Note on Cut-Elimination

Chung Tong Lee

September 3, 2007

Abstract

Cut-elimination for first-order logic is shown by proof-theoretic argument in Tait-style system (copied from [2] with minor notation differences) and by model-theoretic argument for completeness in Gentzen's sequent-based system (adapted from [1]).

1 Introduction

Hilbert-style proof system for first order logic is probably the most concise one. However, important information is lost during the Modus Ponens inference. Other equivalent proof systems which preserve information in a better way are more suitable for computational complexity application.

2 Tait-style System

In the Tait-style calculus, the form of formulas is different from that of classical logic:

- Connectives for formulas are $\wedge$, $\vee$ and $\neg$. $\wedge$ and $\vee$ take their usual meanings;
- Negations ($\neg$) are allowed only in front of *atomic* formulas. (Thus, $\neg\varphi$, where φ is an arbitrary formula, stands for the equivalent formula with negations pushed in front of the atomic subformulas.)
- $\rightarrow$ and $\leftrightarrow$ are just abbreviations for the corresponding equivalent formulas built up with just $\wedge$ and $\vee$ from (negated) atomic formulas.

One can show, by induction on formula complexity, that there is a Tait-style equivalence for every well-formed formula in classical first order logic, and vice versa. $\wedge$, $\vee$ and the restricted $\neg$ form a functionally complete set. Restrictions on the use of $\neg$ does not reduce the expressive power. On the other hand, the complexity of system is reduced.

Definition 1. *The length $|\varphi|$ of a formula φ is defined inductively as:*

- $|\varphi| = |\neg\varphi| = 0$ *for atomic φ;*

- $|\varphi \wedge \psi| = |\varphi \vee \psi| = sup(|\varphi|, |\psi|) + 1$;
- $|\forall x\varphi(x)| = |\exists x\varphi(x)| = |\varphi(x)| + 1$.

Finite sets of formulas are derived in this calculus. Such sets are denoted by capital Greek letters, .e.g. Γ. The intended meaning of Γ is the disjunction of all formulas in Γ. The notation Γ, Δ stands for $\Gamma \cup \Delta$ and Γ, φ stands for $\Gamma \cup \{\varphi\}$. The inference rules of Tait-style system are as follows:

(A) $\Gamma, \varphi, \neg\varphi$ (φ is atomic)

$(\vee)$ $\dfrac{\Gamma, \varphi_0}{\Gamma, \varphi \vee \varphi_1}$, $\dfrac{\Gamma, \varphi_1}{\Gamma, \varphi_0 \vee \varphi_1}$

$(\forall)$ $\dfrac{\Gamma, \varphi(a)}{\Gamma, \forall x\varphi(x)}$ (a is not free in any formula in Γ)

$(\exists)$ $\dfrac{\Gamma, \varphi(t)}{\Gamma, \exists x\varphi(x)}$ (t is a term in the underlying language)

(C) $\dfrac{\Gamma, \varphi \qquad \Gamma, \neg\varphi}{\Gamma}$ (φ is an arbitrary formula)

In general, inference rules are of the form

$(*)$ $\dfrac{\Gamma, \varphi_i \quad \text{for all } i < k}{\Gamma, \Phi}$

where $(0 \leq k \leq 2)$, Φ consists of the principal formulas (p.f.) of $(*)$; φ_i is the minor formula (m.f.) in the i^{th} premiss of $(*)$; Γ are the side formulas (s.f.).

	Principal Formula (p.f.)	Minor Formula (m.f.)
(A)	$\varphi, \neg\varphi$	no m.f
$(\wedge)$	$\varphi_0 \wedge \varphi_1$	φ_0, φ_1
$(\vee)$	$\varphi_0 \vee \varphi_1$	φ_0, φ_1
$(\forall)$	$\forall x\varphi(x)$	$\varphi(x)$
$(\exists)$	$\exists x\varphi(x)$	$\varphi(t)$
(C)	no p.f.	$\varphi, \neg\varphi$

We write $\vdash_d \Gamma$ if d is a derivation of Γ. If x is one of the free variables in Γ, we sometimes use the notation $d(x)$ for the derivation and $\Gamma(x)$ for the conclusion where it is understood that there may be some other free variables. Derivations are built up in tree form. Given $\vdash_{d_i} \Gamma, \varphi_i$, d is a derivation of Γ, Δ built from d_i using the inference$(*)$ as above. Then:

Definition 2.

