Proof Complexity of Intuitionistic Propositional Logic

Alexander Hertel & Alasdair Urquhart *

November 29, 2006

Abstract

We explore the proof complexity of intuitionistic propositional logic (IPL). The problem of determining whether or not an intuitionistic formula is valid is $\mathcal{PSPACE}$-Complete via a reduction from QBF. In view of this reduction (due to Statman), it is natural to compare the proof-theoretic strength of a standard axiomatic system for IPL with a similar proof system for classical quantified Boolean logic. In fact, the intuitionistic system seems to be weak in comparison with the latter, in the following sense – unless a variant of Gentzen's proof system LK is super (and therefore $\mathcal{NP} = co\mathcal{NP}$), Statman's reduction from QBF to IPL cannot even translate trivial classical instances of the law of excluded middle into intuitionistic formulas having proofs with polynomial size upper bounds. An immediate implication of this result is that unless the variant of LK for classical logic is a super proof system, there is a superpolynomial separation between it and a similar variant of Gentzen's system LJ for Intuitionistic Logic.

1 Introduction

Intuitionistic propositional logic (IPL) is perhaps the best-studied non-classical logic. The validity problem for intuitionistic logic appears to be intrinsically more complex than the corresponding problem for classical logic. A well-known paper by Statman [Sta79] shows via a natural reduction from QBF that the problem of determining whether a formula is intuitionistically valid is $\mathcal{PSPACE}$-Complete.

The purpose of this paper is to explore the proof complexity of IPL, considered as a proof system for a $\mathcal{PSPACE}$-Complete set. Intuitionistic logic is weaker than classical logic, since it disallows proof by contradiction; hence it is reasonable to conjecture that there are classical tautologies that are also provable intuitionistically, but whose shortest constructive proofs are super-polynomially longer than their minimal proofs in classical logic. From a complexity-theoretic perspective, IPL is in some regards more tractable than classical logic. For example, it has feasible interpolation [Pud99, BP01]. Hence, it is reasonable to conjecture that IPL lower bounds might be easier to establish than for classical logic. The main theorem of this paper provides some evidence for this view. Although we do not show IPL to be weak in an absolute sense, we show it to be weak relative to Statman's translation. More specifically, we show that unless the system $\mathsf{LK}[\vec{E_S}]$ is super (and therefore $\mathcal{NP} = co\mathcal{NP}$), Statman's reduction from QBF to IPL cannot translate trivial classical instances of the law of excluded middle into intuitionistic formulas with polynomially-bounded proofs. Since Statman's translation is the obvious and natural reduction to use, this result shows that if a more feasible reduction exists, then it must be complicated. An immediate implication of this result is that unless the variant of Gentzen's system LK for classical logic is a super proof system, there is a superpolynomial separation between it and an extended form of his system LJ for Intuitionistic Logic.

*This research supported by NSERC and the University of Toronto Department of Computer Science

The overview of this paper is fairly straightforward. Section 7 contains the main result, and the sections preceding it contain various theorems and lemmas which are needed to prove it.

In Section 2, we describe LJ, the standard sequent calculus system formulated by Gentzen for IPL. Section 3 describes Statman's translation from QBF to IPL and introduces the proof system LJ$[\vec{E_S}]$, an augmented form of LJ which has been strengthened by including the extension variables from the translation as axioms. This is the proof system that we use in our main result.

In Section 4 we show that it is possible to take any proof in an extended form of LJ and eliminate all cuts not involving extension axioms (thereby producing a new potentially exponential proof) without affecting the closure of the proof. In Section 5 we show that any sequent in the closure of a proof can be derived efficiently. Together, Sections 4 and 5 constitute an important proof technique, since they allow us to wander into the realm of exponentially-large proofs, take advantage of the reasoning which is possible only there, and then extract what was learned back in the polynomially-bounded realm.

Section 6 contains two critical lemmas which are closely related to the 'Disjunction Property' of IPL.

As already stated, Section 7 contains the main result. This is followed by the main theorem's immediate implications, described in Section 8, and open problems, described in 9.

2 The System LJ

For the purposes of this paper, we are dealing with a sequent calculus for IPL in the style of Gentzen [Sza69]. We will use capital letters $A, B, C, ...$ to denote complex formulas, lower-case letters $x, y, z, ...$ to denote atomic formulas, and Greek letters $\Gamma, \Delta, \Theta, ...$ to denote sets.

