Kai Brünnler

Cut Elimination inside a Deep Inference System for Classical Predicate Logic

Abstract. Deep inference is a natural generalisation of the one-sided sequent calculus where rules are allowed to apply deeply inside formulas, much like rewrite rules in term rewriting. This freedom in applying inference rules allows to express logical systems that are difficult or impossible to express in the cut-free sequent calculus and it also allows for a more fine-grained analysis of derivations than the sequent calculus. However, the same freedom also makes it harder to carry out this analysis, in particular it is harder to design cut elimination procedures. In this paper we see a cut elimination procedure for a deep inference system for classical predicate logic. As a consequence we derive Herbrand's Theorem, which we express as a factorisation of derivations.

Keywords: cut elimination, deep inference, first-order predicate logic

1. Introduction

This work is part of a broader research effort which aims to develop and exploit a structural proof theory that is richer than the one provided by traditional formalisms like the sequent calculus or natural deduction. It is based on the formalism named *calculus of structures*, which is due to Guglielmi [8] and has the distinguishing feature of *deep inference*, meaning that inference rules apply deeply inside formulas. Deep inference systems so far have been studied for linear logic [14], non-commutative variants of linear logic [11, 7], classical logic [5] and several modal logics [12].

The need for a richer proof theory comes mainly from computer science. It is well-known that the logical systems requested by computer scientists stretch the limits of expressivity of the traditional proof theoretical formalisms. The absence of cut-free sequent systems for some modal logics like S5, for many temporal and also for intermediate logics bears witness to that. Numerous extensions of the sequent calculus have been proposed in order to cope with some of these problems, such as the display calculus [3], hypersequent systems [1] or labelled deduction [2], just to name three approaches. The sequent calculus is also challenged by a very simple logical

Special Issue "Cut-elimination in Classical and Nonclassical Logic"
Edited by **Alexander Leitsch**

Studia Logica (2006) 82: 51–71
DOI: 10.1007/s11225-006-6605-4

system wich is called system BV [8] and which is of relevance to computer science because its connectives resemble those of a process algebra. There is evidence that it can not be expressed in an inference system that does not employ deep inference [16] and thus the calculus of structures had to be developed in order to express this logic. One of the aims of the calculus of structures, namely expressing more logics than the cut-free sequent calculus, is shared with the extensions mentioned above. However, its approach differs significantly from the approaches of these other formalisms. Rather than enriching the set of structural connectives with respect to the sequent calculus, the calculus of structures gets rid of them: by simply using the logical connectives instead.

Deep inference systems for classical predicate logic were introduced in [5]. Cut admissibility for these systems is proved externally, namely by translating a proof into the sequent calculus, eliminating the cut in the sequent calculus, and translating back the cut-free proof. So the question arises whether there is a direct procedure for eliminating the cut, a procedure that does not make the detour via the sequent calculus. For the propositional fragment, there is such a direct cut elimination procedure, cf. [4]. However, in contrast to the situation in the sequent calculus this procedure does not trivially scale to predicate logic. Deep inference allows the cut rule to apply inside the scope of an existential quantifier, which turns out to be problematic for cut elimination. In the sequent calculus this situation does not occur, which is the reason why first-order quantifiers do not make much of a difference with respect to the difficulty of proving cut admissibility. However, in deep inference they constitute a problem, and a solution to this problem is the main contribution of this paper: a direct cut-elimination procedure for a deep inference system for classical predicate logic.

Since the sequent calculus is already very successful in the proof-theoretic analysis of classical predicate logic a fair question is: why study it in a new formalism? One motivation is that, in order to use deep inference to study extensions of classical logic that are not expressible in the cut-free sequent calculus, of course one should first understand the proof theory of deep inference systems for classical logic. But the main motivation is that the structural proof theory of deep inference systems for classical logic already differs significantly from that of the sequent calculus and thus deserves study as a new perspective on the important concept of cut elimination, or, more generally, on the normalisation of proofs.

Some desirable features of deep inference systems for classical logic are that they allow for shorter cut-free proofs than the sequent calculus [9], that they allow to faithfully embed resolution derivations as cut-free proofs [10], that they allow to observe the symmetry between cut and identity axiom and that they allow to decompose inference rules like cut and contraction into more primitive rules [8, 5]. A less desirable feature is that proving cut elimination becomes a significant challenge due to the loss of the main connective, which plays a crucial role in the sequent calculus. And that is the problem that I address in the present work.

The plan of the paper is as follows: I first introduce a deep inference system for predicate logic, then give a cut elimination procedure for that system and then derive Herbrand's Theorem as a result.

2. Basic Definitions

DEFINITION 2.1. *Variables* are denoted by x and y and *terms* are denoted by τ, possibly subscripted. A finite sequence of terms such as $\tau_1, \ldots, \tau_n$ is denoted by $\vec{\tau}$. Let p be a predicate symbol of arity n. Expressions of the form $p(\vec{\tau})$ and their negations $\overline{p(\vec{\tau})}$ are *atoms*. Atoms are denoted by a, b, c and so on. *Formulas* are generated by

$$S ::= \mathsf{f} \mid \mathsf{t} \mid a \mid [\,S, S\,] \mid (\,S, S\,) \mid \exists x S \mid \forall x S \quad ,$$

where f and t are the units *false* and *true*, $[S_1, S_2]$ is a *disjunction* and (S_1, S_2) is a *conjunction*. Note that units are not atoms. Formulas are denoted by S, P, Q, R, T, U and V. A *formula context*, denoted by $S\{\ \}$, is a formula in the language extended by the symbol $\{\ \}$, the *empty context* or *hole*, with exactly one occurrence of the hole. $S\{R\}$ denotes the formula obtained by filling the hole in $S\{\ \}$ with R. We drop the curly braces when they are redundant: for example, $S[R, T]$ is short for $S\{[R, T]\}$. A *propositional context* is a context in which the hole is not in the scope of a quantifier.

The sequent calculus has two types of objects to deduce over, namely formulas and sequents. The inference systems that we will see will have just one type of objects, namely formulas. Since formulas have to play the role of sequents it turns out that the chosen outfix notation for connectives is more convenient than the standard infix notation.

DEFINITION 2.2. We define $\bar{S}$, the *negation* of the formula S, as follows:

$$\begin{array}{llll} \bar{\mathsf{f}} = \mathsf{t} & \overline{[R, T]} = (\bar{R}, \bar{T}) & \overline{\exists x R} = \forall x \bar{R} & \\ \bar{\mathsf{t}} = \mathsf{f} & \overline{(R, T)} = [\bar{R}, \bar{T}] & \overline{\forall x R} = \exists x \bar{R} & \overline{\overline{p(\vec{\tau})}} = p(\vec{\tau}) \end{array} \quad .$$

Definition 2.3. An *inference rule* is written

$$\rho \frac{S\{R\}}{S\{T\}} \quad ,$$

where ρ is the *name* of the rule, $S\{R\}$ is its *premise* and $S\{T\}$ is its *conclusion*. R and T are formulas that may contain schematic formulas, schematic atoms and schematic contexts. An *instance of an inference rule* is obtained by replacing all schematic formulas, schematic atoms and schematic contexts by formulas, atoms and contexts, respectively. In an instance of an inference rule the formula taking the place of R is its *redex*, the formula taking the place of T is its *contractum* and the context taking the place of $S\{\ \}$ is its *context*. A (*deductive*) *system* $\mathcal{S}$ is a set of inference rules.

