Cut-Elimination: Experiments with CERES[⋆]

Matthias Baaz[1], Stefan Hetzl[2], Alexander Leitsch[2], Clemens Richter[2], and Hendrik Spohr[2]

[1] Institute of Discrete Mathematics and Geometry (E104),
Vienna University of Technology, Wiedner Hauptstraße 8-10,
1040 Vienna, Austria
baaz@logic.at

[2] Institute of Computer Languages (E185),
Vienna University of Technology, Favoritenstraße 9,
1040 Vienna, Austria
{hetzl|leitsch|richter|spohr}@logic.at

Abstract. Cut-elimination is the most prominent form of proof transformation in logic. The elimination of cuts in formal proofs corresponds to the removal of intermediate statements (lemmas) in mathematical proofs. The cut-elimination method CERES (cut-elimination by resolution) works by constructing a set of clauses from a proof with cuts. Any resolution refutation of this set can then serve as a skeleton of a proof with only atomic cuts.

In this paper we present a systematic experiment with the implementation of CERES on a proof of reasonable size and complexity. It turns out that the proof with cuts can be transformed into two *mathematically* different proofs of the theorem. In particular, the application of positive and negative hyperresolution yield different mathematical arguments. As an unexpected side-effect the derived clauses of the resolution refutation proved particularly interesting as they can be considered as meaningful universal lemmas.

Though the proof under investigation is intuitively simple, the experiment demonstrates that new (and relevant) mathematical information on proofs can be obtained by computational methods. It can be considered as a first step in the development of an experimental culture of *computer-aided proof analysis* in mathematics.

1 Introduction

Proof analysis is a central mathematical activity which proved crucial to the development of mathematics. Indeed many mathematical concepts such as the notion of group or the notion of probability were introduced by analyzing existing arguments. In some sense the analysis and synthesis of proofs form the very core of mathematical progress[7,8].

Cut-elimination introduced by Gentzen [4] is the most prominent form of proof transformation in logic and plays an important role in automatizing the

[⋆] supported by the Austrian Science Fund (project no. P16264-N05)

F. Baader and A. Voronkov (Eds.): LPAR 2004, LNAI 3452, pp. 481–495, 2005.

analysis of mathematical proofs. The removal of cuts corresponds to the elimination of intermediate statements (lemmas) from proofs resulting in a proof which is analytic in the sense, that all statements in the proof are subformulas of the result. Therefore, the proof of a combinatorial statement is converted into a purely combinatorial proof. Cut-elimination is therefore an essential tool for the analysis of proofs, especially to make implicit parameters explicit. Cut free derivations allow for

- the extraction of Herbrand disjunctions, which can be used to establish bounds on existential quantifiers (e.g. Luckhardt's analysis of the Theorem of Roth [6]).
- the construction of interpolants, which allow for the replacement of implicit definitions by explicit definitions according to Beth's Theorem.
- the calculation of generalized variants of the end formula.

In a formal sense Girard's analysis of van der Waerden's proof [5] is the application of cut-elimination to the proof of Fürstenberg/Weiss with the "perspective" of obtaining van der Waerden's proof. Indeed an application of a complex proof transformation like cut-elimination by humans requires a goal oriented strategy. In contrast, as we demonstrate in this paper, the application of purely computational methods on existing proofs may produce new interesting proofs. Note that cut-elimination is *non-unique*, i.e. there is no single cut-free proof which represents *the* analytic version of a proof with lemmas. Indeed, it is non-uniqueness which makes computational experiments with cut-elimination interesting. The experiments can be considered as a source for a base of proofs in formal format which provide different mathematical and computational information.

CERES [2] is a cut-elimination method that is based on resolution. The method roughly works as follows: The structure of the proof containing cuts is mapped to a clause term which evaluates to an unsatisfiable set of clauses C (the *characteristic clause set*). A resolution refutation of C, which is obtained using a first-order theorem prover, serves as a skeleton for the new proof which contains only atomic cuts. In a final step also these atomic cuts can be eliminated, provided the (atomic) axioms are valid sequents; but this step is of minor mathematical interest only. In the system CERES[3] this method of cut-elimination has been implemented. The system is capable of dealing with formal proofs in **LK**, among them also very large ones.

In this paper we present a systematic experiment with CERES on a proof defined in [9]. It turns out that the proof with cuts is transformed into two *mathematically* different proofs of the theorem. In particular, the application of positive and negative hyperresolution yield different mathematical arguments. As the core of the method is resolution, which works on the characteristic clause set, it is worthwhile to investigate also the resolution proof itself. In fact the derived clauses of the proof can be considered as universal lemmas, which are eventually instantiated in the procedure. As an unexpected side-effect also these

[3] available at `http://www.logic.at/ceres/`

lemmas proved particularly interesting in the experiment. Though the proof under investigation is intuitively simple, the experiment demonstrates that new (and relevant) mathematical information on proofs can be obtained by computational methods. It can be considered as a first step in the development of an experimental culture of *computer-aided proof analysis* in mathematics.

2 The System CERES

The system CERES is an implementation of the cut-elimination method CERES which will be roughly explained below. Also a short description of the behavior of the system will be given including some implementational details.

2.1 Short Description of the Method via an Example

The cut-elimination method by resolution (CERES) is demonstrated in this paper by the following example. You can find an in-depth explanation of the method itself and the underlying **LK** in [3], [2] and on the CERES web page[4].

To simplify the understanding of the method all the premises (the auxiliary formulas of the inferences) are put in bold face, the conclusions are underlined and the ancestors of cut-formulas are marked with an asterisk in the following input proof.