The proof system that we will be using is essentially the same as Gentzen's system LJ. Our proofs are tree-like, meaning that each sequent can be an input for at most one inference rule. Furthermore, each sequent has singular right side; that is, there is at most one formula on the right-hand side of each sequent. In addition, all axioms are of the form $x \mapsto x$, where x is an atomic sentence letter. LJ is formulated as follows:

axiom:

$$\frac{}{x \mapsto x} \qquad \text{where } x \text{ is atomic}$$

weakening:

$$\textbf{left} \quad \frac{\Gamma \mapsto A}{B, \Gamma \mapsto A} \qquad \text{and} \qquad \textbf{right} \quad \frac{\Gamma \mapsto}{\Gamma \mapsto B}$$

exchange:

$$\textbf{left} \quad \frac{\Gamma_1, A, B, \Gamma_2 \mapsto C}{\Gamma_1, B, A, \Gamma_2 \mapsto C}$$

contraction:

$$\textbf{left} \quad \frac{\Gamma_1, A, A, \Gamma_2 \mapsto B}{\Gamma_1, A, \Gamma_2 \mapsto B}$$

cut:

$$\frac{\Gamma \mapsto A \quad A, \Delta \mapsto B}{\Gamma, \Delta \mapsto B}$$

$\bot$ introduction:

$$\textbf{left}\ \frac{}{\bot \mapsto} \qquad \text{and} \qquad \textbf{right}\ \frac{\Gamma \mapsto}{\Gamma \mapsto \bot}$$

$\neg$ introduction:

$$\textbf{left}\ \frac{\Gamma \mapsto A}{\neg A, \Gamma \mapsto} \qquad \text{and} \qquad \textbf{right}\ \frac{A, \Gamma \mapsto}{\Gamma \mapsto \neg A}$$

$\vee$ introduction:

$$\textbf{left}\ \frac{A, \Gamma \mapsto C \quad B, \Delta \mapsto C}{A \vee B, \Gamma, \Delta \mapsto C} \qquad \text{and} \qquad \textbf{right}\ \frac{\Gamma \mapsto A}{\Gamma \mapsto A \vee B} \quad \text{as well as} \quad \frac{\Gamma \mapsto A}{\Gamma \mapsto B \vee A}$$

$\wedge$ introduction:

$$\textbf{left}\ \frac{A, B, \Gamma \mapsto C}{A \wedge B, \Gamma \mapsto C} \qquad \text{and} \qquad \textbf{right}\ \frac{\Gamma \mapsto A \quad \Delta \mapsto B}{\Gamma, \Delta \mapsto A \wedge B}$$

$\supset$ introduction:

$$\textbf{left}\ \frac{\Gamma \mapsto A \quad B, \Delta \mapsto C}{A \supset B, \Gamma, \Delta \mapsto C} \qquad \text{and} \qquad \textbf{right}\ \frac{A, \Gamma \mapsto B}{\Gamma \mapsto A \supset B}$$

One important point of note is that if you remove the requirement of having only a single formula on the right-hand side of a sequent, then the system LJ becomes the system LK for classical logic [Sza69]. In other words, every LJ proof is also an LK proof. Technically speaking, LK is a formulation for first-order logic, but we are only interested in its propositional fragment. There has been some effort to distinguish this propositional part of LK by calling it PK, but we shall use the more standard form, since it is clear that we are dealing with propositional logic.

3 Statman's Translation & LJ$[\vec{E_S}]$

In this section we define an augmented form of LJ that includes extension axioms. We then review Statman's reduction from QBF to IPL. Next we define a specific form of LJ augmented with Statman's extension axioms. Finally, we prove that it is easy to extract the ultimate y extension axiom from Statman's translation.

3.1 Extended LJ

Definition 3.1. *We shall use* LJ$[\vec{E}]$ *to denote an augmented form of* LJ *where* $\vec{E} = E_1, ..., E_n$*, and each* E_i *contains the pair of sequents* $A \circ B \mapsto y_{A \circ B}$ *and* $y_{A \circ B} \mapsto A \circ B$ *with the restriction that* $A \circ B$ *must be a complex formula, and* $y_{A \circ B}$ *is an atom not appearing in* $E_1, ..., E_{i-1}$*. We allow the sequents in* $\vec{E}$ *to be used as axioms in any* LJ$[\vec{E}]$ *proof. We refer to* LJ$[\vec{E}]$ *as an Extended Sequent Calculus, and the two sequents in each* E_i *are referred to as extension axioms. This same augmentation can be defined for the system* LK*.*

3.2 Statman's Translation

In his paper showing that IPL is $\mathcal{PSPACE}$-Complete [Sta79], Statman proves this result by providing a reduction from QBF to IPL. This translation is very important to our main result, and proceeds as follows:

1. As input take a QBF formula $F_{QBF} = Q_n x_n, ..., Q_1 x_1 B_0$, where B_0 is a quantifier-free prenex formula, and each $Q_i = \forall$ or $\exists$.

