\overset{P_4}{\langle \chi, \Gamma_2 \Rightarrow \Delta_2, \varphi \rangle} \quad \overset{P_2}{\langle \varphi, \Pi \Rightarrow \Lambda \rangle}}{\langle \chi, \Pi, \Gamma_2 \Rightarrow \Delta, \Lambda \rangle}}{\langle \psi \to \chi, \Pi, \Gamma_0 \Rightarrow \Delta, \Lambda \rangle}$$

By IH both premises of L$\to$ have derivations with strictly lower cut rank than P and depth bounded by $k-1$, hence $\mathrm{d}(P') \leq k$.

2.6 R$\to$

In the original proof P

$$\dfrac{\dfrac{\overset{P_3}{\langle \psi, \Gamma \Rightarrow \chi \rangle}}{\langle \Gamma \Rightarrow \Delta_1, \psi \to \chi, \varphi \rangle} \quad \overset{P_2}{\langle \varphi, \Pi \Rightarrow \Lambda \rangle}}{\langle \Pi, \Gamma \Rightarrow \Delta_1, \Lambda, \psi \to \chi \rangle},$$

with $\{\psi \to \chi\} \cup \Delta_1 = \Delta$, the formula φ has been introduced by R$\to$ in the left premise of the cut and eliminated in the next step, so the easiest way to get the original endsequent is by not deriving φ.

$$\dfrac{\overset{P_3}{\langle \psi, \Gamma \Rightarrow \chi \rangle}}{\langle \Gamma \Rightarrow \psi \to \chi \rangle},$$

and the rest follows by W.

2.7 L$\forall$

The cut in the original derivation P

$$\dfrac{\dfrac{\overset{P_3}{\langle \psi_x(t), \Gamma_1 \Rightarrow \Delta, \varphi \rangle}}{\langle \forall x \psi, \Gamma_1 \Rightarrow \Delta, \varphi \rangle} \quad \overset{P_2}{\langle \varphi, \Pi \Rightarrow \Lambda \rangle}}{\langle \forall x \psi, \Pi, \Gamma_1 \Rightarrow \Delta, \Lambda \rangle}$$

is permuted over L$\forall$

$$\dfrac{\dfrac{\overset{P_3}{\langle \psi_x(t), \Gamma_1 \Rightarrow \Delta, \varphi \rangle} \quad \overset{P_2}{\langle \varphi, \Pi \Rightarrow \Lambda \rangle}}{\langle \psi_x(t), \Pi, \Gamma_1 \Rightarrow \Delta, \Lambda \rangle}}{\langle \forall x \psi, \Pi, \Gamma_1 \Rightarrow \Delta, \Lambda \rangle},$$

and IH followed by the application of L∀ derives the original endsequent.

2.8 R∀

$$\dfrac{\dfrac{\overset{P_3}{\langle\, \Gamma \Rightarrow \psi_x(y) \,\rangle}}{\langle\, \Gamma \Rightarrow \Delta_1, \forall x\psi, \varphi \,\rangle} \qquad \overset{P_2}{\langle\, \varphi, \Pi \Rightarrow \Lambda \,\rangle}}{\langle\, \Pi, \Gamma \Rightarrow \Delta_1, \Lambda, \forall x\psi \,\rangle}$$

As in 2.6 the introduction of φ was redundant.

$$\dfrac{\overset{P_3}{\langle\, \Gamma \Rightarrow \psi_x(y) \,\rangle}}{\langle\, \Gamma \Rightarrow \forall x\psi \,\rangle},$$

the rest follows by W.

2.9 L∃

$$\dfrac{\dfrac{\overset{P_3}{\langle\, \psi_x(y), \Gamma_1 \Rightarrow \Delta, \varphi \,\rangle}}{\langle\, \exists x\psi, \Gamma_1 \Rightarrow \Delta, \varphi \,\rangle} \qquad \overset{P_2}{\langle\, \varphi, \Pi \Rightarrow \Lambda \,\rangle}}{\langle\, \exists x\psi, \Pi, \Gamma_1 \Rightarrow \Delta, \Lambda \,\rangle}$$

This is the first place where we use the regularity of P. In particular this assumption guarantees that $y \notin \mathrm{Var}\{\Pi \cup \Lambda\}$, and therefore adding Π and Λ as contexts does not disrupt the subsequent application of L∃.

$$\dfrac{\dfrac{\overset{P_3}{\langle\, \psi_x(y), \Gamma_1 \Rightarrow \Delta, \varphi \,\rangle} \qquad \overset{P_2}{\langle\, \varphi, \Pi \Rightarrow \Lambda \,\rangle}}{\langle\, \psi_x(y), \Pi, \Gamma_1 \Rightarrow \Delta, \Lambda \,\rangle}}{\langle\, \exists x\psi, \Pi, \Gamma_1 \Rightarrow \Delta, \Lambda \,\rangle}$$

2.10 R∃ Similar to 2.7.

2.11 Cut
Assume that the cut formula φ occurs in both succedents in the premises of the upper cut inference.

$$\dfrac{\dfrac{\overset{P_3}{\langle\, \Gamma_1 \Rightarrow \Delta_1, \varphi, \psi \,\rangle} \qquad \overset{P_4}{\langle\, \psi, \Gamma_2 \Rightarrow \Delta_2, \varphi \,\rangle}}{\langle\, \Gamma \Rightarrow \Delta, \varphi \,\rangle} \qquad \overset{P_2}{\langle\, \varphi, \Pi \Rightarrow \Lambda \,\rangle}}{\langle\, \Pi, \Gamma \Rightarrow \Delta, \Lambda \,\rangle},$$

$\Gamma_1 \cup \Gamma_2 = \Gamma$ and $\Delta_1 \cup \Delta_2 = \Delta$. The pattern is the same, the lower cut inference gets duplicated as it is permuted over the upper one, but both of its instances have lower level afterwards.

$$\dfrac{\dfrac{\overset{P_3}{\langle \Gamma_1 \Rightarrow \Delta_1, \varphi, \psi \rangle} \qquad \overset{P_2}{\langle \varphi, \Pi \Rightarrow \Lambda \rangle}}{\langle \Pi, \Gamma_1 \Rightarrow \Delta_1, \Lambda, \psi \rangle} \qquad \dfrac{\overset{P_4}{\langle \psi, \Gamma_2 \Rightarrow \Delta_2, \varphi \rangle} \qquad \overset{P_2}{\langle \varphi, \Pi \Rightarrow \Lambda \rangle}}{\langle \psi, \Pi, \Gamma_2 \Rightarrow \Delta_2, \Lambda \rangle}}{\langle \Pi, \Gamma \Rightarrow \Delta, \Lambda \rangle}$$

Both premises of the lower cut inference have derivations of strictly lower cut rank than P by IH, and these derivations have depths bounded by $k - 1$, hence the conditions of the lemma are satisfied.

Case 3. φ is principal in the left premise and not principal in the right premise. The cases in which R is an unrestricted inference rule are symmetric to the previous cases. For illustration I give one example, R is L$\rightarrow$.