An inference rule is best thought of as a rewrite rule known from term rewriting. For example, the rule ρ from the previous definition seen top-down corresponds to a rewrite rule $R \to T$.

Since formulas will have to play the role of sequents it will be convenient to equip them with an equivalence that is usually implicit in the notion of sequent:

Definition 2.4. The *syntactic equivalence relation* is the smallest congruence relation on formulas induced by commutativity and associativity of conjunction and disjunction, the capture-avoiding renaming of bound variables as well as the following equations:

$$\begin{array}{lll} [R, \mathsf{f}] = R & [\mathsf{t}, \mathsf{t}] = \mathsf{t} & \exists x \mathsf{f} = \mathsf{f} = \forall x \mathsf{f} \\ (R, \mathsf{t}) = R & (\mathsf{f}, \mathsf{f}) = \mathsf{f} & \forall x \mathsf{t} = \mathsf{t} = \exists x \mathsf{t} \quad . \end{array}$$

Definition 2.5. A *derivation* Δ in a certain deductive system is either a pair of syntactically equivalent formulas or a finite nonempty sequence of instances of inference rules in the system, where inference rules are applied modulo the syntactic equivalence. They are written respectively as follows:

$$= \frac{R}{T} \qquad \text{and} \qquad \rho \frac{\rho' \dfrac{\vdots}{V}}{T} \quad \text{with top} \quad \pi' \frac{\pi \dfrac{R}{U}}{\vdots} \quad .$$

The topmost formula in a derivation is called the *premise* of the derivation, and the formula at the bottom is called its *conclusion*. The *length* of the derivation is the number of instances of inference rules. A *proof* is a derivation whose premise is the unit t. A derivation Δ from R to T in $\mathcal{S}$ and a proof Π of T in $\mathcal{S}$ are respectively denoted by

$$\begin{array}{c} R \\ \Delta \Big\| \mathcal{S} \\ T \end{array} \qquad \text{and} \qquad \begin{array}{c} \Pi \Big\| \mathcal{S} \\ T \end{array} \quad .$$

NOTATION 2.6. We use $[R, T, U]$ to abbreviate $[R, [T, U]]$ and $[[R, T], U]$, and likewise for an arbitrary number of formulas in a disjunction. We do the same for conjunction. Given an inference rule ρ and a natural number n, ρ^n denotes n instances of ρ and ρ^* denotes n instances of ρ for some $n \geq 0$.

Given two derivations such that the conclusion of the first is the premise of the second, we can compose these two derivations vertically in the obvious way. In addition we will also compose derivations horizontally, as follows.

DEFINITION 2.7. Given a derivation Δ and a context $S\{\ \}$, the derivation $S\{\Delta\}$ is obtained by replacing each formula U in Δ by $S\{U\}$. Given two derivations, Δ_1 from R_1 to T_1 and Δ_2 from R_2 to T_2, we define (Δ_1, Δ_2) as the vertical composition of (R_1, Δ_2) and (Δ_1, T_2), and likewise for $[\Delta_1, \Delta_2]$.

DEFINITION 2.8. A rule ρ is *derivable* for a system $\mathcal{S}$ if for every instance of ρ with premise R and conclusion T there is a derivation from R to T in $\mathcal{S}$. A rule ρ is *admissible* for a system $\mathcal{S}$ if for every instance of ρ with premise R and conclusion T the existence of a proof of R in $\mathcal{S}$ implies the existence of a proof of T in $\mathcal{S}$. Two systems $\mathcal{S}$ and $\mathcal{S}'$ are *strongly equivalent* if for every derivation from R to T in $\mathcal{S}$ there is a derivation from R to T in $\mathcal{S}'$, and vice versa. Two systems $\mathcal{S}$ and $\mathcal{S}'$ are (*weakly*) *equivalent* if for every proof of S in $\mathcal{S}$ there is a proof of S in $\mathcal{S}'$, and vice versa.

DEFINITION 2.9. Our inference system for classical predicate logic is named *system* KSgr, and it is shown in Figure 1. The names of the rules from upper-left to lower-right are *identity*, *weakening*, *contraction*, *switch*, *retract* and *instantiate*. The substitution in $\mathsf{n}\!\downarrow$ is capture-avoiding (in the standard sense, meaning that variables in τ may be captured by quantifiers in $S\{\ \}$). The context $P\{\ \}$ in $\mathsf{r}\!\downarrow$ is a propositional context in which x does not occur. The propositional fragment of the system, namely the system without the retract and instantiate rules, is named *system* KSg. The letter K in KSgr is for *klassisch*, the letter S is for *structures* as in *calculus of structures*, the letter g is for *general*, meaning that all rules are defined for general formulas, and not restricted to atoms, and the letter r says that it contains a *retract* rule. To maintain the same naming conventions with previous papers, the system name should also contain a q for *quantifiers*, but since the presence of the retract rule only makes sense in the presence of quantifiers, we drop the letter q.

DEFINITION 2.10. The *dual* of an inference rule is obtained by exchanging premise and conclusion and replacing each connective by its De Morgan dual. A system of inference rules is called *symmetric* if for each of its rules it also contains the dual rule. The *dual* of a derivation is obtained by turning it upside-down, replacing each atom by its negation and by replacing each connective by its de Morgan dual and each rule name by the name of its dual.

EXAMPLE 2.11. The identity rule and its dual:

$$\mathsf{i}{\downarrow}\,\frac{S\{\mathsf{t}\}}{S[R,\bar{R}]} \qquad\qquad \mathsf{i}{\uparrow}\,\frac{S(R,\bar{R})}{S\{\mathsf{f}\}} \quad .$$

The duality between the two is well-known under the name *contrapositive*.

System SKSgr is the symmetric closure of KSgr, i.e. it contains each rule from KSgr and the dual of each rule in KSgr. The collection of rules with an up-arrow are called *up-fragment*, their names are the names of their duals suffixed by "-up". The rule $\mathsf{i}{\uparrow}$ is also called *cut*. Note that a symmetric system that contains the identity rule by definition contains the cut rule as well, so in general we can read "symmetric" as "contains cut". The notion of cut admissibility in deep inference is the admissibility of up-rules: in our case the admissibility of the rules $\mathsf{i}{\uparrow}, \mathsf{w}{\uparrow}, \mathsf{c}{\uparrow}, \mathsf{r}{\uparrow}$ and $\mathsf{n}{\uparrow}$ for system KSgr.

$$\mathsf{i}{\downarrow}\,\frac{S\{\mathsf{t}\}}{S[R,\bar{R}]} \qquad \mathsf{w}{\downarrow}\,\frac{S\{\mathsf{f}\}}{S\{R\}} \qquad \mathsf{c}{\downarrow}\,\frac{S[R,R]}{S\{R\}}$$

$$\mathsf{s}\,\frac{S([R,T],U)}{S[(R,U),T]} \qquad \mathsf{r}{\downarrow}\,\frac{S\{\forall x P\{R\}\}}{S\{P\{\forall x R\}\}} \qquad \mathsf{n}{\downarrow}\,\frac{S\{R[x/t]\}}{S\{\exists x R\}}$$