1. $(*)$ *is called the* last inference *of* d*;*
2. *The* d_i *are called* direct subderivations *of* d*;*
3. *The length* $|d|$ *of* d *is given by* $|d| = sup(|d_i| + 1)$*, in particular,* $|d| = 0$ *when* d *does not have any subderivation.;*

4. The cut-rank $\rho(d)$ of d is given by

$$\rho(d) = \begin{cases} sup\,(|\varphi_0| + 1, sup(\rho(d_i))) & \text{if } (*) \text{ is } (C) \\ sup(\rho(d_i)) & \text{otherwise} \end{cases}.$$

and $\rho(d) = 0$ when d is cut-free.

3 Cut-elimination in Tait-style System

Let d, Γ be obtained from a derivation d by adding Γ to the side formulas of all inference rules in d with appropriate changes of bound variables, we have:

Lemma 1. Weakening Lemma
If $\vdash_d \Delta$, then $\vdash_{d,\Gamma} \Gamma, \Delta$ with $|d, \Gamma| = |d|$ and $\rho(d, \Gamma) = \rho(d)$

Let $d(t)$ denote the result of substituting a term t for all free occurrences of x in $d(x)$, (with some changes of bound variables if necessary). Then we obviously have:

Lemma 2. Substitution Lemma
If $\vdash_{d(x)} \Gamma(x)$, then $\vdash_{d(t)} \Gamma(t)$ with $|d(t)| = |d|$ and $\rho(d) = \rho(d(t))$.

Lemma 3. Inversion Lemmas

$\vee$-inversion *If $\vdash_d \Gamma, \varphi_0 \vee \varphi_1$, then there is a derivation d', s.t. $\vdash_{d'} \Gamma, \varphi_0, \varphi_1$ with $|d'| \leq |d|$ and $\rho(d') \leq \rho(d)$.*

$\wedge$-inversion *If $\vdash_d \Gamma, \varphi_0 \wedge \varphi_1$, then there are derivations, denoted by d_i, s.t. $\vdash_{d_i} \Gamma, \varphi_i$ with $|d_i| \leq |d|$ and $\rho(d_i) \leq \rho(d)$ for $i = 0, 1$;*

$\forall$-inversion *If $\vdash_d \Gamma, \forall x \varphi(x)$, then there is a derivation d' s.t. $\vdash_{d_0} \Gamma, \varphi(x)$ with $|d'| \leq |d|$ and $\rho(d') \leq \rho(d)$.*

Proof. The proofs of these lemmas are almost identical. All can be shown by induction on $|d|$ and we will demonstrate the last case.

Base Case - $|d| = 0$: The only inference in d is (A) and Γ must contain some atomic formula ψ and its negation $\neg\psi$. Hence $\Gamma, \varphi(x)$ is also an instance of (A).

Induction Step :

Case 1 - $\forall x \varphi(x)$ is not the p.f. of the last inference of d:
Then last inference is of the form

$$\frac{\Lambda, \forall x \varphi(x), \psi_i \quad \text{for all } i < k}{\Lambda, \forall x \varphi(x), \Psi}$$

with m.f ψ_i, p.f. Ψ and s.f. Λ, and $\Gamma = \Lambda, \Psi$. Let us denote d_i the direct subderivations of d. By hypothesis, there are derivations d'_i s.t. $\vdash_{d'_i} \Lambda, \varphi(x), \psi_i$ with $|d'_i| \leq |d_i| < |d|$ and $\rho(d_i) \leq \rho(d_i) \leq \rho(d)$. The required derivation d' can be constructed using d'_i as direct subderivations.

