Hitotsubashi Journal of Arts and Sciences 32 (1991) 79-85. © The Hitotsubashi Academy

# ON THE LENGTH OF PROOFS AFTER ELIMINATING ATOMIC CUT INFERENCES

NORIKO HONDA

In this paper, we prove that Gentzen system without cut and Gentzen system with atomic (inessential) cut $p$-simulate each other when we assume that it is only polynomial speed-up from Gentzen system with atomic cut in the tree format to the one in the linear format.

## I. *Introduction*

It is well-known that cut-elimination for propositional calculus assures the existance of exponential functions. Naturally, there rises the following question: does cut-elimination procedure still insures the existence of exponential functions when the system is restricted to have cut formulas with their complexity less than $k$, where $k$ is an integer?

There is another interesting issue in the field of computational complexity. When we adopt a tree format such as Gentzen system to demonstrate a proof, we sometimes have to demonstrate the same subproofs over and over, which is time and space consuming. Now, it is rather common to assume that either the proof is written in a linear format or in an acyclic digraph, so that once an intermediate sequent in the proof has been derived, we do not need to derive it again even if it is used more than twice, although we do not yet know how much we can save time and space by this modification. (It is known that we can get an exponential speed-up in the case of Gentzen system without cut. (1))

In this paper, we relate above-mentioned two problems; when we assume that it is only polynomial speed-up from tree formats to linear formats, cut-elimination can be done in polynomial time in the case of Gentzen system with atomic cut.

## II. *Syntax and Rules of Propositional Calculus*

Languages:
1) Propositional variables; $p_1, p_2, p_3, \ldots$
2) Propositional connectives; $\wedge$, $\vee$, $\supset$ and $\neg$
3) Parenthesis
4) Sequent connective; $\rightarrow$
5) Comma

*Formulas* are defined as usual.

Propositional variables are also called *atomic formulas*. A series of formulas separated by comma is called a *cedent*. If $\Gamma$ and $\Delta$ are cedents, then $\Gamma \rightarrow \Delta$ is a *sequent*. $\Gamma$ and $\Delta$ are called *antecedent* and *succedent* of $\Gamma \rightarrow \Delta$ respectively.
An *inference* is the deduction of a sequent from a set of sequents. An inference is denoted pictorially by

$$\frac{B}{A} \quad \text{or} \quad \frac{B \quad C}{A}$$

The rules of propositional calculus are listed below. $\Gamma$, $\Pi$, $\Lambda$ and $\Delta$ are used to denote cedents, and $A$ and $B$ are arbitrary formulas.

1) (Weakening)

$$\frac{\Gamma \longrightarrow \Delta}{\Gamma, \Pi \longrightarrow \Lambda, \Delta}$$

2) (Contraction; Left)

$$\frac{A, A, \Gamma \longrightarrow \Delta}{A, \Gamma \longrightarrow \Delta}$$

3) (Contraction; Right)

$$\frac{\Gamma \longrightarrow \Delta, A, A}{\Gamma \longrightarrow \Delta, A}$$

4) (Exchange; Left)

$$\frac{\Gamma, A, B, \Delta \longrightarrow \Pi}{\Gamma, B, A, \Delta \longrightarrow \Pi}$$

5) (Exchange; Right)

$$\frac{\Gamma \longrightarrow \Delta, A, B, \Pi}{\Gamma \longrightarrow \Delta, B, A, \Pi}$$

6) ($\neg$; Left)

$$\frac{\Gamma \longrightarrow \Delta, A}{\neg A, \Gamma \longrightarrow \Delta}$$

7) ($\neg$; Right)

$$\frac{A, \Gamma \longrightarrow \Delta}{\Gamma \longrightarrow \Delta, \neg A}$$

8) ($\wedge$; Left)

$$\frac{A, \Gamma \longrightarrow \Delta}{A \wedge B, \Gamma \longrightarrow \Delta}$$

and

$$\frac{B, \Gamma \longrightarrow \Delta}{A \wedge B, \Gamma \longrightarrow \Delta}$$

9) ($\wedge$; Right)

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

10) ($\vee$ ; Left)

$$\frac{A, \Gamma \longrightarrow \Delta \qquad B, \Gamma \longrightarrow \Delta}{A \vee B, \Gamma \longrightarrow \Delta}$$

11) ($\vee$ ; Right)

$$\frac{\Gamma \longrightarrow \Delta, A}{\Gamma \longrightarrow \Delta, A \vee B}$$

and

$$\frac{\Gamma \longrightarrow \Delta, B}{\Gamma \longrightarrow \Delta, A \vee B}$$

