Article Submitted to Journal of Symbolic Computation

Towards a Clausal Analysis of Cut-Elimination*

Matthias Baaz[1] and Alexander Leitsch[2]

[1]*Institut für Algebra und Computermathematik, TU-Vienna, Wiedner Hauptstraße 8-10, 1040 Vienna, Austria*

[2]*Institut für Computersprachen, TU-Vienna, Favoritenstraße 9, 1040 Vienna, Austria*

Abstract

In this paper we show that a large class of cut-elimination methods can be analyzed by clause terms representing sets of characteristic clauses extractable from the original proof. Every reduction step of a cut-elimination procedure defines an operation on the corresponding clause term. Using this formal framework we prove that the methods of Gentzen and Tait and, more generally, every method based on a specific set of cut-reduction rules $\mathcal{R}$, yield a resolution proof which is subsumed by a resolution proof of the characteristic clause set. As a consequence we obtain that CERES (a resolution based method of cut-elimination) is never inferior to *any* method based on $\mathcal{R}$. On the other hand we show that CERES is not optimal in general; instead there exist cut-reduction rules which efficiently simplify the set of characteristic clauses and thus produce much shorter proofs. Further improvements and pruning methods could thus be obtained by a structural (syntactic) analysis of the characteristic clause terms.

1. Introduction

Cut elimination introduced by Gentzen (6) is one of the most famous procedures of logic. The removal of cuts corresponds to the elimination of intermediate statements (lemmas) from proofs rendering a proof which is analytic in the sense, that all statements in the proof are subformulas of the result. Therefore, the proof of a combinatorial statement is converted into a purely combinatorial proof. In this way, Girard has shown, that the Fürstenberg-Weiss proof of Van der Waerden's theorem on partitions can be transformed into Van der Waerden's original

*supported by the Austrian Research Fund (FWF), proj.nr. P16264-N05

elementary proof. Cut elimination is therefore an essential tool for the analysis of proofs, especially to make implicit parameters explicit. Cut free derivations allow for

- the extraction of Herbrand disjunctions, which can be used to establish bounds on existential quantifiers (e.g. Luckhardt's analysis of the Theorem of Roth (9)).
- the construction of interpolants, which allow for the replacement of implicit definitions by explicit definitions according to Beth's Theorem.
- the calculation of generalized variants of the end formula (5).

This paper focuses on the computational properties of cut elimination in first order logic and is based on the method of cut-elimination by resolution (CERES), which is designed to refute the ancestral formulas of the cut formulas directly by resolution and to compose the resolution derivation and the remaining proof parts to a derivation with atomic cuts (4) (the presence of atomic cuts is not harmful to the constructions mentioned in the last paragraph, and atomic cuts can be eliminated with at most exponential expense). The main result of the paper is the theorem, that cut-elimination by resolution provides a lower bound for cut-elimination methods based on stepwise reduction of the cut formulas (e.g. the well-known original method of Gentzen and the method of Schütte-Tait). The method of proof consists in a symbolic representation of the ancestral clauses of the cut formulas , and it is shown, that the clause set of the original derivation subsumes the clause sets of the derivations with stepwise reduced cuts.

2. Definitions and Notation

Definition 2.1 (position): *We define the positions within terms inductively:*

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

Positions serve the purpose to locate subterms in a term and to perform replacements on subterms. A subterm s of t is just a term with $t.\nu = s$ for some position ν in t. Let $t.\nu = s$; then $t[r]_\nu$ is the term t after replacement of s on position ν by r, in particular $t[r]_\nu.\nu = r$. Let P be a set of positions in t; then $t[r]_P$ is defined from t by replacing all $t.\nu$ with $\nu \in P$ by r.
Positions in formulas can be defined in the same way (the simplest way is to consider all formulas as terms).
Substitutions are defined as usual (functions from the set of variables to the set of terms). If σ is a substitution with $\sigma(x_i) = t_i$ for $x_i \neq t_i$ $(i = 1, \ldots, n)$ and $\sigma(v) = v$ for $v \notin \{x_1, \ldots, x_n\}$ then we denote σ by $\{x_1 \leftarrow t_1, \ldots, x_n \leftarrow t_n\}$. Substitutions are written in postfix, i.e. we write $F\sigma$ instead of $\sigma(F)$.

Let F be a term or a formula. We write $F(x)$ to indicate (potential) free occurrences of the variable x in F. Let t be an arbitrary term, then $F(x/t)$ stands for $F[t]_P$ where $P = \{\nu \mid F.\nu = x\}$.

Definition 2.2 (complexity of formulas): *If F is a formula in* PL *then the complexity $comp(F)$ is the number of logical symbols occurring in F. Formally we define*

$comp(F) = 0$ *if F is an atomic formula,*

$comp(F) = 1 + comp(A) + comp(B)$ *if $F \equiv A \circ B$ for $\circ \in \{\wedge, \vee, \rightarrow\}$,*

$comp(F) = 1 + comp(A)$ *if $F \equiv \neg A$ or $F \equiv (Qx)A$ for $Q \in \{\forall, \exists\}$.*

Definition 2.3 (sequent): *A sequent is an expression of the form $\Gamma \vdash \Delta$ where Γ and Δ are finite multisets of* PL*-formulas (i.e. two sequents $\Gamma_1 \vdash \Delta_1$ and $\Gamma_2 \vdash \Delta_2$ are considered equal if the multisets represented by Γ_1 and by Γ_2 are equal and those represented by Δ_1, Δ_2 are also equal). $\vdash$ is called the* empty *sequent.* $\sharp$

Multiset union within the sequents is just denoted by comma: if $S = \Gamma \vdash \Delta$ where Γ is the multiset union of Γ_1, Γ_2 and Δ is the multiset union of Δ_1, Δ_2 then we write $S = \Gamma_1, \Gamma_2 \vdash \Delta_1, \Delta_2$. If A is a formula then A^n denotes the multiset containing A n-times. E.g. we may write $\vdash A^3$ for $\vdash A, A, A$.

Definition 2.4 (composition of sequents): *If $S = \Gamma \vdash \Delta$ and $S' = \Pi \vdash \Lambda$ we define the composition of S and S' by $S \circ S'$, where $S \circ S' = \Gamma, \Pi \vdash \Delta, \Lambda$.* $\sharp$

Definition 2.5 (subsequent): *Let S, S' be sequents. We define $S' \sqsubseteq S$ if there exists a sequent S'' with $S' \circ S'' = S$ and call S' a* subsequent *of S.* $\sharp$

Definition 2.6 (the calculus **LK**): *In the rules of* **LK** *we always mark the auxiliary formulas (i.e. the formulas in the premiss(es) used for the inference) and the principal (i.e. the inferred) formula using different marking symbols. Thus, in our definition, $\wedge$-introduction to the right takes the form*

$$\frac{\Gamma \vdash A^+, \Delta \qquad \Gamma \vdash \Delta, B^+}{\Gamma \vdash A \wedge B^*, \Delta}$$

We usually avoid markings by putting the auxiliary formulas at the leftmost position in the antecedent of sequents and in the rightmost position in the consequent of sequents. The principal formula mostly is identifiable by the context. Thus the rule above will be written as

$$\frac{\Gamma \vdash \Delta, A \qquad \Gamma \vdash \Delta, B}{\Gamma \vdash \Delta, A \wedge B}$$

Basically we use Gentzen's version of **LK** *(see (6)) adapted to the multiset structure for sequents. For simplification we do not include implication: as we consider classical logic only there exists a polynomial cut-homomorphic transformation translating arbitrary* **LK***-proofs into proofs in negation normal form (see (3)). By the definition via multisets we do not need the exchange rules.*

- *The logical rules for $\wedge$-introduction:*

$$\frac{A, \Gamma \vdash \Delta}{A \wedge B, \Gamma \vdash \Delta} \wedge : l1 \qquad \frac{B, \Gamma \vdash \Delta}{A \wedge B, \Gamma \vdash \Delta} \wedge : l2 \qquad \frac{\Gamma \vdash A \quad \Gamma \vdash B}{\Gamma \vdash \Delta, A \wedge B} \wedge : r$$

- *The logical rules for $\vee$-introduction:*

$$\frac{A, \Gamma \vdash \Delta \quad B, \Gamma \vdash \Delta}{A \vee B, \Gamma \vdash \Delta} \vee : l \qquad \frac{\Gamma \vdash \Delta, A}{\Gamma \vdash \Delta, A \vee B} \vee : r1 \qquad \frac{\Gamma \vdash \Delta, B}{\Gamma \vdash \Delta, A \vee B} \vee : r2$$

- *The logical rules for $\neg$-introduction:*

$$\frac{\Gamma \vdash \Delta, A}{\neg A, \Gamma \vdash \Delta} \neg : l \qquad \frac{A, \Gamma \vdash \Delta}{\Gamma \vdash \Delta, \neg A} \neg : r$$

- *The logical rules for $\forall$-introduction:*

$$\frac{A(x/t), \Gamma \vdash \Delta}{(\forall x)A(x), \Gamma \vdash \Delta} \forall : l_{\dagger} \qquad \frac{\Gamma \vdash \Delta, A(x/y)}{\Gamma \vdash \Delta, (\forall x)A(x)} \forall : r_{\ddagger}$$

- *T conditions for $\exists : l$ are these for $\forall : r$, and similarly for $\exists : r$ and $\forall : l$):*

$$\frac{A(x/y), \Gamma \vdash \Delta}{(\exists x)A(x), \Gamma \vdash \Delta} \exists : l \qquad \frac{\Gamma \vdash \Delta, A(x/t)}{\Gamma \vdash \Delta, (\exists x)A(x)} \exists : r$$

- *The structural rules of* weakening *(Π is an arbitrary multiset of formulas):*

$$\frac{\Gamma \vdash \Delta}{\Gamma \vdash \Delta, \Pi} w : r \qquad \frac{\Gamma \vdash \Delta}{\Pi, \Gamma \vdash \Delta} w : l$$

- *The structural rules of* contraction *(for $i \in \{1, \ldots, k\}$ the A_i are formulas and $n_i \geq 2$):*

$$\frac{A_1^{n_1}, \ldots A_k^{n_k}, \Gamma \vdash \Delta}{A_1, \ldots A_k, \Gamma \vdash \Delta} c : l \qquad \frac{\Gamma \vdash \Delta, A_1^{n_1}, \ldots A_k^{n_k}}{\Gamma \vdash \Delta, A_1, \ldots A_k,} c : r$$

- *Let A be a formula and $n, m \geq 1$. Then the* cut rule *is defined as*

$$\frac{\Gamma \vdash \Delta, A^m \quad A^n, \Pi \vdash \Lambda}{\Gamma, \Pi \vdash \Delta, \Lambda} cut(A)$$

  If A does not occur in Π, Δ then the cut is called a mix.

$^\dagger t$ is an arbitrary term containing only free variables.

$^\ddagger y$ is a free variable which may not occur in Γ, Δ. y is called an eigenvariable.

Definition 2.7 (LK-derivation): *An* **LK**-derivation *is defined as a directed tree where the nodes are occurrences of sequents and the edges are defined according to the rule applications in* **LK** *(they are directed from the root to the leaves). The root is the occurrence of the end-sequent. The leaves must be occurrences of* atomic *sequents. Note that we do not require the leaves being axiom sequents of the form $A \vdash A$.*
Let $\mathcal{A}$ be the set of sequents occurring at the leaves of an **LK***-derivation ψ and S be the sequent occurring at the root (called the* end-sequent*). Then we say that ψ is an* **LK**-derivation of S from $\mathcal{A}$ *(notation $\mathcal{A} \vdash_{LK} S$). Note that, in general, complete cut-elimination is only possible in* **LK***-proofs, where the leaves are axioms. But this causes no troubles as we are not interested in the elimination of atomic cuts.*
We write

$$\begin{array}{c}(\psi)\\ S\end{array}$$

to express that ψ is a derivation with end sequent S. ♯

Paths in an **LK**-derivation ψ, connecting sequent occurrences in ψ, are defined in the traditional way; a *branch* in ψ is a path starting in the end sequent. We use the terms "predecessor" and "successor" in the intuitive sense (i.e. contrary to the direction of edges in the tree): If there exists a path from S_1 to S_2 then S_2 is called a *predecessor* of S_1. The successor relation is defined in a analogous way. E.g. every initial sequent is a predecessor of the end sequent.

Definition 2.8 (subderivation): *A* position ν *in an* **LK***-derivation is defined in the same way as for terms (formally we may consider a derivation as a term). Here the positions can be identified with the nodes in the derivation tree. If there exists a position ν with $\varphi.\nu = \psi$ then we call ψ a* subderivation *of φ. In the same way we write $\varphi[\rho]_\nu$ for the deduction φ after the replacement of $\varphi.\nu$ by ρ on the position ν in φ. The sequent occurring at the position ν is denoted by $S(\nu)$.*♯

The *depth* of a position ν (denoted by depth(ν)) is defined as the depth of the node ν in the derivation tree.

Definition 2.9 (regularity): *An* **LK***-derivation φ is called* regular *if the following condition holds: Let μ, ν two independent nodes (i.e. both are not predecessors of each other) and y be an eigenvariable occurring in $\varphi.\mu$; then y does not occur in $\varphi.\nu$.* ♯

There exists a straightforward transformation from **LK**-derivations into regular ones. From now on we assume, without mentioning the fact explicitly, that all **LK**-derivations we consider are regular.
The formulas in sequents on the branch of a deduction tree are connected by a so-called *ancestor relation.* Indeed if A occurs in a sequent S and A is marked as

principal formula of an, let us say a binary, inference on the sequents S_1, S_2, then the auxiliary formulas in S_1, S_2 are *immediate ancestors* of A (in S). If A occurs in S_1 and is not an auxiliary formula of an inference then A occurs also in S; in this case A in S_1 is also an immediate ancestor of A in S. The case of unary rules is analogous. General ancestors are defined via reflexive and transitive closure of the relation.

Let ν be a node in φ and let S' be a subsequent of $S(\mu)$ for a successor μ of ν). Then we write $S(\nu,(S',\mu))$ for the subsequent of S consisting of formulas which are ancestors of formulas in S' (at μ). Let Ω be a set of (S',μ) with $S' \sqsubseteq S(\mu)$ for successors μ of ν; then $S(\nu,\Omega)$ is the composition of all $S(\nu,\omega)$ for $\omega \in \Omega$. $S(\nu,\Omega)$ is just the subsequent of S consisting of ancestors of some of the formulas in some successors μ.
If Ω consists just of the mix formulas of mixes which occur "below" ν then $S(\nu,\Omega)$ is the subsequent consisting of all formulas which are ancestors of a mix. These subsequents are crucial for the definition of the characteristic set of clauses and of the method CERES in Section 5.

