Comparing The Complexity of Cut-Elimination Methods

Matthias Baaz[1] and Alexander Leitsch[2]

[1] Institut für Computermathematik (E-118),
TU-Vienna, Wiedner Hauptstraße 8-10,
1040 Vienna, Austria
baaz@logic.at

[2] Institut für Computersprachen (E-185),
TU-Vienna, Favoritenstraße 9,
1040 Vienna, Austria
leitsch@logic.at

Abstract. We investigate the relative complexity of two different methods of cut-elimination in classical first-order logic, namely the methods of Gentzen and Tait. We show that the methods are incomparable, in the sense that both can give a nonelementary speed-up of the other one. More precisely we construct two different sequences of LK-proofs with cuts where cut-elimination for one method is elementary and nonelementary for the other one. Moreover we show that there is also a nonelementary difference in complexity for different deterministic versions of Gentzen's method.

1 Introduction

Gentzen's fundamental paper introduced cut-elimination as a fundamental procedure to extract proof theoretic information from given derivations such as Herbrand's Theorem, called Mid-Sequent Theorem in this context. In traditional proof theory, the general *possibility* to extract such informations is stressed, but there is less interest in applying the procedures in concrete cases. This, however, becomes essential if proof theory is considered as a basis for an *automated analysis of proofs*, which becomes important in connection with the development of effective program solving software for mathematical applications such as MATHEMATICA.

In this paper we compare the two most prominent cut-elimination procedures for *classical* logic: Gentzen's procedure and Tait's procedure; we avoid to call them "algorithms" because of their highly indeterministic aspects. From a procedural point of view, they are characterized by their different *cut-selection rule*: Gentzen's procedure selects a highest cut, while Tait's procedure selects a largest one (w.r.t. the number of connectives and quantifiers). The most important logical feature of Gentzen's procedure is, that – contrary to Tait's method – it transforms intuitionistic proofs into intuitionistic proofs (within **LK**) and there is no possibility to take into account classical logic when intended. Tait's procedure, on the other hand, *does not change the inner connections of the derivation*, it replaces cuts by smaller ones without reordering them.

In this paper, we use the sequence γ_n of LK-proofs corresponding to Statman's worst-case sequence to compare Gentzen's and Tait's procedure. The sequence γ_n is transformed twice: first into a sequence ψ_n where Tait's method speeds up Gentzen's nonelementarily, and second into a sequence ϕ_n giving the converse effect. As a complexity measure we take the total number of symbol ocurrences in reduction sequences of cut-elimination (i.e. all symbol occurrences in all proofs occurring during the cut-elimination procedure are measured). Both methods are nondeterministic in nature. But also different deterministic versions of one and the same method may differ quite strongly: we show that even two different deterministic versions of Gentzen's method differ nonelementarily (w.r.t. the total lengths of the corresponding reduction sequences).

Finally we would like to emphasize that the main goal of this paper is to give a comparison of different *cut-elimination methods*. It is not our intention to investigate, at the same time, the efficiency of calculi; for this reason we do not work with improved or computationally optimized versions of **LK**, but rather take a version of **LK** which is quite close to the original one.

2 Definitions and Notation

Definition 1 (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 atom 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 (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).*

Definition 3 (the calculus LK). *The initial sequents are $A \vdash A$ for PL-formulas A. In the rules of* **LK** *we always mark the auxiliary formulas (i.e. the formulas in the premis(ses) used for the inference) and the principal (i.e. the infered) formula using different marking symbols. Thus, in our definition, $\wedge$-introduction to the right takes the form*

$$\frac{\Gamma_1 \vdash A^+, \Delta \qquad \Gamma_2 \vdash \Delta_2, B^+}{\Gamma_1, \Gamma_2 \vdash \Delta_1, A \wedge B^*, \Delta_2}$$

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

$$\frac{\Gamma_1 \vdash \Delta_1, A \qquad \Gamma_2 \vdash \Delta_2, B}{\Gamma_1, \Gamma_2 \vdash \Delta_1, \Delta_2, A \wedge B}$$

Unlike Gentzen's version of **LK** *(see [5]) ours does not contain any "automatic" contractions (in this paper we do not consider intuitionistic logic). Instead we use the additive version of* **LK** *as in the book of Girard [6], combined with multiset structure for the sequents (this is exactly the version of* **LK** *used in [3]) By the definition of sequents over multisets we don't need the exchange rules. In our notation* Γ, Δ, Π *and* Λ *serve as metavariables for multisets of formulas;* $\vdash$ *is the separation symbol. For a complete list of the rules we refer to [3]; we only give three logical and three structural rules here.*

- *The logical rule* $\vee$*-introduction left:*

$$\frac{A, \Gamma \vdash \Delta \qquad B, \Pi \vdash \Lambda}{A \vee B, \Gamma, \Pi \vdash \Delta, \Lambda} \vee : l$$

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

$$\frac{\Gamma \vdash \Delta, A}{\Gamma \vdash \Delta, A \vee B} \vee : r1$$

$$\frac{\Gamma \vdash \Delta, B}{\Gamma \vdash \Delta, A \vee B} \vee : r2$$

- *The structural rules weakening left and right:*

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

- *The cut rule:*

$$\frac{\Gamma \vdash \Delta, A \qquad A, \Pi \vdash \Lambda}{\Gamma, \Pi \vdash \Delta, \Lambda} \, cut$$

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**. Let $\mathcal{A}$ be the set of sequents occurring at the leaf nodes 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 $\mathcal{A} \vdash_{LK} S$). If $\mathcal{A}$ is a set of initial sequents then we call ψ an **LK**-proof of S. Note that, in general, cut-elimination is only possible in **LK**-proofs.

We write

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

to express that ψ is a proof 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 4. *The length of a proof ω is defined by the number of symbol occurrences in ω and is denoted by $l(\omega)$.*

The famous proof of the cut-elimination property of LK is based on a double induction on rank and grade of a modified form of cut, namely the mix.

Definition 5 (mix). *Let $\Gamma \vdash \Pi$ and $\Delta \vdash \Lambda$ two sequents and A be a formula which occurs in Π and in Δ; let Π^*, Δ^* be Π, Δ without occurrences of A. Then the rule*

$$\frac{\Gamma \vdash \Pi \quad \Delta \vdash \Lambda}{\Gamma, \Delta^* \vdash \Pi^*, \Lambda}\ mix$$

is called a mix on A. Frequently we label the rule by $mix(A)$ to indicate that the mix is on A.

Definition 6. *Let ϕ be an* **LK***-proof and ψ be a subderivation of the form*

$$\frac{\begin{matrix}(\psi_1) \\ \Gamma_1 \vdash \Delta_1\end{matrix} \quad \begin{matrix}(\psi_2) \\ \Gamma_2 \vdash \Delta_2\end{matrix}}{\Gamma_1, \Gamma_2^* \vdash \Delta_1^*, \Delta_2}\ mix(A)$$

Then we call ψ a mix-derivation in ϕ; if the mix is a cut we speak about a cut-derivation. We define the grade *of ψ as $comp(A)$; the* left-rank *of ψ 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 *is defined in an analogous way. The* rank *of ψ is the sum of right-rank and left-rank.*

The cut-elimination method of Gentzen can be formalized as a reduction method consisting of rank- and grade reductions on **LK**-proofs. But also Tait's method can be defined in this way, but with another selection of a cut-derivation in the proof. In a slight abuse of language we speak about cut-reduction, even if the cuts are actually mixes.