12) ($\supset$; Left)

$$\frac{\Gamma \longrightarrow \Delta, A \qquad B, \Gamma \longrightarrow \Delta}{A \supset B, \Gamma \longrightarrow \Delta}$$

13) ($\supset$; Right)

$$\frac{A, \Gamma \longrightarrow \Delta, B}{\Gamma \longrightarrow \Delta, A \supset B}$$

14) (Cut)

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

where $A$ is an atomic formula and $A$ is not one of the formulas in $\Gamma$, $\Delta$, $\Pi$ or $\Lambda$. $A$ is called the *cut-formula* of this cut.

Note: The original formulation of Gentzen system is different from the above. It is easily proved that the system presented above and those appear in (2) and (3), restricted to have only atomic cut, p-simulate each other. In each inference, the formulas of the interest (denoted by $A$ and $B$ in the most cases) appearing in the upper sequents are called *auxiliary formulas*, and the one appearing in the lower sequent is called the *principal formula*. Only cut inferences do not have any principal formulas.

A *proof* is a rooted tree of sequents written so that the root of the tree is at the bottom. The leaves of the tree are called *initial sequents* which must be in the form $A \rightarrow A$, where $A$ is an arbitrary atomic formula. Every other sequent in the tree together with the sequents immediately above it must form a valid inference. The root of the tree is called the end sequent, which is what we prove by the proof.

The *length* of a proof is the number of the sequents different from each other in the proof, which is equal to the number of all the sequents appearing in the proof in a linear format. The *size* of a proof is the number of the symbols appearing in the sequents different from each other in the proof.

A proof is *atomic cut free* when no cut inference appears in the proof. A part of a proof which itself forms a proof is called a *subproof*, while the rest part of the proof is called a *stump*.

Ancestors and descendants are defined inductively:

*Definition:* Suppose $C$ is a formula which appears in a given sequent in a proof. The *successor* of $C$ is a formula in the sequent directly below the sequent $C$ appears in. The *successor* of $C$ is defined according to the following cases:

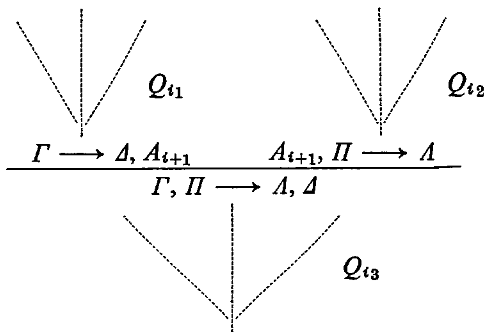
1) If $C$ in the end sequent of the proof or if $C$ is a cut formula of a cut inference, then $C$ has no successor.
2) If $C$ is the auxiliary formula of an inference, then the principal formula of the inference is the successor of $C$.
3) If $C$ is one of the formulas $A$ or $B$ in an exchange inference, the successor of $C$ is the formula denoted by the same letter in the lower sequent of the inference.
4) If $C$ is the $k$-th formula in a sub-cedent $\Gamma$, $\Delta$, $\Pi$ or $\Lambda$ of the upper sequent of an inference, then the successor of $C$ is the $k$-th formula in the corresponding sub-cedent of the lower sequent of the inference.

*Definition:* Let $C$ and $D$ be occurrences of formulas appearing in a proof. Then $C$ is an *ancestor* of $D$ if there are occurrences $C_1, C_2, \ldots, C_n$ of formulas in the proof such that $C_1$ is $C$, each $C_{i+1}$ is the successor of $C_i$ and $D$ is the successor of $C_n$.
If $C$ is an ancestor of $D$, then $D$ is a *descendent* of $C$.

## III. *Cut-Eliminating Algorithm*

Let $P_0$ be a proof in which $k$ cut inferences appear.