Case 2 - $\forall x \varphi(x)$ is the p.f. of the last inference of d:
With weakening, we can assume the last inference is of the form

$$\frac{\Gamma, \forall x \varphi(x), \varphi(x)}{\Gamma, \forall x \varphi(x)}$$

The length of the subderivation of $\Gamma, \forall x \varphi(x), \varphi(x)$ is strictly smaller $|d|$ and by hypothesis, we have the required d'.

Note: Application of inference is valid regardless if the p.f. is already one of the s.f (by weakening). The conclusions are the same because they are finite sets. □

Lemma 4. Reduction Lemma
Let $\vdash_{d_0} \Gamma, \varphi$ *and* $\vdash_{d_1} \Delta, \neg\varphi$ *and* $\rho(d_0), \rho(d_1) \leq |\varphi|$. *Then there is a derivation* d *s.t.* $\vdash_d \Gamma, \Delta$ *with* $|d| \leq |d_0| + |d_1|$ *and* $\rho(d) \leq |\varphi|$.

Of course we could derive Γ, Δ by an application of the cut-rule, but the resulting derivation would then have cut-rank $|\varphi| + 1$.

Proof. We proceed by induction on $|d_0| + |d_1|$:

Case 1 Either φ is not a p.f. in the last inference of d_0 or else $\neg\varphi$ is not a p.f. in the last inference of d_1. By symmetry, we can assume the former. Then the last inference of d_0 is of the form

$$\frac{\Lambda, \varphi, \psi_i \quad \text{for all } i < k}{\Lambda, \varphi, \Psi}$$

with $\Gamma = \Lambda, \Psi$. By hypothesis, there are d'_i s.t. $\vdash_{d'_i} \Lambda, \Delta, \psi_i$ with $|d'_i| < |d_0| + |d_1|$ and $\rho(d'_i) \leq |\varphi|$. Using d'_i as direct subderivations, we can construct the required derivation d.

Case 2 Both φ and $\neg\varphi$ are the p.f. of the last inference of respective derivations.

Case 2.1 φ is atomic.
Then the last inference is (A) and hence Γ, Δ is also an instance of (A).

Case 2.2 φ is $\varphi_0 \vee \varphi_1$
By weakening with φ, the last inference of d_0, φ is of the form

$$\frac{\Gamma, \varphi, \varphi_i}{\Gamma, \varphi}$$

and by hypothesis, there is a d'_0 s.t. $\vdash_{d'_0} \Gamma, \Delta, \varphi_i$ with $|d'_0| < |d_0| + |d_1|$ and $\rho(d'_0) \leq |\varphi|$. Now consider $\neg\varphi$ is $\neg\varphi_0 \wedge \neg\varphi_1$. By inversion, there are two derivations, denoted by d'_{1_i} for $i = 0, 1$, s.t. $\vdash_{d'_{1_i}} \Delta, \neg\varphi_i$. Applying (C) with d'_0 on d'_{1_i}, Γ yields the required derivation d with $\rho(d) = sup(|\varphi_i| + 1, sup(\rho(d_0), \rho(d_1)))$. As $|\varphi_i| < |\varphi|$ and $\rho(d_0), \rho(d_1) \leq |\varphi|$, $\rho(d) \leq |\varphi|$.

Case 2.3 φ is $\exists x \psi(x)$

Again, there is derivation d_0, φ s.t. its last inference is of the form

$$\frac{\Gamma, \varphi, \psi(t)}{\Gamma, \varphi}$$

and by hypothesis, there is d_0' s.t. $\vdash_{d_0'} \Gamma, \Delta, \psi(t)$. Now consider $\neg\varphi$ is $\forall x \psi(x)$ and by inversion and substitution where is a $d_1'(t)$ s.t. $\vdash \Delta, \neg\psi(t)$. Again by weakening and cut using $\psi(t)$ as m.f. we get the required derivation d.

□

Theorem 5. Cut-Elimination Theorem
If $\vdash_d \Gamma$ and $\rho(d) > 0$, then there is a derivation d' s.t. $\rho(d') < \rho(d)$ and $|d'| \leq 2^{|d|}$.