Definition 7 (cut-reduction rule). *In Gentzen's proof a mix-derivation ψ is selected in an* **LK***-proof ϕ and replaced by a derivation ψ' (with the same end-sequent) s.t. the corresponding mix-derivation(s) in ψ' has either lower grade or lower rank than ψ. These replacements can be interpreted as a reduction relation on* **LK***-proofs. Following the lines of Gentzen's proof of the cut-elimination property in [5] we give a formal definition of the relation $>$ on* **LK***-proofs in the Appendix.*

Using $>$ we can define two proof reduction relations, $>_G$ for Gentzen reduction *and $>_T$ for* Tait reduction. *Let ϕ be a proof and let ψ be a mix-derivation in ϕ occurring at position λ (we write $\phi = \phi[\psi]_\lambda$); assume that $\psi > \psi'$.*

If λ is an occurrence of an uppermost mix *in ϕ then we define $\phi[\psi]_\lambda >_G \phi[\psi']_\lambda$. If λ is an occurrence of a* mix with maximal grade *in ϕ then we define $\phi[\psi]_\lambda >_T \phi[\psi']_\lambda$.*

Definition 8 (cut-reduction sequence). *Let $>_x$ be one of the reduction relations $>_T, >_G$ and ϕ be an* **LK***-proof. Then a sequence $\eta:\ \phi_1, \ldots, \phi_n$ is called a* cut-reduction sequence *on ϕ w.r.t. $>_x$ if the following conditions are fulfilled*

- $\phi_1 = \phi$ *and*
- $\phi_k >_x \phi_{k+1}$ *for* $k = 1, ..n-1$.

If ϕ_n is cut-free then η is called a cut-elimination sequence *on ϕ w.r.t. $>_x$.*

Note that $>_G$ is more liberal than the direct intepretation of Gentzen's induction proof as a nondeterministic algorithm. But in the speed-up by Gentzen's over Tait's procedure we use the "traditional" Gentzen procedure (where one uppermost cut is eliminated before other uppermost cuts are transformed); this makes our results even stronger. $>_T$ had to be adapted anyway, as the calculus in Tait's paper [9] is not **LK**.

Definition 9. *Let $e : \mathbb{N}^2 \to \mathbb{N}$ be the following function*

$$e(0, m) = m$$
$$e(n+1, m) = 2^{e(n,m)}.$$

A function $f :\ \mathbb{N}^k \to \mathbb{N}^m$ for $k, m \geq 1$ is called elementary *if there exists an $n \in \mathbb{N}$ and a Turing machine T computing f s.t. the computing time of T on input $(l_1, \ldots, l_n)$ is less or equal $e(n, |(l_1, \ldots, l_k)|)$ where $|\ |$ denotes the maximum norm on $\mathbb{N}^k$.*

The function $s : \mathbb{N} \to \mathbb{N}$ is defined as $s(n) = e(n, 1)$ for $n \in \mathbb{N}$.

Note that the functions e and s are nonelementary.

Definition 10 (NE-improvement). *Let η be a cut-elimination sequence. We denote by $\|\eta\|$ the number of all symbol occurrences in η (i.e. the symbolic length of η). Let $>_x$ and $>_y$ be two cut-reduction relations (e.g. $>_T$ and $>_G$). We say that $>_x$ NE-improves $>_y$ (NE stands for nonelementarily) if there exists a sequence of LK-proofs $(\gamma_n)_{n \in \mathbb{N}}$ with the following properties:*

1. *There exists an elementary function f s.t. for all n there exists a cut-elimination sequence η_n on γ_n w.r.t. $>_x$ with $\|\eta_n\| < f(n)$,*
2. *For all elementary functions g there exists an $m \in \mathbb{N}$ s.t. for all n with $n > m$ and for all cut-elimination sequences θ on γ_n w.r.t. $>_y$: $\|\theta\| > g(n)$.*

3 The Proof Sequence of R. Statman

In [8] Richard Statman proved the remarkable result, that there are sequences of formulas having short proofs, but a nonelementarily increasing Herbrand complexity. More formally, there exists a sequence $(S_n)_{n \in \mathbb{N}}$ of sequents having **LK**-proofs $(\gamma_n)_{n \in \mathbb{N}}$ with $l(\gamma_n) \leq n2^{an}$ for some constant a, but the Herbrand complexity of S_n is $> s(n)/2$ (see Definition 9). As S_n consists of prenex formulas (in fact of universal closures), Herbrand complexity is a lower bound on the length of a cut-free proof. Putting things together we obtain a sequence $(\gamma_n)_{n\in\mathbb{N}}$ of **LK**-proofs of $(S_n)_{n\in\mathbb{N}}$ with the following properties:

- There exist a constant a with $l(\gamma_n) \leq n2^{an}$,
- for all cut-free proofs ψ of S_n $l(\psi) > \frac{1}{2}s(n)$.

This yields that *every* method of cut-elimination on $(\gamma_n)_{n\in\mathbb{N}}$ must be of nonelementary expense. Our aim is to use γ_n in constructing new proofs ϕ_n in which the γ_n are in some sense "redundant"; this redundancy may be "detected" by one cut-elimination method (behaving elementarily on ϕ_n), but not by the other one (having thus nonelementary expense on ϕ_n).

For his result on Herbrand complexity Statman defines a sequence of problems in combinatory logic (expressing iterated exponentiation) together with an elegant sequence of short proofs. The detailed formalization of the short proofs in **LK** can be found in [1]. What we need here is the overall structure of the proofs γ_n, in particular the positions of the cuts; on the other hand we do not need atomic initial sequents like in [1].

The sequence S_n is of the form $\Delta_n \vdash D_n$ where Δ_n consists of a fixed sequence of closed equations + a linear number of equality axioms of at most exponential length. Instances of the formulas H_i defined below form the cut formulas in γ_n, where p is a constant symbol:

$$H_1(y) \equiv (\forall x_1)px_1 = p(yx_1),$$
$$H_{i+1}(y) \equiv (\forall x_{i+1})(H_i(x_{i+1}) \rightarrow H_i(yx_{i+1})).$$

From the definition of the H_i it is easy to derive $l(H_i) \leq 2^{bi}$ for all i and some constant b; moreover $comp(H_i) < 2^{i+1}$.

The sequence γ_n is of the form

$$\dfrac{H_1(q) \vdash H_1(q) \qquad \dfrac{\delta_n \qquad \overset{(\chi_1)}{H_2(\mathbf{T}_n), H_1(q) \vdash H_1(\mathbf{T}_n q)}}{\Gamma_{2n+1}, H_1(q) \vdash H_1(\mathbf{T}_n q)}\ cut}{\Delta_n \vdash H_1(\mathbf{T}_n q)}\ cut$$

where δ_n is

$$\dfrac{\overset{(\psi_2)}{\Gamma_{2n-1} \vdash H_2(\mathbf{T})} \qquad \dfrac{\overset{(\delta_{n-1})}{\Gamma_{2(n-1)} \vdash H_3(\mathbf{T}_{n-1})} \qquad \overset{(\chi_2)}{H_3(\mathbf{T}_{n-1}), H_2(\mathbf{T}) \vdash H_2(\mathbf{T}_n)}}{\Gamma_{2(n-1)}, H_2(\mathbf{T}) \vdash H_2(\mathbf{T}_n)}\ cut}{\Gamma_{2n} \vdash H_2(\mathbf{T}_n)}\ cut$$

$\mathbf{T}$ is a combinator defined from the basic combinators $\mathbf{S}, \mathbf{B}, \mathbf{C}, \mathbf{I}$ by $\mathbf{T} \equiv (\mathbf{SB})((\mathbf{CB})\mathbf{I})$ and $\mathbf{T}_{i+1} = \mathbf{T}_i\mathbf{T}$; $\mathbf{T}$ fulfils the equation $(\mathbf{T}y)x = y(yx)$. q is a constant symbol and $\Gamma_1, \Gamma_2, \Gamma_3$ are subsequences of Δ_n. For details we refer to [1]. The proofs ψ_i and χ_i are cutfree proofs with a fixed number of sequents; their length, however, is exponential as the formulas H_i grow in size exponentially. The cuts are hierarchically located in a linear structure: there is only one uppermost cut, i.e. all cuts occur below this cut; every cut has all other cuts as his predecessor or as his successor.