Definition 2.10: *The* length *of a proof ω is defined by the number of nodes in ω and is denoted by $l(\omega)$.* $\sharp$

Gentzen's famous proof of the cut-elimination property of **LK** is based on a double induction on rank and grade of mixes.

Definition 2.11 (cut/mix derivation): *Let ψ be an **LK**-derivation of the form*

$$\frac{\begin{array}{cc}(\psi_1) & (\psi_2)\\ \Gamma_1 \vdash \Delta_1 & \Gamma_2 \vdash \Delta_2\end{array}}{\Gamma_1, \Gamma_2^* \vdash \Delta_1^*, \Delta_2}\ cut(A)$$

Then ψ is called a cut-derivation*. If the cut is a mix we speak about a* mix-derivation*. Let ψ be a mix-derivation. Then we define the* grade *of ψ as $comp(A)$; the* left-rank *of ψ ($\mathrm{rank}_l(\psi)$) is the maximal number of nodes in a branch in ψ_1 s.t. A occurs in the consequent of a predecessor of $\Gamma_1 \vdash \Delta_1$. If A is "produced" in the last inference of ψ_1 then the left-rank of ψ is 1. The* right-rank *($\mathrm{rank}_r(\psi)$) is defined in an analogous way. The* rank *of ψ is the sum of right-rank and left-rank, i.e.* $\mathrm{rank}(\psi) = \mathrm{rank}_l(\psi) + \mathrm{rank}_r(\psi)$. $\sharp$

Definition 2.12 (clause): *A* clause *is an atomic sequent, i.e. a sequent of the form $\Gamma \vdash \Delta$, where Γ and Δ are multisets of atoms.* $\sharp$

Definition 2.13 (resolvent): *Let $C = \Gamma \vdash \Delta, A^m$ and $D = B^n, \Pi \vdash \Lambda$ s.t. C and D are variable-disjoint, $n, m \geq 1$, and σ be a most general unifier of $\{A, B\}$ (i.e. a most general substitution with $A\sigma = B\sigma$). Then the clause*

$\Gamma\sigma, \Pi\sigma \vdash \Delta\sigma, \Lambda\sigma$

is called a resolvent *of C and D.* $\sharp$

Definition 2.14 (p-resolvent): *Let* $C = \Gamma \vdash \Delta, A^m$ *and* $D = A^n, \Pi \vdash \Lambda$ *with* $n, m \geq 1$. *Then the clause*

$\Gamma, \Pi \vdash \Delta, \Lambda$

is called a p-resolvent *of* C *and* D. $\sharp$

Remark: Note that the p-resolution rule is nothing else than atomic cut. $\sharp$

Definition 2.15 (resolution deduction): *A deduction tree having clauses as leaves and resolution, contraction and weakening as rules is called a* resolution deduction. *If, instead of resolution, we have p-resolution as (the only binary) rule then we call the deduction a* p-resolution deduction. $\sharp$

Remark: A p-resolution deduction γ is an **LK**-deduction with atomic sequents and structural rules only, i.e., the only rules in γ are cut, contraction and weakening. $\sharp$

Let Γ be a multiset of atoms; then set(Γ) denotes the set of atoms occurring in Γ.

Definition 2.16 (subsumption): *Let* $C{:}\,\Gamma \vdash \Delta$ *and* $D{:}\,\Pi \vdash \Lambda$ *be clauses. Then* C subsumes D *(*$C \leq_{ss} D$*) if there exists a substitution* θ *s.t.*

$$\begin{array}{rcl} \mathrm{set}(\Gamma)\theta & \subseteq & \mathrm{set}(\Pi) \ \ \textit{and} \\ \mathrm{set}(\Delta)\theta & \subseteq & \mathrm{set}(\Lambda).\ \sharp \end{array}$$

We extend the relation $\leq_{ss}$ *to sets of clauses* $\mathcal{C}$, $\mathcal{D}$ *in the following way:* $\mathcal{C} \leq_{ss} \mathcal{D}$ *if for all* $D \in \mathcal{D}$ *there exists a* $C \in \mathcal{C}$ *s.t.* $C \leq_{ss} D$. $\sharp$

The subsumption relation can also be extended to resolution deductions.

Definition 2.17: *Let* γ *and* δ *be resolution deductions. We define* $\gamma \leq_{ss} \delta$ *by induction on the number of nodes in* δ*:*

If δ *consists of a single node labelled with a clause* D *then* $\gamma \leq_{ss} \delta$ *if* γ *consists of a single node labelled with* C *and* $C \leq_{ss} D$.

Let δ *be*

$$\frac{\begin{array}{cc}(\delta_1) & (\delta_2) \\ D_1 & D_2\end{array}}{D}\ R$$

and γ_1 *be a deduction of* C_1 *with* $\gamma_1 \leq_{ss} \delta_1$, γ_2 *be a deduction of* C_2 *with* $\gamma_2 \leq_{ss} \delta_2$. *Then we distinguish the following cases:*

$C_1 \leq_{ss} D$*: then* $\gamma_1 \leq_{ss} \delta$.

$C_2 \leq_{ss} D$*: then* $\gamma_2 \leq_{ss} \delta$.

Otherwise let C be resolvent of C_1 and C_2 and $\gamma =$

$$\frac{\begin{array}{cc}(\gamma_1) & (\gamma_2) \\ C_1 & C_2\end{array}}{C}\ R$$

Then $\gamma \leq_{ss} \delta$. ♯

PROPOSITION 2.1: *Let $\mathcal{C}, \mathcal{D}$ be sets of clauses with $\mathcal{C} \leq_{ss} \mathcal{D}$ and let δ be a resolution deduction from $\mathcal{D}$. Then there exists a resolution deduction γ from $\mathcal{C}$ s.t. $\gamma \leq_{ss} \delta$.*

Proof: By Lemma 4.2.1 in (8) and by Definition 2.17. ◇

3. Cut-Reduction Rules

Traditional cut-elimination methods, like those of Gentzen (6) and Tait (14), can be formalized as a reduction method consisting of rank- and grade reductions on **LK**-deductions. The methods of Gentzen and Tait essentially differ in the selection of a sub-derivation to be reduced. Both methods can be formalized as refinements of a proof rewriting system based on a set of reduction rules $\mathcal{R}$ defined in the Appendix. The set $\mathcal{R}$ is extracted from Gentzen's proof of cut-elimination. A refinement of $\mathcal{R}$ can be defined simply as a sub-relation of $\mathcal{R}$. Mathematically $\mathcal{R}$ is a set of pairs of **LK**-derivations.
As in Gentzen's proof we assume that all cuts in a derivation are actually mixes. This assumption does not affect the generality of our results. Indeed there is a simple (and linear) transformation of cuts into mixes (a cut can be simulated by a mix and at most two weakenings), which can be applied prior to cut-elimination.

DEFINITION 3.1: *Let $>$ be a binary relation on* **LK***-derivations. We say that $>$ is* based on $\mathcal{R}$ *if $> \subseteq \mathcal{R}$. For $(\psi, \chi) \in \mathcal{R}$ we write $\psi >_{\mathcal{R}} \chi$ and $\psi > \chi$ for $(\psi, \chi) \in >$. ♯*

DEFINITION 3.2 (REDUCTION): *Let ψ, χ be* **LK***-derivations s.t. $\psi >_{\mathcal{R}} \chi$ for the set of Rules $\mathcal{R}$ defined in the Appendix. Let φ be an* **LK***-derivation with $\varphi.\nu = \psi$ for a node ν in φ. Then we define $\varphi >_{\mathcal{R}} \varphi[\chi]_\nu$ (i.e. $>_{\mathcal{R}}$ is closed under contexts). ♯*

The reduction relation defined by Gentzen's proof is a a subrelation of $\mathcal{R}$. Indeed only mix-derivations which do not contain other non-atomic mixes may be reduced. Note that Gentzen's and Tait's methods are modified, as only non-atomic mixes are eliminated.

DEFINITION 3.3 (GENTZEN REDUCTION): *We define $\psi >_G \chi$ if $\psi >_{\mathcal{R}} \chi$ and ψ is a mix-derivation with a single non-atomic mix only – which is the last inference. $>_G$ is extended like $>_{\mathcal{R}}$: $\varphi >_G \varphi'$ if $\varphi' = \varphi[\chi]_\nu$ and $\varphi.\nu >_G \chi$. ♯*

Obviously $>_G$ is based on $\mathcal{R}$. In case of Tait reduction only sub-derivations with formulas of maximal complexity may be reduced.

Definition 3.4 (Tait reduction): *We define $\varphi >_T \varphi'$ if the following conditions are fulfilled:*

(1) There exists a node ν in φ s.t. $\varphi.\nu$ is a mix-derivation with a maximal mix formula (i.e. if the mix formula of the last mix in $\varphi.\nu$ is A then $comp(B) \leq comp(A)$ for all other mix formulas B in φ).

(2) $\varphi' = \varphi[\chi]_\nu$ for an **LK***-derivation χ with $\varphi.\nu >_{\mathcal{R}} \chi$.* $\sharp$

Like $>_G$ also $>_T$ is based on $\mathcal{R}$. The end products of cut-reduction are **LK**-derivations with atomic mixes only. These derivations are our normal forms.

Definition 3.5 (atomic-cut normal form): *Let $>$ be a cut-reduction relation based on $\mathcal{R}$. Then an* **LK***-deduction ψ is in* atomic-cut normal form *(ACNF) w.r.t. $>$ if there exists no χ s.t. $\psi > \chi$. Let $>^*$ be the reflexive and transitive closure of $>$. We say that ψ is an ACNF of φ if ψ is in ACNF and $\varphi >^* \psi$. Any method which transforms* **LK***-proofs into ACNFs is called an* AC-normalization. $\sharp$

It is easy to verify that for $>_{\mathcal{R}}$, $>_G$ and $>_T$ all normal forms are **LK**-proofs without non-atomic cuts.

Remark: Let ψ be an **LK**-derivation of a sequent S from a set of sequents $\mathcal{A}$ and ψ be in ACNF. If the set $\mathcal{A}$ is closed under cut then there exists also a cut-free derivation of S from $\mathcal{A}$. $\sharp$

4. Clause Terms

In (4) we defined the concept of characteristic clause set corresponding to an **LK**-proof. This set is the central tool for defining cut-elimination by resolution (CERES). In our analysis in Section 6 we do not only need the set of clauses, but also the way it is constructed. This leads us to the definition of clause terms representing sets of clauses.

Definition 4.1 (clause term):

- *(Finite) sets of clauses are clause terms.*
- *If X, Y are clause terms then $X \oplus Y$ is a clause term.*
- *If X, Y are clause terms then $X \otimes Y$ is a clause term.* $\sharp$

Definition 4.2: *We define a mapping $|\;|$ from clause terms to sets of clauses in the following way:*

$$\begin{array}{rcl} |\mathcal{C}| & = & \mathcal{C} \ \ \textit{for sets of clauses } \mathcal{C}, \\ |X \oplus Y| & = & |X| \cup |Y|, \\ |X \otimes Y| & = & |X| \times |Y|, \end{array}$$

where $\mathcal{C} \times \mathcal{D} = \{C \circ D \mid C \in \mathcal{C}, D \in \mathcal{D}\}$. ♯

We define clause terms to be equivalent if the corresponding sets of clauses are equal, i.e. $X \sim Y$ iff $|X| = |Y|$.
Clause terms are binary trees whose nodes are finite sets of clauses. Therefore term occurrences are defined in the same way as for ordinary terms. When speaking about occurrences in clause terms we only consider nodes in this term tree, but not occurrences within the sets of clauses on the leaves. In contrast we consider the internal structure of leaves in the concept of substitution:

Definition 4.3: *Let θ be a substitution. We define the application of θ to clause terms as follows:*

$$\begin{array}{rcl} X\theta & = & \mathcal{C}\theta \ \ \textit{if } X = \mathcal{C} \textit{ for sets of clauses } \mathcal{C}, \\ (X \oplus Y)\theta & = & X\theta \oplus Y\theta, \\ (X \otimes Y)\theta & = & X\theta \otimes Y\theta. \ \sharp \end{array}$$

There are four binary relations on clause terms which will play a important role in the proof of our main result on cut-reduction.

Definition 4.4: *Let X, Y be clause terms. We define*

$X \subseteq Y$ iff $|X| \subseteq |Y|$,

$X \sqsubseteq Y$ iff for all $C \in |Y|$ there exists a $D \in |X|$ s.t. $D \sqsubseteq C$,

$X \leq_s Y$ iff there exists a substitution θ with $X\theta = Y$, [§]

$X \leq_{ss} Y$ iff $|X| \leq_{ss} |Y|$. ♯

The operators $\oplus$ and $\otimes$ are compatible with the relations $\subseteq$ and $\sqsubseteq$. This is formally proved in the following lemmas.

Lemma 4.1: *Let X, Y, Z be clause terms and $X \subseteq Y$. Then*

(1) $X \oplus Z \subseteq Y \oplus Z$,

(2) $Z \oplus X \subseteq Z \oplus Y$,

(3) $X \otimes Z \subseteq Y \otimes Z$,

(4) $Z \otimes X \subseteq Z \otimes Y$.

[§]Note that $\leq_s$ is defined directly on the syntax of clause terms, and not via the semantics.

Proof: (2) follows from (1) because $\oplus$ is commutative, i.e. $X \oplus Z \sim Z \oplus X$. The cases (3) and (4) are analogous. Thus we only prove (1) and (3).

(1) $|X \oplus Z| = |X| \cup |Z| \subseteq |Y| \cup |Z| = |Y \oplus Z|$.

(3) Let $C \in |X \otimes Z|$. Then there exist clauses D, E with $D \in |X|$, $E \in |Z|$ and $C = D \circ E$. Clearly D is also in $|Y|$ and thus $C \in |Y \otimes Z|$. $\diamond$

LEMMA 4.2: *Let X, Y, Z be clause terms and $X \sqsubseteq Y$. Then*

(1) $X \oplus Z \sqsubseteq Y \oplus Z$,

(2) $Z \oplus X \sqsubseteq Z \oplus Y$,

(3) $X \otimes Z \sqsubseteq Y \otimes Z$,

(4) $Z \otimes X \sqsubseteq Z \otimes Y$,

Proof: (1) and (2) are trivial, (3) and (4) are analogous. Thus we only prove (4): Let $C \in |Z \otimes Y|$. Then $C \in |Z| \times |Y|$ and there exist $D \in |Z|$ and $E \in |Y|$ s.t. $C = D \circ E$. By definition of $\sqsubseteq$ there exists an $E' \in |X|$ with $E' \sqsubseteq E$. This implies $D \circ E' \in |Z \otimes X|$ and $D \circ E' \sqsubseteq D \circ E$. So $Z \otimes X \sqsubseteq Z \otimes Y$. $\diamond$

We are now able to show that replacing subterms in a clause term preserves the relations $\subseteq$ and $\sqsubseteq$.