3.1 L$\rightarrow$
The original derivation P

$$\dfrac{\overset{P_1}{\langle \Gamma \Rightarrow \Delta, \varphi \rangle} \qquad \dfrac{\overset{P_3}{\langle \varphi, \Pi_1 \Rightarrow \Lambda_1, \psi \rangle} \qquad \overset{P_4}{\langle \chi, \varphi, \Pi_2 \Rightarrow \Lambda_2 \rangle}}{\langle \varphi, \psi \rightarrow \chi, \Pi_0 \Rightarrow \Lambda \rangle}}{\langle \psi \rightarrow \chi, \Pi_0, \Gamma \Rightarrow \Delta, \Lambda \rangle},$$

with $\Pi_1 \cup \Pi_2 = \Pi_0$, $\Lambda_1 \cup \Lambda_2 = \Lambda$ and $\{\psi \rightarrow \chi\} \cup \Pi_0 = \Pi$, is first transformed into

$$\dfrac{\dfrac{\overset{P_1}{\langle \Gamma \Rightarrow \Delta, \varphi \rangle} \qquad \overset{P_3}{\langle \varphi, \Pi_1 \Rightarrow \Lambda_1, \psi \rangle}}{\langle \Pi_1, \Gamma \Rightarrow \Delta, \Lambda_1, \psi \rangle} \qquad \dfrac{\overset{P_1}{\langle \Gamma \Rightarrow \Delta, \varphi \rangle} \qquad \overset{P_4}{\langle \chi, \varphi, \Pi_2 \Rightarrow \Lambda_2 \rangle}}{\langle \chi, \Pi_2, \Gamma \Rightarrow \Delta, \Lambda_2 \rangle}}{\langle \psi \rightarrow \chi, \Pi_0, \Gamma \Rightarrow \Delta, \Lambda \rangle},$$

and IH guarantees the existence of suitable derivations of the premises of L$\rightarrow$.

Now we get to the part of the proof of cut elimination which is characteristic of intuitionistic logic. The restricted rules do not allow for a straightforward application of the general method of cases 2 and 3 which works for the rest of the rules, i.e. permuting cut over R. This is because if cut was applied before R$\rightarrow$ (R$\forall$), a possibly nonempty context Δ (coming from the left premise of the original cut) might appear in the succedent of the conclusion of this cut inference, and therefore R$\rightarrow$ (R$\forall$) would no longer be applicable. The idea is to use the assumption that φ is now principal in the left premise of the cut. This means that φ is not an atomic formula and we can ask which rule has introduced it into the left premise. There are five possibilities for both rules and the simpler ones are R$\rightarrow$ and R$\forall$. In these two cases it suffices to modify the last inference of P_1 so that there is no Δ in the succedent of its conclusion. This is a permitted step, because by definition of the restricted rules Δ may be empty. Afterwards it suffices to proceed as above, permuting cut over R$\rightarrow$ (R$\forall$) and using IH to obtain an appropriate derivation of the premise of R→ (R∀). The remaining three cases, R∧, R∨ and R∃ are slightly more complex, because here we cannot eliminate the context Δ, and therefore we cannot permute cut over R. Now we apply the method which is characteristic of case 4, first use a suitable inversion on the right premise of the cut, and then replace the cut by one or two cuts on simpler formulas.[7] I give all five cases for R∀ and these can be basically copied for R→.

3.2 R∀

3.2.1 $\varphi \equiv \gamma \to \delta$.

$$\dfrac{\dfrac{\overset{P_3}{\langle \gamma, \Gamma \Rightarrow \delta \rangle}}{\langle \Gamma \Rightarrow \Delta, \gamma \to \delta \rangle} \qquad \dfrac{\overset{P_4}{\langle \gamma \to \delta, \Pi \Rightarrow \psi_x(y) \rangle}}{\langle \gamma \to \delta, \Pi \Rightarrow \Lambda_1, \forall x \psi \rangle}}{\langle \Pi, \Gamma \Rightarrow \Delta, \Lambda_1, \forall x \psi \rangle}$$

First derive the left premise of the cut without Δ (call the derivation P_1'), then use the standard transformation:

$$\dfrac{\dfrac{\overset{P_1'}{\langle \Gamma \Rightarrow \gamma \to \delta \rangle} \qquad \overset{P_4}{\langle \gamma \to \delta, \Pi \Rightarrow \psi_x(y) \rangle}}{\langle \Pi, \Gamma \Rightarrow \psi_x(y) \rangle}}{\langle \Pi, \Gamma \Rightarrow \Delta, \Lambda_1, \forall x \psi \rangle}$$

By IH there is a derivation of the sequent $\langle \Pi, \Gamma \Rightarrow \psi_x(t) \rangle$ with decreased cut rank and with depth bounded by $k-1$. Also by the regularity of P the variable y does not occur in $\Gamma \cup \Delta$, hence R∀ remains applicable on the endsequent of this derivation.

3.2.2 $\varphi \equiv \forall v \chi$

$$\dfrac{\dfrac{\overset{P_3}{\langle \Gamma \Rightarrow \chi_v(t) \rangle}}{\langle \Gamma \Rightarrow \Delta, \forall v \chi \rangle} \qquad \dfrac{\overset{P_4}{\langle \forall v \chi, \Pi \Rightarrow \psi_x(y) \rangle}}{\langle \forall v \chi, \Pi \Rightarrow \Lambda_1, \forall x \psi \rangle}}{\langle \Pi, \Gamma \Rightarrow \Delta, \Lambda_1, \forall x \psi \rangle}$$

$$\dfrac{\dfrac{\overset{P_1'}{\langle \Gamma \Rightarrow \forall v \chi \rangle} \qquad \overset{P_4}{\langle \forall v \chi, \Pi \Rightarrow \psi_x(y) \rangle}}{\langle \Pi, \Gamma \Rightarrow \psi_x(y) \rangle}}{\langle \Pi, \Gamma \Rightarrow \Delta, \Lambda_1, \forall x \psi \rangle}$$

Again, after applying IH and obtaining a suitable derivation of the sequent $\langle \Pi, \Gamma \Rightarrow \psi_x(y) \rangle$, the rule R∀ remains applicable because P is regular.

[7] This shows the necessity of inversions for L∧, L∨ and L∃.

3.2.3 $\varphi \equiv \gamma \wedge \delta$

$$\dfrac{\dfrac{\overset{P_3}{\langle \Gamma \Rightarrow \Delta, \gamma \rangle} \qquad \overset{P_4}{\langle \Gamma \Rightarrow \Delta, \delta \rangle}}{\langle \Gamma \Rightarrow \Delta, \gamma \wedge \delta \rangle} \qquad \overset{P_2}{\langle \gamma \wedge \delta, \Pi \Rightarrow \Lambda_1, \forall x\psi \rangle}}{\langle \Pi, \Gamma \Rightarrow \Delta, \Lambda_1, \forall x\psi \rangle}$$

Use inversion lemma on the endsequent of P_2 and the formula $\gamma \wedge \delta$, obtain a derivation P_2' of the sequent $\langle \gamma, \delta, \Pi \Rightarrow \Lambda_1, \forall x\psi \rangle$. P' is constructed as follows:

$$\dfrac{\overset{P_3}{\langle \Gamma \Rightarrow \Delta, \gamma \rangle} \qquad \dfrac{\overset{P_4}{\langle \Gamma \Rightarrow \Delta, \delta \rangle} \qquad \overset{P_2'}{\langle \delta, \gamma, \Pi \Rightarrow \Lambda_1, \forall x\psi \rangle}}{\langle \gamma, \Pi, \Gamma \Rightarrow \Delta, \Lambda_1, \forall x\psi \rangle}}{\langle \Pi, \Gamma \Rightarrow \Delta, \Lambda_1, \forall x\psi \rangle}$$

As there is no appeal to IH, $\mathrm{d}(P') \leq \max\{\mathrm{d}(P_1)+1, \mathrm{d}(P_2)+1\} \leq k$.