Now, let φ be the proof

$$\dfrac{\varphi_l \quad \varphi_r}{(\forall x)(\forall y)(P(x,y) \supset Q(x,y)) \vdash (\exists x)(\exists y)(\neg Q(x,y) \supset \neg P(x,y))}\ \text{cut}$$

where φ_l is

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\boldsymbol{P(z,a)}^* \vdash P(z,a)}{\vdash \underline{\neg \boldsymbol{P(z,a)}}^*, P(z,a)}\ \neg:\mathrm{r}}{\vdash \underline{\neg P(z,a) \lor Q(z,a)}^*, \boldsymbol{P(z,a)}}\ \lor:\mathrm{r}_1 \qquad \dfrac{Q(z,a) \vdash \boldsymbol{Q(z,a)}^*}{\boldsymbol{Q(z,a)} \vdash \underline{\neg P(z,a) \lor Q(z,a)}^*}\ \lor:\mathrm{r}_2}{\underline{\boldsymbol{P(z,a)} \supset \boldsymbol{Q(z,a)}} \vdash \neg P(z,a) \lor Q(z,a)^*}\ \supset:\mathrm{l}}{\underline{\boldsymbol{(\forall y)(P(z,y) \supset Q(z,y))}} \vdash \neg P(z,a) \lor Q(z,a)^*}\ \forall:\mathrm{l}}{\underline{(\forall x)(\forall y)(P(x,y) \supset Q(x,y))} \vdash \boldsymbol{\neg P(z,a)} \lor \boldsymbol{Q(z,a)}^*}\ \forall:\mathrm{l}}{(\forall x)(\forall y)(P(x,y) \supset Q(x,y)) \vdash \underline{\boldsymbol{(\exists y)(\neg P(z,y) \lor Q(z,y))}}^*}\ \exists:\mathrm{r}}{(\forall x)(\forall y)(P(x,y) \supset Q(x,y)) \vdash \underline{\boldsymbol{(\forall x)(\exists y)(\neg P(x,y) \lor Q(x,y))}}^*}\ \forall:\mathrm{r}$$

[4] The documentation and an online version of the system CERES are available at `http://www.logic.at/ceres/`.

and φ_r is

$$\begin{prooftree}
\AxiomC{$P(b,v) \vdash \boldsymbol{P(b,v)}^*$}
\RightLabel{$\neg:\mathrm{l}$}
\UnaryInfC{$\underline{\neg P(b,v)^*}, \boldsymbol{P(b,v)} \vdash$}
\RightLabel{$\neg:\mathrm{r}$}
\UnaryInfC{$\boldsymbol{\neg P(b,v)}^* \vdash \underline{\neg P(b,v)}$}
\AxiomC{$Q(b,v)^* \vdash \boldsymbol{Q(b,v)}$}
\RightLabel{$\neg:\mathrm{l}$}
\UnaryInfC{$\underline{\neg Q(b,v)}, \boldsymbol{Q(b,v)}^* \vdash$}
\RightLabel{$\vee:\mathrm{l}'$}
\BinaryInfC{$\boldsymbol{\neg Q(b,v)}, \underline{\neg P(b,v) \vee Q(b,v)^*} \vdash \boldsymbol{\neg P(b,v)}$}
\RightLabel{$\supset:\mathrm{r}$}
\UnaryInfC{$\neg P(b,v) \vee Q(b,v)^* \vdash \underline{\boldsymbol{\neg Q(b,v) \supset \neg P(b,v)}}$}
\RightLabel{$\exists:\mathrm{r}$}
\UnaryInfC{$\neg P(b,v) \vee Q(b,v)^* \vdash \underline{\boldsymbol{(\exists y)(\neg Q(b,y) \supset \neg P(b,y))}}$}
\RightLabel{$\exists:\mathrm{r}$}
\UnaryInfC{$\boldsymbol{\neg P(b,v) \vee Q(b,v)}^* \vdash \underline{(\exists x)(\exists y)(\neg Q(x,y) \supset \neg P(x,y))}$}
\RightLabel{$\exists:\mathrm{l}$}
\UnaryInfC{$\underline{\boldsymbol{(\exists y)(\neg P(b,y) \vee Q(b,y))}}^* \vdash (\exists x)(\exists y)(\neg Q(x,y) \supset \neg P(x,y))$}
\RightLabel{$\forall:\mathrm{l}$}
\UnaryInfC{$\underline{\boldsymbol{(\forall x)(\exists y)(\neg P(x,y) \vee Q(x,y))}}^* \vdash (\exists x)(\exists y)(\neg Q(x,y) \supset \neg P(x,y))$}
\end{prooftree}$$

The extraction of the characteristic clause term happens top down starting with those parts of the initial sequents that are marked as ancestors of cut formulas which are now interpreted as sets. At every occurrence of a binary rule the two clause terms resulting from the premises are connected by a binary operator. Depending whether the auxiliary formulas of the inference were ancestors of cut formulas or not the operator will either be $\oplus$ or $\otimes$. All unary inference rules have no influence on the clause term and hence it remains unchanged.

For the example this yields the following characteristic clause term

$$\Theta(\varphi) = ((\{P(z,a) \vdash\} \otimes \{\vdash Q(z,a)\}) \oplus (\{\vdash P(b,v)\} \oplus \{Q(b,v) \vdash\}))$$

which characterizes those parts of the axiom sequents which have been used to derive the cut formula (on both sides).

The operator $\oplus$ of the clause term is interpreted as union and the operator $\otimes$ as merge, i.e. the antecedens and consequent parts of different sequents are exchanged such that only one part is exchanged at once.

Hence by evaluation of $\Theta(\varphi)$ for the characteristic clause set $|\Theta(\varphi)|$ of φ we obtain

$$\begin{aligned}
|\Theta(\varphi)| = \{P(z,a) &\vdash Q(z,a), && (C_2)\\
&\vdash P(b,v), && (C_1)\\
Q(b,v) &\vdash\}. && (C_3)
\end{aligned}$$

The characteristic clause set of an **LK** derivation is always unsatisfiable. Therefore one can always find a resolution refutation of the characteristic clause set.