LEMMA 4.3: *Let λ be an occurrence in a clause term X and $Y \preceq X.\lambda$ for $\preceq \in \{\subseteq, \sqsubseteq\}$. Then $X[Y]_\lambda \preceq X$.*

Proof: We proceed by induction on the term-complexity (i.e. number of nodes) of X.
If X is a set of clauses then λ is the top position and $X.\lambda = X$. Consequently $X[Y]_\lambda = Y$ and thus $X[Y]_\lambda \preceq X$.
Let X be $X_1 \odot X_2$ for $\odot \in \{\oplus, \otimes\}$. If λ is the top position in X then the lemma trivially holds. Thus we may assume that λ is a position in X_1 or in X_2. We consider the case that λ is in X_1 (the other one is completely symmetric): then there exists a position μ in X_1 s.t. $X.\lambda = X_1.\mu$. By induction hypothesis we get $X_1[Y]_\mu \preceq X_1$. By the lemmas 4.1 and 4.2 we obtain

$$X_1[Y]_\mu \odot X_2 \quad \preceq \quad X_1 \odot X_2.$$

But

$$X_1[Y]_\mu \odot X_2 = (X_1 \odot X_2)[Y]_\lambda = X[Y]_\lambda$$

and therefore $X[Y]_\lambda \preceq X$. $\diamond$

We will see in Section 6 that the relations $\subseteq, \sqsubseteq$ and $\leq_s$ are preserved under cut-reduction steps. Together they define a relation $\rhd$:

DEFINITION 4.5: *Let X and Y two clause terms. We define $X \rhd Y$ if (at least) one of the following properties is fulfilled:*

(a) $Y \subseteq X$ *or*

(b) $X \sqsubseteq Y$ *or*

(b) $X \leq_s Y$. $\sharp$

Remark: In general $Y \leq_s Z$ does not imply $X[Y]_\lambda \leq_s X[Z]_\lambda$, i.e. $\leq_s$ is not compatible with $\oplus$ and $\otimes$. Consider, for example, the terms

$$Y = \{\vdash P(x)\},\ Z = \{\vdash P(f(x))\} \text{ and}$$
$$X = \{\vdash Q(x)\} \otimes \{\vdash R(x)\},\ X.\lambda = \{\vdash Q(x)\}.$$

Clearly $Y \leq_s Z$. By replacement and evaluation we obtain

$$|X[Y]_\lambda| = \{\vdash P(x), R(x)\},\ |X[Z]\lambda| = \{\vdash P(f(x)), R(x)\}.$$

Obviously $X[Y]_\lambda \not\leq_s X[Z]_\lambda$. $\sharp$

The transitive closure $\rhd^*$ of $\rhd$ can be considered as a weak form of subsumption:

PROPOSITION 4.1: *Let X and Y be clause terms s.t. $X \rhd^* Y$. Then $X \leq_{ss} Y$.*

Proof: As the relation $\leq_{ss}$ is reflexive and transitive it suffices to show that $\rhd$ is a subrelation of $\leq_{ss}$.

a. $Y \subseteq X$: $X \leq_{ss} Y$ is trivial.

b. $X \sqsubseteq Y$: For all $C \in |Y|$ there exists a $D \in |X|$ with $D \sqsubseteq C$. But then also $D \leq_{ss} C$. The definition of the subsumption relation for sets yields $X \leq_{ss} Y$.

c. $X \leq_s Y$: $X \leq_{ss} Y$ is trivial. $\diamond$

5. The Method CERES

In (4) we defined a method of cut-elimination which is based on specific clause terms representing the derivation of the cut formulas in **LK**-proofs. Roughly speaking we compute a clause term from an **LK**-proof φ of S which corresponds to an unsatisfiable set of formulas, compute a resolution refutation of this set, and finally construct an ACNF of φ. The method in (4) is very general and is capable also of eliminating so-called pseudo-cuts. In this paper we are interested in ordinary cuts and mixes only and thus give a slightly simplified version of the method defined in (4). In particular we avoid the transformation of **LK**-proofs into cut-free proofs with sequent extensions and define the clause term directly. We restrict AC-normalization to derivations with skolemized end-sequents. It is always possible to construct derivations of skolemized end-sequents from the original ones without increase of length (see (1)). After AC-normalization the derivation can be transformed into a derivation of the original (unskolemized) sequent.

DEFINITION 5.1: *Let $\mathcal{SK}$ be the set of all* **LK**-*derivations with skolemized end-sequents.* $\mathcal{SK}_\emptyset$ *is the set of all cut-free proofs in* $\mathcal{SK}$ *and, for all* $i \geq 0$, $\mathcal{SK}^i$ *is the subset of* $\mathcal{SK}$ *containing all derivations with cut-formulas of formula complexity* $\leq i$. $\sharp$

Our goal is to transform a derivation in $\mathcal{SK}$ into a derivation in $\mathcal{SK}^0$. The first step in the corresponding procedure consists in the definition of a clause term corresponding to the sub-derivations of an **LK**-derivation ending in a cut. In particular we focus on derivations of the cut formulas themselves, i.e. on the derivation of formulas having no successors in the end-sequent.

DEFINITION 5.2 (CHARACTERISTIC TERM): *Let* φ *be an* **LK**-*derivation of* S *and let* Ω *be the set of all occurrences of cut formulas in* φ. *We define the* characteristic (clause) term $\Theta(\varphi)$ *inductively:*

Let ν *be the occurrence of an initial sequent* S' *in* φ. *Let* S'' *be the subsequent of* S' *consisting of all atoms which are ancestors of an occurrence in* Ω, *i.e.* $S'' = S(\nu, \Omega)$. *Then* $\Theta(\varphi)/\nu = \{S''\}$.

Let us assume that the clause terms $\Theta(\varphi)/\nu$ *are already constructed for all sequent–occurrences* ν *in* φ *with* $\mathrm{depth}(\nu) \leq k$. *Now let* ν *be an occurrence with* $\mathrm{depth}(\nu) = k + 1$. *We distinguish the following cases:*

(a) ν *is the consequent of* μ, *i.e. a unary rule applied to* μ *gives* ν. *Here we simply define* $\Theta(\varphi)/\nu = \Theta(\varphi)/\mu$.

(b) ν *is the consequent of* μ_1 *and* μ_2, *i.e. a binary rule* X *applied to* μ_1 *and* μ_2 *gives* ν.

- **(b1)** *The auxiliary formulas of* X *are ancestors of* Ω, *i.e. the formulas occur in* $S(\mu_1, \Omega), S(\mu_2, \Omega)$. *Then* $\Theta(\varphi)/\nu = \Theta(\varphi)/\mu_1 \oplus \Theta(\varphi)/\mu_2$.
- **(b2)** *The auxiliary formulas of* X *are not ancestors of* Ω. *In this case we define* $\Theta(\varphi)/\nu = \Theta(\varphi)/\mu_1 \otimes \Theta(\varphi)/\mu_2$.

Note that, in a binary inference, either both auxiliary formulas are ancestors of Ω *or none of them.*
Finally the characteristic term $\Theta(\varphi)$ *is defined as* $\Theta(\varphi)/\nu$ *where* ν *is the occurrence of the end-sequent.* $\sharp$

Remark: If φ is a cut-free proof then there are no occurrences of cut formulas in φ and $|\Theta(\varphi)| = \{\vdash\}$. $\sharp$

DEFINITION 5.3 (CHARACTERISTIC CLAUSE SET): *Let* φ *be an* **LK**-*derivation and* $\Theta(\varphi)$ *be the characteristic term of* φ. *Then* $\mathrm{CL}(\varphi)$, *for* $\mathrm{CL}(\varphi) = |\Theta(\varphi)|$, *is called the* characteristic clause set *of* φ. $\sharp$

EXAMPLE 5.1: *Let* φ *be the derivation (for* u, v *free variables,* a *a constant symbol)*

$$\frac{\varphi_1 \qquad \varphi_2}{(\forall x)(P(x) \to Q(x)) \vdash (\exists y)Q(y)}\ cut$$

where φ_1 is the **LK***-derivation:*

$$
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{P(u)^\star \vdash Q(u)^\star, P(u)}{\neg P(u), P(u)^\star \vdash Q(u)^\star}\ \neg : r \qquad Q(u), P(u)^\star \vdash Q(u)^\star}{P(u)^\star, \neg P(u) \vee Q(u) \vdash Q(u)^\star}\ \vee : l}{\neg P(u) \vee Q(u) \vdash \neg P(u)^\star, Q(u)^\star}\ \neg : r}{\neg P(u) \vee Q(u) \vdash (\neg P(u) \vee Q(u))^\star, Q(u)^\star}\ \vee : r}{\neg P(u) \vee Q(u) \vdash (\neg P(u) \vee Q(u))^\star, (\neg P(u) \vee Q(u))^\star}\ \vee : r}{\neg P(u) \vee Q(u) \vdash (\neg P(u) \vee Q(u))^\star}\ c : r}{\neg P(u) \vee Q(u) \vdash (\exists y)(\neg P(u) \vee Q(y))^\star}\ \exists : r}{(\forall x)(\neg P(x) \vee Q(x)) \vdash (\exists y)(\neg P(u) \vee Q(y))^\star}\ \forall : l}{(\forall x)(\neg P(x) \vee Q(x)) \vdash (\forall x)(\exists y)(\neg P(x) \vee Q(y))^\star}\ \forall : r
$$

and φ_2 is:

$$
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\vdash Q(v), P(a)^\star}{\neg P(a)^\star \vdash Q(v)}\ \neg : l \qquad Q(v)^\star \vdash Q(v)}{(\neg P(a) \vee Q(v))^\star \vdash Q(v)}\ \vee : l}{(\neg P(a) \vee Q(v))^\star \vdash (\exists y)Q(y)}\ \exists : r}{(\exists y)(\neg P(a) \vee Q(y))^\star \vdash (\exists y)Q(y)}\ \exists : l}{(\forall x)(\exists y)(\neg P(x) \vee Q(y))^\star \vdash (\exists y)Q(y)}\ \forall : l
$$

Let Ω be the set of the two occurrences of the cut formula in φ. The ancestors of Ω are marked by $\star$. We compute the characteristic clause term $\Theta(\varphi)$:

From the $\star$-marks in φ we first get the clause terms corresponding to the initial sequents:

$$X_1 = \{P(u) \vdash Q(u)\},\ X_2 = \{P(u) \vdash Q(u)\},\ X_3 = \{\vdash P(a)\},\ X_4 = \{Q(v) \vdash\}.$$

The leftmost-uppermost inference in φ_1 is unary and thus the clause term X_1 corresponding to this position does not change. The first binary inference in φ_1 (it is $\vee : l$) takes place on non-ancestors of Ω – the auxiliary formulas of the inference are not marked by $\star$. Consequently we obtain the term

$$Y_1 = \{P(u) \vdash Q(u)\} \otimes \{P(u) \vdash Q(u)\}.$$

The following inferences in φ_1 are all unary and so we obtain

$$\Theta(\varphi)/\nu_1 = Y_1$$

for ν_1 being the position of the end sequent of φ_1 in φ.
Again the uppermost-leftmost inference in φ_2 is unary and thus X_3 does not change. The first binary inference in φ_2 takes place on ancestors of Ω (the auxiliary formulas are $\star$-ed) and we have to apply the $\oplus$ to X_3, X_4. So we get

$$Y_2 = \{\vdash P(a)\} \oplus \{Q(v) \vdash\}.$$

Like in φ_1 all following inferences in φ_2 are unary leaving the clause term unchanged. Let ν_2 be the occurrence of the end-sequent of φ_2 in φ. Then the corresponding clause term is

$$\Theta(\varphi)/\nu_2 = Y_2.$$

The last inference (cut) in φ takes place on ancestors of Ω and we have to apply $\oplus$ again. This eventually yields the characteristic term

$$\begin{aligned}\Theta(\varphi) &= Y_1 \oplus Y_2 = \\ &\quad (\{P(u) \vdash Q(u)\} \otimes \{P(u) \vdash Q(u)\}) \oplus (\{\vdash P(a)\} \oplus \{Q(v) \vdash\}).\end{aligned}$$

For the characteristic clause set we obtain

$$\mathrm{CL}(\varphi) = |\Theta(\varphi)| = \{P(u), P(u) \vdash Q(u), Q(u);\ \vdash P(a);\ Q(v) \vdash\}.\ \sharp$$

It is easy to verify that the set of characteristic clauses $\mathrm{CL}(\varphi)$ constructed in the example above is unsatisfiable. This is not merely a coincidence, but a general principle expressed in the next proposition.

Proposition 5.1: *Let φ be an* **LK***-derivation. Then* $\mathrm{CL}(\varphi)$ *is unsatisfiable.*

Proof: In (4). $\diamond$

Let φ be a deduction of $S\colon \Gamma \vdash \Delta$ and $\mathrm{CL}(\varphi)$ be the characteristic clause set of φ. Then $\mathrm{CL}(\varphi)$ is unsatisfiable and, by the completeness of resolution (see (11), (8)), there exists a resolution refutation γ of $\mathrm{CL}(\varphi)$. By applying a ground projection to γ we obtain a ground resolution refutation γ' of $\mathrm{CL}(\varphi)$; by our definition of resolution γ' is also an AC-deduction of $\vdash$ from (ground instances of) $\mathrm{CL}(\varphi)$. This deduction γ' may serve as a skeleton of an AC-deduction ψ of $\Gamma \vdash \Delta$ itself. The construction of ψ from γ' is based on *projections* replacing φ by cut-free deductions $\varphi(C)$ of $\bar{P}, \Gamma \vdash \Delta, \bar{Q}$ for clauses $C : \bar{P} \vdash \bar{Q}$ in $\mathrm{CL}(\psi, \alpha)$. We merely give an informal description of the projections, for details we refer to (4). Roughly speaking, the projections of the proof φ are obtained by skipping all the inferences leading to a cut. As a "residue" we obtain a characteristic clause in the end sequent. Thus a projection is a cut-free derivation of the end sequent S + some atomic formulas in S. For the application of projections it is vital to have a skolemized end sequent, otherwise eigenvariable conditions could be violated.