We give some informal description of the proofs ψ_i and χ_i, for greater detail we refer to [1].

First we describe the proof ψ_{i+2} for $i \geq 1$; the case ψ_2 is quite similar. By definition of the formulas H_i we have to show

$$(\forall x_{i+2})(H_{i+1}(x_{i+2}) \to H_{i+1}(\mathbf{T}x_{i+2})).$$

Using the definition of $H_{i+1}(y)$ for some variable y we first prove (in a fixed number of steps) from $H_{i+1}(y)$ that

$$(\forall x_{i+1})(H_i(x_{i+1}) \to H_i(y(yx_{i+1})))$$

holds. Then use an equational proof of $y(yx_{i+1}) = (\mathbf{T}y)x_{i+1}$ and insert it into the consequent of the upper implication via appropriate equational axioms. This way we obtain the formula

$$(\forall x_{i+1})(H_i(x_{i+1}) \to H_i((\mathbf{T}y)x_{i+1}))$$

which is just $H_{i+1}(Ty)$. So we have a derivation of

$$H_{i+1}(y) \to H_{i+1}(\mathbf{T}y).$$

By using y as eigenvariable and applying $\forall$-introduction we obtain

$$(\forall x_{i+2})(H_{i+1}(x_{i+2}) \to H_{i+1}(\mathbf{T}x_{i+2}))$$

which is $H_{i+2}(\mathbf{T})$. This eventually gives the proof ψ_{i+2}.

The proofs χ_i:
We prove the sequent

$$H_{i+1}(\mathbf{T}_{n-i+1}), H_i(\mathbf{T}) \vdash H_i(\mathbf{T}_{n-i+2})$$

by

$$\dfrac{\dfrac{H_i(\mathbf{T}) \vdash H_i(\mathbf{T}) \quad H_i(\mathbf{T}_{n-i+2}) \vdash H_i(\mathbf{T}_{n-i+2})}{H_i(\mathbf{T}) \to H_i(\mathbf{T}_{n-i+2}), H_i(\mathbf{T}) \vdash H_i(\mathbf{T}_{n-i+2})}\ \to: l}{(\forall x_{i+1})(H_i(x_{i+1}) \to H_i(\mathbf{T}_{n-i+1}x_{i+1})), H_i(\mathbf{T}) \vdash H_i(\mathbf{T}_{n-i+2})}\ \forall: l$$

which consists of 4 sequents only; the length of χ_i is exponential in i.

Roughly described, the elimination of the uppermost cut is only exponential and the elimination of a further cut below increases the degree of exponentiation. As there are linearly many cut formulas

$$H_1(q), H_2(\mathbf{T}), \ldots, H_{n+2}(\mathbf{T}) \text{ and } H_n(\mathbf{T}_2), \ldots, H_2(\mathbf{T}_n)$$

the resulting total expense is nonelementary.

4 Comparing the Methods of Gentzen and Tait

Our first result expresses the fact that a cut-elimination method selecting maximal cuts can be nonelementarily faster than methods selecting an uppermost cut.

Theorem 1. *Tait's method can give a nonelementary speed-up of Gentzen's method or more formally: $>_T$ NE-improves $>_G$.*

Proof. Let γ_n be Statman's sequence defined in Section 3. We know that the maximal complexity of cut formulas in γ_n is less than 2^{n+3}. Let $g(n) = 2^{n+3}$ and the formulas A_i be defined as

$$A_0 = A \text{ for an atom formula } A$$
$$A_{i+1} = \neg A_i \text{ for } i \in \mathbb{N}.$$

For every $n \in \mathbb{N}$ we set $E_n \equiv A_{g(n)}$. Then clearly $comp(E_n) = g(n)$ and thus is greater than the cut-complexity of γ_n. We will build E_n into a more complex formula, making this formula the main formula of a cut. For every $n \in \mathbb{N}$ let ψ_n be the **LK**-proof:

$$\dfrac{\dfrac{E_n \vdash E_n \quad \begin{matrix}(\gamma_n)\\ \Delta_n \vdash D_n\end{matrix}}{E_n, \Delta_n \vdash D_n \wedge E_n}\wedge : r \qquad \dfrac{\dfrac{E_n \vdash E_n \quad A \vdash A}{E_n \to A, E_n \vdash A}\to: l}{D_n \wedge E_n, E_n \to A \vdash A}\wedge : l}{E_n, \Delta_n, E_n \to A \vdash A}\ cut$$

By definition of γ_n the proofs γ_n and ψ_n contain only a linear number of sequents, where the size of each sequent is less or equal than $n2^{cn}$ for some constant independent of n. Consequently there exists a constant d s.t. $l(\psi_n) \leq n^2 2^{dn}$ for all n.

We now construct a cut-elimination sequence on ψ_n based on Tait's cut-reduction $>_T$. As $comp(E_n)$ is greater than the cut-complexity of γ_n, and $comp(D_n \wedge E_n) > comp(E_n)$, the most complex cut formula in ψ_n is $D_n \wedge E_n$. This formula is selected by Tait's method and we obtain $\psi_n >_T \psi'_n$ (via rule 3.113.31 in the appendix) for the proof ψ'_n below

$$\dfrac{\dfrac{E_n \vdash E_n \quad \dfrac{E_n \vdash E_n \quad A \vdash A}{E_n, E_n \to A \vdash A}\to: l}{E_n, E_n \to A \vdash A}\ cut}{E_n, \Delta_n, E_n \to A \vdash A}\ w : l^*$$

ψ'_n contains only one single cut with cut formula E_n. Now the left hand side of the cut consists of an atomic sequent only making rule 3.111. applicable. Moreover the cut with formula E_n is the only one in ψ'_n. So reduction can be applied via $>_T$ and we obtain $\psi'_n >_T \psi''_n$ for ψ''_n:

$$\dfrac{\dfrac{E_n \vdash E_n \quad A \vdash A}{E_n, E_n \to A \vdash A}\to: l}{E_n, \Delta_n, E_n \to A \vdash A}\ w : l^*$$

Therefore $\eta_n: \ \psi_n, \psi'_n, \psi''_n$ is a cut-elimination sequence based on $>_T$. It is easy to see that the lengths of the proofs decrease in every reduction step. So we obtain

$$\|\eta_n\| \leq n^2 2^{dn+2}.$$

In the second part of the proof we show that *every* cut-elimination sequence on ψ_n based on the relation $>_G$ is of nonelementary length in n.