Proof. The proof is by induction on $|d|$. We may assume that the last inference of d is a cut

$$\frac{\Gamma, \varphi \qquad \Gamma, \neg\varphi}{\Gamma}$$

with $|\varphi| + 1 = \rho(d)$, for otherwise the result follows by the induction hypothesis (making use of the fact that our rules all have finitely many premisses). So assume this. Let $\vdash_{d_0} \Gamma, \varphi$ and $\vdash_{d_1} \Gamma, \neg\varphi$. By hypothesis, there are derivations d_0' and d_1' s.t. $\vdash_{d_0'} \Gamma, \varphi$ and $\vdash_{d_1'} \Gamma, \neg\varphi$ and $\rho(d_i') \leq |\varphi|, |d_i'| \leq |d_i|$ for $i = 0, 1$. The result then follows, by the Reduction Lemma, since $|d_0'| + |d_1'| \leq 2^{sup(|d_0|,|d_1|)+1} = 2^{|d|}$. □

Let $2_0^x = x, 2_{y+1}^x = 2^{2_y^x}$

Corollary 6. *If $\vdash_d \Gamma$, then there is a cut-free derivation d^* s.t. $\vdash_{d^*} \Gamma$ and $|d^*| \leq 2_{\rho(d)}^{|d|}$.*

4 Sequent Calculus

Gentzen's sequent style calculus is another commonly used system for first-order logic. The basic unit is a sequent of the form $\Gamma \longrightarrow \Delta$ where Γ (the antecedent) and Δ (the succedent) are finite sets (possibly empty) of formulas.The intended meaning of a sequent $\Gamma \longrightarrow \Delta$ is

$$\bigwedge_{\gamma \in \Gamma} \gamma \rightarrow \bigvee_{\delta \in \Delta} \delta$$

where $\rightarrow$ is the usual symbol for logical implication. Thus $\longrightarrow \varphi$ means φ and $\varphi \longrightarrow$ means $\neg\varphi$. The empty sequent means contradiction.

There are *left*- and *right*-introduction rules corresponds to each symbol for connectives and quantifiers. Any functionally complete set of connectives can be used and we choose $\{\neg, \vee, \wedge\}$ (similar to Tait-style system). The quantifiers, of course, are the usual ones. The inference rules are as follows:

Rules	Left	Right
¬-introduction	$\dfrac{\Gamma \longrightarrow \Delta, \varphi}{\neg\varphi, \Gamma \longrightarrow \Delta}$	$\dfrac{\varphi, \Gamma \longrightarrow \Delta}{\Gamma \longrightarrow \Delta, \neg\varphi}$
∧-introduction	$\dfrac{\varphi, \psi, \Gamma \longrightarrow \Delta}{\varphi \wedge \psi, \Gamma \longrightarrow \Delta}$	$\dfrac{\Gamma \longrightarrow \Delta, \varphi \qquad \Gamma \longrightarrow \Delta, \psi}{\Gamma \longrightarrow \Delta, \varphi \wedge \psi}$
∨-introduction	$\dfrac{\varphi, \Gamma \longrightarrow \Delta \qquad \psi, \Gamma \longrightarrow \Delta}{\varphi \vee \psi, \Gamma \longrightarrow \Delta}$	$\dfrac{\Gamma \longrightarrow \Delta, \varphi, \psi}{\Gamma \longrightarrow \Delta, \varphi \vee \psi}$
∀-introduction	$\dfrac{\varphi(t), \Gamma \longrightarrow \Delta}{\forall x\varphi(x), \Gamma \longrightarrow \Delta}$	$\dfrac{\Gamma \longrightarrow \Delta, \varphi(b)}{\Gamma \longrightarrow \Delta, \forall x\varphi(x)}$
∃-introduction	$\dfrac{\varphi(b), \Gamma \longrightarrow \Delta}{\exists x\varphi(x), \Gamma \longrightarrow \Delta}$	$\dfrac{\Gamma \longrightarrow \Delta, \varphi(t)}{\Gamma \longrightarrow \Delta, \exists x\varphi(x)}$

where t is any term not involving any bound variables and b is the eigenvariable which must not occur in the conclusion.

Other inference rules correspond to

- the logical axiom $\varphi \vee \neg\varphi$,
- properties of finite sets and,
- weakening lemma and (C) in Tait-style calculus.