Due to "automatic" contractions of side formulas in our version of **LK** the clauses in $\mathrm{CL}(\varphi)$ and those appearing in the projections may differ in the multiplicity of their atoms. This effect it inessential in the construction of the resolution proofs and the corresponding ACNFs (indeed only the number of contracted atom occurrences may differ).

Definition 5.4: *A sequent $\bar{P}' \vdash \bar{Q}'$ is called a contraction variant of $\bar{P} \vdash \bar{Q}$ if* $\mathrm{set}(\bar{P}') = \mathrm{set}(\bar{P})$ *and* $\mathrm{set}(\bar{Q}') = \mathrm{set}(\bar{Q})$ *(i.e. the sequents would be equal if defined via sets instead of multisets).* $\sharp$

LEMMA 5.1: *Let* φ *be a deduction in* $\mathcal{SK}$ *of a sequent* $S : \Gamma \vdash \Delta$. *Let* $C\colon \bar{P} \vdash \bar{Q}$ *be a clause in* $\mathrm{CL}(\varphi)$. *Then there exists a deduction* $\varphi(C)$ *of* $\bar{P}', \Gamma \vdash \Delta, \bar{Q}'$ *s.t.* $\bar{P}' \vdash \bar{Q}'$ *is a contraction variant of* $\bar{P} \vdash \bar{Q}$, $\varphi(C) \in \mathcal{SK}_\emptyset$ *and* $l(\varphi(C)) \leq l(\varphi)$.

Proof: In (4). $\diamond$

The construction of $\varphi(C)$ is illustrated below.

EXAMPLE 5.2: *Let* φ *be the proof of the sequent*

$$S : (\forall x)(P(x) \rightarrow Q(x)) \vdash (\exists y)Q(y)$$

as defined in Example 5.1. We have shown that

$$\mathrm{CL}(\varphi) = \{P(u), P(u) \vdash Q(u), Q(u);\ \ \vdash P(a);\ \ Q(v) \vdash\}.$$

We now define $\varphi(C_1)$, *the "projection" of* φ *to* $C_1\colon P(u), P(u) \vdash Q(u), Q(u)$: *The problem can be reduced to a projection in* φ_1 *because the last inference in* φ *is a cut and*

$$\Theta(\varphi)/\nu_1 \quad = \quad \{P(u), P(u) \vdash Q(u), Q(u)\}.$$

By skipping all inferences in φ_1 *leading to the cut formulas we obtain the deduction*

$$\dfrac{\dfrac{\dfrac{P(u) \vdash P(u), Q(u)}{\neg P(u), P(u) \vdash Q(u)}\ \neg : l \qquad Q(u), P(u) \vdash Q(u)}{P(u), \neg P(u) \vee Q(u) \vdash Q(u)}\ \vee : l}{P(u), (\forall x)(\neg P(x) \vee Q(x)) \vdash Q(u)}\ \forall : l$$

In order to obtain the end sequent we only need an additional weakening and $\varphi(C_1) =$

$$\dfrac{\dfrac{\dfrac{\dfrac{P(u) \vdash P(u), Q(u)}{\neg P(u), P(u) \vdash Q(u)}\ \neg : l \qquad Q(u), P(u) \vdash Q(u)}{P(u), \neg P(u) \vee Q(u) \vdash Q(u)}\ \vee : l}{P(u), (\forall x)(\neg P(x) \vee Q(x)) \vdash Q(u)}\ \forall : l}{P(u), (\forall x)(\neg P(x) \vee Q(x)) \vdash (\exists y)Q(y), Q(u)}\ w : r$$

For $C_2 = \vdash P(a)$ *we obtain the projection* $\varphi(C_2)$:

$$\dfrac{\dfrac{\vdash P(a), Q(v)}{\vdash P(a), (\exists y)Q(y)}\ \exists : r}{(\forall x)(\neg P(x) \vee Q(x)) \vdash (\exists y)Q(y), P(a)}\ w : l$$

Similarly we obtain $\varphi(C_3)$:

$$\dfrac{\dfrac{Q(v) \vdash Q(v)}{Q(v) \vdash (\exists y)Q(y)}\ \exists : r}{(\forall x)(\neg P(x) \vee Q(x)), Q(v) \vdash (\exists y)Q(y)}\ w : l$$

$\sharp$

We have seen that, in the projections, only inferences on nonancestors of cuts are performed. If the auxiliary formulas of a binary rule are ancestors of cuts we have to apply weakening in order to obtain the required formulas from the second premise.
Let φ be a proof of S s.t. $\varphi \in \mathcal{SK}$ and let γ be a resolution refutation of the (unsatisfiable) set of clauses $\mathrm{CL}(\varphi)$. Then γ can be transformed into a deduction $\varphi(\gamma)$ of S s.t. $\varphi(\gamma) \in \mathcal{SK}^0$. $\varphi(\gamma)$ is a proof with atomic cuts, thus an AC-normal form of φ. $\varphi(\gamma)$ is constructed from γ simply by replacing the resolution steps by the corresponding proof projections. The construction of $\varphi(\gamma)$ is the essential part of the method CERES (the final elimination of atomic cuts is inessential). The resolution refutation γ can be considered as the characteristic part of $\varphi(\gamma)$ representing the essential result of AC-normalization. Below we give an example of a construction of $\varphi(\gamma)$, for details we refer to (4) again.

EXAMPLE 5.3: *Let φ be the proof of*

$$S\colon (\forall x)(P(x) \to Q(x)) \vdash (\exists y)Q(y)$$

as defined in Example 5.1 and in Example 5.2. Then

$$\mathrm{CL}(\varphi) = \{C_1 : P(u), P(u) \vdash Q(u), Q(u);\ C_2 : \vdash P(a);\ C_3 : Q(v) \vdash\}.$$

First we define a resolution refutation δ of $\mathrm{CL}(\varphi)$*:*

$$\dfrac{\dfrac{\vdash P(a) \quad P(u), P(u) \vdash Q(u), Q(u)}{\vdash Q(a), Q(a)}\,R \qquad Q(v) \vdash}{\vdash}\,R$$

and a corresponding ground refutation γ:

$$\dfrac{\dfrac{\vdash P(a) \quad P(a), P(a) \vdash Q(a), Q(a)}{\vdash Q(a), Q(a)}\,R \qquad Q(a) \vdash}{\vdash}\,R$$

The ground substitution defining the ground projection is

$\sigma : \{u \leftarrow a, v \leftarrow a\}$.

Let $\chi_1 = \varphi(C_1)\sigma$, $\chi_2 = \varphi(C_2)\sigma$ and $\chi_3 = \varphi(C_3)\sigma$. Moreover let us write B for $(\forall x)(P(x) \to Q(x))$ and C for $(\exists y)(P(a) \to Q(y))$.
Then $\varphi(\gamma)$ is of the form

$$\dfrac{\dfrac{\dfrac{\dfrac{\begin{matrix}(\chi_2)\\ B \vdash C, P(a)\end{matrix} \quad \begin{matrix}(\chi_1)\\ P(a), B \vdash C, Q(a)\end{matrix}}{B, B \vdash C, C, Q(a)}\,cut \qquad \begin{matrix}(\chi_3)\\ Q(a), B \vdash C\end{matrix}}{B, B, B \vdash C, C, C}\,cut}{B \vdash C, C, C}\,c:l}{B \vdash C}\,c:r$$

$\sharp$

If ψ a deduction in AC-normal form then there exists a "canonic" resolution refutation $\mathrm{RES}(\psi)$ of the set of clauses $\mathrm{CL}(\psi)$. $\mathrm{RES}(\psi)$ is "the" resolution proof corresponding to ψ. Indeed, as ψ is a deduction with atomic cuts only, the part of ψ ending in the cut formulas is nothing else than a resolution refutation. For the construction of $\mathrm{RES}(\psi)$ we need some technical definitions:

Definition 5.5: *Let γ be a p-resolution deduction of a clause C from a set of clauses $\mathcal{C}$ and let D be a clause. We define a p-resolution deduction $\gamma(D)$ of $D \circ C$ from $\{D\} \times \mathcal{C}$ in the following way:*

(1) construct a deduction γ' by replacing all initial clauses S in γ by $D \circ S$, and leave the inference nodes unchanged.

(2) Apply contractions and weakenings to the end clause of γ' (if necessary) in order to obtain a deduction $\gamma(D)$ of $D \circ C$ from $\{D\} \times \mathcal{C}$. $\sharp$

Remark: Contractions may become necessary as the occurrence of D in clauses may be multiplied by resolutions in γ'. Weakenings are required if atoms in D are cut out by resolutions in γ'. $\sharp$

Definition 5.6: *Let γ be a p-resolution deduction of C from $\mathcal{C}$ and let δ be a p-resolution deduction of D from $\mathcal{D}$. We define a p-resolution deduction $\gamma \odot \delta$ of $C \circ D$ from $\mathcal{C} \times \mathcal{D}$ in the following way:*

(1) construct a deduction η by replacing all initial clauses S in γ by the deductions $\delta(S)$ of $D \circ S$, and leave the inference nodes in γ unchanged.

(2) Apply contractions and weakenings to the end clause of η (if necessary) in order to obtain the deduction $\gamma \odot \delta$ of $D \circ C$. $\sharp$

Remark: $\gamma \odot \delta$ is indeed a p-deduction from $\mathcal{C} \times \mathcal{D}$ as the initial clauses are of the form $S \circ S'$ for $S \in \mathcal{C}$ and $S' \in \mathcal{D}$. $\sharp$

If ψ is in ACNF then there exists something like a *canonic* resolution refutation of $\mathrm{CL}(\psi)$. The definition of this refutation follows the steps of the definition of the characteristic clause term.

Definition 5.7: *Let ψ be an LK-derivation in ACNF, Ω be the set of occurrences of the (atomic) cut formulas in ψ and $\mathcal{C} = \mathrm{CL}(\psi)$. For comfort we write $\mathcal{C}/\nu$ for the set of clauses $|\Theta(\psi)/\nu|$ defined by the characteristic terms as in Definition 5.2. Clearly $\mathcal{C} = \mathcal{C}/\nu_0$ for the root node ν_0 in ψ.*
We proceed inductively and define a p-resolution deduction γ_ν for every deduction node ν in ψ s.t. γ_ν is a deduction of $S(\nu, \Omega)$ from $\mathcal{C}/\nu$.
If ν is a leaf in ψ then we define γ_ν as $S(\nu, \Omega)$. By definition of $\mathcal{C}$ we have $\mathcal{C}_\nu = S(\nu, \Omega)$. Clearly γ_ν is p-resolution deduction of $S(\nu, \Omega)$ from $\mathcal{C}_\nu$.

(1) Let γ_μ be already defined for a node in μ in ψ s.t. γ_μ is a p-resolution deduction of $S(\mu, \Omega)$ from $\mathcal{C}/\mu$. Moreover let ξ be a unary inference in ψ with premiss μ and conclusion ν. We distinguish two cases:

(1a) The auxiliary formulas of ξ are in $S(\mu,\Omega)$.
Then ξ is a weakening or a contraction (note that the cuts are atomic!) and we define $\gamma_\nu =$

$$\frac{\gamma_\mu}{S(\nu,\Omega)}\ \xi.$$

(b) The auxiliary formulas of ξ are not in $S(\mu,\Omega)$.
Then we define $\gamma_\nu = \gamma_\mu$.

In both cases γ_ν is a p-resolution deduction of $S(\nu,\Omega)$ from $\mathcal{C}_\mu$. But by definition of the characteristic clause term we have $\mathcal{C}_\nu = \mathcal{C}_\mu$.

(2) Assume that γ_{μ_i} are p-resolution deductions of $S(\mu_i,\Omega)$ from $\mathcal{C}_{\mu_i}$ for $i = 1,2$. Let ν be an inference node in ψ with premisses μ_1,μ_2 and the corresponding binary rule ξ. Again we distinguish two cases:

(2a) The auxiliary formulas of ξ are in $S(\mu_1,\Omega)$ and $S(\mu_2,\Omega)$.
Then ξ must be a cut (there are no other binary inferences leading to Ω) and we define $\gamma_\nu =$

$$\frac{\gamma_{\mu_1} \quad \gamma_{\mu_2}}{S(\nu,\Omega)}\ cut$$

By definition γ_ν is a p-resolution deduction of $S(\nu,\Omega)$ from $\mathcal{C}_{\mu_1} \cup \mathcal{C}_{\mu_2}$. By definition of the characteristic term we have $\mathcal{C}_\nu = \mathcal{C}_{\mu_1} \cup \mathcal{C}_{\mu_2}$ and, therefore, γ_ν is a p-resolution deduction of $S(\nu,\Omega)$ from $\mathcal{C}_\nu$.

(2b) The auxiliary formulas of ξ are not in $S(\mu_1,\Omega)$ and $S(\mu_2,\Omega)$.
In this case we define

$$\gamma_\nu = \gamma_{\mu_1} \odot \gamma_{\mu_2}.$$

By definition of $\odot$ the deduction γ_ν is a p-resolution deduction of $S(\mu_1,\Omega) \circ S(\mu_2,\Omega)$ from $\mathcal{C}_{\mu_1} \times \mathcal{C}_{\mu_2}$. But $S(\nu,\Omega) = S(\mu_1,\Omega) \circ S(\mu_2,\Omega)$ and, by definition of the characteristic term, $\mathcal{C}_\nu = \mathcal{C}_{\mu_1} \times \mathcal{C}_{\mu_2}$.

Finally we define $\mathrm{RES}(\psi) = \gamma_{\nu_0}$ *where ν_0 is the root node in ψ.* $\sharp$

Remark: The root node does not contain any ancestors of cut occurrences Ω, i.e. $S(\nu_0,\Omega) = \vdash$ and γ_{ν_0} as defined above is also a *refutation* of $\mathrm{CL}(\psi)$. $\sharp$

For an AC-deduction ψ the number of nodes in $\mathrm{RES}(\psi)$ may be exponential in the number of nodes of ψ. But note that, in general, resolution refutations of $\mathrm{CL}(\psi)$ are of nonelementary length (see (4)). Thus the proofs $\mathrm{RES}(\psi)$ for AC-deductions ψ can be considered as "small".