Note that every cut in γ_n lies *above* the cut with cut formula $D_n \wedge E_n$. Therefore, in Gentzen's method, we have to eliminate all cuts in γ_n before eliminating the cut with $D_n \wedge E_n$. So every cut-elimination sequence on ψ_n based on $>_G$ must contain a proof of the form

$$\dfrac{\dfrac{E_n \vdash E_n \quad \overset{(\gamma_n^*)}{\Delta_n \vdash D_n}}{E_n, \Delta_n \vdash D_n \wedge E_n}\wedge : r \qquad \dfrac{\dfrac{E_n \vdash E_n \quad A \vdash A}{E_n \to A, E_n \vdash A}\to : l}{D_n \wedge E_n, E_n \to A \vdash A}\wedge : l}{E_n, \Delta_n, E_n \to A \vdash A}\mathit{cut}$$

where γ_n^* is a cut-free proof of $\Delta_n \vdash D_n$. But according to Statman's result we have $l(\gamma_n^*) > \frac{s(n)}{2}$. Clearly the length of γ_n^* is a lower bound on the length of every cut-elimination sequence on ψ_n based on $>_G$. Thus for all cut-elimination sequences θ on ψ_n w.r.t. $>_G$ we obtain

$$\|\theta\| > \frac{s(n)}{2}.$$

◇

A nonelementary speed-up is possible also the other way around. In this case it is an advantage to select the cuts from upwards instead by formula complexity.

Theorem 2. *Gentzen's method can give a nonelementary speed-up of Tait's method or more formally: $>_G$ NE-improves $>_T$.*

Proof. Consider Statman's sequence γ_n defined in Section 3. Locate the uppermost proof δ_1 in γ_n; note that δ_1 is identical to ψ_{n+1}. In γ_n we first replace the proof δ_1 (or ψ_{n+1}) of $\Gamma_{n+1} \vdash H_{n+1}(\mathbf{T})$ by the proof $\hat{\delta}_1$ below:

$$\dfrac{\dfrac{\overset{(\omega)}{P \wedge \neg P}}{P \wedge \neg P \vdash Q}w : r \qquad \dfrac{\overset{(\psi_{n+1})}{\Gamma_1 \vdash H_{n+1}(\mathbf{T})}}{Q, \Gamma_1 \vdash H_{n+1}(\mathbf{T})}w : l}{P \wedge \neg P, \Gamma_1 \vdash H_{n+1}(\mathbf{T})}\mathit{cut}$$

The subproof ω is a proof of $P \wedge \neg P \vdash$ of constant length. Furthermore we use the same inductive definition in defining $\hat{\delta}_k$ as that of δ_k in Section 3. Finally we obtain a proof ϕ_n in place of γ_n. Note that ϕ_n differs from γ_n only by an additional (atomic) cut and the formula $P \wedge \neg P$ in the antecedents of sequents. Clearly

$$l(\phi_n) \leq l(\gamma_n) + cn$$

for some constant c.

Our aim is to define a cut-elimination sequence on ϕ_n w.r.t. $>_G$ which is of elementary complexity. Let S_k be the end sequent of the proof $\hat{\delta}_k$. We first investigate cut-elimination on the proof $\hat{\delta}_n$; the remaining two cuts are eliminated in a similar way. To this aim we prove by induction on k:

(∗) There exists a cut-elimination sequence $\hat{\delta}_{k,1}, \ldots, \hat{\delta}_{k,m}$ of $\hat{\delta}_k$ w.r.t. $>_G$ with the following properties:
 (1) $m \leq l(\hat{\delta}_k)$,
 (2) $l(\hat{\delta}_{k,i}) \leq l(\hat{\delta}_k)$ for $i = 1, \ldots, m$,
 (3) $\hat{\delta}_{k,m}$ is of the form

$$\cfrac{\begin{array}{c}(\omega)\\ P \wedge \neg P \vdash\end{array}}{S_k}\; w:^{*}$$

Induction basis $k = 1$:
In $\hat{\delta}_1$ there is only one cut (with the formula Q) where the cut formula is introduced by weakening. Thus by definition of $>_G$, using the rule 3.113.1, we get $\hat{\delta}_1 >_G \hat{\delta}_{1,2}$ where $\hat{\delta}_{1,2}$ is the proof

$$\cfrac{\begin{array}{c}(\omega)\\ P \wedge \neg P \vdash\end{array}}{P \wedge \neg P, \Gamma_{n+1} \vdash H_{n+1}(\mathbf{T})}\; w:^{*}$$

Clearly $2 \leq l(\hat{\delta}_1)$ and $l(\hat{\delta}_{1,2}) \leq l(\hat{\delta}_1)$. Moreover $\hat{\delta}_{1,2}$ is of the form (3). This gives (∗) for $k = 1$.

(IH) Assume that (∗) holds for k.

By definition, $\hat{\delta}_{k+1}$ is of the form

$$\cfrac{\begin{array}{c}(\psi_{n-k+1})\\ \Gamma_{2k+1} \vdash H_{n-k+1}(\mathbf{T})\end{array} \quad \rho_k}{P \wedge \neg P, \Gamma_{2(k+1)} \vdash H_{n-k+1}(\mathbf{T}_{k+1})}\; cut$$

for ρ_k:

$$\cfrac{\begin{array}{c}(\hat{\delta}_k)\\ P \wedge \neg P, \Gamma_{2k} \vdash H_{n-k+2}(\mathbf{T}_k)\end{array} \quad \begin{array}{c}(\chi_{n-k+1})\\ H_{n-k+2}(\mathbf{T}_k), H_{n-k+1}(\mathbf{T}) \vdash H_{n-k+1}(\mathbf{T}_{k+1})\end{array}}{P \wedge \neg P, \Gamma_{2k}, H_{n-k+1}(\mathbf{T}) \vdash H_{n-k+1}(\mathbf{T}_{k+1})}\; cut$$

By (IH) there exists a cut-elimination sequence $\hat{\delta}_{k,1}, \ldots, \hat{\delta}_{k,m}$ on $\hat{\delta}_k$ w.r.t. $>_G$ fulfilling (1), (2) and (3). In particular we have $l(\hat{\delta}_{k,m}) \leq l(\hat{\delta}_k)$ and $\hat{\delta}_{k,m}$ is of the form

$$\cfrac{\begin{array}{c}(\omega)\\ P \wedge \neg P\end{array}}{S_k}\; w:^{*}$$

All formulas in S_k, except $P \wedge \neg P$, are introduced by weakening in $\hat{\delta}_{k,m}$. In particular this holds for the formula $H_{n-k+2}(\mathbf{T}_k)$ which is a cut formula in $\hat{\delta}_{k+1}$. After cut-elimination on $\hat{\delta}_k$ the proof ρ_k is transformed (via $>_G$) into a proof $\hat{\rho}_k$:

$$\frac{\begin{array}{cc}(\hat{\delta}_{k,m}) & (\chi_{n-k+1}) \\ P \wedge \neg P, \Gamma_{2k} \vdash H_{n-k+2}(\mathbf{T}_k) & H_{n-k+2}(\mathbf{T}_k), H_{n-k+1}(\mathbf{T}) \vdash H_{n-k+1}(\mathbf{T}_{k+1})\end{array}}{P \wedge \neg P, \Gamma_{2k}, H_{n-k+1}(\mathbf{T}) \vdash H_{n-k+1}(\mathbf{T}_{k+1})}\ cut$$

Now the (only) cut in $\hat{\rho}_k$ is with the cut formula $H_{n-k+2}(\mathbf{T}_k)$ which is introduced by $w : r$ in $\hat{\delta}_{k,m}$. By using iterated reduction of left-rank via the symmetric versions of 3.121.21 and 3.121.22 in the appendix, the cut is eliminated and the proof χ_{n-k+1} "disappears" and the result is again of the form

$$\frac{\begin{array}{c}(\omega) \\ P \wedge \neg P\end{array}}{P \wedge \neg P, \Gamma_{2k}, H_{n-k+1}(\mathbf{T}) \vdash H_{n-k+1}(\mathbf{T}_{k+1})}\ w :^*$$