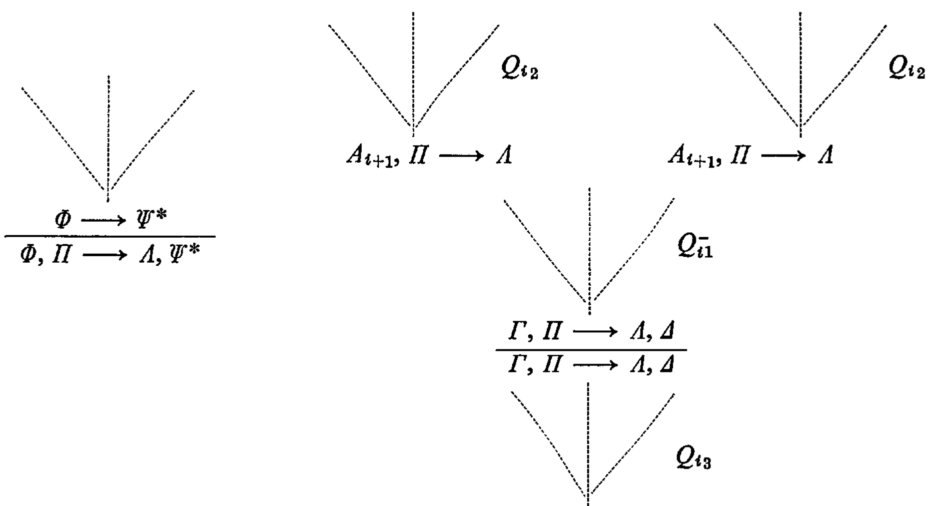
*Definition:* (Priority among cut inferences)
When a cut inference appears above another cut inference, the former one has *priority over* the latter. If neither one is above the other, then the one appearing left to the other has priority.
Name the cut inference of the $i$-th priority cut no. $i$ ($i=1,2,\ldots,k$)
Let the cut formula of cut no. $i$ be $A_i$. ($A_i$ might be the same formula with $A_j$ for some $j \neq i$)
Let $P_i$ denote the proof obtained from $P_0$ by eliminating cut no. 1 to cut no. $i$ ($i=1,2,\ldots,k$). In the course of eliminating cut no. $(i+1)$, every occurrence of sequent in $P_i$ generates some occurrences of sequents in $P_{i+1}$.
$P_i$ is in the following figure;



$Q_{i1}$ and $Q_{i2}$ are subproofs up to the left upper sequent and the right upper sequent of cut no. $(i+1)$, respectively. $Q_{i3}$ is the stump obtained by deleting the subproof up to the lower sequent of cut no. $(i+1)$ from $P_i$.

Construct $P_{i+1}$ from $P_i$ as follows. (For the details, it must be carried out by the induction on the number of the sequents appearing.) The part of $Q_{i3}$ is conserved as it is. The each sequent in old $Q_{i3}$ generates the same sequent in the same place in new $Q_{i3}$. Above the sequent $\Gamma, \Pi \rightarrow \Lambda, \Delta$, write the same sequent, $\Gamma, \Pi \rightarrow \Lambda, \Delta$ and above it, place the stump $Q^{-}{}_{i1}$ obtained from $Q_{i1}$ as follows: If a sequent $\Phi \rightarrow \Psi$ contains a formula $A_{i+1}$ which is an ancestor of the $A_{i+1}$ in the left upper sequent of cut no. $(i+1)$, then rewrite it by $\Phi, \Pi \rightarrow \Lambda, \Psi^*$, where $\Psi^*$ is the cedent obtained by deleting all the ancestors of the $A_{i+1}$ in the left upper sequent of cut no $(i+1)$. The original sequent, $\Phi \rightarrow \Psi$ generates the new sequent $\Phi, \Pi \rightarrow \Lambda, \Psi^*$. In particular, if the original sequent is an initial sequent, $A_{i+1} \rightarrow A_{i+1}$, then note that it is rewritten by $A_{i+1}, \Pi \rightarrow \Lambda$. Above each sequent of this kind, place the subproof $Q_{i2}$. Each sequent in the old $Q_{i2}$ generates the same sequents in the same places in each new $Q_{i2}$. If there is no initial sequent, $A_{i+1} \rightarrow A_{i+1}$, which contains an ancestor of $A_{i+1}$ in the left upper sequent of cut no. $(i+1)$ in $Q_{i1}$, then any sequent in $Q_{i2}$ does not generate anything. Other sequents in $Q_{i1}$ are conserved as it is in the same place. Each of them generate the same sequent in the same place.
$P_{i+1}$, constructed as above is again a proof (proof left to the reader).