PROPOSITION 5.2: *Let ψ be an LK-derivation in ACNF. Then*

$$l(\mathrm{RES}(\psi)) \ \leq \ l(\psi) * 2^{2*l(\psi)}.$$

Proof: We show first that

$$l(\mathrm{RES}(\psi)) \;\; \leq \;\; 2^{l(\psi)} * |\mathrm{CL}(\psi)|.$$

To this aim we proceed by induction on the definition of the γ_ν in Definition 5.7, i.e. we prove that for all nodes ν in ψ

$$(*)\ l(\gamma_\nu) \;\; \leq \;\; 2^{l(\psi.\nu)} * |\mathcal{C}_\nu|.$$

For leaves ν we have $l(\gamma_\nu) = 1$ and $(*)$ is trivial.
So let us assume that $(*)$ holds for the node μ and ν is the conclusion of a unary inference with premiss μ. Then by definition of γ_ν:

$$\begin{array}{rcl} l(\gamma_\nu) & = & l(\gamma_\mu) + 1, \\ \mathcal{C}_\nu & = & \mathcal{C}_\mu, \\ l(\psi.\nu) & = & l(\psi.\mu) + 1 \text{ and by assumption on } \nu \end{array}$$

$$l(\gamma_\nu) \leq 2^{l(\psi.\mu)} * |\mathcal{C}_\mu| + 1 \leq 2^{l(\psi.\nu)} * |\mathcal{C}_\nu|.$$

Assume that $(*)$ holds for nodes μ_1, μ_2 and ν is the conclusion of a binary inference with premisses μ_1, μ_2.
If the inference takes place on ancestors of Ω then

$$\begin{array}{rcl} l(\gamma_\nu) & = & l(\gamma_{\mu_1}) + l(\gamma_{\mu_2}) + 1, \\ \mathcal{C}_\nu & = & \mathcal{C}_{\mu_1} \cup \mathcal{C}_{\mu_2}, \\ l(\psi.\nu) & = & l(\psi.\mu_1) + l(\psi.\mu_2) + 1. \end{array}$$

By the assumptions on μ_1, μ_2 we have

$$\begin{array}{r} l(\gamma_{\mu_1}) \leq 2^{l(\psi.\mu_1)} * |\mathcal{C}_{\mu_1}|, \\ l(\gamma_{\mu_2}) \leq 2^{l(\psi.\mu_2)} * |\mathcal{C}_{\mu_2}| \end{array}$$

and therefore

$$\begin{array}{rcl} l(\gamma_\nu) & = & l(\gamma_{\mu_1}) + l(\gamma_{\mu_2}) + 1 \\ & \leq & 2^{l(\psi.\mu_1)} * |\mathcal{C}_{\mu_1}| + 2^{l(\psi.\mu_2)} * |\mathcal{C}_{\mu_2}| + 1 \\ & \leq & 2^{l(\psi.\mu_1)+l(\psi.\mu_2)} * (|\mathcal{C}_{\mu_1}| + |\mathcal{C}_{\mu_2}|) + 1 \\ & \leq & 2^{l(\psi.\nu)} * |\mathcal{C}_\nu|. \end{array}$$

If the inference takes place on non-ancestors of Ω then

$$\begin{array}{rcl} l(\gamma_\nu) & \leq & 2 * l(\gamma_{\mu_1}) * l(\gamma_{\mu_2}), \\ \mathcal{C}_\nu & = & \mathcal{C}_{\mu_1} \times \mathcal{C}_{\mu_2}, \\ l(\psi.\nu) & = & l(\psi.\mu_1) + l(\psi.\mu_2) + 1. \end{array}$$

and, by the assumptions on μ_1, μ_2,

$$\begin{array}{rcl} l(\gamma_\nu) & \leq & 2 * l(\gamma_{\mu_1}) * l(\gamma_{\mu_2}) \\ & \leq & 2 * 2^{l(\psi.\mu_1)} * |\mathcal{C}_{\mu_1}| * 2^{l(\psi.\mu_2)} * |\mathcal{C}_{\mu_2}| \\ & = & 2^{l(\psi.\mu_1)+l(\psi.\mu_2)+1} * |\mathcal{C}_{\mu_1}| * |\mathcal{C}_{\mu_2}| \\ & = & 2^{l(\psi.\nu)} * |\mathcal{C}_\nu|. \end{array}$$

Thus by induction and choosing the root node for ν we obtain

$$(I)\ l(\mathrm{RES}(\psi)) \ \leq \ 2^{l(\psi)} * |\mathrm{CL}(\psi)|.$$

In (4) we have shown that

$$(II)\ |\mathrm{CL}(\psi)| \ \leq \ l(\psi) * 2^{l(\psi)}.$$

Putting (I) and (II) together we eventually obtain

$$(I)\ l(\mathrm{RES}(\psi)) \ \leq \ l(\psi) * 2^{2*l(\psi)}. \ \diamond$$

6. Characteristic Terms and Cut-Reduction

In this section we are proving our main result. The key lemma below shows that a cut-reduction step on a derivation (based on the set $\mathcal{R}$ defined in the Appendix) corresponds to a reduction step (w.r.t. $\rhd$) on the corresponding clause term. As the set $\mathcal{R}$ is a reduction set for mixes, we assume throughout this section that all cuts in the derivations are also mixes.

LEMMA 6.1: *Let φ, φ' be* **LK**-*derivations with $\varphi >_{\mathcal{R}} \varphi'$ for a cut reduction relation $>_{\mathcal{R}}$ based on $\mathcal{R}$. Then $\Theta(\varphi) \rhd \Theta(\varphi')$.*

Proof: We construct a proof by cases on the definition of $>_{\mathcal{R}}$. To this aim we consider sub-derivations ψ of φ of the form

$$\frac{\begin{array}{cc} (\rho, X) & (\sigma, Y) \\ \Gamma \vdash \Delta & \Pi \vdash \Lambda \end{array}}{\Gamma, \Pi^* \vdash \Delta^*, \Lambda}\ mix(A)$$

where $X = \Theta(\varphi)/\lambda$ for the occurrence λ corresponding to the deduction ρ and $Y = \Theta(\varphi)/\mu$ for the occurrence μ corresponding to σ. By ν we denote the occurrence of ψ in φ. That means we do not only indicate the sub-derivations ending in the mix, but also the corresponding clause terms. Note that by definition of the characteristic term we have $\Theta(\varphi)/\nu = X \oplus Y$.
If $\psi >_{\mathcal{R}} \chi$ then, by definition of the reduction relation $>_{\mathcal{R}}$, we get $\varphi = \varphi[\psi]_\nu >_{\mathcal{R}} \varphi[\chi]_\nu$. For the remaining part of the proof we denote $\varphi[\chi]_\nu$ by φ'. Our aim is to prove that $\Theta(\varphi) \rhd \Theta(\varphi')$.

(I) $\mathrm{rank}(\psi) = 2$:

(Ia) ψ is of the form

$$\dfrac{\dfrac{\begin{matrix}(\rho',X)\\ \Gamma\vdash\Delta\end{matrix}}{\Gamma\vdash\Delta,A}\;w:r \qquad \begin{matrix}(\sigma,Y)\\ \Pi\vdash\Lambda\end{matrix}}{\Gamma,\Pi^*\vdash\Delta,\Lambda}\;mix(A)$$

By definition of $\mathcal{R}$ we have $\psi >_{\mathcal{R}} \chi$ for $\chi =$

$$\dfrac{\begin{matrix}(\rho',X)\\ \Gamma\vdash\Delta\end{matrix}}{\Gamma,\Pi^*\vdash\Delta,\Lambda}\;w:l,r$$

Therefore also $\varphi[\psi]_\nu >_{\mathcal{R}} \varphi[\chi]_\nu$, i.e. $\varphi >_{\mathcal{R}} \varphi'$. But $\Theta(\varphi')/\nu = X$ and $\Theta(\varphi)/\nu = X \oplus Y$. Clearly $X \oplus Y \rhd X$ and, by Lemma 4.3, $\Theta(\varphi) \rhd \Theta(\varphi')$.

(Ib) $A = \neg B$ and ψ is of the form

$$\dfrac{\dfrac{\begin{matrix}(\rho',X)\\ B,\Gamma\vdash\Delta\end{matrix}}{\Gamma\vdash\Delta,\neg B}\;\neg:r \qquad \dfrac{\begin{matrix}(\sigma',Y)\\ \Pi\vdash\Lambda,B\end{matrix}}{\neg B,\Pi\vdash\Lambda}\;\neg:l}{\Gamma,\Pi\vdash\Delta,\Lambda}\;mix(A)$$

Then $\psi >_{\mathcal{R}} \chi$ for $\chi =$

$$\dfrac{\dfrac{\begin{matrix}(\sigma',Y)\\ \Pi\vdash\Lambda,B\end{matrix} \qquad \begin{matrix}(\rho',X)\\ B,\Gamma\vdash\Delta\end{matrix}}{\Gamma^*,\Pi\vdash\Delta,\Lambda^*}\;mix(B)}{\Gamma,\Pi\vdash\Delta,\Lambda}\;w:l,r$$

Here we have

$$\begin{array}{rcl}\Theta(\varphi)/\nu &=& X\oplus Y,\\ \Theta(\varphi')/\nu &=& Y\oplus X.\end{array}$$

Clearly $X \oplus Y \rhd Y \oplus X$ (we even have $X \oplus Y \sim Y \oplus X$) and by Lemma 4.3 we obtain $\Theta(\varphi) \rhd \Theta(\varphi')$.

(Ic) $A = B \wedge C$ and ψ is of the form

$$\dfrac{\dfrac{\begin{matrix}(\rho_1,X_1)\\ \Gamma\vdash\Delta,B\end{matrix} \quad \begin{matrix}(\rho_2,X_2)\\ \Gamma\vdash\Delta,C\end{matrix}}{\Gamma\vdash\Delta,B\wedge C}\;\wedge:r \qquad \dfrac{\begin{matrix}(\sigma',Y)\\ B,\Pi\vdash\Lambda\end{matrix}}{B\wedge C,\Pi\vdash\Lambda}\;\wedge:l}{\Gamma,\Pi\vdash\Delta,\Lambda}\;mix(A)$$

Then $\psi >_{\mathcal{R}} \chi$ for $\chi =$

$$\dfrac{\dfrac{\begin{matrix}(\rho_1,X_1)\\ \Gamma\vdash\Delta,B\end{matrix} \qquad \begin{matrix}(\sigma',Y)\\ B,\Pi\vdash\Lambda\end{matrix}}{\Gamma,\Pi^*\vdash\Delta^*,\Lambda}\;mix(B)}{\Gamma,\Pi\vdash\Delta,\Lambda}\;w:l,r$$

In this case we have

$$\begin{array}{rcl}\Theta(\varphi)/\nu & = & (X_1 \oplus X_2) \oplus Y, \\ \Theta(\varphi')/\nu & = & X_1 \oplus Y.\end{array}$$

Clearly $X_1 \oplus Y \subseteq (X_1 \oplus X_2) \oplus Y$ and thus $(X_1 \oplus X_2) \oplus Y \rhd X_1 \oplus Y$. By application of Lemma 4.3 we obtain $\Theta(\varphi) \rhd \Theta(\varphi')$.

The case where $B \wedge C$ is inferred from C is completely symmetric.

(Ic) $A = B \vee C$: symmetric to (Ib).

(Id) $A = (\forall x)B$. Then ψ is of the form

$$\dfrac{\dfrac{\begin{array}{c}(\rho'(x/y), X(x/y))\\ \Gamma \vdash \Delta, B(x/y)\end{array}}{\Gamma \vdash \Delta, (\forall x)B(x)}\ \forall : r \qquad \dfrac{\begin{array}{c}(\sigma', Y)\\ B(x/t), \Pi \vdash \Lambda\end{array}}{(\forall x)B(x), \Pi \vdash \Lambda}\ \forall : l}{\Gamma, \Pi \vdash \Delta, \Lambda}\ mix(A)$$

$\psi >_{\mathcal{R}} \chi$ for

$$\dfrac{\dfrac{\begin{array}{c}(\rho'(x/t), X(x/t))\\ \Gamma \vdash \Delta, B(t)\end{array} \qquad \begin{array}{c}(\sigma', Y)\\ B(x/t), \Pi \vdash \Lambda\end{array}}{\Gamma, \Pi^* \vdash \Delta^*, \Lambda}\ mix(B(x/t)}{\Gamma, \Pi \vdash \Delta, \Lambda}\ w : l, r$$

By definition of the characteristic terms we have

$$\begin{array}{rcl}\Theta(\varphi)/\nu & = & X(x/y) \oplus Y, \\ \Theta(\varphi')/\nu & = & X(x/t) \oplus Y.\end{array}$$

By assumption φ is regular and the variable y only occurs in the sub-derivation ρ. Therefore

$$\begin{array}{rcl}\Theta(\varphi')/\nu & = & (X(x/y) \oplus Y)\{y \leftarrow t\} \text{ and even} \\ \Theta(\varphi') & = & \Theta(\varphi)\{y \leftarrow t\}.\end{array}$$

But this means $\Theta(\varphi) \leq_s \Theta(\varphi')$ and therefore $\Theta(\varphi) \rhd \Theta(\varphi')$.

(Ie) $A = (\exists x)B$: symmetric to (Id).

(II) $\mathrm{rank}(\psi) > 2$.

We assume that $\mathrm{rank}_r(\psi) > 1$ (the case $\mathrm{rank}_l(\psi) > 1$ is symmetric).

(IIa) A occurs in Γ. Then $\psi >_{\mathcal{R}} \chi$ for $\chi =$

$$\dfrac{\begin{array}{c}(\sigma, Y)\\ \Pi \vdash \Lambda\end{array}}{\Gamma, \Pi^* \vdash \Delta^*, \Lambda}\ w : l, r; c : l$$

In this case

$$\begin{array}{rcl} \Theta(\varphi)/\nu & = & X \oplus Y, \\ \Theta(\varphi')/\nu & = & Y. \end{array}$$

Clearly $X \oplus Y \rhd Y$ and by Lemma 4.3 $\Theta(\varphi) \rhd \Theta(\varphi')$.

(IIb) A does not occur in Γ.

(IIb.1) ξ is one of the inferences $w : l$ or $c : l$ and ψ is of the form:

$$\dfrac{\begin{array}{c}(X,\rho)\\ \Gamma \vdash \Delta\end{array} \quad \dfrac{\begin{array}{c}(Y,\sigma')\\ \Sigma \vdash \Lambda\end{array}}{\Pi \vdash \Lambda}\,\xi}{\Gamma, \Pi^* \vdash \Delta^*, \Lambda}\,mix(A)$$

Then $\psi >_{\mathcal{R}} \chi$ for $\chi =$

$$\dfrac{\dfrac{\begin{array}{c}(X,\rho)\\ \Gamma \vdash \Delta\end{array} \quad \begin{array}{c}(Y,\sigma')\\ \Sigma \vdash \Lambda\end{array}}{\Gamma, \Sigma^* \vdash \Delta^*, \Lambda}\,mix(A)}{\Gamma, \Pi^* \vdash \Delta^*, \Lambda}\,(\xi)$$

It is obvious that $\Theta(\varphi) = \Theta(\varphi')$ and so $\Theta(\varphi) \rhd \Theta(\varphi')$.