2. Let $y_0, ..., y_n$ be entirely new variables not appearing in F_{QBF}. These are extension variables that are necessary to keep our translation from growing exponentially.

3. Define a series of $B_i^{\checkmark}$ formulas as follows (note that these are not extension variables, but rather are just shorthand for the purposes of this reduction; similarly, $A \leftrightarrow B$ is shorthand for $(A \supset B) \wedge (B \supset A)$):

   - $B_0^{\checkmark} = \neg\neg B_0 \leftrightarrow y_0$
   - $B_k^{\checkmark} = ((x_k \vee \neg x_k) \supset y_{k-1}) \leftrightarrow y_k$ if $Q_k = \forall$
   - $B_k^{\checkmark} = ((x_k \supset y_{k-1}) \vee (\neg x_k \supset y_{k-1})) \leftrightarrow y_k$ if $Q_k = \exists$

4. Output $F_{IPL} = B_0^{\checkmark} \supset (B_1^{\checkmark} \supset (...(B_{n-1}^{\checkmark} \supset (B_n^{\checkmark} \supset y_n))...))$

3.3 Proof of Correctness

A full proof of correctness showing that F_{QBF} is true if and only if F_{IPL} is intuitionistically valid does not add to the understanding of this paper, so we shall refer the interested reader to [Sta79]. However, part of the proof is of interest. Specifically, Statman's proof of correctness implicitly proves that a cut-free tree-like sequent calculus formulation of LJ p-simulates Boolean Truth Trees (BTT), a brute-force tree-like system for QBF.

3.3.1 Boolean Truth Trees

The BTT proof system is defined as follows:

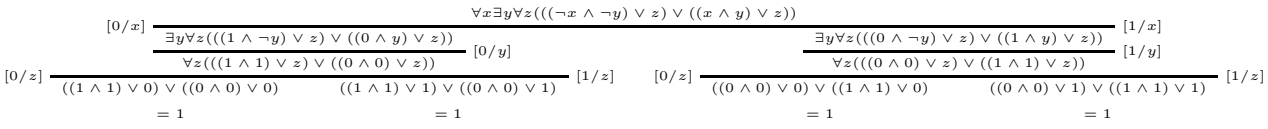
- Every BTT proof is a tree in which every node contains a fully-quantified QBF formula. These formulas may contain constants 0 and 1. The root contains the formula which is to be proven true, and each leaf contains a formula in which all variables have been replaced by constants, and it consequently evaluates to 1, thereby showing that the formula is true. (It is most convenient to picture the proof with the root at the top and the leaves at the bottom).

- Every internal node v in the proof tree has either one or two children. If the outermost quantifier in the formula F at v is $\exists x$, then v has one child containing F', which is F with $\exists x$ removed, and every instance of x replaced by an appropriate constant such that F' is true. If the outermost quantifier in the formula F at v is $\forall x$, then v has two children, one containing F_0, where all instances of x have been replaced by 0, and the other containing F_1, where all instances of x have been replaced by the value 1.

It is easy to see that BTT is sound; if a BTT tree can be built for a QBF formula F, then F is true. It is also easy to see that BTT is complete; every true QBF formula has a BTT proof. This is a brute-force proof system, because it is extremely inefficient; the size of a BTT proof is exponential in the number of $\forall$ quantifiers contained in the formula to be proved true.

3.3.2 BTT Example

Consider the formula $\forall x \exists y \forall z(((\neg x \wedge \neg y) \vee z) \vee ((x \wedge y) \vee z))$. The following BTT proof shows that it is a true QBF formula:



3.3.3 P-Simulation Result

We shall now make explicit the implicit p-simulation result in Statman's proof of correctness; this provides a deeper insight into the mechanisms at play underlying the reduction.

If we forget about the extension variables, then Statman's translation converts the formula F_{QBF} to an intermediate formula A^+ as follows:

- $B_0^+ = \neg\neg B_0$
- $B_k^+ = (x_k \vee \neg x_k) \supset B_{k-1}^+$ if $Q_k = \forall$
- $B_k^+ = (x_k \supset B_{k-1}^+) \vee (\neg x_k \supset B_{k-1}^+)$ if $Q_k = \exists$

Note that $A^+ = B_n^+$. Statman's proof of correctness proceeds in two parts. He first shows that F_{QBF} is true if and only if A^+ is intuitionistically provable. Next he proves that A^+ is intuitionistically provable if and only if F_{IPL} is. We are only interested in the forward direction of the first part; the implicit p-simulation result is given by the proof that F_{QBF} being true implies that A^+ is intuitionistically provable.