3.2.4 $\varphi \equiv \gamma \vee \delta$

$$\dfrac{\dfrac{\overset{P_3}{\langle \Gamma \Rightarrow \Delta, \gamma \rangle}}{\langle \Gamma \Rightarrow \Delta, \gamma \vee \delta \rangle} \qquad \overset{P_2}{\langle \gamma \vee \delta, \Pi \Rightarrow \Lambda_1, \forall x\psi \rangle}}{\langle \Pi, \Gamma \Rightarrow \Delta, \Lambda_1, \forall x\psi \rangle}$$

Inversion lemma on the right premise of the cut and the formula $\gamma \vee \delta$ gives a derivation P_2' of the sequent $\langle \gamma, \Pi \Rightarrow \Lambda_1, \forall x\psi \rangle$.

$$\dfrac{\overset{P_3}{\langle \Gamma \Rightarrow \Delta, \gamma \rangle} \qquad \overset{P_2'}{\langle \gamma, \Pi \Rightarrow \Lambda_1, \forall x\psi \rangle}}{\langle \Pi, \Gamma \Rightarrow \Delta, \Lambda_1, \forall x\psi \rangle}$$

3.2.5 $\varphi \equiv \exists x\chi$

$$\dfrac{\dfrac{\overset{P_3}{\langle \Gamma \Rightarrow \Delta, \chi_v(t) \rangle}}{\langle \Gamma \Rightarrow \Delta, \exists v\chi \rangle} \qquad \overset{P_2}{\langle \exists v\chi, \Pi \Rightarrow \Lambda_1, \forall x\psi \rangle}}{\langle \Pi, \Gamma \Rightarrow \Delta, \Lambda_1, \forall x\psi \rangle}$$

Since P is regular, no variable of t is generalized or quantified in P_2, and hence we can apply the inversion lemma on the right premise of the cut and the formula $\exists v\chi$. This produces a derivation P_2' of the sequent $\langle \chi_v(t), \Pi \Rightarrow \Lambda_1, \forall x\psi \rangle$.

$$\dfrac{\overset{P_3}{\langle \Gamma \Rightarrow \Delta, \chi_v(t) \rangle} \qquad \overset{P_2'}{\langle \chi_v(t), \Pi \Rightarrow \Lambda_1, \forall x\psi \rangle}}{\langle \Pi, \Gamma \Rightarrow \Delta, \Lambda_1, \forall x\psi \rangle}$$

Case 4. φ is principal in both premises. There are five possibilities according to the outermost symbol of φ. Each time I assume the case in which the principal formula also appears in the premise(s) of the respective rule. The general method is to use inversions to get rid of the complex formula, and then to apply cuts only on formulas of lower depths.

4.1 $\varphi \equiv \psi \wedge \chi$

$$\dfrac{\dfrac{\overset{P_3}{\langle \Gamma \Rightarrow \Delta, \psi \wedge \chi, \psi \rangle} \qquad \overset{P_4}{\langle \Gamma \Rightarrow \Delta, \psi \wedge \chi, \chi \rangle}}{\langle \Gamma \Rightarrow \Delta, \psi \wedge \chi \rangle} \qquad \dfrac{\overset{P_5}{\langle \psi, \psi \wedge \chi, \Pi \Rightarrow \Lambda \rangle}}{\langle \psi \wedge \chi, \Pi \Rightarrow \Lambda \rangle}}{\langle \Pi, \Gamma \Rightarrow \Delta, \Lambda \rangle}$$

Inversion lemmas on P_3, P_4 and P_5 give derivations P_3', P_4' and P_5' of the sequents $\langle \Gamma \Rightarrow \Delta, \psi \rangle, \langle \Gamma \Rightarrow \Delta, \chi \rangle$ and $\langle \psi, \chi, \Pi \Rightarrow \Lambda \rangle$, respectively. These are arranged to form the final proof P' satisfying both conditions required by the lemma:

$$\dfrac{\overset{P_3'}{\langle \Gamma \Rightarrow \Delta, \psi \rangle} \qquad \dfrac{\overset{P_4'}{\langle \Gamma \Rightarrow \Delta, \chi \rangle} \qquad \overset{P_5'}{\langle \psi, \chi, \Pi \Rightarrow \Lambda \rangle}}{\langle \psi, \Pi, \Gamma \Rightarrow \Delta, \Lambda \rangle}}{\langle \Pi, \Gamma \Rightarrow \Delta, \Lambda \rangle}$$

4.2 $\varphi \equiv \psi \vee \chi$

$$\dfrac{\dfrac{\overset{P_3}{\langle \Gamma \Rightarrow \Delta, \psi \vee \chi, \psi \rangle}}{\langle \Gamma \Rightarrow \Delta, \psi \vee \chi \rangle} \qquad \dfrac{\overset{P_4}{\langle \psi, \psi \vee \chi, \Pi \Rightarrow \Lambda \rangle} \qquad \overset{P_5}{\langle \chi, \psi \vee \chi, \Pi \Rightarrow \Lambda \rangle}}{\langle \psi \vee \chi, \Pi \Rightarrow \Lambda \rangle}}{\langle \Pi, \Gamma \Rightarrow \Delta, \Lambda \rangle}$$

Use inversions on the endsequents of P_3,P_4 and P_5 to obtain derivations of these sequents without the occurrences of $\psi \vee \chi$. Then apply two cuts on the endsequents of those derivations as follows:

$$\dfrac{\dfrac{\overset{P_3'}{\langle \Gamma \Rightarrow \Delta, \psi, \chi \rangle} \qquad \overset{P_4'}{\langle \psi, \Pi \Rightarrow \Lambda \rangle}}{\langle \Pi, \Gamma \Rightarrow \Delta, \Lambda, \chi \rangle} \qquad \overset{P_5'}{\langle \chi, \Pi \Rightarrow \Lambda \rangle}}{\langle \Pi, \Gamma \Rightarrow \Delta, \Lambda \rangle}$$

4.3 $\varphi \equiv \psi \to \chi$

This is one of the two most involved cases of the entire lemma, as it combines using both the inversion lemma and IH.

$$\dfrac{\dfrac{\overset{P_3}{\langle \Gamma, \psi \Rightarrow \chi \rangle}}{\langle \Gamma \Rightarrow \Delta, \psi \to \chi \rangle} \qquad \dfrac{\overset{P_4}{\langle \psi \to \chi, \Pi_1 \Rightarrow \Lambda_1, \psi \rangle} \quad \overset{P_5}{\langle \chi, \psi \to \chi, \Pi_2 \Rightarrow \Lambda_2 \rangle}}{\langle \psi \to \chi, \Pi \Rightarrow \Lambda, \psi \rangle}}{\langle \Pi, \Gamma \Rightarrow \Delta, \Lambda \rangle}$$