Figure 1. System KSgr

In the sequent calculus, the identity axiom usually can be replaced by its atomic form without a change of derivability. The same is true for the identity rule and the weakening rule in our system, and by duality, also for their duals. We define the following inference rules, *atomic identity* and *atomic weakening*:

$$\mathsf{ai}{\downarrow}\,\frac{S\{\mathsf{t}\}}{S[a,\bar{a}]} \qquad \text{and} \qquad \mathsf{aw}{\downarrow}\,\frac{S\{\mathsf{f}\}}{S\{a\}}$$

The following proposition will allow us to conveniently assume that instances of the rule $\mathsf{i}{\downarrow}$ and $\mathsf{w}{\downarrow}$ are atomic:

PROPOSITION 2.12. *The rules* $\mathsf{i}{\downarrow}$ *and* $\mathsf{w}{\downarrow}$ *are derivable for* $\{\mathsf{ai}{\downarrow}, \mathsf{s}, \mathsf{r}{\downarrow}, \mathsf{n}{\downarrow}\}$ *and* $\{\mathsf{aw}{\downarrow}, \mathsf{s}\}$*, respectively. Dually, the rules* $\mathsf{i}{\uparrow}$ *and* $\mathsf{w}{\uparrow}$ *are derivable for* $\{\mathsf{ai}{\uparrow}, \mathsf{s}, \mathsf{r}{\uparrow}, \mathsf{n}{\uparrow}\}$ *and* $\{\mathsf{aw}{\uparrow}, \mathsf{s}\}$*, respectively.*

Similarly to the sequent calculus, the reduction to atomic form is achieved by inductively replacing an instance of the rule by instances on smaller formulas, details are in [5].

Soundness, completeness and cut admissibility for system KSgr can be obtained by translating back-and-forth between its derivations and derivations in some one-sided sequent system. A detailed proof for system KSgq can be found in [5] and can be easily adapted for KSgr:

THEOREM 2.13 (Cut Elimination). *The rules* $\mathsf{i}{\uparrow}$, $\mathsf{r}{\uparrow}$, $\mathsf{w}{\uparrow}$, $\mathsf{n}{\uparrow}$ *and* $\mathsf{c}{\uparrow}$ *are admissible for system* KSgr. *Put differently, the systems* SKSgr *and* KSgr *are equivalent.*

Notice that they are not strongly equivalent, since the cut rule is clearly not derivable in KSgr. Our main goal in the next section is now to prove this theorem again, but this time without resorting to the sequent calculus.

3. Cut Elimination

The cut rule in the sequent calculus serves the purpose of composing proofs (when seen top-down) and the purpose of splitting proof obligations (when seen bottom-up). The cut rule in the calculus of structures is different. Here, the familiar sequent calculus cut is broken into smaller pieces, as shown in Figure 2.

Notice that the crux of the sequent calculus cut is isolated in the rule $\mathsf{i}{\uparrow}$: when seen bottom-up it introduces a formula A out of thin air. Notice also that the deep inference rules can be composed in a more flexible way than rules in the sequent calculus. For example we know that A and $\bar{A}$ in the sequent calculus proof will never interact because they are in different

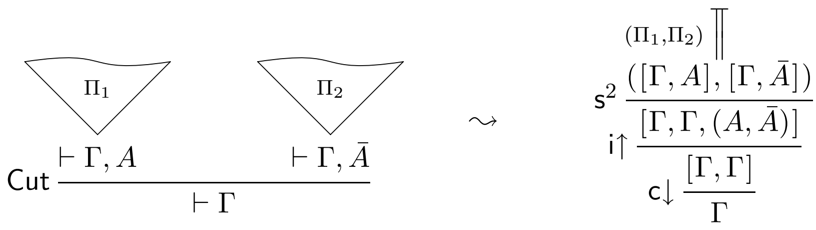

Figure 2. Dissecting a cut

branches. This is not true in the proof in the calculus of structures, because the rule $\mathsf{i}\!\uparrow$ does not force a splitting of proof obligations. Also, the rule $\mathsf{i}\!\uparrow$ can introduce the cut formula together with its negation anywhere deep inside a formula, for example in such a way that an existential quantifier in the context captures a variable in the cut formula. This also is impossible in the sequent calculus.

This freedom in applying inference rules in the calculus of structures is a significant challenge for cut elimination. While a proof in the sequent calculus decomposes a formula starting from the main connective, a proof in the calculus of structures is more like a myriad of interacting particles, atoms and quantified formulas, swimming in a soup of propositional connectives. During cut elimination, the sequent calculus allows to get into the crucial situation where on one branch a logical rule applies to the main connective of the cut formula and on the other branch the corresponding rule applies to the dual connective of the dual cut formula. Since rules in the calculus of structures are not restricted to main connectives, Gentzen's technique of permuting up the (generalised) cut does not apply. For example, one cannot permute the cut over the switch rule.

A cut elimination procedure for the propositional fragment SKS of SKSgr has been presented in [4]. It uses the fact that the cut rule trivially reduces to atomic form, a standard feature of systems in the calculus of structures, in order to give an especially simple cut elimination procedure. In particular, it does not involve an induction on the cut rank. The problem of the greater freedom in applying inference rules is dealt with by splitting the proof above the cut into two separate proofs. Once this is done, the procedure is very similar to normalisation in natural deduction. It works like Tait-style cut elimination [15]: given a cut in the sequent calculus, as in the picture above, the left proof Π_1 says that $\bar{A}$ implies Γ and the right proof Π_2 says that A implies Γ. To obtain a proof of $\vdash \Gamma, \Gamma$ and thus of Γ we take Π_1 and replace A by Γ everywhere inside it. This process of replacement will break the proof at certain places, but wherever that happens we can fix the proof by plugging in Π_2. In Taits procedure, which works in the sequent calculus, the proof breaks and has to be fixed in several cases, since rules apply to the cut formula. The procedure in [4] is so simple because there is just one place where the proof breaks: when the replacement reaches an identity.

The interesting question now is how this procedure for the propositional system scales to predicate logic. This question is nontrivial, mainly because existential quantifiers in the context of a cut prevent the splitting of the proof above into two separate proofs. In a nutshell, the solution we adopt here is to get rid of such existential quantifiers by trading them for bigger cuts.

Definition 3.1. A context $S\{\ \}$ is *splittable* if the hole is not in the scope of an existential quantifier. A *splittable cut*, denoted $\mathsf{si}\!\uparrow$, is a cut inside of a splittable context. A cut is called *solid* if the main connective of its cut formula is not propositional, i.e. if it is either a quantifier or if the cut formula is atomic. The *quantifier nesting* of a formula is defined as follows:

$$\begin{array}{l} qn(a) = qn(\mathsf{t}) = qn(\mathsf{f}) = 0 \\ qn(\forall x R) = qn(\exists x R) = qn(R) + 1 \\ qn([R,T]) = qn((R,T)) = max(qn(R), qn(T)) \quad . \end{array}$$

Given an instance of the cut rule with cut formula A, we define its *cut rank* as $qn(A)+1$. The *cut rank* of a derivation is the supremum of the cut ranks of its cuts. For $r \geq 0$ the inference rule $\mathsf{si}_r\!\uparrow$ is $\mathsf{si}\!\uparrow$ with the proviso that its cut rank is at most r.

