Cut-Elimination by Resolution*

M. Baaz† A. Leitsch‡

Cut-elimination is a key technique in proof theory. In his famous paper [2] Gentzen defined an (algorithmic) procedure to transform LK-proofs with cuts into cut-free LK-proofs. Cut-elimination is not only a tool to prove several important logical theorems (e.g. the interpolation theorem) but is also of relevance to computer science; indeed, after elimination of cuts, bounds and even programs can be extracted from the resulting proofs [5]. For these applications to computer science and mathematics *efficient* cut-elimination algorithms are crucial. Unfortunately, as Statman [4] ond Orevkov [3] independently showed, the complexity of cut-elimination is nonelementary; i.e. the length of cut-free proofs cannot be bounded by the length of the original proofs with cut via an elementary function. This clearly implies that *every* cut-elimination procedure is of nonelementary complexity. But even in presence of this horrible worst-case bound it pays out to focus on good algorithms which – at least for some types of problems and cuts – are fast and yield short cut-free proofs. In [1] we presented a cut-elimination method based on a projection technique, which (on a class of proofs called QMON) strongly outperforms Gentzen's method. Here we introduce a method based on *resolution*. Roughly it works in the following way: Consider the cut-application in the proof ω below (ω_1, ω_2 are cut-free)

$$\dfrac{\begin{array}{cc}\vdots\,(\omega_1) & \vdots\,(\omega_2)\\ \Gamma_1 \vdash \Delta_1, A & A, \Gamma_2 \vdash \Delta_2\end{array}}{\Gamma_1, \Gamma_2 \vdash \Delta_1, \Delta_2}$$

and assume that ω is an LK-proof with atomic initial sequents. Then, formally, the set of initial sequents is a set of clauses of the form $S : P(\bar{t}) \vdash P(\bar{t})$ (where P is a predicate symbol and $\bar{t}$ is a term tuple). Either the left or the right occurrence or none of them is a predecessor of the cut formula A (in ω_1 or in ω_2). Thus one of the form $\vdash$, $P(\bar{t}) \vdash$, $\vdash P(\bar{t})$ or $P(\bar{t}) \vdash P(\bar{t})$ characterizes the connection of the initial sequent with the (occurrence of the) cut in ω. By following the rule applications in ω we construct a set of clauses $\mathcal{C}$ out of the "characteristic" atomic ancestors s.t. $\mathcal{C}$ is unsatisfiable. In the next step we construct a resolution refutation γ of $\mathcal{C}$ which is transformed into a ground refutation γ'. We may consider γ' as an LK-derivation using contraction and atomic cut. By inserting into γ' appropriate subproofs of ω we obtain a proof of $\Gamma_1, \Gamma_2 \vdash \Delta_1, \Delta_2$ which is not cut-free, but having only atomic cuts. In the last step the atomic cuts are eliminated and a cut-free proof is obtained. We illustrate the procedure by a simple example (α, β are variables, a is a constant symbol):

*Supported by the Austrian Research Fund (FWF Projekt P11934-MAT)

†Institut für Algebra und diskrete Mathematik, E118/2, Technische Universität Wien

‡Institut für Computersprachen, E185/2, Technische Universität Wien

ψ_1 :

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{P(\alpha)^\star \vdash P(\alpha) \quad Q(\alpha) \vdash Q(\alpha)^\star}{P(\alpha)^\star, P(\alpha) \to Q(\alpha) \vdash Q(\alpha)^\star}\,{\to}{:}l}{P(\alpha) \to Q(\alpha) \vdash (P(\alpha) \to Q(\alpha))^\star}\,{\to}{:}r}{P(\alpha) \to Q(\alpha) \vdash (\exists y)(P(\alpha) \to Q(y))^\star}\,\exists{:}r}{(\forall x)(P(x) \to Q(x)) \vdash (\exists y)(P(\alpha) \to Q(y))^\star}\,\forall{:}l}{(\forall x)(P(x) \to Q(x)) \vdash (\forall x)(\exists y)(P(x) \to Q(y))^\star}\,\forall{:}r$$

ψ_2 :

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{P(a) \vdash P(a)^\star \quad Q(\beta)^\star \vdash Q(\beta)}{P(a), (P(a) \to Q(\beta))^\star \vdash Q(\beta)}\,{\to}{:}l}{(P(a) \to Q(\beta))^\star \vdash P(a) \to Q(\beta)}\,{\to}{:}r}{(P(a) \to Q(\beta))^\star \vdash (\exists y)(P(a) \to Q(y))}\,\exists{:}r}{(\exists y)(P(a) \to Q(y))^\star \vdash (\exists y)(P(a) \to Q(y))}\,\exists{:}l}{(\forall x)(\exists y)(P(x) \to Q(y))^\star \vdash (\exists y)(P(a) \to Q(y))}\,\forall{:}l$$

$$\dfrac{\begin{array}{cc}\vdots\ \psi_1 & \vdots\ \psi_2\end{array}}{S : (\forall x)(P(x) \to Q(x)) \vdash (\exists y)(P(a) \to Q(y))}\ \text{cut}$$