The proof above is the result of a cut-elimination sequence $\hat{\rho}_{k,1}, \ldots, \hat{\rho}_{k,p}$ on $\hat{\rho}_k$ w.r.t. $>_G$. But then also δ_{k+1} is further reduced to a proof where $\hat{\rho}_k$ is replaced by $\hat{\rho}_{k,p}$; in this proof there is only one cut left (with the formula $H_{n-k+1}(\mathbf{T})$ and we may play the "weakening game" once more. Finally we obtain a proof $\hat{\delta}_{k,r}$ of the form

$$\frac{\begin{array}{c}(\omega) \\ P \wedge \neg P\end{array}}{P \wedge \neg P, \Gamma_{2(k+1)} \vdash H_{n-k+1}(\mathbf{T}_{k+1})}\ w :^*$$

The conditions (1) and (2) are obviously fulfilled. This eventually gives $(*)$.

After the reduction of ϕ_n to $\phi_n[\hat{\delta}_{n,s}]_\lambda$, where λ is the position of $\hat{\delta}_n$ and $\hat{\delta}_{n,s}$ is the result of a Gentzen cut-elimination sequence on $\hat{\delta}_n$, there are only two cuts left. Again these cuts are swallowed by the proofs beginning with ω and followed by a sequence of weakenings. Putting things together we obtain a cut-elimination sequence

$$\eta_n :\ \phi_{n,1}, \ldots, \phi_{n,q}$$

on ϕ_n w.r.t. $>_G$ with the properties:

(1) $l(\phi_{n,i}) \leq l(\phi_n)$ and
(2) $q \leq l(\phi_n)$.

But then

$$\|\eta_n\| \leq l(\phi_n)^2 \leq n^4 2^{cn}.$$

for an appropriate constant c. Therefore η_n is a Gentzen cut-elimination sequence on ϕ_n of elementary complexity.

For the other direction consider Tait's reduction method on the sequence ϕ_n. The cut formulas in ϕ_n fall into two categories;

- the new cut formula Q with $comp(Q) = 0$ and

– the old cut formulas from γ_n.

Now let η be an arbitrary cut-elimination sequence on ϕ_n w.r.t. $>_T$. By definition of $>_T$ only cuts with maximal cut formulas can be selected in a reduction step w.r.t. $>_T$. Therefore, there exists a constant k s.t. $k \geq 1$ and η contains a proof ψ with cut-complexity k. As the new cut in ϕ_n with cut formula Q is of complexity 0, it is still present in ψ.

A straightforward proof transformation gives a proof χ s.t. $\gamma_n >_T^* \chi$, the cut-complexity of χ is k, and $l(\chi) < l(\psi)$ (in some sense the Tait procedure does not "notice" the new atomic cut). But every cut-free proof of γ_n has a length $> \frac{s(n)}{2}$ and cut-elimination of cuts with (fixed) complexity k is elementary [7].

More precisely there exists an elementary function f and a cut-elimination sequence θ on χ w.r.t. $>_T$ s.t. $\|\theta\| \leq f(l(\chi))$. This is only possible if there is no elementary bound on $l(\chi)$ in terms of n (otherwise we would get cut-free proofs of γ_n of length elementarily in n). But then there is no elementary bound on $l(\psi)$ in terms of n. Putting things together we obtain that for every elementary function f and for *every* cut-elimination sequence η on ϕ_n

$$\|\eta\| > f(n) \text{ almost everywhere .}$$

$\diamond$

Theorem 2 shows that there exist cut elimination sequences η_n on ϕ_n w.r.t. $>_G$ s.t. $\|\eta_n\|$ is elementarily bounded in n; however this does not mean that *every* cut-elimination sequence on ϕ_n w.r.t. $>_G$ is elementary. In fact $>_G$ is highly "unstable" in its different deterministic versions. Consider the subproof $\hat{\delta}_1$ in the proof of Theorem 2:

$$\dfrac{\dfrac{\begin{matrix}(\omega)\\ P \wedge \neg P\end{matrix}}{P \wedge \neg P \vdash Q}\ w:r \qquad \dfrac{\begin{matrix}(\psi_{n+1})\\ \Gamma_{n+1} \vdash H_{n+1}(\mathbf{T})\end{matrix}}{Q, \Gamma_{n+1} \vdash H_{n+1}(\mathbf{T})}\ w:l}{P \wedge \neg P, \Gamma_{n+1} \vdash H_{n+1}(\mathbf{T})}\ cut$$

If, in $>_G$, we focus on the weakening $(w : l)$ in the right part of the cut and apply rule 3.113.2 (appendix) we obtain $\hat{\delta}_1 >_G \mu$, where μ is the proof

$$\dfrac{\begin{matrix}(\psi_{n+1})\\ \Gamma_{n+1} \vdash H_{n+1}(\mathbf{T})\end{matrix}}{P \wedge \neg P, \Gamma_{n+1} \vdash H_{n+1}(\mathbf{T})}\ w:l$$

But μ contains the whole proof ψ_{n+1}. In the course of cut-elimination ψ_{n+1} is built into the produced proofs exactly as in the cut-elimination procedure on γ_n itself. The resulting cut-free proof is in fact longer than γ_n^* (the corresponding cut-free proof of the n-th element of Statman's sequence) and thus is of nonelementary length! This tells us that there are different deterministic versions α_1 and α_2 of $>_G$ s.t. α_1 gives a nonelementary speed-up of α_2 on the input set $(\phi_n)_{n \in \mathbb{N}}$.

In the introduction of additional cuts into Statman's proof sequence we use the weakening rule. Similar constructions can be carried out in versions of the Gentzen calculus without weakening. What we need is just a sequence of short **LK**-proofs of valid sequents containing "simple" redundant (in our case atomic) formulas on both sides serving as cut formulas. Note that **LK** without any redundancy (working with minimally valid sequents only) is not complete.

5 Conclusion

The main results of this paper hint to a more general theory of algorithmic cut elimination encompassing not only algorithmic specifications of Gentzen's and Tait's procedures but also approaches as cut projection [2] and the resolution based method CERES [4] . From a more proof theoretic point of view, the sequences of proofs arising from the different stages of cut-elimination can be considered as a specific analysis of the proof which extends the information obtainable from the cut-free final stage. Different cut-elimination algorithms stress different aspects of the proof, e.g. constructive content (Gentzen's procedure) or connectivity (Tait's procedure).