First use inversion lemma on the endsequent of P_5 and the formula $\psi \to \chi$ to obtain a derivation P_5' of $\langle \chi, \Gamma_2 \Rightarrow \Delta_2 \rangle$. Now transform P so that the cut on the complex formula $\psi \to \chi$ has lower level (i.e. so that IH becomes applicable). This is done as follows:

$$\dfrac{\dfrac{\overset{P_1}{\langle \Gamma \Rightarrow \Delta, \psi \to \chi \rangle} \quad \overset{P_4}{\langle \psi \to \chi, \Pi_1 \Rightarrow \Lambda_1, \psi \rangle}}{\langle \Pi_1, \Gamma \Rightarrow \Delta, \Lambda_1, \psi \rangle} \qquad \dfrac{\overset{P_3}{\langle \Gamma, \psi \Rightarrow \chi \rangle} \quad \overset{P_5'}{\langle \chi, \Pi_2 \Rightarrow \Lambda_2 \rangle}}{\langle \psi, \Pi_2, \Gamma \Rightarrow \Lambda_2 \rangle}}{\langle \Pi, \Gamma \Rightarrow \Delta, \Lambda \rangle}$$

By IH the sequent $\langle \Pi_1, \Gamma \Rightarrow \Delta, \Lambda, \psi \rangle$ has a derivation of depth at most $k-1$, so the resulting proof P' still fits the bound. Also the cut on the complex formula has been replaced by two cuts on simpler formulas, and hence $\mathrm{r}(P') < \mathrm{r}(P)$.

4.4 $\varphi \equiv \forall x \psi$

$$\dfrac{\dfrac{\overset{P_3}{\langle \Gamma \Rightarrow \psi_x(y) \rangle}}{\langle \Gamma \Rightarrow \Delta, \forall x \psi \rangle} \qquad \dfrac{\overset{P_4}{\langle \psi_x(t), \forall x \psi, \Pi \Rightarrow \Lambda \rangle}}{\langle \forall x \psi, \Pi \Rightarrow \Lambda \rangle}}{\langle \Pi, \Gamma \Rightarrow \Delta, \Lambda \rangle}$$

Since P is regular, no variable in t is quantified or generalized in P_3 and also y is not generalized in P_3. So by substitution lemma there is a derivation $P_3[y/t]$ of the sequent $\langle \Gamma \Rightarrow \psi_x(t) \rangle$. The transformation proceeds as follows:

$$\dfrac{\overset{P_3[y/t]}{\langle \Gamma \Rightarrow \psi_x(t) \rangle} \qquad \dfrac{\overset{P_1}{\langle \Gamma \Rightarrow \Delta, \forall x \psi \rangle} \quad \overset{P_4}{\langle \psi_x(t), \forall x \psi, \Pi \Rightarrow \Lambda \rangle}}{\langle \psi_x(t), \Pi, \Gamma \Rightarrow \Delta, \Lambda \rangle}}{\langle \Pi, \Gamma \Rightarrow \Delta, \Lambda \rangle}$$

The right premise of the lower cut has a derivation with decreased cut rank by IH, whose depth is bounded by $k-1$.

4.5 $\varphi \equiv \exists x\psi$

Same pattern as with $\to$: use inversion lemma followed by rearranging the complex cut inference, so that it has lower level, and in the last step eliminate the simpler formula.

$$\dfrac{\dfrac{\overset{P_3}{\langle\, \Gamma \Rightarrow \Delta, \exists x\psi, \psi_x(t)\,\rangle}}{\langle\, \Gamma \Rightarrow \Delta, \exists x\psi\,\rangle} \qquad \overset{P_2}{\langle\, \exists x\psi, \Pi \Rightarrow \Lambda\,\rangle}}{\langle\, \Pi, \Gamma \Rightarrow \Delta, \Lambda\,\rangle}$$

By the regularity of P no variable in t is quantified or generalized in P_2, and so we can consider the following two steps: first use inversion lemma on the endsequent of P_2 and the formula $\exists x\psi$. We can choose any term s satisfying the condition of the inversion lemma, so let s be t. This yields a derivation P_2' of the sequent $\langle\, \psi_x(t), \Pi \Rightarrow \Lambda\,\rangle$. Now eliminate the complex formula $\exists x\psi$ one level above the original cut:

$$\dfrac{\dfrac{\overset{P_3}{\langle\, \Gamma \Rightarrow \Delta, \exists x\psi, \psi_x(t)\,\rangle} \qquad \overset{P_2}{\langle\, \exists x\psi, \Pi \Rightarrow \Lambda\,\rangle}}{\langle\, \Pi, \Gamma \Rightarrow \Delta, \Lambda, \psi_x(t)\,\rangle} \qquad \overset{P_2'}{\langle\, \psi_x(t), \Pi \Rightarrow \Lambda\,\rangle}}{\langle\, \Pi, \Gamma \Rightarrow \Delta, \Lambda\,\rangle}$$

IH guarantees the existence of a suitable derivation of the left premise of the lower cut. □

Lemma 2.2.2 (Iteration of cut reduction)**.** *Let P be a regular proof of S with $\mathrm{r}(P) > 0$. Then there is a proof P' of S such that* $\mathrm{r}(P') < \mathrm{r}(P)$ *and* $\mathrm{d}(P') \le 2^{\mathrm{d}(P)}$.

Proof. The idea is to successively decrease the cut rank of all subderivations whose last inference is a cut on a formula of maximal complexity among cut formulas in P, proceeding from the top to the bottom, in each case using reduction lemma. The proof proceeds by induction on $\mathrm{d}(P)$.

Let the last inference R of P be a binary rule with conclusion S and premises S_1 and S_2, whose respective subderivations are P_1 and P_2, and denote $\mathrm{d}(P_1) = l$, $\mathrm{d}(P_2) = k$. By IH S_1 and S_2 also have derivations P_1' and P_2' with $\mathrm{r}(P_1) < \mathrm{r}(P)$, $\mathrm{r}(P_2) < \mathrm{r}(P)$ and $\mathrm{d}(P_1') \le 2^k$, $\mathrm{d}(P_2') \le 2^l$.

If R is not a maximal-rank cut inference, P' is obtained from P_1' and P_2' by R, $\mathrm{d}(P') = 1 + \max\{2^l, 2^k\} < 2^{\mathrm{d}(P)}$ and $\mathrm{r}(P') < \mathrm{r}(P)$. The same holds if R is a unary rule.

Let R be a cut on a formula φ of maximal depth. Then both P_1' and P_2' have cut rank at most $\mathrm{d}(\varphi)$. Since P is a regular proof, S is a regular sequent, so we can assume by Lemma 2.2.1 that the proof P_0 obtained from P_1' and P_2' by applying cut on φ is also regular. By reduction lemma there is a derivation P' of S with $\mathrm{r}(P') < \mathrm{r}(P_0) = \mathrm{r}(P)$ and

$$\mathrm{d}(P') \le \mathrm{d}(P_1') + \mathrm{d}(P_2') \le 2^l + 2^k \le 2^{\mathrm{d}(P)},$$

because $\mathrm{d}(P) = 1 + \max\{l, k\}$. □

The previous lemma shows that the cut rank of any derivation can be decreased by 1. This procedure is iterated until the derivation contains no instances of cut.