The cut formula and its ancestors are marked by $(\star)$. We start with the left subproof ψ_1; initially the clauses are $P(\alpha) \vdash$ and $\vdash Q(\alpha)$. The rule $\to{:}l$ merges the clauses into $C_1 : P(\alpha) \vdash Q(\alpha)$. To the sequence of unary rules following this application we always assign the clause C_1. In ψ_2 the rule $\to{:}l$ is applied to ancestors of the cut-formula (in contrast to ψ_1); here we construct the union of the clauses $\vdash P(a)$ and $Q(\beta) \vdash$. The unary rules down to the cut don't change the set of clauses $\{\vdash P(a), Q(\beta) \vdash\}$. Eventually the cut rule generates the union of both sets resulting in

$$\mathcal{C} : \{P(\alpha) \vdash Q(\alpha),\ \vdash P(a),\ Q(\beta) \vdash\}$$

From $\mathcal{C}$ we construct a resolution refutation γ and the ground refutation γ':

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

To $\vdash P(a)$ and $Q(a) \vdash$ we assign the LK-proofs ψ_{21} and ψ_{22} (extracted from ψ_2).

(ψ_{21})

$$\dfrac{\dfrac{\dfrac{P(a) \vdash P(a)^\star}{P(a) \vdash Q(\beta), P(a)^\star}\,W{:}r}{\vdash P(a) \to Q(\beta), P(a)^\star}\,{\to}{:}r}{\vdash (\exists y)(P(a) \to Q(y)), P(a)^\star}\,\exists{:}r$$

(ψ_{22})

$$\dfrac{\dfrac{\dfrac{Q(a)^\star \vdash Q(a)}{P(a), Q(a)^\star \vdash Q(a)}\,W{:}l}{Q(a)^\star \vdash P(a) \to Q(a)}\,{\to}{:}r}{Q(a)^\star \vdash (\exists y)(P(a) \to Q(y))}\,\exists{:}r$$

To $P(a) \vdash Q(a)$ we assign the proof ψ_{11} (extracted from ψ_1).

$$\dfrac{\dfrac{\dfrac{P(a)^\star \vdash P(a) \quad Q(a) \vdash Q(a)^\star}{P(a)^\star, P(a) \to Q(a) \vdash Q(a)^\star}\,{\to}{:}l}{P(a)^\star, (\forall x)(P(x) \to Q(x)) \vdash Q(a)^\star}\,\forall{:}l$$

By replacing the initial sequents in γ' by the proofs ψ_{21}, ψ_{11}, ψ_{22} we obtain the following LK-proofs with atomic cuts only:

$$\dfrac{\dfrac{\begin{array}{c}\vdots\ \psi_{21}\\ \vdash (\exists y)(P(a)\to Q(y)), P(a)\end{array} \quad \begin{array}{c}\vdots\ \psi_{11}\\ P(a), (\forall x)(P(x)\to Q(x)) \vdash Q(a)\end{array}}{(\forall x)(P(x)\to Q(x)) \vdash (\exists y)(P(a)\to Q(y)), Q(a)}\ \text{cut} \quad \begin{array}{c}\vdots\ \psi_{22}\\ Q(a) \vdash (\exists y)(P(a)\to Q(y))\end{array}}{(\forall x)(P(x)\to Q(x)) \vdash (\exists y)(P(a)\to Q(y))}\ \text{cut}$$

The last step, the elimination of the atomic cuts, is not very interesting and we skip it.

Once we have the ground projection γ' of the resolution refutation γ of $\mathcal{C}$ the construction of the LK-proof corresponding to γ' is relatively easy: For every clause $\bar{P} \vdash \bar{Q}$ in $\mathcal{C}$ we construct a proof of $\bar{P}, \Gamma_1 \vdash \Delta_1, \bar{Q}$ or $\bar{P}, \Gamma_2 \vdash \Delta_2, \bar{Q}$, respectively. For this purpose we simply suppress the inferences on successors of $\bar{P} \vdash \bar{Q}$ in ω. This is only possible if no eigenvariable conditions are violated; thus in order to guarantee the correctness of this transformation we have to skolemize the proof and reskolemize after cut-elimination. Skolemization does not increase the length of the proof and enables the clauses $\bar{P} \vdash \bar{Q}$ to be propagated downwards. The most subtle point is the construction of the corresponding set of clauses $\mathcal{C}$. In general the procedure works as follows:

1. Construct an initial set $\mathcal{C}_0$ by selecting the atoms in the initial sequents which are ancestors of the cut occurrences.