In particular, we define a resolution refutation δ of $|\Theta(\varphi)|$:

$$\dfrac{Q(b,v) \vdash \qquad \dfrac{\vdash P(b,v) \quad P(z,a) \vdash Q(z,a)}{\vdash Q(b,a)}}{\vdash}$$

and a corresponding ground refutation γ of δ, i. e. $\gamma = \delta\sigma$:

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

with the ground substitution $\sigma = \{v \mapsto a, z \mapsto b\}$.

Now we have to reduce φ to projections of the clauses used as initial clauses in the resolution refutation of $|\Theta(\varphi)|$. A projection of φ w.r.t. a clause in $|\Theta(\varphi)|$ is defined by skipping all inferences going into cuts, which leads to cut-free proof of (a subsequent of) the end sequent extended by C. Projections may be understood as projection schemes of the clauses in question modulo a corresponding ground substitution.

Again, we start at the initial sequents (without those parts marked as ancestors of cut formulas and not necessary for the creation of the clause in question) and apply all inference rules not operating on ancestors of cut formulas until all such binary rules have been applied and at least one formula also occurring in the end sequent has been composed.

The projection scheme of φ corresponding to the clause C_1 is:

$\varphi(C_1) =$

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

and let the ground projection $\chi_1 = \varphi(C_1)\sigma$.

The projection scheme of φ corresponding to the clause C_2 is:

$\varphi(C_2) =$

$$\dfrac{\dfrac{\dfrac{P(z,a) \vdash P(z,a) \quad Q(z,a) \vdash Q(z,a)}{P(z,a) \supset Q(z,a), P(z,a) \vdash Q(z,a)}\ \supset : l}{(\forall y)(P(z,y) \supset Q(z,y)), P(z,a) \vdash Q(z,a)}\ \forall : l}{(\forall x)(\forall y)(P(x,y) \supset Q(x,y)), P(z,a) \vdash Q(z,a)}\ \forall : l$$

and let the ground projection $\chi_2 = \varphi(C_2)\sigma$.

The projection scheme of φ corresponding to the clause C_3 is:

$\varphi(C_3) =$

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

and let the ground projection $\chi_3 = \varphi(C_3)\sigma$.

Finally the ground projections can be composed to a cut-free proof of φ, i.e. a proof of φ containing only atomic cuts, using its resolution refutation as a skeleton.

$$\dfrac{\dfrac{\overset{(\chi_1)}{\vdash Y, P(b,a)} \quad \overset{(\chi_2)}{P(b,a), X \vdash Q(b,a)}}{X \vdash Y, Q(b,a)}\;\text{cut} \qquad \overset{(\chi_3)}{Q(b,a) \vdash Y}}{X \vdash Y}\;\text{cut}$$

where $X = (\forall x)(\forall y)(P(x,y) \supset Q(x,y))$ and $Y = (\exists x)(\exists y)(\neg Q(x,y) \supset \neg P(x,y))$.

2.2 Description of the Program

The cut-elimination program CERES is written in ANSI-C++[5]. There are two main tasks. On the one hand to compute an unsatisfiable set of clauses $\mathcal{C}$ characterising the cut formulas. This is done by automatically extracting the characteristic clause term and computation of the resulting characteristic clause set. On the other hand to evaluate the resolution refutation gained from an external theorem prover[6] and to compute the necessary projection schemes which are properly instantiated and concatenated using the resolution refutation as a skeleton of the cut-free proof, i.e. a proof without non-atomic cuts.

The input format and the output format are following the *proof* style[7] of LaTeX with some extensions, and are translatable by any LaTeX compiler. This feature allows an easier input of proofs and reading of the output. Nevertheless new approaches are planned (see section 4 for details).

[5] The C++ Programming Language following the International Standard 14882:1998 approved as an American National Standard (see `http://www.ansi.org`).

[6] The current version of CERES uses the automated theorem prover Otter (see `http://www-unix.mcs.anl.gov/AR/otter/`), but any refutational theorem prover may be used.

[7] see `http://research.nii.ac.jp/~tatsuta/proof-sty.html`

3 Experiments with Resolution Refinements

The use of the resolution refutation of the characteristic clause set as a skeleton for the cut-free proof makes it possible to change the mathematical character of the resulting proof via different resolution refutations, e.g. using different resolution refinements. Within these refutations *universal lemmas*, i.e. clauses containing variables representing universal formulas, appear which do neither occur in the original proof nor in the cut-eliminated proofs, where they are already instantiated.

Now we are doing exactly such an interesting experiment using an input proof already analyzed and defined as an **LK**-derivation in [9] with the program CERES.

The proof deals with the following situation: We are given an infinite tape where each cell contains either ‘0’ or ‘1’. We prove that on this tape there are two cells with the same value. The contents of a cell of the tape is denoted by f, s is the sucessor function and $m^{x,y}$ is the maximum of x and y.

Within this section the following formula abbreviations are used:

$$\begin{aligned}
M_1 &= (\forall y)(\forall x) x \leq m^{x,y}\\
M_2 &= (\forall y)(\forall x) y \leq m^{x,y}\\
S &= (\forall x)(\forall y)(s(x) \leq y \supset x < y)\\
T &= (\forall i)(\forall x)(\forall y)((f(x) = i \wedge f(y) = i) \supset f(x) = f(y))\\
A &= (\forall x)(f(x) = 0 \vee f(x) = 1)\\
P &= (\exists p)(\exists q)(p < q \wedge f(p) = f(q))\\
\infty_0 &= (\forall n)(\exists k)(n \leq k \wedge f(k) = 0)\\
\infty_1 &= (\forall m)(\exists l)(m \leq l \wedge f(l) = 1)
\end{aligned}$$

moreover 1 is an abbreviation for $s(0)$.

Then, let the proof φ be defined as follows.

$$\varphi = \dfrac{\dfrac{\overset{(\tau)}{M_1, M_2, A \vdash \infty_0, \infty_1} \qquad \overset{(\epsilon_1)}{\infty_1, S, T \vdash P}}{M_1, M_2, S, T, A \vdash P, \infty_0}\,\text{cut} \qquad \overset{(\epsilon_0)}{\infty_0, S, T \vdash P}}{M_1, M_2, S, T, A \vdash P}\,\text{cut}$$

For the subproofs of τ, ϵ_0 and ϵ_1 please see [9] and the appendix.