Theorem 2.2.3 (Cut elimination for smG3i)**.** *Cut is admissible for smG3i. If a regular sequent S has a proof P containing instances of cut, there is a cut-free proof P' of S whose depth is bounded by $2^{\mathrm{d}(P)}_{\mathrm{r}(P)}$.*

Proof. By induction on $\mathrm{r}(P)$. If $\mathrm{r}(P) = 0$, the derivation is cut-free by definition. Otherwise by Lemma 2.2.2 there is a proof P_0 of S with $\mathrm{r}(P_0) < \mathrm{r}(P)$ and $\mathrm{d}(P_0)) \leq 2^{\mathrm{d}(P)}$. We can apply IH on P_0 to obtain a cut-free P' with

$$\mathrm{d}(P') \leq 2^{\mathrm{d}(P_0))}_{\mathrm{r}(P_0)} \leq 2^{2^{\mathrm{d}(P)}}_{\mathrm{r}(P)-1} = 2^{\mathrm{d}(P)}_{\mathrm{r}(P)}.$$

□

3. Aspects of cut elimination

The final chapter discusses a few selected topics directly connected to the procedure of cut elimination. I briefly mention a different method of proof than the one used in the previous chapter. After some general remarks, I move on to give one particular application of cut-free systems, the Midsequent theorem, which is already considered by Gentzen in [10].

3.1 Eliminating topmost cuts

The method of proving cut elimination used in the previous chapter is due to Tait's [19], the first paper to explicitly state an upper bound for the increase of derivation during the procedure in classical logic. Schwichtenberg in [16] works with a simplified calculus, follows the same strategy and also gives an upper bound. Since Tait's result is a corollary of a corresponding theorem stated for infinitary derivations, Schwichtenberg's paper is rather easier to follow. Besides the fact that the proof for his calculus requires fewer cases to consider, because not all connectives are considered primitive, its structure is basically the same as the proof for smG3i. The extension to full classical logic is given in [18] and the proof from the previous section is a slight modification for intuitionistic logic.

Recall that the strategy of eliminating topmost maximal-rank cuts consists in successively decreasing the cut rank of the subderivations of cuts with maximal rank from the top to the bottom. This procedure is iterated until the derivation is cut-free, so the main induction is on the cut rank. There is another more common method of eliminating cuts that goes back to Gentzen's seminal paper. In this approach we first mark *all* cut inferences in the given derivation and select a topmost one. By induction on the cut rank with a subinduction on the level of the cut inference it is shown that the cut can be removed altogether. Replace the original subderivation of the cut by the respective cut-free derivation of the same endsequent and process another topmost cut in the same way. This time the outer induction is on the number of cut inferences in the whole derivation, but the central lemma of the proof is rather similar to the cut reduction lemma for the former strategy. Let me call these two approaches of proving cut elimination Tait's and Gentzen's method, respectively.

For either strategy there are preliminaries of the *Hauptsatz* that are needed when the calculus is based on sequences. Since contraction is no longer implicit, it must either be assumed as a rule of the calculus or the logical rules must be reformulated to absorb its properties. If it is taken as an explicit rule one must somehow deal with the case when the right premise of the cut was derived by contraction on the cut formula. To this end the authors of calculi similar to LK (e.g. Kleene's G2 systems) eliminate multicut instead of cut. Another option is to ensure the calculus is closed under contraction, that each premise of the logical rules either contains the principal formula or the rule is invertible with respect to this premise. This possibility is discussed by Kleene in [12] (G3-systems) and Curry in [6] (formulation III of his LA and LC systems).[1]

[1] See chapter 1.

Under Tait's strategy of eliminating topmost maximal-rank cuts for a calculus C we have to ensure that contraction is admissible for $C +$ Cut. This is because the appeal to closure under contraction during the proof of the central lemma is conditioned by the absorption of contraction also in the rule of cut, due to the possible occurrences of less complex cut inferences in the subderivation. This is not necessary if we first fully eliminate a topmost cut before processing the others, that is if we follow Gentzen's method. Under this strategy the particular formulation of cut that I have been using throughout is perfectly sufficient. But notice that it does not conform to the above condition (by Curry) on contraction-free systems because the contexts may be different in each premise. Hence under Tait's strategy for contraction-free calculi what must be eliminated is the following context-sharing version of cut, Cut_{cs}:

$$\frac{\langle \Gamma \Rightarrow \Delta, \varphi \rangle \quad \langle \varphi, \Gamma \Rightarrow \Delta \rangle}{\langle \Gamma \Rightarrow \Delta \rangle}$$

Instances of Cut can be replaced by instances of Cut_{cs} without increasing the depth of the derivation, so the elimination of Cut is a consequence of the elimination of Cut_{cs}. The various initial settings of the *Hauptsatz* for G3-calculi, namely what strategy is followed and what version of cut is eliminated are discussed in [20] (p.100–101). In [15] all proofs of admissibility proceed by eliminating the topmost instance of the respective rule.

3.2 Miscellaneous remarks

The procedure of cut elimination differs to some extent in systems with explicit contraction and in those based on sets or those having contraction admissible. The difference is most notably manifested in case 4 in which the cut formula is principal in both premises of the cut. If the outermost symbol of the cut formula is $\rightarrow$ or $\exists$ (subcases 4.3 and 4.5) the solution requires more than the application of cut to formulas of lower depth, which is a local modification of the derivation. Since contraction is implicit in calculi based on sets the option needs to be considered when the cut formula also appears in the respective places in the premises of both the left and the right premise of the cut. A straightforward application of the method characteristic of this stage of the proof, the replacement of the complex cut by cuts on simpler formulas, does not remove the occurrences of the cut formula in these sequents. Therefore what is needed is a more global modification of their subderivations and the elimination of the complex formula is achieved by means of inversion lemmas. A similar phenomenon occurs when the system is based on sequences and contraction is absorbed in the logical rules (G3-systems).

When contraction is considered a rule of the calculus this duplication of the cut formula above the cut premises does not occur in case of logical inferences, and thus the transformations of this stage only consist of rearranging the derivation so that cuts on simpler formulas suffice to produce the endsequent. The problem appears with the cut formula being introduced into the right premise of cut by (left) contraction and the solution (besides using multicut) is still to use global inversion lemmas on suitable sequents in the subderivation of the right premise.

The most important complexity measure of proofs used in the procedure is cut rank. This was defined as the *maximal* depth among cut formulas (+1) in the derivation. Regardless whether we take the complexity of formulas to be their depth or size (the number of logical symbols, possibly plus the number of atomic formulas), the advantage of the corresponding notion of cut rank defined this way is that it is global and can be used as the central induction measure for Tait's strategy. Gentzen's original definition is restricted to the subderivation of an uppermost instance of cut (see [9], p.197). This is perfectly fine for his approach, but since this definition does not apply to the whole derivation containing multiple cut inferences, it needs to be extended in case we follow Tait's method. This is simply because the outer induction on the cut rank presupposes that this notion applies globally.

The notion of cut rank which applies to entire derivations (and is not local to particular cut inferences) that is used in the previous chapter is assumed in [16], [18] and [20]. A slight variant is also used in [15], where the depth of formulas is replaced by the weight of formulas that assigns 0 to $\perp$, 1 to atomic formulas and the weight of a complex formula is the sum of the weights of its constituents. Gentzen in [9] and Kleene in [12] define the complexity of formulas as the number of logical symbols and the cut rank (*Grad, grade*) is assigned only to one uppermost cut instance at a time.