2. Assume that $S_1 : \Lambda_1, \Gamma_1 \vdash \Delta_1, \Pi_1$ and $S_2 : \Lambda_2, \Gamma_2 \vdash \Delta_2, \Pi_2$ are derived by the proofs ψ_1 and ψ_2 and that $\mathcal{C}_1$ and $\mathcal{C}_2$ are the sets of clauses corresponding to ψ_1 and ψ_2 (representing the ancestors of the subsequents $\Lambda_i \vdash \Pi_i$).

   (a) A binary rule applies to the parts $\Gamma_1 \vdash \Delta_1$ and $\Gamma_2 \vdash \Delta_2$ – and not to the (cut-relevant) parts $\Lambda_i \vdash \Pi_i$ – giving a sequent $S : \Lambda_1, \Lambda_2, \Gamma_1', \Gamma_2' \vdash \Delta_1', \Delta_2', \Pi_1, \Pi_2$. Then we define the set of clauses $\mathcal{C}$ corresponding to S by $\mathcal{C}_1 \otimes \mathcal{C}_2$, where

   $$\{\bar{P}_1 \vdash \bar{Q}_1, \ldots, \bar{P}_m \vdash \bar{Q}_m\} \otimes \{\bar{R}_1 \vdash \bar{T}_1, \ldots, \bar{R}_n \vdash \bar{T}_n\} = \\ = \{\bar{P}_i, \bar{R}_j \vdash \bar{Q}_i, \bar{T}_j | i \leq m, j \leq n\}.$$

   Before applying $\otimes$ the variables in $\mathcal{C}_1, \mathcal{C}_2$ must be renamed s.t. $\mathcal{C}_1$ and $\mathcal{C}_2$ don't share variables, which correspond to $\forall$:r, $\exists$:l introductions.

   (b) A binary rule applies to the ancestors of the cut, i.e. to $\Lambda_1 \vdash \Pi_1$ and $\Lambda_2 \vdash \Pi_2$ giving a sequent $S : \Lambda_1', \Lambda_2', \Gamma_1, \Gamma_2 \vdash \Delta_1, \Delta_2, \Pi_1', \Pi_2'$. The set of clauses $\mathcal{C}$ corresponding to S is defined as $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2$.

3. A unary rule is applied: The set of clauses is left unchanged.

If there are no binary rules, then the initial set of clauses remains unchanged. Therefore the crucial operations in the construction of the characteristic set of clauses are union and product. Of course, iteration of $\otimes$ and $\cup$ may result in a set of clauses of exponential size. However, by an appropriate transformation of Statman's sequence we can show that cut-elimination by resolution can give a nonelementary speed-up over Gentzen's method. The speed-up is obtained by the analysis of the internal structure of the cut formula by resolution. While Gentzen's method works from outside in (it always applies to the upmost

logical operation of the cut formula), the resolution method uses the whole *derivation* of the cut formula and thus also applies to the inner structure of the formula. In some sense, Gentzen's method is "context-free" as the reduction of rank and grade of cuts does only depend on the occurrence of the cut formula A in the proof and on its upmost operator, but is largely independent of the proof of A. In particular Gentzen's method cannot eliminate inner redundancies in the cut formula, but resolution can: Within the set of clauses $\mathcal{C}$ tautology deletion and subsumption can be applied in order to reduce the size of $\mathcal{C}$. By using one of the efficient resolution refinements, the real construction of the resolution proof and, eventually, of the cut-free proof becomes more realistic.

References

[1] M. Baaz and A. Leitsch. Fast cut-elimination by projection. In *Proc. of the CSL'96*, LNCS 1258, 18-33, Springer, 1997.

[2] Gerhard Gentzen. Untersuchungen über das logische Schließen I–II. *Math. Z.*, 39:176–210, 405–431, 1934.

[3] V.P. Orevkov. Lower bounds for increasing complexity of derivations after cut elimination. *Journal of Soviet Mathematics*, pages 2337–2350, 1982. Translation.

[4] R. Statman. Lower bounds on Herbrand's theorem. *Proc. Math. Soc.*, 75:104–107, 1979.

[5] A.S. Troelstra and H. Schwichtenberg. *Basic Proof Theory.* Cambridge University Press, 1996.