This transformation allows us to replace up-rules by splittable cuts:

$$\rho\!\uparrow \frac{S\{T\}}{S\{R\}} \quad \rightsquigarrow \quad \mathsf{si}\!\uparrow \frac{\rho\!\downarrow \frac{\mathsf{s} \frac{\mathsf{i}\!\downarrow \frac{= \frac{S\{T\}}{(S\{T\},\mathsf{t})}}{(S\{T\},[S\{R\},\bar{S}\{\bar{R}\}])}}{[S\{R\},(S\{T\},\bar{S}\{\bar{R}\})]}}{[S\{R\},(S\{T\},\bar{S}\{\bar{T}\})]}}{S\{R\}} \quad ,$$

so we have

Lemma 3.2. *For each proof* $\begin{array}{c}\| \mathsf{SKSgr} \\ T\end{array}$ *there is a proof* $\begin{array}{c}\| \mathsf{KSgr} \cup \{\mathsf{si}\!\uparrow\} \\ T\end{array}$.

This transformation allows us to inductively replace splittable cuts by solid splittable cuts:

$$
\mathsf{si}_r\!\uparrow \frac{S(\bar{R},\bar{T},[R,T])}{S\{\mathsf{f}\}}
\quad\leadsto\quad
\mathsf{si}_r\!\uparrow \frac{\mathsf{si}_r\!\uparrow \dfrac{\mathsf{s}\dfrac{\mathsf{s}\dfrac{S(\bar{R},\bar{T},[R,T])}{S(\bar{R},[(\bar{T},T),R])}}{S[(\bar{T},T),(\bar{R},R)]}}{S(\bar{R},R)}}{S\{\mathsf{f}\}}
\quad ,
$$

so we have

Lemma 3.3. *The rule* $\mathsf{si}_r\!\uparrow$ *is derivable for solid* $\mathsf{si}_r\!\uparrow$ *and switch.*

Definition 3.4. A rule ρ is *length-preserving admissible* for a system $\mathcal{S}$ if for every instance of ρ with premise R and conclusion T for all n the existence of a proof of length n of R in $\mathcal{S}$ implies the existence of a proof of length n of T in $\mathcal{S}$. *Cut-rank-preserving admissible* is defined in the same way, replacing length by cut-rank.

3.1. Splitting

During cut elimination in the sequent calculus one has access to two proofs above the cut such that the cut formula is in the conclusion of one proof and the dual of the cut formula is in the conclusion of the other proof. In the calculus of structures, we just have one proof above the cut and its conclusion contains both, the cut formula and its dual. This subsection is devoted to gaining access to two proofs as in the sequent calculus.

In a cut-free proof of a formula $S(R,T)$ rules can apply in many different chaotic ways. We now see a lemma, which tells us that for each such proof there is one in which inference rules apply in a certain orderly fashion. In fact, it can be *split* into two proofs, one containing R and one containing T:

$$
\begin{array}{c} \| \\ S(R,T) \end{array}
\quad\leadsto\quad
\begin{array}{c}
\mathsf{s}^2;\mathsf{c}\!\downarrow \dfrac{\forall\vec{x}(\,\begin{array}{c}\|\\ {[U,R]}\end{array}\,,\;\begin{array}{c}\|\\ {[U,T]}\end{array}\,)}{\forall\vec{x}[U,(R,T)]} \\ \| \\ S(R,T)
\end{array}
\quad .
$$

During cut elimination, the splitting lemma will be applied to the proof above the cut with R being the cut formula and T being the dual cut formula.

This will make available a situation more comparable to the sequent calculus, where a cut splits the proof.

The splitting lemma presented here is inspired by a similar one used by Guglielmi for a substructural logic in [8]. However, the proof is very different. Guglielmi not only splits the proof, but also the context. In the example above this means that U is split into two formulas: one that goes into the proof with R and another that goes into the proof with T. In classical logic we have contraction at our disposal, which means that instead of having to split U into two parts, which requires some work, we can simply duplicate it. Before we state the splitting lemma, we need two more lemmas.

LEMMA 3.5. *The weakening-up rule* $\mathsf{w}\!\uparrow$ *is cut-rank-preserving admissible for system* $\mathsf{KSgr} \cup \{\mathsf{si}\!\uparrow\}$.

PROOF. By Proposition 2.12 we it suffices to prove the lemma for atomic weakening-up. Consider a proof

$$\mathsf{aw}\!\uparrow\frac{\begin{array}{c}\Pi\,\Big\Vert\;{\scriptstyle \mathsf{KSgr}\cup\{\mathsf{si}\uparrow\}}\\ T\{a\}\end{array}}{T\{\mathsf{t}\}} \quad .$$

Starting with the conclusion of Π, going up in the proof, in each formula we replace the atom a, and its copies that are produced by contractions, and their instances that are produced by instantiations, by the unit t. Replacements inside the context of any rule instance leave this rule instance intact. Instances of all the rules in $\mathsf{KSgr} \cup \{\mathsf{si}\!\uparrow\}$ remain intact also in the case that atom occurrences are replaced by t inside redex and contractum, except for $\mathsf{ai}\!\downarrow$. We replace them by weakenings:

$$\mathsf{ai}\!\downarrow\frac{S\{\mathsf{t}\}}{S[a,\bar{a}]} \quad \rightsquigarrow \quad \mathsf{aw}\!\downarrow\frac{=\dfrac{S\{\mathsf{t}\}}{S[\mathsf{t},\mathsf{f}]}}{S[\mathsf{t},\bar{a}]} \quad .$$

■

LEMMA 3.6. *The instantiation-up rule* $\mathsf{n}\!\uparrow$ *is length- and cut-rank-preserving admissible for system* $\mathsf{KSgr} \cup \{\mathsf{si}\!\uparrow\}$.

PROOF. We proceed by induction on the length of the proof in $\mathsf{KSgr} \cup \{\mathsf{si}\!\uparrow\}$. The base case is easy: if the premise of $\mathsf{n}\!\uparrow$ is syntactically equivalent to t then so is its conclusion. To prove the induction step, consider a proof in $\mathsf{KSgr} \cup \{\mathsf{si}\!\uparrow\}$ above an instance of $\mathsf{n}\!\uparrow$. Let ρ be the inference rule above $\mathsf{n}\!\uparrow$. We do a case analysis on the position of the redex of $\mathsf{n}\!\uparrow$ with respect to the contractum of ρ. If the redex is inside the context of ρ then $\mathsf{n}\uparrow$ trivially permutes up and the lemma follows from the induction hypothesis. Consider the case that it is inside a schematic formula of the contractum of ρ. Then ρ is one of $\mathsf{s}, \mathsf{c}\downarrow, \mathsf{r}\downarrow, \mathsf{n}\downarrow$. In the case of $\mathsf{c}\downarrow$ we push up $\mathsf{n}\uparrow$ to obtain two instances of $\mathsf{n}\uparrow$ and apply the induction hypothesis twice. The case of s is trivial and so is $\mathsf{r}\downarrow$, where we possibly have to rename bound variables in order to respect the proviso of $\mathsf{r}\downarrow$. The somewhat tedious case is permuting $\mathsf{n}\uparrow$ up over $\mathsf{n}\downarrow$, where we have to check the variable conditions in the derivation on the right:

$$\mathsf{n}\uparrow \dfrac{\mathsf{n}\downarrow \dfrac{S\{R\{\forall yT\}[x/\tau_1]\}}{S\{\exists xR\{\forall yT\}\}}}{S\{\exists xR\{T[y/\tau_2]\}\}} \quad \leadsto \quad = \dfrac{\mathsf{n}\uparrow \dfrac{= \dfrac{S\{R\{\forall yT\}[x/\tau_1]\}}{S\{R[x/\tau_1]\{\forall yT[x/\tau_1]\}\}}}{S\{R[x/\tau_1]\{T[x/\tau_1][y/\tau_2[x/\tau_1]]\}\}}}{\mathsf{n}\downarrow \dfrac{S\{R\{T[y/\tau_2]\}[x/\tau_1]\}}{S\{\exists xR\{T[y/\tau_2]\}\}}} \quad .$$

We can safely assume that differently bound variables have different names, so in particular we have that no variable from τ_1 occurs bound in $R\{\forall yT\}$ and that no variable from τ_2 occurs bound in T. From that we conclude the validity of the equalities and the instances of $\mathsf{n}\uparrow$ and $\mathsf{n}\downarrow$: $\tau_2[x/\tau_1]$ is free for y in $T[x/\tau_1]$ and τ_1 is free for x in $R\{T[y/\tau_2]\}$.