References

1. M. Baaz, A. Leitsch: On skolemization and proof complexity, *Fundamenta Informaticae*, 20(4), pp. 353–379, 1994.
2. M. Baaz, A. Leitsch: Fast Cut-Elimination by Projection, Proc. CSL'96, *Lecture Notes in Computer Science* 1258, pp. 18–33, Springer, Berlin, 1997.
3. M. Baaz, A. Leitsch: Cut normal forms and proof complexity, *Annals of Pure and Applied Logic*, 97, pp. 127-177, 1999.
4. M. Baaz, A. Leitsch: Cut-Elimination and Redundancy-Elimination by Resolution, *Journal of Symbolic Computation*, 29, pp. 149-176, 2000.
5. G. Gentzen: Untersuchungen über das logische Schließen, *Mathematische Zeitschrift* 39, pp. 405–431, 1934–1935.
6. J.Y. Girard: Proof Theory and Logical Complexity, in *Studies in Proof Theory*, Bibliopolis, Napoli, 1987.
7. H. Schwichtenberg: Proof Theory: Some Applications of Cut-Elimination, in *Handbook of Mathematical Logic*, ed. by J. Barwise, North Holland, Amsterdam, pp. 867–895, 1989.
8. R. Statman: Lower bounds on Herbrand's theorem, in *Proc. of the Amer. Math. Soc.* 75, pp. 104–107, 1979.
9. W.W. Tait: Normal derivability in classical logic, in *The Syntax and Semantics of Infinitary Languages*, ed. by J. Barwise, Springer, Berlin, pp. 204–236, 1968.
10. G. Takeuti: Proof Theory, North-Holland, Amsterdam, 2nd edition, 1987.

6 Appendix

Below we list the transformation rules used in Gentzen's proof of cut-elimination in [5]. Thereby we use the same numbers for labelling the subcases. Note that our rules slightly differ from that of Gentzen as we use the purely additive version of **LK**. If a mix-derivation ψ is transformed to ψ' then we define $\psi > \psi'$; remember that the relation $>$ is the crucial tool in defining Gentzen- and Tait reduction. In all reductions below ψ is a mix-derivation of the form

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

3.11. rank = 2.

3.111. $\psi_1 = A \vdash A$:

$$\dfrac{\begin{array}{cc} & (\psi_2)\\ A \vdash A & \Delta \vdash \Lambda\end{array}}{A, \Delta^* \vdash \Lambda}\ mix(A)$$

transforms to

$$\dfrac{\begin{array}{c}(\psi_2)\\ \Delta \vdash \Lambda\end{array}}{A, \Delta^* \vdash \Lambda}\ c:l^*$$

3.112. $\psi_2 = A \vdash A$: analogous to 3.111.

3.113.1. the last inference in ψ_1 is $w:r$:

$$\dfrac{\dfrac{\begin{array}{c}(\chi_1)\\ \Gamma \vdash \Delta\end{array}}{\Gamma \vdash \Delta, A}\ w:r \qquad \begin{array}{c}(\psi_2)\\ \Pi \vdash \Lambda\end{array}}{\Gamma, \Pi^* \vdash \Delta, \Lambda}\ mix(A)$$

transforms to

$$\dfrac{\begin{array}{c}(\chi_1)\\ \Gamma \vdash \Delta\end{array}}{\Gamma, \Pi^* \vdash \Delta, \Lambda}\ w:l^*$$

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

The last inferences in ψ_1, ψ_2 are logical ones and the mix-formula is the principal formula of these inferences:

3.113.31.

$$\dfrac{\dfrac{\begin{array}{cc}(\chi_1) & (\chi_2)\\ \Gamma_1 \vdash \Theta_1, A & \Gamma_2 \vdash \Theta_2, B\end{array}}{\Gamma_1, \Gamma_2 \vdash \Theta_1, \Theta_2, A \wedge B}\ \wedge:r \qquad \dfrac{\begin{array}{c}(\chi_3)\\ A, \Gamma_3 \vdash \Theta_3\end{array}}{A \wedge B, \Gamma_3 \vdash \Theta_3}\ \wedge:l}{\Gamma_1, \Gamma_2, \Gamma_3 \vdash \Theta_1, \Theta_2, \Theta_3}\ mix(A \wedge B)$$

transforms to

$$\dfrac{\dfrac{\begin{array}{cc}(\chi_1) & (\chi_3)\\ \Gamma_1 \vdash \Theta_1, A & A, \Gamma_3 \vdash \Theta_3\end{array}}{\Gamma_1, \Gamma_3^* \vdash \Theta_1^*, \Theta_3}\ mix(A)}{\Gamma_1, \Gamma_2, \Gamma_3 \vdash \Theta_1, \Theta_2, \Theta_3}\ w:^*$$

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

3.113.32. The last inferences of ψ_1, ψ_2 are $\vee : r, \vee : l$: symmetric to 3.113.31.

3.113.33.

$$\dfrac{\dfrac{\begin{matrix}(\chi_1[\alpha])\\ \Gamma_1 \vdash \Theta_1, B^x_\alpha\end{matrix}}{\Gamma_1 \vdash \Theta_1, (\forall x)B}\,\forall : r \qquad \dfrac{\begin{matrix}(\chi_2)\\ B^x_t, \Gamma_2 \vdash \Theta_2\end{matrix}}{(\forall x)B, \Gamma_2 \vdash \Theta_2}\,\forall : l}{\Gamma_1, \Gamma_2 \vdash \Theta_1, \Theta_2}\, mix((\forall x)B)$$

transforms to

$$\dfrac{\dfrac{\begin{matrix}(\chi_1[t])\\ \Gamma_1 \vdash \Theta_1, B^x_t\end{matrix} \quad \begin{matrix}(\chi_2)\\ B^x_t, \Gamma_2 \vdash \Theta_2\end{matrix}}{\Gamma_1, \Gamma_2^* \vdash \Theta_1^*, \Theta_2}\, mix(B^x_t)}{\Gamma_1, \Gamma_2 \vdash \Theta_1, \Theta_2}\, w :^*$$

3.113.34. The last inferences in ψ_1, ψ_2 are $\exists : r, \exists : l$: symmetric to 3.113.33.

3.113.35

$$\dfrac{\dfrac{\begin{matrix}(\chi_1)\\ A, \Gamma_1 \vdash \Theta_1\end{matrix}}{\Gamma_1 \vdash \Theta_1, \neg A}\,\neg : r \qquad \dfrac{\begin{matrix}(\chi_2)\\ \Gamma_2 \vdash \Theta_2, A\end{matrix}}{\neg A, \Gamma_2 \vdash \Theta_2}\,\neg : l}{\Gamma_1, \Gamma_2 \vdash \Theta_1, \Theta_2}\, mix(\neg A)$$

reduces to

$$\dfrac{\dfrac{\begin{matrix}(\chi_2)\\ \Gamma_2 \vdash \Theta_2, A\end{matrix} \quad \begin{matrix}(\chi_1)\\ A, \Gamma_1 \vdash \Theta_1\end{matrix}}{\Gamma_1^*, \Gamma_2 \vdash \Theta_1, \Theta_2^*}\, mix(A)}{\Gamma_1, \Gamma_2 \vdash \Theta_1, \Theta_2}\, w :^*$$

3.113.36.

$$\dfrac{\dfrac{\begin{matrix}(\chi_1)\\ A, \Gamma_1 \vdash \Theta_1, B\end{matrix}}{\Gamma_1 \vdash \Theta_1, A \to B}\,\to: r \qquad \dfrac{\begin{matrix}(\chi_2)\\ \Gamma \vdash \Theta, A\end{matrix} \quad \begin{matrix}(\chi_3)\\ B, \Delta \vdash \Lambda\end{matrix}}{A \to B, \Gamma, \Delta \vdash \Theta, \Lambda}\,\to: l}{\Gamma_1, \Gamma, \Delta \vdash \Theta_1, \Theta, \Lambda}\, mix(A \to B)$$

reduces to

$$\dfrac{\dfrac{\begin{matrix}(\chi_2)\\ \Gamma \vdash \Theta, A\end{matrix} \quad \dfrac{\begin{matrix}(\chi_1)\\ A, \Gamma_1 \vdash \Theta_1, B\end{matrix} \quad \begin{matrix}(\chi_3)\\ B, \Delta \vdash \Lambda\end{matrix}}{A, \Gamma_1, \Delta^* \vdash \Theta_1^*, \Lambda}\, mix(B)}{\Gamma_1^+, \Gamma, \Delta^{*+} \vdash \Theta_1^*, \Theta^+, \Lambda}\, mix(A)}{\Gamma_1, \Gamma, \Delta \vdash \Theta_1, \Theta, \Lambda}\, w :^*$$