The distinction between the notions of cut elimination and cut admissibility should by now be more transparent. The former implies the latter, but the admissibility of cut is commonly established semantically, as a consequence of the completeness of the calculus in question with respect to suitable semantics (because cut is sound both in classical and in intuitionistic logic). The term "(structural) cut elimination" is usually reserved for a procedure which shows how to gradually transform a derivation into a cut-free one.[2] This is stronger than cut admissibility because certain computational content can be extracted from this procedure, and in our case this content is manifested by establishing an upper bound for cut-free derivations. Another aspect of cut elimination which may be considered advantageous when compared to more common proofs of cut admissibility is that it consists of certain syntactic transformations of proofs and does not require the introduction of (possibly rather complicated) semantic notions.

3.3 Applications – Midsequent theorem

The following topic is an application of the *Hauptsatz* that was already considered by Gentzen in [10]. Although he took it as such, the Midsequent theorem is also a consequence of the admissibility of cut. This is because what is assumed is (only) that certain sequents have cut-free proofs. But since in contrast to cut admissibility the procedure of cut elimination provides a particular (non-deterministic) algorithm, if considered a consequence of the *Hauptsatz* the Midsequent theorem is of certain computational interest. Still the following applications are generally to be understood as applications of cut-free systems, that is systems for which cut is admissible.

[2] See [20], p.92–93.

Example 3.3.1. Let me first illustrate another application on a rather self-contained example, one that is also already considered by Gentzen (§ 2, 1.3), namely the unprovability of the law of excluded middle in LJ. The merit of demonstrating that this law does not hold in intuitionistic logic by showing that the construction of its cut-free derivation must fail lies in that the process is rather straightforward.[3] Assume that the sequent $S : \langle \Rightarrow \varphi \vee \neg\varphi \rangle$ is derivable in LJ and let P be its cut-free proof. By virtue of the subformula property every formula in P is a subformula of $\varphi \vee \neg\varphi$. Since every succedent must contain at most one formula there are only two rules that may have been used in the last step of P, weakening and R$\vee$. The three possibilities are thus the following:

$$\dfrac{\overset{P_1}{\langle \Rightarrow \rangle}}{\langle \Rightarrow \varphi \vee \neg\varphi \rangle} \qquad \dfrac{\overset{P_2}{\langle \Rightarrow \varphi \rangle}}{\langle \Rightarrow \varphi \vee \neg\varphi \rangle} \qquad \dfrac{\overset{P_3}{\langle \Rightarrow \neg\varphi \rangle}}{\langle \Rightarrow \varphi \vee \neg\varphi \rangle}$$

The empty sequent $S_1 : \langle \Rightarrow \rangle$ is underivable (see chapter 1) and so the first option is ruled out. But since φ is arbitrary the two sequents $S_2 : \langle \Rightarrow \varphi \rangle$ and $S_3 : \langle \Rightarrow \neg\varphi \rangle$ are clearly not derivable either. The formula φ could not have been introduced by weakening into S_1 and other options are excluded on account of φ being any well-formed formula (in particular it may be an atom p). Similarly for S_2 weakening could not have been used in the last step, so what remains is R$\neg$. But then its premise would be $\langle \varphi \Rightarrow \rangle$ and the pattern repeats itself.

The main topic of this section is another important consequence of the admissibility of cut, the Midsequent theorem. As the proof of this theorem is again essentially constructive it is natural to regard it as an extension of cut elimination. Its statement and proof are given by Gentzen in [10], § 2.

The theorem is given only for classical logic. This is because there is a necessary assumption in its formulation that all formulas in a certain derivable sequent are in prenex normal form. And although this assumption is unrestrictive for classical logic, in intuitionistic logic there is no way of equivalently rewriting formulas this way.[4]

Let the underlying calculus be smG3i with the following classical versions of the two critical rules R$\to$ and R$\forall$

$$\dfrac{\langle \varphi, \Gamma \Rightarrow \Delta, \psi \rangle}{\langle \Gamma \Rightarrow \Delta, \varphi \to \psi \rangle} \qquad \dfrac{\langle \Gamma \Rightarrow \Delta, \varphi_x(y) \rangle}{\langle \Gamma \Rightarrow \Delta, \forall x \varphi \rangle}$$

and the context-independent variants of R$\wedge$ and L$\vee$ [5]

$$\dfrac{\langle \Gamma \Rightarrow \Delta, \varphi \rangle \quad \langle \Pi \Rightarrow \Lambda, \psi \rangle}{\langle \Pi, \Gamma \Rightarrow \Delta, \Lambda, \varphi \wedge \psi \rangle} \qquad \dfrac{\langle \varphi, \Gamma \Rightarrow \Delta \rangle \quad \langle \psi, \Pi \Rightarrow \Lambda \rangle}{\langle \varphi \vee \psi, \Pi, \Gamma \Rightarrow \Delta, \Lambda \rangle},$$

[3] This is debatable for more complicated formulas when contraction is explicit because at any point in the bottom-up construction of a derivation it must be considered that an occurrence of a formula gets duplicated one level above.

[4] For this see e.g. [15], chapter 4.

[5] These are inter-derivable with the context-sharing versions. Their advantage in the proof of the Midsequent theorem lies in that after a permutation of two successive rule applications the contexts in the premise of the second inference in the transformed derivation may well be different. See the examples below.

Theorem 3.3.2 (Midsequent theorem)**.** *Let S be a classically derivable regular sequent whose formulas are in prenex form. Then there is a cut-free proof P of S such that all propositional inferences precede all quantifier inferences in P. There is a sequent S' in P such that all inferences above S' are propositional, all those below S' are quantifier and all formulas in S' are quantifier-free substitution instances of formulas in S. S' is called the midsequent (Mittelsequenz) of P.*

Since each quantifier inference has a single premise the part between the midsequent S' and the endsequent S is a linear thread and the derivation P can first branch above S'. Also note that both assumptions of the theorem are essential.[6] The regularity of S is necessary because the demonstration of the theorem requires a regular derivation of S and the condition that all formulas in S are in prenex form is needed to ensure that all quantifier inferences can come after all propositional inferences.

Proof. The central method of the proof is the permutation of lower propositional inferences over quantifier inferences so as to arrive at the required derivation with the lower part consisting of all the original quantifier inferences only. These permutations, the reordering of rule applications, can be correctly demonstrated on a sufficiently simplified derivation, i.e. a cut-free one. So let P be a regular cut-free derivation of S. In order to formally capture the process let us define the following measure of how many propositional inferences are yet to be permuted upwards: the order of a quantifier inference R is the number of propositional inferences between the conclusion of R and the endsequent S. The *order* $\mathrm{o}(P)$ of the derivation P is the sum of the orders of all quantifier inferences in P. The proof of the theorem is by induction on $\mathrm{o}(P)$.