Consider now the case that the contractum of ρ is inside the redex of $\mathsf{n}\uparrow$. Then it clearly has to be inside the schematic formula, so $\mathsf{n}\uparrow$ trivially permutes up over ρ, except when $\rho = \mathsf{r}\downarrow$, when we possibly have to rename a bound variable, and when $\rho = \mathsf{n}\downarrow$ when we have to check variable conditions, but this case is dual to the one that we considered above.

The only remaining case is that the active universal quantifier in the redex of $\mathsf{n}\uparrow$ matches an active universal quantifier in the contractum of ρ. This can only happen when ρ is $\mathsf{r}\downarrow$ and we apply the following transformation in order to apply the induction hypothesis:

$$\mathsf{n}\uparrow \dfrac{\mathsf{r}\downarrow \dfrac{S\{\forall xP\{R\}\}}{S\{P\{\forall xR\}\}}}{S\{P\{R[x/\tau]\}\}} \quad \leadsto \quad \mathsf{n}\uparrow \dfrac{S\{\forall xP\{R\}\}}{S\{P\{R[x/\tau]\}\}} \quad .$$

■

Lemma 3.7 (Splitting). *Let $S\{\ \}$ be a splittable context and let $\forall\vec{x}$ be the sequence of all its universal quantifiers that have the hole in their scope. Then for each proof* $\begin{array}{c} \Pi \Big\| \mathsf{KSgr} \cup \{\mathsf{si}\uparrow\} \\ S(R,T) \end{array}$ *there are a formula U and proofs*

$$\begin{array}{c} \Big\Vert \mathsf{KSgr} \cup \{\mathsf{si}{\uparrow}\} \\ [U, R] \end{array} \quad \textit{and} \quad \begin{array}{c} \Big\Vert \mathsf{KSgr} \cup \{\mathsf{si}{\uparrow}\} \\ [U, T] \end{array} \quad \textit{and a derivation} \quad \begin{array}{c} \forall \vec{x} U \\ \Big\Vert \{\mathsf{r}{\downarrow}\} \\ S\{\mathsf{f}\} \end{array}$$

such that the cut ranks of both proofs are smaller than or equal to the cut rank of Π.

Proof. Let $U = S'\{\mathsf{f}\}$, where $S'\{\ \}$ is obtained from $S\{\ \}$ by removing all universal quantifiers that have the hole in their scope. We obtain the proofs and the derivation as follows:

$$\begin{array}{c} \Big\Vert \mathsf{KSgr} \cup \{\mathsf{si}{\uparrow}\} \\ \mathsf{w}{\uparrow} \dfrac{S(R, T)}{\mathsf{n}{\uparrow} \dfrac{S\{R\}}{\mathsf{s}^* \dfrac{S'\{R\}}{[S'\{\mathsf{f}\}, R]}}} \end{array} \quad , \quad \begin{array}{c} \Big\Vert \mathsf{KSgr} \cup \{\mathsf{si}{\uparrow}\} \\ \mathsf{w}{\uparrow} \dfrac{S(R, T)}{\mathsf{n}{\uparrow} \dfrac{S\{T\}}{\mathsf{s}^* \dfrac{S'\{T\}}{[S'\{\mathsf{f}\}, T]}}} \end{array} \quad , \quad \begin{array}{c} \forall \vec{x} S'\{\mathsf{f}\} \\ \Big\Vert \{\mathsf{r}{\downarrow}\} \\ S\{\mathsf{f}\} \end{array} \quad ,$$

where $\mathsf{w}{\uparrow}$ and $\mathsf{n}{\uparrow}$ are eliminated by Lemma 3.5 and Lemma 3.6. ■

3.2. Eliminating Atomic Cuts

The cut elimination procedure we are after will first reduce cuts to atomic cuts and then eliminate the atomic cuts. However, I present cut reduction after the elimination of atomic cuts. I find it interesting that quantifiers behave like atoms and both of them behave differently from propositional connectives. So the cut reduction for quantified formulas is the same as the elimination of an atomic cut, with just one additional difficulty: rules can apply inside the quantified formula, while rules cannot apply inside an atom. Since elimination of an atomic cut is the simpler case, I present it first.

Lemma 3.8 (Atomic Cut Elimination).

For each proof $\begin{array}{c} \Pi \Big\Vert \mathsf{KSgr} \\ \mathsf{si}{\uparrow} \dfrac{T(a, \bar{a})}{T\{\mathsf{f}\}} \end{array}$ *there is a proof* $\begin{array}{c} \Big\Vert \mathsf{KSgr} \\ T\{\mathsf{f}\} \end{array}$.

Proof. We apply the splitting lemma to Π in order to obtain

$$\begin{array}{c} \Pi_1 \Big\Vert \mathsf{KSgr} \\ [U, a] \end{array} \quad , \quad \begin{array}{c} \Pi_2 \Big\Vert \mathsf{KSgr} \\ [U, \bar{a}] \end{array} \quad \text{and} \quad \begin{array}{c} \forall \vec{x} U \\ \Delta \Big\Vert \{\mathsf{r}{\downarrow}\} \\ T\{\mathsf{f}\} \end{array} \quad .$$

Note that Π_2 proves that a implies U. We thus replace a inside Π_1 by U in order to obtain a proof of $[U, U]$ and thus of U. Starting with the conclusion, going up in proof Π_1, in each formula we replace the atom a, and its copies that are produced by contractions, by the formula U.

Replacements inside the context of any rule instance leave the rule instance intact. Instances of the rules $\mathsf{s}, \mathsf{c}{\downarrow}$ and $\mathsf{w}{\downarrow}$ remain intact, also in the case that atom occurrences are replaced inside the contractum and redex. The same is true for $\mathsf{r}{\downarrow}$, where we possibly have to rename the universally bound variable in order not to violate the proviso. No replacement happens inside the contractum of a $\mathsf{n}{\downarrow}$ rule because in Π_1 no copy of a is in the scope of an existential quantifier. The interesting case is $\mathsf{ai}{\downarrow}$. We replace its instances by $S\{\Pi_2\}$:

$$\mathsf{ai}{\downarrow}\frac{S\{\mathsf{t}\}}{S[a,\bar{a}]} \quad \leadsto \quad \begin{array}{c} S\{\mathsf{t}\} \\ {\scriptstyle S\{\Pi_2\}}\Big\| {\scriptstyle \mathsf{KSgr}} \\ S[U,\bar{a}] \end{array} \quad .$$

The result of this process of substituting Π_2 into Π_1 is a proof Π_3, from which we build

$$\begin{array}{c} \\ {\scriptstyle \forall\vec{x}\,\Pi_3}\Big\| {\scriptstyle \mathsf{KSgr}} \\ \mathsf{c}{\downarrow}\dfrac{\forall\vec{x}[U,U]}{\forall\vec{x}U} \\ {\scriptstyle \Delta}\Big\| {\scriptstyle \{\mathsf{r}{\downarrow}\}} \\ T\{\mathsf{f}\} \end{array} \quad .$$

■

3.3. Cut Reduction

Cut reduction is very similar to the elimination of an atomic cut, except that replacing a compound cut formula of the form $\exists xR$ is a bit more involved than replacing an atom, because inference rules apply inside R. We will accomplish this replacement by pushing up a special inference rule which keeps track of these inference rules.