The characteristic clause term $\Theta(\varphi)$ extracted from φ is

$$\begin{aligned}
\Theta(\varphi) =& (((\{\vdash v \leq m^{u,v}\} \oplus (\{\vdash u \leq m^{u,v}\} \oplus (\{\vdash f(m^{u,v}) = 0\} \otimes \{\vdash f(m^{u,v}) = 1\})))\\
&\oplus ((\{s(u) \leq v \vdash\} \otimes \{\vdash\}) \otimes ((\{f(u) = 1 \vdash\} \otimes \{f(v) = 1 \vdash\}) \otimes \{\vdash\})))\\
&\oplus ((\{s(u) \leq v \vdash\} \otimes \{\vdash\}) \otimes ((\{f(u) = 0 \vdash\} \otimes \{f(v) = 0 \vdash\}) \otimes \{\vdash\})))
\end{aligned}$$

and the corresponding characteristic clause set $|\Theta(\varphi)|$ obtained from $\Theta(\varphi)$ is

$$\begin{aligned}
|\Theta(\varphi)| = \{ & \vdash v \leq m^{u,v}, && (C_1)\\
& \vdash u \leq m^{u,v}, && (C_2)\\
& \vdash f(m^{u,v}) = 0, f(m^{u,v}) = 1, && (C_3)\\
& s(u) \leq v, f(u) = 1, f(v) = 1 \vdash, && (C_4)\\
& s(u) \leq v, f(u) = 0, f(v) = 0 \vdash \} && (C_5)
\end{aligned}$$

The projection schemes obtained from φ for the five clauses above are the following:

$\varphi(C_1) =$

$$\dfrac{\dfrac{v \leq m^{u,v} \vdash v \leq m^{u,v}}{(\forall x)v \leq m^{x,v} \vdash v \leq m^{u,v}}\ \forall:\mathrm{l}}{(\forall y)(\forall x)y \leq m^{x,y} \vdash v \leq m^{u,v}}\ \forall:\mathrm{l}$$

$\varphi(C_2) =$

$$\dfrac{\dfrac{u \leq m^{u,v} \vdash u \leq m^{u,v}}{(\forall x)x \leq m^{x,v} \vdash u \leq m^{u,v}}\ \forall:\mathrm{l}}{(\forall y)(\forall x)x \leq m^{x,y} \vdash u \leq m^{u,v}}\ \forall:\mathrm{l}$$

$\varphi(C_3) =$

$$\dfrac{\dfrac{f(m^{u,v}) = 0 \vdash f(m^{u,v}) = 0 \quad f(m^{u,v}) = 1 \vdash f(m^{u,v}) = 1}{f(m^{u,v}) = 0 \vee f(m^{u,v}) = 1 \vdash f(m^{u,v}) = 0, f(m^{u,v}) = 1}\ \vee:\mathrm{l}}{(\forall x)(f(x) = 0 \vee f(x) = 1) \vdash f(m^{u,v}) = 0, f(m^{u,v}) = 1}\ \forall:\mathrm{l}$$

$\varphi(C_4) = \psi_1$
$\varphi(C_5) = \psi_0$

where ψ_j is defined:

$\psi_j =$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{s(u) \leq v \vdash s(u) \leq v \quad u < v \vdash u < v}{s(u) \leq v \supset u < v, s(u) \leq v \vdash u < v}\ \supset:\mathrm{l}}{(\forall y)(s(u) \leq y \supset u < y), s(u) \leq v \vdash u < v}\ \forall:\mathrm{l}}{(\forall x)(\forall y)(s(x) \leq y \supset x < y), s(u) \leq v \vdash u < v}\ \forall:\mathrm{l} \quad \psi'_j}{S, s(u) \leq v, T, f(u) = j, f(v) = j \vdash u < v \wedge f(u) = f(v)}\ \wedge:\mathrm{r}}{S, s(u) \leq v, T, f(u) = j, f(v) = j \vdash (\exists q)(u < q \wedge f(u) = f(q))}\ \exists:\mathrm{r}}{S, s(u) \leq v, T, f(u) = j, f(v) = j \vdash (\exists p)(\exists q)(p < q \wedge f(p) = f(q))}\ \exists:\mathrm{r}$$

$\psi'_j =$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{f(u) = j \vdash f(u) = j \quad f(v) = j \vdash f(v) = j}{f(u) = j, f(v) = j \vdash f(u) = j \wedge f(v) = j}\ \wedge:\mathrm{r} \quad f(u) = f(v) \vdash f(u) = f(v)}{(f(u) = j \wedge f(v) = j) \supset f(u) = f(v), f(u) = j, f(v) = j \vdash f(u) = f(v)}\ \supset:\mathrm{l}}{(\forall y)((f(u) = j \wedge f(y) = j) \supset f(u) = f(y)), f(u) = j, f(v) = j \vdash f(u) = f(v)}\ \forall:\mathrm{l}}{(\forall x)(\forall y)((f(x) = j \wedge f(y) = j) \supset f(x) = f(y)), f(u) = j, f(v) = j \vdash f(u) = f(v)}\ \forall:\mathrm{l}}{(\forall i)(\forall x)(\forall y)((f(x) = i \wedge f(y) = i) \supset f(x) = f(y)), f(u) = j, f(v) = j \vdash f(u) = f(v)}\ \forall:\mathrm{l}$$

The resolution refutations yielding two mathematically different proofs of φ are demonstrated in the following two subsections. The resulting cut-free proofs have been ommited because of their sizes.