First assume that P has been brought to the desired form and that $\mathrm{o}(P) = 0$. Follow the branch from S upwards until the conclusion S'' of the first propositional inference is arrived at. S'' may still contain quantified formulas because they can get implicitly contracted below. But since there are no quantifier inferences above S'' these quantified formulas must occur as contexts in at least some initial sequents (not necessarily in all of them because binary rule do not have shared contexts) and none of them is either principal or active in the subderivation P_1 of S''. So remove all occurrences of these formulas from P_1. Afterwards all propositional rules remain correctly applied and the endsequent of P_1 is the required midsequent S' containing only quantifier-free substitution instances of formulas from the endsequent S of the whole derivation P. Adjoin the lower part of this original derivation below S'. Quantifier inferences remain applicable as their active formulas remain in S'. The upper part of the resulting derivation contains only propositional inferences, S' is the borderline and the lower part contains quantifier inferences only.

In the inductive step the goal is to reduce the order of P. This stage includes several possibilities of permutations, but these are somewhat limited by the assumption of the theorem that the formulas in S are in prenex form. Firstly we only permute a propositional inference R_2 over a quantifier inference R_1 at a time, and secondly it cannot happen that the principal formula of R_1 is an active formula of R_2 because the respective principal formula of R_2 would contain a quantifier within the scope of a propositional symbol and such a formula cannot

[6] The third assumption that S is *classically* derivable is a prerequisite for the last one.

occur in the required derivation because of the subformula property. Another important thing to notice is that we start with a *regular* derivation of S. This is a necessary preparatory step which guarantees that the whenever a propositional inference is permuted upwards over a generalization rule, this rule remains applicable in the modified derivation.

Take a lowest quantifier inference R_1 in P which is immediately followed by a propositional inference R_2. Let R_1 be R$\forall$ and R_2 be R$\vee$.

$$\dfrac{\dfrac{\begin{matrix} P_1 \\ \langle \Gamma \Rightarrow \Delta, \varphi_x(y), \psi \rangle \end{matrix}}{\langle \Gamma \Rightarrow \Delta, \forall x \varphi, \psi \rangle}}{\langle \Gamma \Rightarrow \Delta, \forall x \varphi, \psi \vee \chi \rangle}$$

These two inferences are easily permuted as follows:

$$\dfrac{\dfrac{\begin{matrix} P_1 \\ \langle \Gamma \Rightarrow \Delta, \varphi_x(y), \psi \rangle \end{matrix}}{\langle \Gamma \Rightarrow \Delta, \varphi_x(y), \psi \vee \chi \rangle}}{\langle \Gamma \Rightarrow \Delta, \forall x \varphi, \psi \vee \chi \rangle}$$

The variable y is not free in χ because the original derivation is regular, and hence the rules can be permuted. Similarly if R_2 is another unary rule, L$\wedge$ or R$\rightarrow$. If R_1 is a different quantifier rule the pattern works as well. The order of R_1 is decreased by 1 and the rest follows from IH.
Let R_2 be R$\wedge$.

$$\dfrac{\dfrac{\begin{matrix} P_1 \\ \langle \Gamma \Rightarrow \Delta, \varphi_x(y), \psi \rangle \end{matrix}}{\langle \Gamma \Rightarrow \Delta, \forall x \varphi, \psi \rangle} \qquad \begin{matrix} P_2 \\ \langle \Pi \Rightarrow \Lambda, \chi \rangle \end{matrix}}{\langle \Pi, \Gamma \Rightarrow \Delta, \Lambda, \forall x \varphi, \psi \wedge \chi \rangle}$$

By the regularity of P the variable y is not free in $\Pi \cup \Lambda \cup \{\chi\}$ and the rules can be permuted as follows:

$$\dfrac{\dfrac{\begin{matrix} P_1 \\ \langle \Gamma \Rightarrow \Delta, \varphi_x(y), \psi \rangle \end{matrix} \qquad \begin{matrix} P_2 \\ \langle \Pi \Rightarrow \Lambda, \chi \rangle \end{matrix}}{\langle \Pi, \Gamma \Rightarrow \Delta, \Lambda, \varphi_x(y), \psi \wedge \chi \rangle}}{\langle \Pi, \Gamma \Rightarrow \Delta, \Lambda, \forall x \varphi, \psi \wedge \chi \rangle}$$

If some occurrences of formulas get contracted during the transformation, if for instance $\varphi_x(y) = \psi \wedge \chi$, the result follows by weakening. A similar pattern works in case R_1 precedes the other premise of L$\rightarrow$, for the other two binary propositional rules L$\vee$ and L$\rightarrow$ and (or) in case R_1 is a different quantifier rule. Again the order of R_1 is decreased and the procedure is iterated until the quantifier rules are below all the propositional rules. $\square$

Gentzen proved the Midsequent theorem for his calculus LK and applied it to a consistency proof of a particular arithmetical theory without induction in § 3. There are some differences in his proof because in contrast to smG3i both weakening and contraction are explicit and principal formulas of initial sequents can be quantified. This last point can present difficulties in the step where all permutations have already been made and the task is to remove the occurrences of the remaining quantified formulas above the "temporary" midsequent. If some of these formulas can be traced back to the principal formulas of some initial sequents, they cannot simply be deleted. Gentzen offers the following solution. When given the regular cut-free derivation, first replace all instances of initial sequents with quantified principal formulas, i.e. sequents of the form $\langle \forall x\varphi \Rightarrow \forall x\varphi \rangle$ (and analogously for $\exists$) by the following:

$$\dfrac{\dfrac{\langle \varphi_x(y) \Rightarrow \varphi_x(y) \rangle}{\langle \forall x\varphi \Rightarrow \varphi_x(y) \rangle}}{\langle \forall x\varphi \Rightarrow \forall x\varphi \rangle}$$

If the principal formula contains more quantifiers (necessarily on the outside) this method is repeated until the initial sequent contains quantifier-free formulas only. A similar problem obtains when a quantified formula is introduced by weakening. Following Gentzen's approach, Curry in [5] suggests the replacement of inferences of the form

$$\dfrac{\begin{matrix} P_1 \\ \langle \Gamma \Rightarrow \Delta \rangle \end{matrix}}{\langle \Gamma \Rightarrow \Delta, \forall x\varphi \rangle} \qquad \dfrac{\begin{matrix} P_1 \\ \langle \Gamma \Rightarrow \Delta \rangle \end{matrix}}{\langle \Gamma \Rightarrow \Delta, \exists x\varphi \rangle},$$

and similarly for left weakening, by the following inferences with the principal formula of the weakening being restricted to one that is quantifier-free.

$$\dfrac{\dfrac{\begin{matrix} P_1 \\ \langle \Gamma \Rightarrow \Delta \rangle \end{matrix}}{\langle \Gamma \Rightarrow \Delta, \varphi_x(y) \rangle}}{\langle \Gamma \Rightarrow \Delta, \forall x\varphi \rangle} \qquad \dfrac{\dfrac{\begin{matrix} P_1 \\ \langle \Gamma \Rightarrow \Delta \rangle \end{matrix}}{\langle \Gamma \Rightarrow \Delta, \varphi_x(y) \rangle}}{\langle \Gamma \Rightarrow \Delta, \exists x\varphi \rangle}$$

In both cases y is a fresh variable not occurring anywhere in the original derivation P. Analogously, whenever there are more nested quantifiers on the outside of the formula introduced by weakening, one can always use weakening only on the respective quantifier-free instance containing fresh variables and apply several quantifier inferences (in a suitable order) afterwards.