Rules	Left	Right
Logical	$\varphi \longrightarrow \varphi$	
Exchange	$\dfrac{\Gamma_0, \varphi, \psi, \Gamma_1 \longrightarrow \Delta}{\Gamma_0, \psi, \varphi, \Gamma_1 \longrightarrow \Delta}$	$\dfrac{\Gamma \longrightarrow \Delta_0, \varphi, \psi, \Delta_1}{\Gamma \longrightarrow \Delta_0, \psi, \varphi, \Delta_1}$
Contraction	$\dfrac{\Gamma, \varphi, \varphi \longrightarrow \Delta}{\Gamma, \varphi \longrightarrow \Delta}$	$\dfrac{\Gamma \longrightarrow \Delta, \varphi, \varphi}{\Gamma \longrightarrow \Delta, \varphi}$
Weakening	$\dfrac{\Gamma, \longrightarrow \Delta}{\Gamma, \varphi \longrightarrow \Delta}$	$\dfrac{\Gamma \longrightarrow \Delta}{\Gamma \longrightarrow \Delta, \varphi}$
Cut	$\dfrac{\Gamma, \longrightarrow \Delta, \varphi \qquad \varphi, \Gamma \longrightarrow \Delta}{\Gamma \longrightarrow \Delta}$	

The system described above is called **LK**.

Note: The conclusions of an inference rule are logic consequences of their respective premisses **except** in the case of ∀-*right* and ∃-*left* (similar in (∀) of Tait-style system). Nonetheless, all rules are validity preserving.

5 Completeness Lemma

Lemma 7. *If* $\Gamma \longrightarrow \Delta$ *is logically valid, then there is a derivation in* ***LK*** *which does not use the cut rule*

Proof. We present an algorithm to construct a cut-free derivation d for any sequent $\Gamma \longrightarrow \Delta$. A sequent in our construction is *active* when it is a leaf and there is no common formula in its antecedent and succedent. We assume that the language is countable so we can enumerate all the terms in a list T. The uncountable case holds but requires different argument. We will restrict our discussion in the countable case.

We denote by d_0 the partial derivation with the target sequent as the only premiss and by T_0 the set of all terms without bound variable in d_0. Further, we use Λ and Ψ to keep track of universal formulas in antecedent and existential formulas in succedent for all premisses in this and all subsequent partial derivations.

At the beginning of stage $i+1$, we set T_{i+1} to be empty. We modify d_i by replacing any active sequent with a derivation of appropriate introduction rule in reverse. Reverse applications of quantifier-related rules need some attention. We demonstrate the $\exists$ cases and the $\forall$ cases are similar:

If the active sequent is of the form $\Gamma', \exists x\varphi(x), \Gamma'' \longrightarrow \Delta'$, we replace it with the derivation

$$\frac{\Gamma', \varphi(b), \Gamma'' \longrightarrow \Delta'}{\Gamma', \exists x\varphi(x), \Gamma'' \longrightarrow \Delta'}$$

where b is a new variable. We add all the terms which involve b and contain no bound variable to T_{i+1}, including b itself.

If the active sequent is of the form $\Gamma' \longrightarrow \Delta', \exists x\varphi(x), \Delta''$, we replace it with the derivation

$$\frac{\Gamma' \longrightarrow \Delta', \varphi(t_0), \ldots, \varphi(t_k), \Delta''}{\Gamma' \longrightarrow \Delta', \exists x\varphi(x), \Delta''}$$

where $t_j \in T_i$ for $0 \leq j \leq k$. The double lines indicate a successive applications of $\exists$-*right* and *contraction* rules for $\exists x\varphi(x)$. We add $\exists x\varphi(x)$ to Ψ. We also add all the terms in $\varphi(t_i)$ to T_{i+1}, provided that they are not in T_j for all $j < i+1$ and they do not involve any bound variable.

The newly added leaves may be active and we repeat the process until none of them are active or until every formula in active sequent is atomic. We can finish every stage after a finite number of iterations as every formulas is of finite length and T_i is finite. Then we add Λ to every antecedent and Ψ to every succedent in the modified d_i to get d_{i+1}.