3.1 Positive Hyperresolution

Derivation of C_6:

$$\dfrac{\dfrac{\overset{(C_4\sigma_1)}{s(u') \leq v', f(u') = 1, f(v') = 1 \vdash} \quad \overset{(C_2\sigma_2)}{\vdash u \leq m^{u,w}}}{f(u') = 1, f(m^{s(u'),w}) = 1 \vdash}\,\sigma_3 \qquad \overset{(C_3)}{\vdash f(m^{u,v}) = 0, f(m^{u,v}) = 1}}{\underbrace{f(m^{s(m^{u,v}),w}) = 1 \vdash f(m^{u,v}) = 0}_{C_X}}\,\sigma_4$$

$$\dfrac{C_X \quad \overset{(C_3\sigma_5)}{\vdash f(m^{u',v'}) = 0, f(m^{u',v'}) = 1}}{\vdash f(m^{u,v}) = 0, f(m^{s(m^{u,v}),w}) = 0}\,\sigma_6 \qquad (C_6)$$

where $\sigma_1 = \{u \mapsto u', v \mapsto v'\}$, $\sigma_2 = \{v \mapsto w\}$, $\sigma_3 = \{u \mapsto s(u'), v' \mapsto m^{s(u'),w}\}$, $\sigma_4 = \{u' \mapsto m^{u,v}\}$, $\sigma_5 = \{u \mapsto u', v \mapsto v'\}$ and $\sigma_6 = \{u' \mapsto s(m^{u,v}), v' \mapsto w\}$.

For arbitrary u, v and w either the cell with index $i = m^{u,v}$ is labelled '0' or the cell with index $m^{i+1,w}$.

Derivation of C_7:

$$\dfrac{\dfrac{\overset{(C_5\sigma_7)}{s(u') \leq v', f(u') = 0, f(v') = 0 \vdash} \quad \overset{(C_1\sigma_8)}{\vdash v \leq m^{u'',v}}}{f(u') = 0, f(m^{u'',s(u')}) = 0 \vdash}\,\sigma_9 \qquad \overset{(C_6)}{\vdash f(m^{u,v}) = 0, f(m^{s(m^{u,v}),w}) = 0}}{\underbrace{f(m^{u'',s(m^{s(m^{u,v}),w})}) = 0 \vdash f(m^{u,v}) = 0}_{C_Y}}\,\sigma_{10}$$

$$\dfrac{\dfrac{C_Y \quad \overset{(C_6\sigma_{11})}{\vdash f(m^{u',v'}) = 0, f(m^{s(m^{u',v'}),w'}) = 0}}{\vdash f(m^{u,v}) = 0, f(m^{u',v'}) = 0}\,\sigma_{12}}{\vdash f(m^{u,v}) = 0}\,\sigma_{13} \qquad (C_7)$$

where $\sigma_7 = \{u \mapsto u', v \mapsto v'\}$, $\sigma_8 = \{u \mapsto u''\}$, $\sigma_9 = \{v \mapsto s(u'), v' \mapsto m^{u'',s(u')}\}$, $\sigma_{10} = \{u' \mapsto m^{s(m^{u,v}),w}\}$, $\sigma_{11} = \{u \mapsto u', v \mapsto v', w \mapsto w'\}$, $\sigma_{12} = \{u'' \mapsto s(m^{u',v'}), w' \mapsto s(m^{s(m^{u,v}),w})\}$ and $\sigma_{13} = \{u' \mapsto u, v' \mapsto v\}$.

For arbitrary u and v the cell with index $i = m^{u,v}$ is labelled ‘0’.

$$\dfrac{\dfrac{\dfrac{\overset{(C_5)}{s(u) \le v, f(u) = 0, f(v) = 0 \vdash} \quad \overset{(C_2\sigma_{14})}{\vdash u' \le m^{u',v'}}}{f(u) = 0, f(m^{s(u),v'}) = 0 \vdash}\;\sigma_{15} \quad \overset{(C_7\sigma_{16})}{\vdash f(m^{u',v}) = 0}}{f(m^{s(m^{u',v}),v'}) = 0 \vdash}\;\sigma_{17} \quad \overset{(C_7\sigma_{18})}{\vdash f(m^{u,v''}) = 0}}{\vdash}\;\sigma_{19}$$

where $\sigma_{14} = \{u \mapsto u', v \mapsto v'\}$, $\sigma_{15} = \{u' \mapsto s(u), v \mapsto m^{s(u),v'}\}$, $\sigma_{16} = \{u \mapsto u'\}$, $\sigma_{17} = \{u \mapsto m^{u',v}\}$, $\sigma_{18} = \{v \mapsto v''\}$ and $\sigma_{19} = \{u \mapsto s(m^{u',v}), v'' \mapsto v'\}$.

For arbitrary u and v where $u < v$ at least one of the cells with index u or v should be labelled ‘1’ but again for arbitrary u' and v' the cell with index $i = m^{u',v'}$ is labelled ‘0’. Hence choosing one time u as u' and one time v as v' leads to a contradiction.

3.2 Negative Hyperresolution

Derivation of C_6':

$$\dfrac{\overset{(C_1\sigma_1)}{\vdash v' \le m^{u,v'}} \quad \overset{(C_4\sigma_2)}{s(v) \le u', f(v) = 1, f(u') = 1 \vdash}}{f(v) = 1, f(m^{u,s(v)}) = 1 \vdash}\;\sigma_3 \qquad (C_6')$$

where $\sigma_1 = \{v \mapsto v'\}$, $\sigma_2 = \{u \mapsto v, v \mapsto u'\}$ and $\sigma_3 = \{u' \mapsto m^{u,s(v)}, v' \mapsto s(v)\}$.

If a cell with index v is labelled ‘1’ then no cell with an index bigger than v is labelled ‘1’.

Derivation of C_7':

$$\dfrac{\overset{(C_2\sigma_4)}{\vdash u' \le m^{u',v}} \quad \overset{(C_5\sigma_5)}{s(u) \le v', f(u) = 0, f(v') = 0 \vdash}}{f(u) = 0, f(m^{s(u),v}) = 0 \vdash}\;\sigma_6 \qquad (C_7')$$

where $\sigma_4 = \{u \mapsto u'\}$, $\sigma_5 = \{v \mapsto v'\}$ and $\sigma_6 = \{u' \mapsto s(u), v' \mapsto m^{s(u),v}\}$.