Curry's paper [5] contains an interesting hint on how to deal with the case when the three binary rules in the underlying calculus have shared contexts. This presents some difficulty in that if the following two inferences

$$\dfrac{\dfrac{\begin{matrix} P_1 \\ \langle \psi \Rightarrow \varphi_x(y) \rangle \end{matrix}}{\langle \psi \Rightarrow \forall x\varphi \rangle} \qquad \begin{matrix} P_2 \\ \langle \chi \Rightarrow \forall x\varphi \rangle \end{matrix}}{\langle \psi \vee \chi \Rightarrow \forall x\varphi \rangle}$$

are permuted the rule L$\vee$ is generally not applicable due to the different contexts in the succedents of its premises. He proposes the following approach (p. 247): first permute all quantifier rules below all one-premise rules until there is no such possibility left. Then select a highest binary rule R_2 with a quantifier rule R_1 above one of its premises. If R_1 precedes R_2 on both sides these two rules can be permuted. (In the example above this corresponds to the case with the last inference of P_2 being R$\forall$.) Otherwise there is an inference R_1' like R_1 on the other side higher up in the derivation (on account of the preliminary solution to the cases when a quantified formula is either principal in an initial sequent or it is introduced by weakening). The inferences between R_1' and R_2 may only be quantifier or structural since otherwise R_2 is not the highest binary rule with the respective property or not all possible permutations of unary propositional rules with quantifier rules have been made. Moreover these intermediate quantifier inferences do not relate to the principal formula of R_1' since otherwise the contexts in the premises of R_2 would be different. Now Curry proposes that we simply permute R_1' below these quantifier inferences down up to the respective premise of R_2 after which R_1 and R_2 can be permuted. This procedure decreases the order of P and its iteration brings P to the required form.

It is only the assumption of the last step in Curry's procedure that is somewhat mysterious. This is that in a regular derivation quantifier inferences relating to different formulas can be permuted. Consider the following derivation consisting of two inferences, L$\forall$ and R$\forall$:

$$\dfrac{\dfrac{\langle\, P(z) \;\Rightarrow\; P(z)\,\rangle}{\langle\, \forall x P(x) \;\Rightarrow\; P(z)\,\rangle}}{\langle\, \forall x P(x) \;\Rightarrow\; \forall y P(y)\,\rangle}$$

This is a regular derivation, no variable occurs both free and bound and the eigenvariable z occurs only above the generalization rule. It is evident that we cannot interchange these two inferences as the eigenvariable condition would be violated. And if L$\forall$ is the rule R_1' that needs to be brought down the derivation over a premise of a binary rule with shared contexts such straightforward interchanges of successive quantifier inferences do not apply in this case, and one needs to come up with a different way to guarantee that all contexts remain the same after some permutations are made. This is also why I chose the context-independent variants of the three binary rules.

Conclusion

The central part of the thesis provides a detailed proof of cut elimination for an intuitionistic multi-succedent calculus based on pairs of sets of formulas. The proof strategy consists in iteratively decreasing the cut rank of a highest cut inference among those with maximal cut rank. One of the advantage of this method of proving Gentzen's *Hauptsatz* is that it can be regarded as a slight modification of the corresponding proof for classical logic whose systems do not need to contain restricted inference rules. Common approaches of proving cut elimination for intuitionistic systems include working with single-succedent systems based on (possibly unordered) sequences and following a different strategy of fully eliminating a topmost cut inference at a time. Calculi that work with sequences of formulas allow for some diversity, most importantly regarding the structural rule of contraction. Some general observations and historical remarks regarding this rule together with comments on the aspects of allowing arbitrary number of formulas in the succedents of sequents are given in the first chapter. In the final chapter I briefly discuss the alternative strategy of proving cut elimination and the second part concerns one of its important corollaries, the Midsequent theorem.

Bibliography

[1] P. Bernays. Reviews - Oiva Ketonen. Untersuchungen zum Prädikatenkalkül. *Journal of Symbolic Logic*, 10(4):127–130, 1945.

[2] Samuel R. Buss. An introduction to proof theory. In Samuel R. Buss, editor, *Handbook of Proof Theory*, number 137 in Studies in Logic and the Foundations of Mathematics, chapter I, pages 1–78. Elsevier, 1998.

[3] H. B. Curry. A note on the reduction of Gentzen's calculus LJ. *Bulletin of the American Mathematical Society*, 45(4):288–293, 1939.

[4] H. B. Curry. *A Theory of Formal Deducibility*. Notre Dame mathematical lectures. University of Notre Dame, 1950.

[5] H. B. Curry. The permutability of rules in the classical inferential calculus. *J. Symb. Log.*, 17:245–248, 1952.

[6] H. B. Curry. *Foundations of Mathematical Logic*. Dover Publications, 1963.

[7] A.G. Dragalin. *Mathematical Intuitionism: Introduction to Proof Theory*. Translations of mathematical monographs. American Mathematical Society, 1988.

[8] R. Dyckhoff. Dragalin's proofs of cut-admissibility for the intuitionistic sequent calculi G3i and G3i'. University of St Andrews Report CS/97/8, 1997.

[9] G. Gentzen. Untersuchungen über das logische Schließen I. *Mathematische Zeitschrift*, 35:176–210, 1935.

[10] G. Gentzen. Untersuchungen über das logische Schließen II. *Mathematische Zeitschrift*, 39:405–431, 1935.

[11] A. Heyting. *Intuitionism: An Introduction*. Studies in logic and the foundations of mathematics. North-Holland, 1956.

[12] S.C. Kleene. *Introduction to Metamathematics*. North-Holland, 1952.

[13] Ivo Kylar. Eliminace řezu v klasické predikátové logice (Cut elimination in classical predicate logic). Master's thesis, Dept. of Logic, School of Arts, Charles Univ. in Prague, 2000.

[14] S. Maehara. Eine Darstellung der intuitionistischen Logik in der klassischen. *Nagoya Mathematical Journal*, 7:45–64, 1954.

[15] S. Negri, J. von Plato, and A. Ranta. *Structural Proof Theory*. Cambridge University Press, 2001.

[16] H. Schwichtenberg. Proof theory: Some applications of cut-elimination. In Jon Barwise, editor, *Handbook of Mathematical Logic*, volume 90 of *Studies in Logic and the Foundations of Mathematics*, pages 867–895. Elsevier, 1977.

[17] K. Schütte. Schlußweisen-Kalküle der Prädikatenlogik. *Mathematische Annalen*, 122:47–65, 1950/51.

[18] V. Švejdar. *Logika: neúplnost, složitost a nutnost* (*Logic: Incompleteness, Complexity, and Necessity*). Academia, Praha, 2002.

[19] W. Tait. Normal derivability in classical logic. In Jon Barwise, editor, *The Syntax and Semantics of Infinitary Languages*. Berlin: Springer, 1968.

[20] A.S. Troelstra and H. Schwichtenberg. *Basic Proof Theory*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 2000.

[21] J. von Plato. A proof of Gentzen's Hauptsatz without multicut. *Archive for Mathematical Logic*, 40(1):9–18, 2001.

[22] J. von Plato. Rereading Gentzen. *Synthese*, 137(1-2):195–209, 2003.