Figure of $P_{i+1}$




*Theorem:* If the number of all the sequents appearing in $P_0$ is $n$, then the length of $P_k$ is less than or equal to $n$.
(proof) Let $s_0$ be an arbitrary (occurrence of a) sequent in $P_0$.
$S_1 = \{s \mid s$ is a sequent generated from $s_0.\}$
$S_{i+1} = \{s \mid s$ is a sequent generated from some $s' \in S_i\}$
Then, for any $1 \leqq i \leqq k$ and any sequents $s$ and $s^*$ in $S_i$, the following three conditions hold;

1) $s$ and $s^*$ are the same sequents (of different occurrences).
2) if $s$ appears in $Q_{i1}$, then $s^*$ appears in $Q_{i1}$,
if $s$ appears in $Q_{i2}$, then $s^*$ appears in $Q_{i2}$, and
if $s$ appears in $Q_{i3}$, then $s^*$ appears in $Q_{i3}$.
3) if the $j$-th formula in the succedent of $s$ is an ancestor of cut formula in the left upper sequent of cut no. $m$ $(i+1 \leqq m \leqq k)$, then so is the $j$-th formula in the succedent of $s^*$.

We prove it by induction on $i$.

Suppose $s,s^* \in S_{i+1}$. Then, there exist $t,t^* \in S_i$ such that $s$ and $s^*$ are generated from $t$ and $t^*$, respectively. From the induction hypothesis, $t$ and $t^*$ are the same sequent of the different occurrences.

Case 1) Suppose that $t$ appears in $Q_{i1}$ and it contains an ancestor of $A_{i+1}$ in the left upper sequent of cut no. $(i+1)$. Then by the induction hypothesis, so does $t^*$. Suppose $t$ and $t^*$ are in the form, $\Phi \rightarrow \Psi$. Then both of them are rewritten by $\Phi$, $\Pi \rightarrow \Lambda$, $\Psi^*$, by the induction hypothesis. Thus, $s$ and $s^*$ are the same sequents. If the $j$-th formula of the succedent of $s$ is an ancestor of the left cut formula of cut no. $m$ for some $m$, and if it is in $\Lambda$, then from the definition of ancestors, obviously so is the $j$-th fornula of $s^*$. If it is in $\Psi^*$, then from the induction hypothesis, the $j$-th formula of the succedent of $s^*$ is also an ancestor of the left cut formula of cut no. $m$ and in $\Psi^*$. Condition 2) obviously holds.

Case 2) Suppose that $t$ appears in $Q_{i1}$ and that it contains no ancestor of the left cut formula of cut no. $(i+1)$. Then, by the induction hypothesis, so does $t^*$. Then both of them remain the same in $P_{i+1}$. Thus, $s$ and $s^*$ are the same. The rest of the proof is obvious.

Case 3) Suppose that $t$ appears in $Q_{i2}$. Then, so does $t^*$. Clearly from the procedure of the algorithm and the induction hypothesis, $s$ and $s^*$ are the same sequents. Since none of the formulas among $\Pi$ and $\Lambda$ plays the role of auxiliary formulas in the stump $Q^-_{i1}$, the condition, the $j$-th formula of $s$ is an ancestor of the left cut formula of cut no. $m$ is the equivalent with the one claiming that the $j$-th formula of $t$ is an ancestor of the left cut formula of cut no. $m$. (For the detail, we must prove it by the induction on the length of the subproof, $Q_{i2}$.) Condition 2) clearly holds.

Case 4) Suppose that $t$ appears in $Q_{i3}$, the proof is obvious.

Thus, any $s,s^* \in S_k$ are the same sequents. The theorem is now proved.

*Corollary:* Suppose that for any proof $P$, there exist polynomial functions $p,q$ and a proof $P^-$ such that the end sequents of $P$ and $P^-$ are the same sequents and the length of $P^- = p$ (the length of $P$) and the number of the sequents in $P^- = q$ (the length of $P^-$). Then, for any proof $P$, there exists a cut free proof $P^*$ such that the end sequents of $P$ and $P^*$ are the same sequents and the number of the length of $P^* = q(p$ (the length of $P$)).

(proof) Obvious from Theorem proved above.

NAGOYA UNIVERSITY

## *REFERENCES*

(1) R. Statman, Bounds for proof-search and speed-up in the predicate calculus, Ann. Math. Logic 15 (1978) 225–287.
(2) S. Buss, Bounded Arithmetic, Bibliopolis, 1986.
(3) G. Takeuti, Proof Theory, North-Holland, 1987.
(4) S. Cook, Feasibly constructive proofs and the propositional calculus, Proc. Seventh ACM Symp. on Theory of Computing (1975) 83–97.
(5) S. Cook and R. Reckhow, The relative efficiency of propositional proof system, J. Symbolic Logic 44 (1979) 36–50.