(IIb.2) ξ is a unary inference, $\xi \notin \{w : l, c : l\}$ and ψ is of the form

$$\dfrac{\begin{array}{c}(X,\rho)\\ \Gamma \vdash \Delta\end{array} \quad \dfrac{\begin{array}{c}(Y,\sigma')\\ B, \Pi \vdash \Lambda_1\end{array}}{C, \Pi \vdash \Lambda_2}\,\xi}{\Gamma, C^*, \Pi^* \vdash \Delta^*, \Lambda_2}\,mix(A)$$

where $C^* = \top$ for $C = A$ and $C^* = C$ for $C \neq A$. First we define a deduction τ:

$$\dfrac{\dfrac{\dfrac{\begin{array}{c}(X,\rho)\\ \Gamma \vdash \Delta\end{array} \quad \begin{array}{c}(Y,\sigma')\\ B, \Pi \vdash \Lambda_1\end{array}}{\Gamma, B^*, \Pi^* \vdash \Delta^*, \Lambda_1}\,mix(A)}{\Gamma, B, \Pi^* \vdash \Delta^*, \Lambda_1}\,(w : l)}{\Gamma, C, \Pi^* \vdash \Delta^*, \Lambda_2}\,\xi$$

It is easy to see that

$$\begin{array}{rcl} \Theta(\varphi[\tau]_\nu)/\nu & = & X \oplus Y \ \text{ and} \\ \Theta(\varphi) & = & \Theta(\varphi[\tau]_\nu). \end{array}$$

Indeed changing the order of unary inferences does not affect characteristic terms. If $A \neq C$ then, by definition of $>_{\mathcal{R}}$, we define $\chi = \tau$ and $\Theta(\varphi) = \Theta(\varphi')$.

If $A = C$ and $A \neq B$ we have $\chi =$

$$\dfrac{\dfrac{\begin{matrix}(X,\rho) & (\tau, X \oplus Y)\\ \Gamma \vdash \Delta & \Gamma, A, \Pi^* \vdash \Delta^*, \Lambda_2\end{matrix}}{\Gamma, \Gamma^*, \Pi^* \vdash \Delta^*, \Delta^*, \Lambda_2}\, mix(A)}{\Gamma, \Pi^* \vdash \Delta^*, \Lambda_2}\, c : l, r$$

Now we have

$$\begin{aligned}\Theta(\varphi)/\nu &= X \oplus Y,\\ \Theta(\varphi')/\nu &= X \oplus (X \oplus Y).\end{aligned}$$

But $X \oplus Y \sim X \oplus (X \oplus Y)$ and thus also $X \oplus Y \rhd X \oplus (X \oplus Y)$. Therefore, using Lemma 4.3 again, we obtain $\Theta(\varphi) \rhd \Theta(\varphi')$.

If $A = B = C$ then $\Lambda_1 \neq \Lambda_2$ and χ is defined as

$$\dfrac{\dfrac{\begin{matrix}(X,\rho) & (Y,\sigma')\\ \Gamma \vdash \Delta & A, \Pi \vdash \Lambda_1\end{matrix}}{\Gamma, \Pi^* \vdash \Delta^*, \Lambda_1}\, mix(A)}{\Gamma, \Pi^* \vdash \Delta^*, \Lambda_2}\, \xi$$

In this case, clearly, $\Theta(\varphi') = \Theta(\varphi)$ and thus $\Theta(\varphi) \rhd \Theta(\varphi')$.

(IIb.3) The last inference in σ is a binary one.

(IIb.3.1) The last inference in σ is $\wedge : r$. Then ψ is of the form

$$\dfrac{\begin{matrix}(\rho, X)\\ \Gamma \vdash \Delta\end{matrix} \quad \dfrac{\begin{matrix}(\sigma_1, Y_1) & (\sigma_2, Y_2)\\ \Pi \vdash \Lambda, B & \Pi \vdash \Lambda, C\end{matrix}}{\Pi \vdash \Lambda, B \wedge C}\, \wedge : r}{\Gamma, \Pi^* \vdash \Delta^*, \Lambda, B \wedge C}\, mix(A)$$

Clearly A occurs in Π and ψ reduces to the following proof χ via cross-cut:

$$\dfrac{\dfrac{\begin{matrix}(\rho, X) & (\sigma_1, Y_1)\\ \Gamma \vdash \Delta & \Pi \vdash \Lambda, B\end{matrix}}{\Gamma, \Pi^* \vdash \Delta^*, \Lambda, B}\, mix(A) \quad \dfrac{\begin{matrix}(\rho, X) & (\sigma_2, Y_2)\\ \Gamma \vdash \Delta & \Pi \vdash \Lambda, C\end{matrix}}{\Gamma, \Pi^* \vdash \Delta^*, \Lambda, C}\, mix(A)}{\Gamma, \Pi^* \vdash \Delta^*, \Lambda, B \wedge C}\, \wedge : r$$

Now we have to distinguish two cases:
case a: $B \wedge C$ is ancestor of (another) mix in φ.
Then

$$\begin{aligned}\Theta(\varphi)/\nu &= X \oplus (Y_1 \oplus Y_2),\\ \Theta(\varphi')/\nu &= (X \oplus Y_1) \oplus (X \oplus Y_2).\end{aligned}$$

Clearly

$$X \oplus (Y_1 \oplus Y_2) \sim (X \oplus Y_1) \oplus (X \oplus Y_2)$$

and therefore $\Theta(\varphi') \sim \Theta(\varphi)$, thus $\Theta(\varphi) \rhd \Theta(\varphi')$.

case b: $B \wedge C$ is not an ancestor of a mix in φ.
Then

$$\begin{array}{rcl} \Theta(\varphi)/\nu & = & X \oplus (Y_1 \otimes Y_2), \\ \Theta(\varphi')/\nu & = & (X \oplus Y_1) \otimes (X \oplus Y_2). \end{array}$$

But by using elementary properties of $\cup$ and $\times$ we obtain

$$X \oplus (Y_1 \otimes Y_2) \;\; \sqsubseteq \;\; (X \oplus Y_1) \otimes (X \oplus Y_2)$$

That means $\Theta(\varphi)/\nu \sqsubseteq \Theta(\varphi')/\nu$ and by application of Lemma 4.3 we again get $\Theta(\varphi) \sqsubseteq \Theta(\varphi')$, thus also $\Theta(\varphi) \rhd \Theta(\varphi')$.

(IIb.3.2) The last inference in σ is $\vee : l$. Then ψ is of the form

$$\dfrac{\begin{array}{c}(\rho, X)\\ \Gamma \vdash \Delta\end{array} \quad \dfrac{\begin{array}{cc}(\sigma_1, Y_1) & (\sigma_2, Y_2)\\ B, \Pi \vdash \Lambda & C, \Pi \vdash \Lambda\end{array}}{B \vee C, \Pi \vdash \Lambda}\ \vee : l}{(B \vee C)^*, \Gamma, \Pi^* \vdash \Delta^*, \Lambda}\ mix(A)$$

Note that A is in Π; for otherwise $A = B \vee C$ and $\mathrm{rank}_r(\psi) = 1$, contradicting the assumption.
We first define the following deduction τ:

$$\dfrac{\dfrac{\dfrac{\begin{array}{cc}(\rho, X) & (\sigma_1, Y_1)\\ \Gamma \vdash \Delta & B, \Pi \vdash \Lambda\end{array}}{B^*, \Gamma, \Pi^* \vdash \Delta^*, \Lambda}\ mix(A)}{B, \Gamma, \Pi^* \vdash \Delta^*, \Lambda}\ w:^* \quad \dfrac{\dfrac{\begin{array}{cc}(\rho, X) & (\sigma_2, Y_2)\\ \Gamma \vdash \Delta & C, \Pi \vdash \Lambda\end{array}}{C^*, \Gamma, \Pi^* \vdash \Delta^*, \Lambda}\ mix(A)}{C, \Gamma, \Pi^* \vdash \Delta^*, \Lambda}\ w:^*}{(B \vee C), \Gamma, \Pi^* \vdash \Delta^*, \Lambda}\ \vee : l$$

As in IIb.3.1 we have to distinguish the case where $B \vee C$ is an ancestor of another mix in φ or not. So if we replace ψ by τ in φ we get either get

$$\begin{array}{rcl} \Theta(\varphi)/\nu & = & X \oplus (Y_1 \oplus Y_2), \\ \Theta(\varphi[\tau]_\nu)/\nu & = & (X \oplus Y_1) \oplus (X \oplus Y_2). \end{array}$$

or

$$\begin{array}{rcl} \Theta(\varphi)/\nu & = & X \oplus (Y_1 \otimes Y_2), \\ \Theta(\varphi[\tau]_\nu)/\nu & = & (X \oplus Y_1) \otimes (X \oplus Y_2). \end{array}$$

Thus the situation is analogous to (IIb.3.1) and we get $\Theta(\varphi) \rhd \Theta(\varphi[\tau]_\nu)$.
If $A \neq B \vee C$ then $\chi = \tau$ and therefore $\Theta(\varphi) \rhd \Theta(\varphi')$.

If $A = B \vee C$ we define $\chi =$

$$\dfrac{\dfrac{\begin{array}{c}(\rho, X)\\ \Gamma \vdash \Delta\end{array} \quad \begin{array}{c}(\tau, (X \oplus Y_1) \otimes (X \oplus Y_2))\\ (B \vee C), \Gamma, \Pi^* \vdash \Delta^*, \Lambda\end{array}}{\Gamma, \Gamma^*, \Pi^* \vdash \Delta^*, \Delta^*, \Lambda}\ mix(A)}{\Gamma, \Pi^* \vdash \Delta^*, \Lambda}\ c:^*$$

In this case either

$$\begin{array}{rcl}\Theta(\varphi)/\nu & = & X \oplus (Y_1 \oplus Y_2),\\ \Theta(\varphi')/\nu & = & X \oplus ((X \oplus Y_1) \oplus (X \oplus Y_2)).\end{array}$$

or

$$\begin{array}{rcl}\Theta(\varphi)/\nu & = & X \oplus (Y_1 \otimes Y_2),\\ \Theta(\varphi')/\nu & = & X \oplus ((X \oplus Y_1) \otimes (X \oplus Y_2)).\end{array}$$

In the first case we obtain

$$\Theta(\varphi)/\nu \sim \Theta(\varphi')/\nu$$

and in the second one

$$\Theta(\varphi)/\nu \sqsubseteq \Theta(\varphi')/\nu$$

Once more Lemma 4.3 gives us $\Theta(\varphi) \rhd \Theta(\varphi')$.

(IIb.3.3) The last inference in σ is a mix. Then ψ is of the form

$$\dfrac{\begin{array}{c}(\rho, X)\\ \Gamma \vdash \Delta\end{array} \quad \dfrac{\begin{array}{c}(\sigma_1, Y_1)\\ \Pi_1 \vdash \Lambda_1\end{array} \quad \begin{array}{c}(\sigma_2, Y_2)\\ \Pi_2 \vdash \Lambda_2\end{array}}{\Pi_1, \Pi_2^+ \vdash \Lambda_1^+, \Lambda_2}\ mix(B)}{\Gamma, \Pi_1^*, \Pi_2^{+*} \vdash \Delta^*, \Lambda_1^+, \Lambda_2}\ mix(A)$$

If A occurs in Π_1 and in Π_2 then $\chi =$

$$\dfrac{\dfrac{\dfrac{\begin{array}{c}(\rho, X)\\ \Gamma \vdash \Delta\end{array} \quad \begin{array}{c}(\sigma_1, Y_1)\\ \Pi_1 \vdash \Lambda_1\end{array}}{\Gamma, \Pi_1^* \vdash \Delta^*, \Lambda_1}\ mix(A) \quad \dfrac{\begin{array}{c}(\rho, X)\\ \Gamma \vdash \Delta\end{array} \quad \begin{array}{c}(\sigma_2, Y_2)\\ \Pi_2 \vdash \Lambda_2\end{array}}{\Gamma, \Pi_2^* \vdash \Delta^*, \Lambda_2}\ mix(A)}{\Gamma, \Gamma^+, \Pi_1^*, \Pi_2^{+*} \vdash \Delta^{*+}, \Delta^*, \Lambda_1^+, \Lambda_2}\ mix(B)}{\Gamma, \Pi_1^*, \Pi_2^{+*} \vdash \Delta^*, \Lambda_1^+, \Lambda_2}\ c:^*, w:^*$$

In this case we have

$$\begin{array}{rcl}\Theta(\varphi)/\nu & = & X \oplus (Y_1 \oplus Y_2),\\ \Theta(\varphi')/\nu & = & (X \oplus Y_1) \oplus (X \oplus Y_2).\end{array}$$

Clearly $X \oplus (Y_1 \oplus Y_2) \sim (X \oplus Y_1) \oplus (X \oplus Y_2)$ and so

$$X \oplus (Y_1 \oplus Y_2) \rhd (X \oplus Y_1) \oplus (X \oplus Y_2).$$

By Lemma 4.3 we get $\Theta(\varphi) \rhd \Theta(\varphi')$.
If A occurs in Π_1 and not in Π_2 then $\chi =$

$$\cfrac{\cfrac{\begin{array}{c}(\rho, X)\\ \Gamma \vdash \Delta\end{array} \quad \begin{array}{c}(\sigma_1, Y_1)\\ \Pi_1 \vdash \Lambda_1\end{array}}{\Gamma, \Pi_1^* \vdash \Delta^*, \Lambda_1}\ mix(A) \quad \begin{array}{c}(\sigma_2, Y_2)\\ \Pi_2 \vdash \Lambda_2\end{array}}{\Gamma, \Pi_1^*, \Pi_2^+ \vdash \Delta^*, \Lambda_1^+, \Lambda_2}\ mix(B)$$

Here we have

$$\begin{array}{rcl}\Theta(\varphi)/\nu & = & X \oplus (Y_1 \oplus Y_2),\\ \Theta(\varphi')/\nu & = & (X \oplus Y_1) \oplus Y_2.\end{array}$$

and $\Theta(\varphi) \rhd \Theta(\varphi')$ is trivial.
The case where A is in Π_2, but not in Π_1 is completely symmetric. $\diamond$

THEOREM 6.1: *Let φ be an* **LK**-*deduction and ψ be an ACNF of φ under a cut reduction relation $>_{\mathcal{R}}$ based on $\mathcal{R}$. Then $\Theta(\varphi) \leq_{ss} \Theta(\psi)$.*

Proof: $\varphi >^*_{\mathcal{R}} \psi$. By Lemma 6.1 we get $\Theta(\varphi) \rhd^* \Theta(\psi)$. By Proposition 4.1 we obtain $\Theta(\varphi) \leq_{ss} \Theta(\psi)$. $\diamond$

THEOREM 6.2: *Let φ be an* **LK**-*derivation and ψ be an ACNF of φ under a cut reduction relation $>_{\mathcal{R}}$ based on $\mathcal{R}$. Then there exists a resolution refutation γ of* $\mathrm{CL}(\varphi)$ *s.t.* $\gamma \leq_{ss} \mathrm{RES}(\psi)$.