Definition 3.9. An *n-context* is a formula with n occurences of $\{\ \}$, and a *splittable n-context* is an n-context in which no hole is in the scope of an existential quantifier. Given a proof Π of $[U,\forall xR]$ in $\mathsf{KSgr} \cup \{\mathsf{si}{\uparrow}\}$ and some $n \geq 1$ we define the inference rule $\mathsf{plug}_{\Pi,n}$ as

$$\mathsf{plug}_{\Pi,n}\frac{S\{\exists xR_1\}\ldots\{\exists xR_n\}}{S\{U\}\ldots\{U\}} \quad ,$$

where $S\{\ \}\ldots\{\ \}$ is a splittable n-context and for all $i \leq n$ there is a derivation Δ_i in KSgr from R_i to $\bar{R}$.

LEMMA 3.10 (Cut Reduction).

For each proof $\mathsf{si}_{r+1}\!\uparrow \dfrac{\overset{\Pi \,\Vert\, \mathsf{KSgr}\cup\{\mathsf{si}_r\uparrow\}}{T(\forall x R, \exists x \bar{R})}}{T\{\mathsf{f}\}}$ *there is a proof* $\overset{\Vert\, \mathsf{KSgr}\cup\{\mathsf{si}_r\uparrow\}}{T\{\mathsf{f}\}}$.

PROOF. Just like in the case of an atomic cut, we apply the splitting lemma on Π to obtain

$$\overset{\Pi_1 \,\Vert\, \mathsf{KSgr}\cup\{\mathsf{si}_r\uparrow\}}{[U,\exists x\bar{R}]} \quad , \quad \overset{\Pi_2 \,\Vert\, \mathsf{KSgr}\cup\{\mathsf{si}_r\uparrow\}}{[U,\forall x R]} \qquad \text{and} \qquad \begin{array}{c} \forall\vec{x}U \\ \Delta \,\Vert\, \{\mathsf{r}\downarrow\} \\ T\{\mathsf{f}\} \end{array} \quad .$$

Note that Π_2 proves that $\exists x\bar{R}$ implies U. The idea is thus to replace $\exists x\bar{R}$ inside Π_1 by U in order to obtain a proof of $[U,U]$ and thus of $T\{\mathsf{f}\}$. More formally, we will obtain a proof of $T\{\mathsf{f}\}$ by eliminating plug from

$$\begin{array}{c} \forall\vec{x}\,\Pi_1 \,\Vert\, \mathsf{KSgr}\cup\{\mathsf{si}_r\uparrow\} \\ \mathsf{plug}_{\Pi_2,1}\dfrac{\forall\vec{x}[U,\exists x\bar{R}]}{\mathsf{c}\!\downarrow \dfrac{\forall\vec{x}[U,U]}{\forall\vec{x}U}} \\ \Delta \,\Vert\, \{\mathsf{r}\downarrow\} \\ T\{\mathsf{f}\} \end{array} \quad .$$

We push plug to the top until it disappears. Pushing it up over the propositional rules and over $\mathsf{r}\downarrow$ and $\mathsf{si}\uparrow$ is easy: they cannot affect the active existential quantifiers in the premise of plug. So either plug trivially permutes up or, if the rule above applies inside one of the R_i, it is added to Δ_i. The interesting case is $\mathsf{n}\downarrow$. We push plug up as follows:

$$\mathsf{plug}_{\Pi_2,n}\dfrac{\mathsf{n}\!\downarrow \dfrac{S'\{R_i[x/\tau]\}}{S'\{\exists x R_i\}}}{S\{U\}} \quad \leadsto \quad \begin{array}{c} \mathsf{plug}_{\Pi_2,n-1}\dfrac{S'\{R_i[x/\tau]\}}{S\{R_i[x/\tau]\}} \\ S(\Delta_i[x/\tau],\Pi_2') \,\Vert\, \mathsf{KSgr}\cup\{\mathsf{si}_r\uparrow\} \\ \mathsf{si}_r\!\uparrow \dfrac{\mathsf{s}\dfrac{S(\bar{R}[x/\tau],[U,R[x/\tau]])}{S[U,(\bar{R}[x/\tau],R[x/\tau])]}}{S\{U\}} \end{array} \quad ,$$

where we obtain a derivation $\Delta_i[x/\tau]$ by applying the substitution $[x/\tau]$ to each formula in Δ_i and a proof Π_2' by applying $\mathsf{n}\uparrow$ to Π_2 and eliminating it by using Lemma 3.6. We can safely assume that all bound variables in Π_2 and in Δ_i are distinct from variables in τ.

Once plug reaches the top, its premise $S\{\exists x R_1\}\ldots\{\exists x R_n\}$ is equivalent to t. Since no atoms can occur in a formula that is equivalent to t, there are two cases to distinguish: 1) all of the R_i are equivalent to f, or 2) at least one of the R_i is equivalent to t. In the first case we can simply replace the instance of plug by instances of weakening and and in the second case we directly build a proof of $T\{\mathsf{f}\}$, respectively as follows:

$$\mathsf{w}\!\downarrow^n \frac{S\{\mathsf{f}\}\ldots\{\mathsf{f}\}}{S\{U\}\ldots\{U\}} \qquad \text{and} \qquad \begin{array}{c} \forall\vec{x}\,(\Pi_2'',\Delta_i)\,\Big\Vert \\ \mathsf{si}_r\!\uparrow \dfrac{\mathsf{s}\,\dfrac{\forall\vec{x}([U,R],\bar{R})}{\forall\vec{x}[U,(R,\bar{R})]}}{\forall\vec{x}U} \\ \Delta\,\Big\Vert \\ T\{\mathsf{f}\} \end{array} \quad ,$$

where we obtain Π_2'' by adding an instance of $\mathsf{n}\!\uparrow$ in the obvious way to Π_2 and eliminating it by using Lemma 3.6. ∎

Now we can give a proof of cut elimination, i.e. of Theorem 2.13, without any reference to the sequent calculus:

Proof. By Lemma 3.2 we just need to show that for each proof $\begin{array}{c}\Big\Vert\,\mathsf{KSgr}\cup\{\mathsf{si}\!\uparrow\}\\ T\end{array}$ there is a proof $\begin{array}{c}\Big\Vert\,\mathsf{KSgr}\\ T\end{array}$. We eliminate instances of $\mathsf{si}\!\uparrow$ in two phases:

Phase 1 By induction on the cut rank of the proof. First replace all cuts by solid cuts by Lemma 3.3. Then, by a subinduction on the number of maximal-rank cuts, choose the topmost and apply the cut reduction lemma. The result is a proof with a cut rank of at most one.