3.12. rank > 2:

3.121. right-rank > 1:

3.121.1. The mix formula occurs in the antecedent of the end-sequent of ψ_1.

$$\dfrac{\begin{matrix}(\psi_1)\\ \Pi \vdash \Sigma\end{matrix} \quad \begin{matrix}(\psi_2)\\ \Delta \vdash \Lambda\end{matrix}}{\Pi, \Delta^* \vdash \Sigma^*, \Lambda}\, mix(A)$$

transforms to

$$\dfrac{\begin{matrix}(\psi_2)\\ \delta \vdash \Lambda\end{matrix}}{\Pi, \Delta^* \vdash \Sigma^*, \Lambda}\; w:^* \; c:*$$

3.121.2. The mix formula does not occur in the antecedent of the end-sequent of ψ_1.

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

$$\dfrac{\begin{matrix}(\psi_1)\\ \Pi \vdash \Sigma\end{matrix} \quad \dfrac{\begin{matrix}(\chi_1)\\ \Delta \vdash \Lambda\end{matrix}}{\Theta \vdash \Lambda}\, r}{\Pi, \Theta^* \vdash \Sigma^*, \Lambda}\; mix(A)$$

transforms to

$$\dfrac{\dfrac{\begin{matrix}(\psi_1)\\ \Pi \vdash \Sigma\end{matrix} \quad \begin{matrix}(\chi_1)\\ \Delta \vdash \Lambda\end{matrix}}{\Pi, \Delta^* \vdash \Sigma^*, \Lambda}\; mix(A)}{\Pi, \Theta^* \vdash \Sigma^*, \Lambda}\; r$$

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

3.121.22. Let r 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

$$\dfrac{\begin{matrix}(\psi_1)\\ \Pi \vdash \Sigma\end{matrix} \quad \dfrac{\begin{matrix}(\chi_1)\\ B, \Gamma \vdash \Omega_1\end{matrix}}{C, \Gamma \vdash \Omega_2}\, r}{\Pi, C^*, \Gamma^* \vdash \Sigma^*, \Omega_2}\; mix(A)$$

Let τ be the proof

$$\dfrac{\dfrac{\dfrac{\begin{matrix}(\psi_1)\\ \Pi \vdash \Sigma\end{matrix} \quad \begin{matrix}(\chi_1)\\ B, \Gamma \vdash \Omega_1\end{matrix}}{\Pi, B^*, \Gamma^* \vdash \Sigma^*, \Omega_2}\; mix(A)}{\Pi, B, \Gamma^* \vdash \Sigma^*, \Omega_2}\; w:^*}{\Pi, C, \Gamma^* \vdash \Sigma^*, \Omega_2}\; r$$

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 r. Then ψ transforms to

$$\dfrac{\dfrac{\begin{matrix}(\psi_1)\\ \Pi \vdash \Sigma\end{matrix} \quad \begin{matrix}(\tau)\\ \Pi, A, \Gamma^* \vdash \Sigma^*, \Omega_2\end{matrix}}{\Pi, \Pi^*, \Gamma^* \vdash \Sigma^*, \Sigma^*, \Omega_2}\; mix(A)}{\Pi, \Gamma^* \vdash \Sigma^*, \Omega_2}\; c:^*$$

3.121.223 $A = B = C$. Then $\Omega_1 \neq \Omega_2$ and ψ transforms to

$$\dfrac{\dfrac{\begin{matrix}(\psi_1)\\ \Pi \vdash \Sigma\end{matrix}\quad\begin{matrix}(\chi_1)\\ A,\Gamma \vdash \Omega_1\end{matrix}}{\Pi,\Gamma^* \vdash \Sigma^*,\Omega_1}\,mix(A)}{\Pi,\Gamma^* \vdash \Sigma^*,\Omega_2}\,r$$

3.121.23. The last inference in ψ_2 is binary:

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

$$\dfrac{\begin{matrix}(\psi_1)\\ \Pi \vdash \Sigma\end{matrix}\quad \dfrac{\begin{matrix}(\chi_1)\\ \Gamma_1 \vdash \Theta_1, B\end{matrix}\quad\begin{matrix}(\chi_2)\\ \Gamma_2 \vdash \Theta_2, C\end{matrix}}{\Gamma_1,\Gamma_2 \vdash \Theta_1,\Theta_2, B\wedge C}\,\wedge : r}{\Pi,\Gamma_1^*,\Gamma_2^* \vdash \Sigma^*,\Theta_1,\Theta_2,B\wedge C}\,mix(A)$$

transforms to

$$\dfrac{\dfrac{\dfrac{\begin{matrix}(\psi_1)\\ \Pi \vdash \Sigma\end{matrix}\quad\begin{matrix}(\chi_1)\\ \Gamma_1 \vdash \Theta_1, B\end{matrix}}{\Pi,\Gamma_1^* \vdash \Sigma^*,\Theta_1,B}\,mix(A)\quad \dfrac{\begin{matrix}(\psi_1)\\ \Pi \vdash \Sigma\end{matrix}\quad\begin{matrix}(\chi_1)\\ \Gamma_2 \vdash \Theta_2, C\end{matrix}}{\Pi,\Gamma_2^* \vdash \Sigma^*,\Theta_2,C}\,mix(A)}{\Pi,\Gamma_1^*,\Gamma_2^* \vdash \Sigma^*,\Sigma^*,\Theta_1,\Theta_2,B\wedge C}\,\wedge : r}{\Pi,\Gamma_1^*,\Gamma_2^* \vdash \Sigma^*,\Theta_1,\Theta_2,B\wedge C}\,c : r^*$$

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

$$\dfrac{\begin{matrix}(\psi_1)\\ \Pi \vdash \Sigma\end{matrix}\quad \dfrac{\begin{matrix}(\chi_1)\\ B,\Gamma_1 \vdash \Theta_1\end{matrix}\quad\begin{matrix}(\chi_2)\\ C,\Gamma_2 \vdash \Theta_2\end{matrix}}{B\vee C,\Gamma_1,\Gamma_2 \vdash \Theta_1,\Theta_2}\,\vee : l}{\Pi,(B\vee C)^*,\Gamma_1^*,\Gamma_2^* \vdash \Sigma^*,\Theta_1,\Theta_2}\,mix(A)$$

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

$$\dfrac{\dfrac{\dfrac{\begin{matrix}(\psi_1)\\ \Pi \vdash \Sigma\end{matrix}\quad\begin{matrix}(\chi_1)\\ B,\Gamma_1 \vdash \Theta_1\end{matrix}}{B^*,\Pi,\Gamma_1^* \vdash \Sigma^*,\Theta_1}\,mix(A)}{B,\Pi,\Gamma_1^* \vdash \Sigma^*,\Theta_1}\,x\quad \dfrac{\dfrac{\begin{matrix}(\psi_1)\\ \Pi \vdash \Sigma\end{matrix}\quad\begin{matrix}(\chi_1)\\ C,\Gamma_2 \vdash \Theta_2\end{matrix}}{C^*,\Pi,\Gamma_2^* \vdash \Sigma^*,\Theta_2}\,mix(A)}{C,\Pi,\Gamma_2^* \vdash \Sigma^*,\Theta_2}\,x}{B\vee C,\Pi,\Pi^*,\Gamma_1^*,\Gamma_2^* \vdash \Sigma^*,\Sigma^*,\Theta_1,\Theta_2}\,\vee : l$$