Proof: By Theorem 6.1 $\Theta(\varphi) \leq_{ss} \Theta(\psi)$ and therefore $\mathrm{CL}(\varphi) \leq_{ss} \mathrm{CL}(\psi)$. By Definition 5.7, $\mathrm{RES}(\psi)$ is a resolution refutation of $\mathrm{CL}(\psi)$; by Proposition 2.1 there exists a resolution refutation γ of $\mathrm{CL}(\varphi)$ s.t. $\gamma \leq_{ss} \mathrm{RES}(\psi)$. $\diamond$

COROLLARY 6.1: *Let φ be an* **LK**-*derivation and ψ be an ACNF of φ under a cut reduction relation $>_{\mathcal{R}}$ based on $\mathcal{R}$. Then there exists a resolution refutation γ of* $\mathrm{CL}(\varphi)$ *s.t.*
$l(\gamma) \leq l(\mathrm{RES}(\psi)) \leq l(\psi) * 2^{2*l(\psi)}$.

Proof: By Theorem 6.1 there exists a resolution refutation γ with $\gamma \leq_{ss} \mathrm{RES}(\psi)$. By definition of subsumption of proofs (see Definition 2.17) we have $l(\gamma) \leq l(\mathrm{RES}(\psi))$. Finally the result follows from Proposition 5.2. $\diamond$

COROLLARY 6.2: *Let φ be an* **LK**-*derivation and ψ be an ACNF of φ under a cut reduction relation $>_{\mathcal{R}}$ based on $\mathcal{R}$. Let χ be an ACNF of φ under CERES. Then*
$l(\chi) \leq l(\varphi) * l(\psi) * 2^{2*l(\psi)} + 2$.

Proof: If γ is a resolution refutation of $\mathrm{CL}(\varphi)$ then a ACNF χ of φ can be obtained by CERES using projection. As the **LK**-derivations in the projections are not longer than φ itself we get $l(\chi) \leq l(\varphi) * l(\gamma) + 2$ (the term "+2" comes from the final contractions $c : l, c : r$). Then the inequality follows from Corollary 6.1. $\diamond$

COROLLARY 6.3: *Let φ be an **LK**-derivation and ψ be an ACNF of φ under Gentzen's or Tait's method. Let χ be an ACNF of φ under CERES. Then*

$$l(\chi) \leq l(\varphi) * l(\psi) * 2^{2*l(\psi)} + 2.$$

Proof: Gentzen's and Tait's methods are reduction methods based on $\mathcal{R}$. $\diamond$

In (4) we have shown that cut-elimination based on CERES may be much faster that Gentzen's and Tait's method. The speed-up one can achieve is given by the complexity of cut-elimination itself, which is nonelementary. On the other hand, Corollary 6.3 shows that the computational expense of CERES is exponentially (and thus elementarily) bounded by that of Gentzen's or Tait's method. This shows that CERES is never "much slower" than the traditional methods, but there are sequences of derivations where it is substantially faster. Indeed, in some sense, Theorem 6.2 indicates that all cut-elimination methods based on $\mathcal{R}$ are *redundant* w.r.t. CERES.

7. Beyond $\mathcal{R}$: Stronger Pruning Methods

At the first glimpse it might appear that all cut-reduction methods based on a set of rules yield characteristic terms which are subsumed by the characteristic term of the original proof. However, Theorem 6.1 and Theorem 6.2 are not valid in general. Below we will define a set of cut-reduction rules $\mathcal{R}'$ for which the theorems above are not valid.

DEFINITION 7.1 ($\mathcal{R}'$): *Let $\mathcal{R}$ be the set of cut-reduction rules defined in the Appendix. With the exception of the rule in case 3.121.232 (right-rank > 1, case $\vee : l$) the rules in $\mathcal{R}'$ are the same as those in $\mathcal{R}$. We only modify the case where the mix formula A is identical to B (which is one of the auxiliary formulas of the $\vee : l$-inference). In this case the derivation ψ in case 3.121.232 is of the form:*

$$\dfrac{\begin{matrix}(\rho)\\ \Gamma \vdash \Delta\end{matrix} \qquad \dfrac{\begin{matrix}(\sigma_1)\\ B, \Pi \vdash \Lambda\end{matrix} \quad \begin{matrix}(\sigma_2)\\ C, \Pi \vdash \Lambda\end{matrix}}{B \vee C, \Pi \vdash \Lambda}\ \vee : l}{\Gamma, B \vee C, \Pi^* \vdash \Delta^*, \Lambda}\ mix(B)$$

We define $\psi >_{\mathcal{R}'} \chi$ for $\chi =$

$$\dfrac{\dfrac{\begin{matrix}(\rho)\\ \Gamma \vdash \Delta\end{matrix} \quad \begin{matrix}(\sigma_1)\\ B, \Pi \vdash \Lambda\end{matrix}}{\Gamma, \Pi^* \vdash \Delta^*, \Lambda}\ mix(B)}{\Gamma, B \vee C, \Pi^* \vdash \Delta^*, \Lambda}\ w : l$$

$\sharp$

Theorem 7.1: *There exists an* **LK***-derivation* φ *s.t. for all ACNFs* ψ *under* $\mathcal{R}'$*:*

(1) $\Theta(\varphi) \not\leq_{ss} \Theta(\psi)$*,*

(2) $\gamma \not\leq_{ss} \mathrm{RES}(\psi)$ *for all resolution refutations* γ *of* $\mathrm{CL}(\varphi)$*.*

Proof: In the **LK**-derivations below we mark all ancestors of mixes by $*$. Let P, Q, R be arbitrary atomic formulas and φ be the derivation

$$\dfrac{\dfrac{\vdash P^* \quad \vdash P^*}{\vdash (P \land P)^*}\ \land : r \qquad \dfrac{\dfrac{\dfrac{P, P^* \vdash P}{P, (P \land P)^* \vdash P}\ \land : l}{P \land P, (P \land P)^* \vdash P}\ \land : l \quad \dfrac{\dfrac{\dfrac{\dfrac{P^* \vdash Q^* \quad Q^* \vdash P}{P^* \vdash P}\ mix(Q)}{R, P^* \vdash P}\ w : l}{R, (P \land P)^* \vdash P}\ \land : l}{(P \land P)^*, (P \land P) \lor R \vdash P}\ \lor : l}{(P \land P) \lor R \vdash P}\ mix(P \land P)$$

Then

$$\begin{array}{rcl} \Theta(\varphi) & = & (\{\vdash P\} \oplus \{\vdash P\}) \oplus ((\{P \vdash\} \otimes (\{P \vdash Q\} \oplus \{Q \vdash\}), \\ \mathrm{CL}(\varphi) & = & \{\vdash P;\ \ P, P \vdash Q;\ \ P, Q \vdash\}. \end{array}$$

There exists only one non-atomic mix in φ. By definition of $\mathcal{R}'$ we get $\varphi >_{\mathcal{R}'} \chi$ (and this is the only one-step reduction) for $\chi =$

$$\dfrac{\dfrac{\dfrac{\vdash P^* \quad \vdash P^*}{\vdash (P \land P)^*}\ \land : r \qquad \dfrac{\dfrac{P^*, P^* \vdash P}{P^*, (P \land P)^* \vdash P}\ \land : l}{(P \land P)^*, (P \land P)^* \vdash P}\ \land : l}{\vdash P}\ mix(P \land P)}{(P \land P) \lor R \vdash P}\ w : l$$

It is easy to see that the only ACNF of χ (under $\mathcal{R}$ and $\mathcal{R}'$) is ψ for $\psi =$

$$\dfrac{\dfrac{\vdash P^* \quad P^*, P^* \vdash P}{\vdash P}\ mix(P)}{(P \land P) \lor R \vdash P}\ w : l$$

But

$$\begin{array}{rcl} \Theta(\psi) & = & \{\vdash P\} \oplus \{P, P \vdash\}, \\ \mathrm{CL}(\psi) & = & \{\vdash P;\ \ P, P \vdash\}. \end{array}$$

There exists no clause $C \in \mathrm{CL}(\varphi)$ with $C \leq_{ss} P, P \vdash$, therefore $\mathrm{CL}(\varphi) \not\leq_{ss} \mathrm{CL}(\psi)$ and $\Theta(\varphi) \not\leq_{ss} \Theta(\psi)$. This proves (1).
By definition of RES we obtain $\mathrm{RES}(\psi) =$

$$\dfrac{\vdash P \quad P, P \vdash}{\vdash}\ cut.$$

As $\mathrm{CL}(\varphi) \not\leq_{ss} \{P, P \vdash\}$ there exists no refutation γ of $\mathrm{CL}(\varphi)$ with $\gamma \leq_{ss} \mathrm{RES}(\psi)$. This proves (2). $\diamond$

Remark: Our choice of $\mathcal{R}'$ was in fact a minimal one, aimed to falsify Theorem 6.1. It is obvious that the principle can be extended to the case where $A = C$, and to the symmetric situation of left-rank > 1 and $\wedge : r$. Indeed there are several simple ways for further improving cut-elimination methods based on $\mathcal{R}$. All these stronger methods of pruning the proof trees during cut-reduction do not fulfil the properties expressed in Theorem 6.1 and in Theorem 6.2. $\sharp$

8. Conclusion

The main technical tool of this paper is the symbolic representation of clauses by terms composed from clauses and the operators $\oplus$ and $\otimes$. This tool enables the incorporation of information about the clauses extracted from proofs exceeding pure extensionality. To deal with various forms of pruning in the clausal framework even more information has to be included, i.e. the set of operators has to be extended. The analysis of cut-elimination via $\oplus$ and $\otimes$ in this paper has much in common with an approach of G. Mints to the construction of interpolants in first-order intuitionistic logic (see (10)). Thus there is some evidence that the use of abstract algebraic structures may lead to substantially new insights in the nature of proofs. In this sense this paper can be considered as a step towards an *algebraic proof theory*, which – like all reasonable algebraic approaches – has to deal with partial representations of the objects, whose interest is *discovered* rather than obvious from the first glance.

Acknowledgements:
We would like to thank Georg Moser and Agata Ciabattoni for their constructive criticism and for several comments which resulted in an improvement of this paper.

References

[1] Baaz, M. and Leitsch, A., On skolemization and proof complexity, *Fundamenta Informaticae*, (1994). **20**, 353–379.

[2] Baaz, M. and Leitsch, A., Fast Cut-Elimination by Projection, *Lecture Notes in Computer Science*, (1997). **1258**, 18–33.

[3] Baaz, M. and Leitsch, A., Cut normal forms and proof complexity, *Annals of Pure and Applied Logic*, (1999). **97**, 127–177.

[4] Baaz, M. and Leitsch, A., Cut-Elimination and Redundancy-Elimination by Resolution, *Journal of Symbolic Computation*, (2000). **29**, 149–176.

[5] Baaz, M. and Zach, R., Generalizing theorems in real closed fields, *Annals of Pure and Applied Logic*, (1995). **75**, 3–23.

[6] Gentzen, G., Untersuchungen über das logische Schließen, *Mathematische Zeitschrift*, (1934). **39**, 405–431.

[7] Girard, J.Y., *Proof Theory and Logical Complexity*, (1987). Bibliopolis, Napoli,

[8] Leitsch, A., *The Resolution Calculus*, (1997). Springer, Berlin Heidelberg New York,

[9] Luckhardt, H., Herbrand-Analysen zweier Beweise des Satzes von Roth: polynomiale Anzahlschranken, *The Journal of Symbolic Logic*, (1989). **54**, 234–263.

[10] Mints, G., Interpolation theorems for intuitionistic predicate logic, *Annals of Pure and Applied Logic*, (2002). **113**, 225–242.

[11] Robinson, J.A., A machine oriented logic based on the resolution principle, *Journal of the ACM*, (1965). **12**, 23–41.

[12] Schwichtenberg, H., Proof Theory: Some Applications of Cut-Elimination, *Handbook of Mathematical Logic*, Barwise, J., (1989). North Holland, 867–895.

[13] Statman, R., Lower bounds on Herbrand's theorem, *Proc. of the Amer. Math. Soc.*, (1979). **75**, 104–107.

[14] Tait, W.W., Normal derivability in classical logic, *The Syntax and Semantics of Infinitary Languages*, Barwise, J., (1968). Springer, 204–236.

[15] Takeuti, G., *Proof Theory*, (1987). North-Holland, Amsterdam, 2nd edition,

9. Appendix

Below we list the transformation rules used in Gentzen's proof of cut-elimination in (6). Thereby we use the same numbers for labelling the subcases. As we do *not* eliminate atomic cuts and our initial sequents are not necessarily of the form $A \vdash A$ some rules can be omitted. Moreover we need not consider the rules for implication as our version of **LK** is $\rightarrow$-free. If a mix-derivation ψ is transformed to ψ' then we define $\psi > \psi'$; note that ψ and ψ' have the same endsequent. Remember that the relation $>_{\mathcal{R}}$ is the crucial tool in defining Gentzen- and Tait reduction. In all reductions below ψ is a mix-derivation of the form

$$\frac{\begin{array}{cc}(\rho) & (\sigma)\\ \Gamma \vdash \Delta & \Pi \vdash \Lambda\end{array}}{\Gamma, \Pi^* \vdash \Delta^*, \Lambda}\ mix(A)$$

where A is a *non-atomic formula* (i.e. $comp(A) > 0$).

3.11. rank(ψ) = 2.