If a cell with index u is labelled ‘0’ then no cell with an index bigger than u is labelled ‘0’.

Derivation of C'_8:

$$\dfrac{\overset{(C_3\sigma_7)}{\vdash f(m^{u',v'}) = 0, f(m^{u',v'}) = 1} \quad \overset{(C'_7)}{f(u) = 0, f(m^{s(u),v}) = 0 \vdash}}{\underbrace{f(u) = 0 \vdash f(m^{s(u),v'}) = 1}_{C'_X}}\ \sigma_8$$

$$\dfrac{C'_X \quad \overset{(C'_6\sigma_9)}{f(v) = 1, f(m^{u',s(v)}) = 1 \vdash}}{f(v) = 1, f(u) = 0 \vdash}\ \sigma_{10} \qquad (C'_8)$$

where $\sigma_7 = \{u \mapsto u', v \mapsto v'\}$, $\sigma_8 = \{u' \mapsto s(u), v \mapsto v'\}$, $\sigma_9 = \{u \mapsto u'\}$ and $\sigma_{10} = \{u' \mapsto s(u), v' \mapsto s(v)\}$.

If a cell with index v is labelled ‘1’ then there is no cell with index u labelled ‘0’, i.e. all cells are either only labelled ‘0’ or only labelled ‘1’.

Derivation of C'_9:

$$\dfrac{\overset{(C_3\sigma_{11})}{\vdash f(m^{u',v'}) = 0, f(m^{u',v'}) = 1} \quad \overset{(C'_7)}{f(u) = 0, f(m^{s(u),v}) = 0 \vdash}}{\underbrace{f(u) = 0 \vdash f(m^{s(u),v}) = 1}_{C'_Y}}\ \sigma_{12}$$

$$\dfrac{\dfrac{C'_Y \quad \overset{(C'_8\sigma_{13})}{f(v') = 1, f(u') = 0 \vdash}}{f(u) = 0, f(u') = 0 \vdash}\ \sigma_{14}}{f(u) = 0 \vdash}\ \sigma_{15} \qquad (C'_9)$$

where $\sigma_{11} = \{u \mapsto u', v \mapsto v'\}$, $\sigma_{12} = \{v' \mapsto v, u' \mapsto s(u)\}$, $\sigma_{13} = \{u \mapsto u', v \mapsto v'\}$, $\sigma_{14} = \{v' \mapsto m^{s(u),v}\}$ and $\sigma_{15} = \{u' \mapsto u\}$.

No cell is labelled ‘0’.

Derivation of C'_{10}:

$$\dfrac{\overset{(C_3\sigma_{16})}{\vdash f(m^{u',v'}) = 0, f(m^{u',v'}) = 1} \quad \overset{(C'_8)}{f(v) = 1, f(u) = 0 \vdash}}{\underbrace{f(v) = 1 \vdash f(m^{u',v'}) = 1}_{C'_Z}}\ \sigma_{17}$$

$$\dfrac{\dfrac{C'_Z \quad \overset{(C'_6\sigma_{18})}{f(v'') = 1, f(m^{u,s(v'')}) = 1 \vdash}}{f(v) = 1, f(v'') = 1 \vdash}\ \sigma_{19}}{f(v) = 1 \vdash}\ \sigma_{20} \qquad (C'_{10})$$

where $\sigma_{16} = \{u \mapsto u', v \mapsto v'\}$, $\sigma_{17} = \{u \mapsto m^{u',v'}\}$, $\sigma_{18} = \{v \mapsto v''\}$, $\sigma_{19} = \{u' \mapsto u, v' \mapsto s(v'')\}$ and $\sigma_{20} = \{v'' \mapsto v\}$.

No cell is labelled ‘1’.

$$\dfrac{\dfrac{\overset{(C_3)}{\vdash f(m^{u,v}) = 0, f(m^{u,v}) = 1} \quad \overset{(C'_9\sigma_{21})}{f(u') = 0 \vdash}}{\vdash f(m^{u,v}) = 1}\ \sigma_{22} \qquad \overset{(C'_{10}\sigma_{23})}{f(v') = 1 \vdash}}{\vdash}\ \sigma_{24}$$

where $\sigma_{21} = \{u \mapsto u'\}$, $\sigma_{22} = \{u' \mapsto m^{u,v}\}$, $\sigma_{23} = \{v \mapsto v'\}$ and $\sigma_{24} = \{v' \mapsto m^{u,v}\}$.

The contradiction follows from the axiom that for arbitrary u and v the cell with the index $m^{u,v}$ is either labelled with ‘0’ or with ‘1’ in combination with the facts that no cell is labelled ‘0’ and no cell is labelled ‘1’.

4 Possible Extensions

We plan to develop the following extensions of CERES:

- Due to the central importance of equality in mathematical proofs an investigation of cut-elimination in proofs with equality is very important to the application of cut-elimination. We intend to use the Gentzen calculus **LK** with the paramodulation rule (we refer to [10]) and to extend CERES to equality.

- As the cut-free proofs are often very large and difficult to interpret, we intend to provide the possibility to analyse certain characteristics of the cut-free proof (which are simpler than the proof itself). An important example are Herbrand sequents which may serve to extract bounds from proofs (see e.g. [6]). We plan to develop algorithms for extracting Herbrand sequents (also from proofs of nonprenex sequents as indicated in [1]) and for computing interpolants.

- A great challenge in the formal analysis of mathematical proofs lies in providing a suitable format for the input and output of proofs. We plan to develop an intermediary proof language connecting the language of mathematical proofs with **LK**. Furthermore we will implement a proof editor with a graphical user interface that allows for convenient input and analysis of the output of CERES.