Phase 2 First reduce all cuts to atomic cuts by Lemma 3.3. Then, by induction on the number of atomic cuts, choose the topmost and apply the atomic cut elimination lemma. ∎

4. Herbrand's Theorem

A weak version of Herbrand's theorem immediately follows from Gentzen's Mid-Sequent Theorem which in turn immediately follows from cut elimination. I will prove the strong version of Herbrand's theorem cf. [6], which also can be proved without difficulties by using cut elimination in the sequent calculus. I will tune the deductive system a bit in order to present a factorisation of proofs from which the strong version of Herbrand's theorem immediately follows, in the same sense as the weak version follows from the Mid-sequent factorisation of proofs in the sequent calculus. This is of course impossible in the sequent calculus, since the restriction of rules to the main connective does not allow to represent the expansion and prenexification phase of a Herbrand proof.

In order to prove Herbrand's theorem one needs to keep track of existentially quantified formulas that are duplicated. In our setting we do so by decomposing contraction, i.e. we inductively replace contraction by the following rules:

$$\mathsf{ac}\!\downarrow \frac{S[a,a]}{S\{a\}} \qquad \mathsf{m}\,\frac{S[(R,U),(T,V)]}{S([R,T],[U,V])}$$
$$\mathsf{qc}\!\downarrow \frac{S[\exists xR,\exists xR]}{S\{\exists xR\}} \qquad \mathsf{m_2}\!\downarrow \frac{S[\forall xR,\forall xT]}{S\{\forall x[R,T]\}} \quad ,$$

which are called *atomic contraction*, *medial*, *contraction-quantified*, and *medial two*, respectively.

PROPOSITION 4.1. *The rule* $\mathsf{c}\!\downarrow$ *is derivable for* $\{\mathsf{ac}\!\downarrow,\mathsf{m},\mathsf{qc}\!\downarrow,\mathsf{m_2}\!\downarrow\}$. *Each rule in* $\{\mathsf{ac}\!\downarrow,\mathsf{m},\mathsf{qc}\!\downarrow,\mathsf{m_2}\!\downarrow\}$ *is derivable for* $\{\mathsf{c}\!\downarrow,\mathsf{w}\!\downarrow\}$.

We define *system* KS as $\{\mathsf{ai}\!\downarrow,\mathsf{aw}\!\downarrow,\mathsf{ac}\!\downarrow,\mathsf{s},\mathsf{m}\}$. It is easy to check that it is strongly equivalent to system KSg, i.e. the propositional fragment of KSgr. For details see [5].

In order to represent the prenexification phase in a Herbrand proof, we define a *generalised retract* rule:

$$\mathsf{gr}\!\downarrow \frac{S\{Q\{P\{R\}\}\}}{S\{P\{Q\{R\}\}\}} \quad ,$$

where $Q\{\ \}$ is a sequence of quantifiers and $P\{\ \}$ is a propositional context such that no variable in $P\{\ \}$ is bound by a quantifier in $Q\{\ \}$ in the premise.

THEOREM 4.2 (Herbrand's Theorem). *For each proof of a formula S in system* SKSgr *there is a is a substitution σ, a propositional formula P, a context $Q\{\ \}$ consisting only of quantifiers and a proof given in Figure 3 at the right.*

$$
\begin{array}{c} \| \, \mathsf{KS}\cup\left\{\begin{smallmatrix}\mathsf{qc}\downarrow,\mathsf{m_2}\downarrow,\\ \mathsf{n}\downarrow,\mathsf{r}\downarrow,\mathsf{ai}\uparrow\end{smallmatrix}\right\} \\ S \end{array}
\overset{1}{\leadsto}
\begin{array}{c} \| \, \mathsf{KS}\cup\left\{\begin{smallmatrix}\mathsf{m_2}\downarrow,\\ \mathsf{n}\downarrow,\mathsf{r}\downarrow,\mathsf{ai}\uparrow\end{smallmatrix}\right\} \\ S' \\ \| \, \mathsf{qc}\downarrow \\ S \end{array}
\overset{2}{\leadsto}
\begin{array}{c} \| \, \mathsf{KS}\cup\{\mathsf{n}\downarrow,\mathsf{ai}\uparrow\} \\ Q\{P\} \\ \| \, \mathsf{gr}\downarrow \\ S' \\ \| \, \mathsf{qc}\downarrow \\ S \end{array}
\overset{3}{\leadsto}
\begin{array}{c} \| \, \mathsf{KS}\cup\{\mathsf{ai}\uparrow\} \\ \forall\vec{x}\,P\sigma \\ \| \, \mathsf{n}\downarrow \\ Q\{P\} \\ \| \, \mathsf{gr}\downarrow \\ S' \\ \| \, \mathsf{qc}\downarrow \\ S \end{array}
$$

Figure 3. The proof of Herbrand's Theorem

Proof. Given the proof in SKSgr, we apply Lemma 3.2 and cut elimination to get a proof in $\mathsf{KSgr} \cup \{\mathsf{ai}\uparrow\}$. The first phase of the procedure is sufficient since atomic cuts make no difference for Herbrand's Theorem. By Proposition 4.1 we decompose contraction to get a proof in $\mathsf{KS}\cup\{\mathsf{qc}\downarrow, \mathsf{m_2}\downarrow, \mathsf{n}\downarrow, \mathsf{r}\downarrow, \mathsf{ai}\uparrow\}$. From here, we get the factorisation of the proof that we are after by three phases that are shown in Figure 3.

Phase 1 We push all instances of $\mathsf{qc}\downarrow$ down to the bottom of the proof starting with the bottommost instance, and proceeding by induction on the number of instances of $\mathsf{qc}\downarrow$. To push down one instance of $\mathsf{qc}\downarrow$ we proceed by induction on the number of rule instances below.

Consider an instance of $\mathsf{qc}\downarrow$ together with one rule instance $\rho \in \mathsf{KS} \cup \{\mathsf{m_2}\downarrow, \mathsf{n}\downarrow, \mathsf{r}\downarrow, \mathsf{ai}\uparrow\}$ below it. If the contractum of $\mathsf{qc}\downarrow$ is inside of a schema of ρ (i.e. a subformula of the schematic context $S\{\ \}$ or of the schematic formulas in the redex), then $\mathsf{qc}\downarrow$ trivially permutes down. Since the contractum of $\mathsf{qc}\downarrow$ cannot overlap with the redex of ρ the only remaining case is that the redex of ρ is inside of the schematic formula in the contractum of $\mathsf{qc}\downarrow$. We apply the following transformation:

$$
\mathsf{qc}\downarrow \frac{S[\exists xT\{R'\}, \exists xT\{R'\}]}{\rho\,\dfrac{S\{\exists xT\{R'\}\}}{S\{\exists xT\{R\}\}}}
\quad \leadsto \quad
\rho^2 \frac{S[\exists xT\{R'\}, \exists xT\{R'\}]}{\mathsf{qc}\downarrow\,\dfrac{S[\exists xT\{R\}, \exists xT\{R\}]}{S\{\exists xT\{R\}\}}} \quad .
$$

Phase 2 We factor the upper proof into a derivation in $\{\mathsf{gr}\downarrow\}$ transforming a formula into prenex normal form and a proof in $\mathsf{KS} \cup \{\mathsf{n}\downarrow, \mathsf{ai}\uparrow\}$ which contains prenex formulas only. In the following the $Q_{1,2,3}\{\ \}$ denote sequences of quantifiers. We assume that differently bound variables have different names and their names are different from the names of free variables. Given a formula S, $\underline{S}_{\mathsf{p}}$ denotes the formula obtained from S by removing all quantifiers.