3.113.1. the last inference in ρ is $w : r$:

$$\dfrac{\dfrac{\begin{matrix}(\rho')\\ \Gamma \vdash \Delta\end{matrix}}{\Gamma \vdash \Delta, A^n}\, w:r \qquad \begin{matrix}(\sigma)\\ \Pi \vdash \Lambda\end{matrix}}{\Gamma, \Pi^* \vdash \Delta, \Lambda}\, mix(A)$$

transforms to

$$\dfrac{\begin{matrix}(\rho')\\ \Gamma \vdash \Delta\end{matrix}}{\Gamma, \Pi^* \vdash \Delta, \Lambda}\, w:l,r$$

3.113.2. the last inference in ψ_2 is $w : l$: symmetric to 3.113.1.

The last inferences in ρ, σ are logical ones and the mix-formula is the principal formula of these inferences:

3.113.31.

$$\dfrac{\dfrac{\begin{matrix}(\rho_1)\\ \Gamma \vdash \Delta, A\end{matrix} \quad \begin{matrix}(\rho_2)\\ \Gamma \vdash \Delta, B\end{matrix}}{\Gamma \vdash \Delta, A \wedge B}\, \wedge:r \qquad \dfrac{\begin{matrix}(\sigma')\\ A, \Pi \vdash \Lambda\end{matrix}}{A \wedge B, \Pi \vdash \Lambda}\, \wedge:l}{\Gamma, \Pi \vdash \Delta, \Lambda}\, mix(A \wedge B)$$

transforms to

$$\dfrac{\dfrac{\begin{matrix}(\rho_1)\\ \Gamma \vdash \Delta, A\end{matrix} \quad \begin{matrix}(\sigma')\\ A, \Pi \vdash \Lambda\end{matrix}}{\Gamma, \Pi^* \vdash \Delta^*, \Lambda}\, mix(A)}{\Gamma, \Pi \vdash \Delta, \Lambda}\, w:l,r$$

For the other form of $\wedge : l$ the transformation is straightforward.

3.113.32. The last inferences of ρ, σ are $\vee : r, \vee : l$: symmetric to 3.113.31.

3.113.33.

$$\dfrac{\dfrac{\begin{matrix}(\rho'(x/y))\\ \Gamma \vdash \Delta, B(x/y)\end{matrix}}{\Gamma \vdash \Delta, (\forall x)B(x)}\, \forall:r \qquad \dfrac{\begin{matrix}(\sigma')\\ B(x/t), \Pi \vdash \Lambda\end{matrix}}{(\forall x)B(x), \Pi \vdash \Lambda}\, \forall:l}{\Gamma, \Pi \vdash \Delta, \Lambda}\, mix((\forall x)B)$$

transforms to

$$\dfrac{\dfrac{\begin{matrix}(\rho'(x/t))\\ \Gamma \vdash \Delta, B(x/t)\end{matrix} \quad \begin{matrix}(\sigma')\\ B(x/t), \Pi \vdash \Lambda\end{matrix}}{\Gamma, \Pi^* \vdash \Delta^*, \Lambda}\, mix(B(x/t))}{\Gamma, \Pi \vdash \Delta, \Lambda}\, w:l,r$$

3.113.34. The last inferences in ρ, σ are $\exists : r, \exists : l$: symmetric to 3.113.33.

3.113.35

$$\dfrac{\dfrac{\begin{array}{c}(\rho')\\ A, \Gamma \vdash \Delta\end{array}}{\Gamma \vdash \Delta, \neg A}\ \neg : r \qquad \dfrac{\begin{array}{c}(\sigma')\\ \Pi \vdash \Lambda, A\end{array}}{\neg A, \Pi \vdash \Lambda}\ \neg : l}{\Gamma, \Pi \vdash \Delta, \Lambda}\ mix(\neg A)$$

reduces to

$$\dfrac{\dfrac{\begin{array}{c}(\sigma')\\ \Pi \vdash \Lambda, A\end{array} \quad \begin{array}{c}(\rho')\\ A, \Gamma \vdash \Delta\end{array}}{\Gamma^*, \Pi \vdash \Delta, \Lambda^*}\ mix(A)}{\Gamma, \Pi \vdash \Delta, \Lambda}\ w : l, r$$

3.12. rank$(\psi) > 2$:

3.121. rank$_r(\psi) > 1$:

3.121.1. The mix formula occurs in the antecedent of the end-sequent of ρ.

$$\dfrac{\begin{array}{c}(\rho)\\ \Gamma \vdash \Delta\end{array} \quad \begin{array}{c}(\sigma)\\ \Pi \vdash \Lambda\end{array}}{\Gamma, \Pi^* \vdash \Delta^*, \Lambda}\ mix(A)$$

transforms to

$$\dfrac{\begin{array}{c}(\sigma)\\ \Pi \vdash \Lambda\end{array}}{\Gamma, \Pi^* \vdash \Delta^*, \Lambda}\ w : l, r; c : l$$

3.121.2. The mix formula does not occur in the antecedent of the end-sequent of ρ.

3.121.21. Let ξ be one of the rules $w : l$ or $c : l$; then

$$\dfrac{\begin{array}{c}(\rho)\\ \Gamma \vdash \Delta\end{array} \quad \dfrac{\begin{array}{c}(\sigma')\\ \Sigma \vdash \Lambda\end{array}}{\Pi \vdash \Lambda}\ \xi}{\Gamma, \Pi^* \vdash \Delta^*, \Lambda}\ mix(A)$$

transforms to

$$\dfrac{\dfrac{\begin{array}{c}(\rho)\\ \Gamma \vdash \Delta\end{array} \quad \begin{array}{c}(\sigma')\\ \Sigma \vdash \Lambda\end{array}}{\Gamma, \Sigma^* \vdash \Delta^*, \Lambda}\ mix(A)}{\Gamma, \Pi^* \vdash \Delta^*, \Lambda}\ \xi$$

Note that ξ may be "degenerated", i.e. it can be skipped if the sequent does not change.

3.121.22. Let ξ be an arbitrary unary rule (different from $c : l, w : l$) and let C^* be empty if $C = A$ and C otherwise. The formulas B and C may be equal or different or simply nonexisting. Let us assume that ψ is of the form

$$\dfrac{\begin{array}{c}(\rho)\\ \Gamma \vdash \Delta\end{array} \quad \dfrac{\begin{array}{c}(\sigma')\\ B, \Pi \vdash \Sigma\end{array}}{C, \Pi \vdash \Lambda}\ \xi}{\Gamma, C^*, \Pi^* \vdash \Delta^*, \Lambda}\ mix(A)$$

Let τ be the proof

$$\dfrac{\dfrac{\dfrac{\begin{matrix}(\rho)\\ \Gamma \vdash \Delta\end{matrix} \quad \begin{matrix}(\sigma')\\ B, \Pi \vdash \Sigma\end{matrix}}{\Gamma, B^*, \Pi^* \vdash \Delta^*, \Sigma}\, mix(A)}{\Gamma, B, \Pi^* \vdash \Delta^*, \Sigma}\, (w:l)}{\Gamma, C, \Pi^* \vdash \Delta^*, \Lambda}\, \xi$$

3.121.221. $A \neq C$: then ψ transforms to τ.

3.121.222. $A = C$ and $A \neq B$: in this case C is the principal formula of ξ. Then ψ transforms to

$$\dfrac{\dfrac{\begin{matrix}(\rho)\\ \Gamma \vdash \Delta\end{matrix} \quad \begin{matrix}(\tau)\\ \Gamma, A, \Pi^* \vdash \Delta^*, \Lambda\end{matrix}}{\Gamma, \Gamma^*, \Pi^* \vdash \Delta^*, \Delta^*, \Lambda}\, mix(A)}{\Gamma, \Pi^* \vdash \Delta^*, \Lambda}\, c:l,r$$

3.121.223 $A = B = C$. Then $\Sigma \neq \Lambda$ and ψ transforms to

$$\dfrac{\dfrac{\begin{matrix}(\rho)\\ \Gamma \vdash \Delta\end{matrix} \quad \begin{matrix}(\sigma')\\ A, \Pi \vdash \Sigma\end{matrix}}{\Gamma, \Pi^* \vdash \Delta^*, \Sigma}\, mix(A)}{\Gamma, \Pi^* \vdash \Delta^*, \Lambda}\, \xi$$

3.121.23. The last inference in σ is binary:

3.121.231. The case $\wedge : r$. Here

$$\dfrac{\begin{matrix}(\rho)\\ \Gamma \vdash \Delta\end{matrix} \quad \dfrac{\begin{matrix}(\sigma_1)\\ \Pi \vdash \Lambda, B\end{matrix} \quad \begin{matrix}(\sigma_2)\\ \Pi \vdash \Lambda, C\end{matrix}}{\Pi, \vdash \Lambda, B \wedge C}\, \wedge : r}{\Gamma, \Pi^* \vdash \Delta^*, \Lambda, B \wedge C}\, mix(A)$$

transforms to

$$\dfrac{\dfrac{\begin{matrix}(\rho)\\ \Gamma \vdash \Delta\end{matrix} \quad \begin{matrix}(\sigma_1)\\ \Pi \vdash \Lambda, B\end{matrix}}{\Gamma, \Pi^* \vdash \Delta^*, \Lambda, B}\, mix(A) \quad \dfrac{\begin{matrix}(\rho)\\ \Gamma \vdash \Delta\end{matrix} \quad \begin{matrix}(\sigma_2)\\ \Pi \vdash \Lambda, C\end{matrix}}{\Gamma, \Pi^* \vdash \Delta^*, \Lambda, C}\, mix(A)}{\Gamma, \Pi^* \vdash \Delta^*, \Lambda, B \wedge C}\, \wedge : r$$

3.121.232. The case $\vee : l$. Then ψ is of the form

$$\dfrac{\begin{matrix}(\rho)\\ \Gamma \vdash \Delta\end{matrix} \quad \dfrac{\begin{matrix}(\sigma_1)\\ B, \Pi \vdash \Lambda\end{matrix} \quad \begin{matrix}(\sigma_2)\\ C, \Pi \vdash \Lambda\end{matrix}}{B \vee C, \Pi \vdash \Lambda}\, \vee : l}{\Gamma, (B \vee C)^*, \Pi^* \vdash \Delta^*, \Lambda}\, mix(A)$$

Again $(B \vee C)^*$ is empty if $A = B \vee C$ and $B \vee C$ otherwise.
We first define the proof τ:

$$\dfrac{\dfrac{\dfrac{\overset{(\rho)}{\Gamma \vdash \Delta} \quad \overset{(\sigma_1)}{B, \Pi \vdash \Lambda}}{B^*, \Gamma, \Pi^* \vdash \Delta^*, \Lambda}\,mix(A)}{B, \Gamma, \Pi^* \vdash \Delta^*, \Lambda}\,\xi \qquad \dfrac{\dfrac{\overset{(\rho)}{\Gamma \vdash \Delta} \quad \overset{(\sigma_2)}{C, \Pi \vdash \Lambda}}{C^*, \Gamma, \Pi^* \vdash \Delta^*, \Lambda}\,mix(A)}{C, \Gamma, \Pi^* \vdash \Delta^*, \Lambda}\,\xi}{B \vee C, \Gamma, \Pi^* \vdash \Delta^*, \Lambda}\,\vee : l$$

Note that, in case $A = B$ or $A = C$, the inference ξ is $w : l$; otherwise ξ is the identical transformation and can be dropped.
If $(B \vee C)^* = B \vee C$ then ψ transforms to τ.
If, on the other hand, $(B \vee C)^*$ is empty (i.e. $B \vee C = A$) then we transform ψ to

$$\dfrac{\dfrac{\overset{(\rho)}{\Gamma \vdash \Delta} \quad \tau}{\Gamma, \Gamma, \Pi^* \vdash \Delta^*, \Delta^*, \Lambda}\,mix(A)}{\Gamma, \Pi^* \vdash \Delta^*, \Lambda}\,c : l, r$$

3.121.234. The last inference in σ is $mix(B)$ for some formula B. Then ψ is of the form

$$\dfrac{\overset{(\rho)}{\Gamma \vdash \Delta} \quad \dfrac{\overset{(\sigma_1)}{\Pi_1 \vdash \Lambda_1} \quad \overset{(\sigma_2)}{\Pi_2 \vdash \Lambda_2}}{\Pi_1, {\Pi_2}^+ \vdash {\Lambda_1}^+, \Lambda_2}\,mix(B)}{\Gamma, {\Pi_1}^*, {\Pi_2}^{+*} \vdash \Delta^*, {\Lambda_1}^+, \Lambda_2}\,mix(A)$$

3.121.234.1 A occurs in Π_1 and in Π_2. Then ψ transforms to

$$\dfrac{\dfrac{\dfrac{\overset{(\rho)}{\Gamma \vdash \Delta} \quad \overset{(\sigma_1)}{\Pi_1 \vdash \Lambda_1}}{\Gamma, {\Pi_1}^* \vdash \Delta^*, \Lambda_1}\,mix(A) \qquad \dfrac{\overset{(\rho)}{\Gamma \vdash \Delta} \quad \overset{(\sigma_2)}{\Pi_2 \vdash \Lambda_2}}{\Gamma, {\Pi_2}^* \vdash \Delta^*, \Lambda_2}\,mix(A)}{\Gamma, \Gamma^+, {\Pi_1}^*, {\Pi_2}^{++} \vdash \Delta^{*+}, \Delta^*, {\Lambda_1}^+, \Lambda_2}\,mix(B)}{\Gamma, {\Pi_1}^*, {\Pi_2}^{+*} \vdash \Delta^*, {\Lambda_1}^+, \Lambda_2}\,c : l, r$$

Note that, for $A = B$, we have $\Pi^{*+} = \Pi^*$ and $\Delta^{*+} = \Delta^*$; $\Pi^{*+} = \Pi^{+*}$ holds in all cases.

3.121.234.2 A occurs in Π_1, but not in Π_2. In this case we have ${\Pi_2}^{+*} = {\Pi_2}^+$ and we transform ψ to

$$\dfrac{\dfrac{\overset{(\rho)}{\Gamma \vdash \Delta} \quad \overset{(\sigma_1)}{\Pi_1 \vdash \Lambda_1}}{\Gamma, {\Pi_1}^* \vdash \Delta^*, \Lambda_1}\,mix(A) \qquad \overset{(\sigma_2)}{\Pi_2 \vdash \Lambda_2}}{\Gamma, {\Pi_1}^*, {\Pi_2}^+ \vdash \Delta^*, {\Lambda_1}^+, \Lambda_2}\,mix(B)$$

3.121.234.3 A is in Π_2, but not in Π_1: symmetric to 3.121.234.2.

3.122. $\text{rank}_r(\psi) = 1$ and $\text{rank}_l(\psi) > 1$: symmetric to 3.121.