- In the present version CERES eliminates all cuts at once. But - for the application to real mathematical proofs - only interesting cuts (i.e. lemmas) deserve to be eliminated, others should be integrated as additional axioms.

5 Conclusion

The computer experiments with CERES described in this paper lead to the following main consequences:

- even in the simple proof under consideration numerous formal variants of cut free proofs condense to relatively few mathematically distinguishable variants.

- On the other hand, the number of mathematically distinguishable variants is *greater than one.* This demonstrates, that the non-confluence of CERES is not just a formality within **LK**.

- CERES does not eliminate the mathematical activity of cut-elimination, it just supports it. In fact it is essential to interprete the resources and results mathematically.

- New features of CERES, concerning the relation of resolution refutations of the characteristic clause set and the proof projections, evolved in the course of the computer experiments.

References

1. M. Baaz, A. Leitsch: On skolemization and proof complexity, *Fundamenta Informaticae*, 20(4), pp. 353–379, 1994.
2. M. Baaz, A. Leitsch: Cut-Elimination and Redundancy-Elimination by Resolution, *Journal of Symbolic Computation*, 29, pp. 149-176, 2000.
3. M. Baaz, A. Leitsch: Towards a Clausal Analysis of Cut-Elimination, *Journal of Symbolic Computation* to appear.
4. G. Gentzen: Untersuchungen über das logische Schließen, *Mathematische Zeitschrift*, 39, pp. 405–431, 1934–1935.
5. J.Y. Girard: Proof Theory and Logical Complexity, in *Studies in Proof Theory*, Bibliopolis, Napoli, 1987.
6. H. Luckhardt: Herbrand-Analysen zweier Beweise des Satzes von Roth: polynomiale Anzahlschranken. *The Journal of Symbolic Logic*, 54, pp. 234–263, 1989.
7. G. Polya: Mathematics and plausible reasoning, Volume I: Induction and Analogy in Mathematics. Princeton University Press, Princeton, New Jersey, 1954.
8. G. Polya: Mathematics and plausible reasoning, Volume II: Patterns of Plausible Inference. Princeton University Press, Princeton, New Jersey, 1954.
9. C. Urban: Classical Logic and Computation. Ph.D. Thesis, University of Cambridge Computer Laboratory, 2000.
10. A. Degtyarev, A. Voronkov: Equality Reasoning in Sequent-Based Calculi, *Handbook of Automated Reasoning*, vol. I, ed. by A. Robinson and A. Voronkov, chapter 10, pp. 611-706, Elsevier Science, 2001.

APPENDIX

Input Proof

This is the proof[8] used for the experiments in section 3. Again all the premises (the auxiliary formulas of the inferences) are put in bold face, the conclusions are underlined and the same formula abbreviations are used.

$p =$

$$\dfrac{\dfrac{\overset{(\tau)}{M_1, M_2, A \vdash \infty_0, \infty_1} \quad \overset{(\epsilon_1)}{\infty_1, S, T \vdash P}}{M_1, M_2, S, T, A \vdash P, \infty_0}\,\text{cut} \qquad \overset{(\epsilon_0)}{\infty_0, S, T \vdash P}}{M_1, M_2, S, T, A \vdash P}\,\text{cut}$$

$\tau =$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\boldsymbol{v \le m^{u,v}} \vdash v \le m^{u,v}}{\underline{\boldsymbol{(\forall x)v \le m^{x,v}}} \vdash v \le m^{u,v}}\,\forall:\mathrm{l}}{\underline{(\forall y)(\forall x)y \le m^{x,y}} \vdash \boldsymbol{v \le m^{u,v}}}\,\forall:\mathrm{l} \qquad \dfrac{\dfrac{\dfrac{\boldsymbol{u \le m^{u,v}} \vdash u \le m^{u,v}}{\underline{\boldsymbol{(\forall x)x \le m^{x,v}}} \vdash u \le m^{u,v}}\,\forall:\mathrm{l}}{\underline{(\forall y)(\forall x)x \le m^{x,y}} \vdash \boldsymbol{u \le m^{u,v}}}\,\forall:\mathrm{l} \quad (\tau')}{M_1, A \vdash \underline{u \le m^{u,v} \land f(m^{u,v}) = 0}, \boldsymbol{f(m^{u,v}) = 1}}\,\land:\mathrm{r}}{M_1, M_2, A \vdash \boldsymbol{u \le m^{u,v} \land f(m^{u,v}) = 0}, \underline{v \le m^{u,v} \land f(m^{u,v}) = 1}}\,\land:\mathrm{r}}{M_1, M_2, A \vdash \underline{(\exists k)(u \le k \land f(k) = 0)}, \boldsymbol{v \le m^{u,v} \land f(m^{u,v}) = 1}}\,\exists:\mathrm{r}}{M_1, M_2, A \vdash \boldsymbol{(\exists k)(u \le k \land f(k) = 0)}, \underline{(\exists l)(v \le l \land f(l) = 1)}}\,\exists:\mathrm{r}}{M_1, M_2, A \vdash \underline{(\forall n)(\exists k)(n \le k \land f(k) = 0)}, \boldsymbol{(\exists l)(v \le l \land f(l) = 1)}}\,\forall:\mathrm{r}}{M_1, M_2, A \vdash \infty_0, \underline{(\forall m)(\exists l)(m \le l \land f(l) = 1)}}\,\forall:\mathrm{r}$$

$\tau' =$

$$\dfrac{\dfrac{\boldsymbol{f(m^{u,v}) = 0} \vdash f(m^{u,v}) = 0 \quad \boldsymbol{f(m^{u,v}) = 1} \vdash f(m^{u,v}) = 1}{\underline{\boldsymbol{f(m^{u,v}) = 0 \lor f(m^{u,v}) = 1}} \vdash f(m^{u,v}) = 0, f(m^{u,v}) = 1}\,\lor:\mathrm{l}}{\underline{(\forall x)(f(x) = 0 \lor f(x) = 1)} \vdash \boldsymbol{f(m^{u,v}) = 0}, f(m^{u,v}) = 1}\,\forall:\mathrm{l}$$