We proceed by induction on the length of the given proof. The induction base is trivial. The induction step is trivial for the propositional rules and for $\mathsf{r}\downarrow$, which is a special case of $\mathsf{gr}\downarrow$. To prove it for the rules involving quantifiers, apply the following transformations:

$$\mathsf{gr}\downarrow^* \frac{Q_1\{\forall x Q_2\{\forall y Q_3\{\underline{S[R,T[x/y]]}_{\mathsf{p}}\}\}\}}{\mathsf{m}_2\downarrow \dfrac{S[\forall x R, \forall y T[x/y]]}{S\{\forall x[R,T]\}}} \quad \leadsto \quad \mathsf{n}\uparrow \frac{Q_1\{\forall x Q_2\{\forall y Q_3\{\underline{S[R,T[x/y]]}_{\mathsf{p}}\}\}\}}{\mathsf{gr}\downarrow^* \dfrac{Q_1\{\forall x Q_2\{Q_3\{\underline{S[R,T]}_{\mathsf{p}}\}\}\}}{S\{\forall x[R,T]\}}}$$

as well as

$$\mathsf{gr}\downarrow^* \frac{Q_1\{Q_2\{\underline{S\{R[x/\tau]\}}_{\mathsf{p}}\}\}}{\mathsf{n}\downarrow \dfrac{S\{R[x/\tau]\}}{S\{\exists x R\}}} \quad \leadsto \quad \mathsf{n}\downarrow \frac{Q_1\{Q_2\{\underline{S\{R[x/\tau]\}}_{\mathsf{p}}\}\}}{\mathsf{gr}\downarrow^* \dfrac{Q_1\{\exists x Q_2\{\underline{S\{R\}}_{\mathsf{p}}\}\}}{S\{\exists x R\}}} \quad .$$

We eliminate the instance of $\mathsf{n}\uparrow$ using the same procedure as in the proof of Lemma 3.6.

Phase 3 To get the final result we now push down instances of $\mathsf{n}\downarrow$. We proceed by induction on proof length. The base case is trivial, as is the induction step since besides $\mathsf{n}\downarrow$ only propositional rules are left and contraction is restricted to atoms. ∎

5. Conclusion

We have seen a cut elimination procedure inside a deep inference system for classical predicate logic. The calculus of structures for classical predicate logic now stands on its own feet, so to speak, as a proof-theoretic formalism: it does not rely on the sequent calculus to prove cut elimination. Since a cut-free deep inference system does not technically have the subformula property, it is a fair question whether it indeed deserves the name "cut-free". The fact that we have easily obtained Herbrand's Theorem from our cut elimination result provides some evidence for a positive answer. Also, the techniques presented here can serve as a basis for native cut elimination procedures in deep inference systems for modal logics like S5. Current cut elimination results for these systems are based on hypersequents [13].

This work does not close the chapter on cut elimination in deep inference for predicate logic. The lemma that turns cuts into splittable cuts makes these cuts shallow at the cost of potentially increasing a lot the cut rank. It is a somewhat unnatural operation in a deep inference system. It would be interesting to make the cut elimination procedure work in the presence of existential quantifiers in the context of a cut. This seems possible and is likely to involve a factorisation as in Herbrand's Theorem as a part of the cut elimination procedure. It would also be interesting to see a cut elimination procedure that works directly on SKSgq.

Proof complexity is a natural direction for future research. As already happens in the propositional case [9], the ability of applying inference rules deep inside of formulas allows for shorter proofs. The question is whether it also leads to a hyperexponential speedup for proofs in predicate logic.

References

[1] Avron, A., 'The method of hypersequents in the proof theory of propositional non-classical logics', in Wilfrid Hodges, Martin Hyland, Charles Steinhorn, and John Truss, (eds.), *Logic: from foundations to applications. Proc. Logic Colloquium, Keele, UK, 1993.* Oxford University Press, New York, 1996, pp. 1–32.

[2] Basin, D., M. D'Agostino, D. M. Gabbay, S. Matthews, and L. Viganò (eds.), *Labelled Deduction*, volume 17 of *Applied Logic Series.* Kluwer Academic Publishers, Dordrecht, 2000.

[3] Belnap, N. D., Jr., 'Display logic', *Journal of Philosophical Logic*, 11:375–417, 1982.

[4] Brünnler, K., 'Atomic cut elimination for classical logic', in M. Baaz and J. A. Makowsky, (eds.), *CSL 2003*, volume 2803 of *Lecture Notes in Computer Science.* Springer-Verlag, 2003, pp. 86–97.

[5] Brünnler, K., *Deep Inference and Symmetry in Classical Proofs.* PhD thesis, Technische Universität Dresden, September 2003.

[6] Buss, S. R., 'On Herbrand's theorem', in *Logic and Computational Complexity*, volume 960 of *Lecture Notes in Computer Science.* Springer-Verlag, 1995, pp. 195–209.

[7] Di Gianantonio, P., 'Structures for multiplicative cyclic linear logic: Deepness vs cyclicity', in J. Marcinkowski and A. Tarlecki, (eds.), *CSL 2004*, volume 3210 of *Lecture Notes in Computer Science.* Springer-Verlag, 2004, pp. 130–144.

[8] Guglielmi, A., 'A system of interaction and structure', Technical Report WV-02-10, Technische Universität Dresden, 2002. To appear in ACM Transactions on Computational Logic.

[9] Guglielmi, A., 'Polynomial size deep-inference proofs instead of exponential size shallow-inference proofs', Manuscript, 2003. http://www.ki.inf.tu-dresden.de/~guglielm/res/notes/AG12.pdf.

[10] Guglielmi, A., 'Resolution in the calculus of structures', Manuscript, 2003. http://www.ki.inf.tu-dresden.de/~guglielm/res/notes/AG10.pdf.

[11] Guglielmi, A., and L.Strassburger, 'Non-commutativity and MELL in the calculus of structures', in L. Fribourg, (ed.), *CSL 2001*, volume 2142 of *Lecture Notes in Computer Science*. Springer-Verlag, 2001, pp. 54–68.

[12] Stewart, Ch., and Ph. Stouppa, 'A systematic proof theory for several modal logics', Technical Report WV-03-08, Technische Universität Dresden, 2003. Accepted at Advances in Modal Logic 2004, to appear in proceedings published by King's College Publications.

[13] Stouppa, P., *The design of modal proof theories: The case of S5*, Master's thesis. Technische Universität Dresden, 2004.

[14] Strassburger, L., *Linear Logic and Noncommutativity in the Calculus of Structures*, PhD thesis. Technische Universität Dresden, 2003.

[15] Tait, W. W., 'Normal derivability in classical logic', in *The Syntax and Semantics of Infinitary Languages*, volume 72 of *Lecture Notes in Mathematics*. Springer, 1968, pp. 204–236.

[16] Tiu, A.F., 'Properties of a Logical System in the Calculus of Structures', Master's thesis. Technische Universität Dresden, 2001.

Kai Brünnler
Institut für angewandte Mathematik und Informatik
Neubrückstr. 10, CH – 3012 Bern, Switzerland
`kai@iam.unibe.ch`