Finally, we check if we should continue next stage.

Case 1 — None of the leaves are active.

We need not continue to the next stage. Instead, we modify all the leaves as followings and get d. The algorithm halts with a success.

An leaf which is not active must be of the form $\Gamma', \varphi, \Gamma'' \longrightarrow \Delta', \varphi, \Delta''$. We replace these leaves with

$$\frac{\varphi \longrightarrow \varphi}{\overline{\Gamma', \varphi, \Gamma'' \longrightarrow \Delta', \varphi, \Delta''}}$$

where the double line means successive applications of weakening and exchange

Case 2 — at least one leaf is active and both Λ and Ψ are empty.

The algorithm halts with a failure.

Case 3 — at least one leaf is active and Λ or Ψ (or both) is not empty.

Then we pick the first $t \in T$ s.t. $t \notin T_j$ for $j \leq i+1$ and add it to T_{i+1} and proceed to next stage. This is the only part that may allow the algorithm to run without a halt. Nonetheless, it allows us to "exhaust" the countable list T.

If the algorithm halts with a failure or never halts, we can construct a model $\mathcal{M}$ that falsifies $\Gamma \longrightarrow \Delta$. The universe M of the model $\mathcal{M}$ contains all free variables in the above derivation d and all constants of the language. Further, M is closed under all functions $f^{\mathcal{M}}$ which is the interpretation of function symbol f. The object assignment maps every variable to itself and $f^{\mathcal{M}}(\vec{t})$ to the term $f(\vec{t})$.

Let β be a branch with an active sequent. The interpretation $\varphi^{\mathcal{M}}$ of a predicate symbol φ is defined by letting $\varphi^{\mathcal{M}}(\vec{t})$ holds iff $\varphi(\vec{t})$ occurs in the antecedent of some sequent in the branch β. By induction on formula complexity, one can show $\mathcal{M} \models \bigwedge_{\gamma \in \Gamma'} \gamma$ and $\mathcal{M} \not\models \bigvee_{\delta \in \Delta'} \delta$ for every sequent $\Gamma' \longrightarrow \Delta'$ in branch β. Hence $\mathcal{M} \not\models \Gamma \longrightarrow \Delta$.

Note that for formulas $\forall x \varphi(x)$ in antecedents and $\exists x \varphi(x)$ in succedents, the argument is based on the assumption that T is enumerable. □

6 Derivations with Equality Axioms and Non-logical Axioms

We need more inference rules for both Tait-style and Gentzen's sequent-style systems in order to handle derivation with equality and non-logical axioms. The inferences of this axioms are similar to the (A) in the former system and *Logical* in the latter. Details can be found in both [1] and [2]. With axioms other than the logical ones, we need the cut rule the get the conclusion. For example, $\varphi \wedge \psi \vdash \varphi$. The conclusion is "shorter" than the premiss, but every rule in Tait-style system increases the length of formulas except (C). Without cut, we cannot achieve derivational completeness [1].

Often, we impose additional restrictions on how the theory and the axioms are specified in order to restrict the length (complexity) of a cut formula. If a theory $\mathbf{T}$ proves a set Γ, by compactness, there is a finite subset $\mathbf{T_0}$ of $\mathbf{T}$ s.t. $\vdash \Phi_0, \Gamma$ where $\Phi_0 = \{\neg\varphi \mid \varphi \in \mathbf{T_0}\}$. After getting the cut-free derivation of Φ_0, Γ, we can apply cut rules successively with cut formula in $\mathbf{T_0}$. By this, complexity of cut formulas are restricted – they must be an instance of non-logical/equality axioms.

References

[1] Stephen Cook and Phuong Nguyen. Foundations of proof complexity: Bounded arithmetic and propositional translations. manuscript in preparation, http://www.cs.toronto.edu/ sacook/csc2429h/book/.

[2] Helmut Schwichtengerg. Proof theory: Some applications of cut-elimination. In J. Barwise, editor, *Handbook of Mathematical Logic*, chapter D.2, pages 867–895. North-Holland, 1977.