$\epsilon_0 =$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\overset{(\epsilon_0')}{0 \le u \land f(u) = 0, \boldsymbol{s(u) \le v \land f(v) = 0}, S, T \vdash P}}{0 \le u \land f(u) = 0, \underline{\boldsymbol{(\exists k)(s(u) \le k \land f(k) = 0)}}, S, T \vdash P}\,\exists:\mathrm{l}}{\boldsymbol{0 \le u \land f(u) = 0}, \underline{(\forall n)(\exists k)(n \le k \land f(k) = 0)}, S, T \vdash P}\,\forall:\mathrm{l}}{\underline{\boldsymbol{(\exists k)(0 \le k \land f(k) = 0)}}, (\forall n)(\exists k)(n \le k \land f(k) = 0), S, T \vdash P}\,\exists:\mathrm{l}}{\underline{\boldsymbol{(\forall n)(\exists k)(n \le k \land f(k) = 0)}}, \boldsymbol{(\forall n)(\exists k)(n \le k \land f(k) = 0)}, S, T \vdash P}\,\forall:\mathrm{l}}{\underline{(\forall n)(\exists k)(n \le k \land f(k) = 0)}, S, T \vdash P}\,\mathrm{c}:\mathrm{l}$$

[8] specified and analyzed by Urban[9]

$\epsilon_1 =$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{(\epsilon_1')}{1 \leq u \wedge f(u) = 1, \boldsymbol{s(u) \leq v \wedge f(v) = 1}, S, T \vdash P}}{1 \leq u \wedge f(u) = 1, \underline{\boldsymbol{(\exists l)(s(u) \leq l \wedge f(l) = 1)}}, S, T \vdash P}\;\exists:\mathrm{l}}{\boldsymbol{1 \leq u \wedge f(u) = 1}, \underline{(\forall m)(\exists l)(m \leq l \wedge f(l) = 1)}, S, T \vdash P}\;\forall:\mathrm{l}}{\underline{\boldsymbol{(\exists l)(1 \leq l \wedge f(l) = 1)}}, (\forall m)(\exists l)(m \leq l \wedge f(l) = 1), S, T \vdash P}\;\exists:\mathrm{l}}{\underline{\boldsymbol{(\forall m)(\exists l)(m \leq l \wedge f(l) = 1)}}, \boldsymbol{(\forall m)(\exists l)(m \leq l \wedge f(l) = 1)}, S, T \vdash P}\;\forall:\mathrm{l}}{\underline{(\forall m)(\exists l)(m \leq l \wedge f(l) = 1)}, S, T \vdash P}\;\mathrm{c}:\mathrm{l}$$

$\epsilon_j' =$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{s(u) \leq v \vdash \boldsymbol{s(u) \leq v} \quad \boldsymbol{u < v} \vdash u < v}{s(u) \leq v, \underline{\boldsymbol{s(u) \leq v \supset u < v}} \vdash u < v}\;\supset:\mathrm{l}}{s(u) \leq v, \underline{\boldsymbol{(\forall y)(s(u) \leq y \supset u < y)}} \vdash u < v}\;\forall:\mathrm{l}}{s(u) \leq v, \underline{(\forall x)(\forall y)(s(x) \leq y \supset x < y)} \vdash \boldsymbol{u < v}}\;\forall:\mathrm{l} \quad (\epsilon_j'')}{f(u) = j, \boldsymbol{s(u) \leq v}, \boldsymbol{f(v) = j}, S, T \vdash \underline{u < v \wedge f(u) = f(v)}}\;\wedge:\mathrm{r}}{\boldsymbol{f(u) = j}, \underline{s(u) \leq v \wedge f(v) = j}, S, T \vdash u < v \wedge f(u) = f(v)}\;\wedge:\mathrm{l}}{\underline{j \leq u \wedge f(u) = j}, s(u) \leq v \wedge f(v) = j, S, T \vdash \boldsymbol{u < v \wedge f(u) = f(v)}}\;\wedge:\mathrm{l}}{j \leq u \wedge f(u) = j, s(u) \leq v \wedge f(v) = j, S, T \vdash \underline{\boldsymbol{(\exists q)(u < q \wedge f(u) = f(q))}}}\;\exists:\mathrm{r}}{j \leq u \wedge f(u) = j, s(u) \leq v \wedge f(v) = j, S, T \vdash \underline{(\exists p)(\exists q)(p < q \wedge f(p) = f(q))}}\;\exists:\mathrm{r}$$

$\epsilon_j'' =$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{f(u) = j \vdash \boldsymbol{f(u) = j} \quad f(v) = j \vdash \boldsymbol{f(v) = j}}{f(u) = j, f(v) = j \vdash \underline{\boldsymbol{f(u) = j \wedge f(v) = j}}}\;\wedge:\mathrm{r} \quad \boldsymbol{f(u) = f(v)} \vdash f(u) = f(v)}{f(u) = j, f(v) = j, \underline{\boldsymbol{((f(u) = j \wedge f(v) = j) \supset f(u) = f(v))}} \vdash f(u) = f(v)}\;\supset:\mathrm{l}}{f(u) = j, f(v) = j, \underline{\boldsymbol{(\forall y)((f(u) = j \wedge f(y) = j) \supset f(u) = f(y))}} \vdash f(u) = f(v)}\;\forall:\mathrm{l}}{f(u) = j, f(v) = j, \underline{\boldsymbol{(\forall x)(\forall y)((f(x) = j \wedge f(y) = j) \supset f(x) = f(y))}} \vdash f(u) = f(v)}\;\forall:\mathrm{l}}{f(u) = j, f(v) = j, \underline{(\forall i)(\forall x)(\forall y)((f(x) = i \wedge f(y) = i) \supset f(x) = f(y))} \vdash \boldsymbol{f(u) = f(v)}}\;\forall:\mathrm{l}$$