Note that, in case $A = B$ or $A = C$, the inference x is $w : l$; otherwise x is the identical transformation and can be dropped.
If $(B \vee C)^* = B \vee C$ then ψ transforms to

$$\dfrac{\tau}{\Pi, B\vee C,\Gamma_1^*,\Gamma_2^* \vdash \Sigma^*,\Theta_1,\Theta_2}\,c :^*$$

If, on the other hand, $(B \vee C)^*$ is empty (i.e. $B \vee C = A$) then we transform ψ to

$$\dfrac{\dfrac{\begin{matrix}(\psi_1)\\ \Pi \vdash \Sigma\end{matrix}\quad \tau}{\Pi,\Pi^*,\Pi^*,\Gamma_1^*,\Gamma_2^* \vdash \Sigma^*,\Sigma^*,\Sigma^*,\Theta_1,\Theta_2}\,mix(A)}{\Pi,\Gamma_1^*,\Gamma_2^* \vdash \Sigma^*,\Theta_1,\Theta_2}\,c :^*$$

3.121.233. The last inference in ψ_2 is $\to: l$. Then ψ is of the form:

$$\dfrac{\begin{matrix}(\psi_1)\\ \Pi \vdash \Sigma\end{matrix} \quad \dfrac{\begin{matrix}(\chi_1)\\ \Gamma \vdash \Theta, B\end{matrix} \quad \begin{matrix}(\chi_2)\\ C, \Delta \vdash \Lambda\end{matrix}}{B \to C, \Gamma, \Delta \vdash \Theta, \Lambda}\, \to: l}{\Pi, (B \to C)^*, \Gamma^*, \Delta^* \vdash \Sigma^*, \Theta, \Lambda}\, mix(A)$$

As in 3.121.232 $(B \to C)^* = B \to C$ for $B \to C \neq A$ and $(B \to C)^*$ empty otherwise.

3.121.233.1. A occurs in Γ and in Δ. Again we define a proof τ:

$$\dfrac{\dfrac{\begin{matrix}(\psi_1)\\ \Pi \vdash \Sigma\end{matrix} \quad \begin{matrix}(\chi_1)\\ \Gamma \vdash \Theta, B\end{matrix}}{\Pi, \Gamma^* \vdash \Sigma^*, \Theta, B}\, mix(A) \quad \dfrac{\dfrac{\begin{matrix}(\psi_1)\\ \Pi \vdash \Sigma\end{matrix} \quad \begin{matrix}(\chi_2)\\ C, \Delta \vdash \Lambda\end{matrix}}{C^*, \Pi, \Delta^* \vdash \Sigma^*, \Lambda}\, mix(A)}{C, \Pi, \Delta^* \vdash \Sigma^*, \Lambda}\, x}{B \to C, \Pi, \Gamma^*, \Pi, \Delta^* \vdash \Sigma^*, \Theta, \Sigma^*, \Lambda}\, \to: l$$

If $(B \to C)^* = B \to C$ then, as in 3.121.232, ψ is transformed to τ + some additional contractions. Otherwise an additional mix with mix formula A is appended.

3.121.233.2 A occurs in Δ, but not in Γ. As in 3.121.233.1 we define a proof τ:

$$\dfrac{\begin{matrix}(\chi_1)\\ \Gamma \vdash \Theta, B\end{matrix} \quad \dfrac{\dfrac{\begin{matrix}(\psi_1)\\ \Pi \vdash \Sigma\end{matrix} \quad \begin{matrix}(\chi_2)\\ C, \Delta \vdash \Lambda\end{matrix}}{C^*, \Pi, \Delta^* \vdash \Sigma^*, \Lambda}\, mix(A)}{C, \Pi, \Delta^* \vdash \Sigma^*, \Lambda}\, x}{B \to C, \Gamma, \Pi, \Delta^* \vdash \Theta, \Sigma^*, \Lambda}\, \to: l$$

Again we distinguish the cases $B \to C = A$ and $B \to C \neq A$ and define the transformation of ψ exactly like in 3.121.233.1.

3.121.233.3 A occurs in Γ, but not in Δ: analogous to 3.121.233.2.

3.121.234. The last inference in ψ_2 is $mix(B)$ for some formula B. Then ψ is of the form

$$\dfrac{\begin{matrix}(\psi_1)\\ \Pi \vdash \Sigma\end{matrix} \quad \dfrac{\begin{matrix}(\chi_1)\\ \Gamma_1 \vdash \Theta_1\end{matrix} \quad \begin{matrix}(\chi_2)\\ \Gamma_2 \vdash \Theta_2\end{matrix}}{\Gamma_1, \Gamma_2^+ \vdash \Theta_1^+, \Theta_2}\, mix(B)}{\Pi, \Gamma_1^*, \Gamma_2^{+*} \vdash \Sigma^*, \Theta_1^+, \Theta_2}\, mix(A)$$

3.121.234.1 A occurs in Γ_1 and in Γ_2. Then ψ transforms to

$$\dfrac{\dfrac{\dfrac{\begin{matrix}(\psi_1)\\ \Pi \vdash \Sigma\end{matrix} \quad \begin{matrix}(\chi_1)\\ \Gamma_1 \vdash \Theta_1\end{matrix}}{\Pi, \Gamma_1^* \vdash \Sigma^*, \Theta_1}\, mix(A) \quad \dfrac{\begin{matrix}(\psi_1)\\ \Pi \vdash \Sigma\end{matrix} \quad \begin{matrix}(\chi_2)\\ \Gamma_2 \vdash \Theta_2\end{matrix}}{\Pi, \Gamma_2^* \vdash \Sigma^*, \Theta_2}\, mix(A)}{\Pi, \Pi^+, \Gamma_1^*, \Gamma_2^{+*} \vdash \Sigma^{*+}, \Sigma^{*+}, \Theta_1^+, \Theta_2}\, mix(B)}{\Pi, \Gamma_1^*, \Gamma_2^{+*} \vdash \Sigma^*, \Theta_1^+, \Theta_2}\, c:^*, w:^*$$

Note that, for $A = B$, we have $\Gamma_2^{*+} = \Gamma_2^*$ and $\Sigma^{*+} = \Sigma^*$; $\Gamma_2^{*+} = \Gamma_2^{+*}$ holds in all cases.

3.121.234.2 A occurs in Γ_1, but not in Γ_2. In this case we have $\Gamma_2^{+*} = \Gamma_2$ and we transform ψ to

$$\dfrac{\dfrac{\begin{matrix}(\psi_1) & (\chi_1)\\ \Pi \vdash \Sigma & \Gamma_1 \vdash \Theta_1\end{matrix}}{\Pi, \Gamma_1^* \vdash \Sigma^*, \Theta_1}\ mix(A) \qquad \begin{matrix}(\chi_2)\\ \Gamma_2 \vdash \Theta_2\end{matrix}}{\Pi, \Gamma_1^*, \Gamma_2^+ \vdash \Sigma^*, \Theta_1^+, \Theta_2}\ mix(B)$$

3.121.234.3 A is in Γ_2, but not in Γ_1: symmetric to 3.121.234.2.

3.122. right-rank $= 1$ and left-rank > 1: symmetric to 3.121.