The overall idea of the simulation is to take the brute-force BTT proof tree T_{BTT} for F_{QBF}, and by way of very local transformations, build an exactly analogous cut-free tree-like LJ proof T_{LJ}. In order to prove this result we shall make use of the following Lemma:

Lemma 3.2. *Let A be any intuitionistically valid formula, let $l_1, ..., l_n$ be a sequence of literals containing all of the variables in A, and let V be the classical truth assignment which sets all of $l_1, ..., l_n$ to true. If A is true under V, then there exists a cut-free, tree-like* LJ *proof of $l_1, ..., l_n \mapsto A$, and if A is false under V, then there exists a cut-free, tree-like* LJ *proof of $l_1, ..., l_n, A \mapsto$; in either case the size of the proof is linear in the number of logical particles contained in A, and therefore has size $O(n)$.*

The proof is by induction on the number of logical particles in A and is completely straightforward; the basis consists of axioms, which contain no logical particles, and the induction step has eight cases to be considered, four for when A is true, and four for when A is false. An interesting corollary to this Lemma is that the double negation in Statman's translation is unnecessary, and in fact is better omitted.

This brings us to our p-simulation result:

Theorem 3.3. *Cut-free tree-like* LJ *p-simulates* BTT.

Proof: First build T_{BTT} as follows:

1. The root contains $\vdash_{\mathsf{BTT}} F_{QBF}$, where $F_{QBF} = Q_n x_n, ..., Q_1 x_1 B_0$.
2. For every tree node containing $\vdash_{\mathsf{BTT}} \forall x_i, Q_{i-1} x_{i-1}, ..., Q_1 x_1 B_0$, create two new children $\vdash_{\mathsf{BTT}} Q_{i-1} x_{i-1}, ..., Q_1 x_1 B_0[0/x_i]$ and $\vdash_{\mathsf{BTT}} Q_{i-1} x_{i-1}, ..., Q_1 x_1 B_0[1/x_i]$.
3. For every tree node containing $\vdash_{\mathsf{BTT}} \exists x_i, Q_{i-1} x_{i-1}, ..., Q_1 x_1 B_0$, create one new child node, either $\vdash_{\mathsf{BTT}} Q_{i-1} x_{i-1}, ..., Q_1 x_1 B_0[0/x_i]$ or $\vdash_{\mathsf{BTT}} Q_{i-1} x_{i-1}, ..., Q_1 x_1 B_0[1/x_i]$, depending on which form makes the formula true.

Using T_{BTT} as a template, build T_{LJ} as follows (note that unlike with most LJ proofs which have the root at the bottom, we are building a proof with the root at the top so as to make the relationship with BTT as clear as possible):

1. The root contains $\mapsto A^+$ (which is Statman's translation of F_{QBF} from step 1 of the T_{BTT} construction above).

2. For every tree node containing $l_n, ..., l_{k+1} \mapsto (x_k \vee \neg x_k) \supset B^+_{k-1}$, introduce the following three new nodes:

$$\begin{array}{c} l_n, ..., l_{k+1} \mapsto (x_k \vee \neg x_k) \supset B^+_{k-1} \\ | \\ l_n, ..., l_{k+1}, x_k \vee \neg x_k \mapsto B^+_{k-1} \\ \diagup \quad \diagdown \\ l_n, ..., l_{k+1}, x_k \mapsto B^+_{k-1} \qquad l_n, ..., l_{k+1}, \neg x_k \mapsto B^+_{k-1} \end{array}$$

   Note that this requires only one application of $\vee$-$Left$, and one application of $\supset$-$Right$ (remember that we are looking at the proof upside-down). This step corresponds exactly to step 2 of the T_{BTT} construction above.

3. For every tree node containing $l_n, ..., l_{k+1} \mapsto (x_k \supset B^+_{k-1}) \vee (\neg x_k \supset B^+_{k-1})$, introduce the following two new nodes:

$$\begin{array}{c} l_n, ..., l_{k+1} \mapsto (x_k \supset B^+_{k-1}) \vee (\neg x_k \supset B^+_{k-1}) \\ | \\ l_n, ..., l_{k+1} \mapsto l_k \supset B^+_{k-1} \\ | \\ l_n, ..., l_{k+1}, l_k \mapsto B^+_{k-1} \end{array}$$

   Note that this requires only one application of $\supset$-$Right$, and one application of $\vee$-$Right$. This step corresponds exactly to step 3 of the T_{BTT} construction above.

This construction halts once we reach the leaves of T_{BTT}. Its 1-leaves correspond to sequents of the form $l_n, ..., l_1 \mapsto \neg\neg B_0$ in T_{LJ}, but $l_n, ..., l_1$ entail $\neg\neg B_0$, so by Lemma 3.2, each of these sequents has a size-$O(n)$ cut-free, tree-like LJ proof. In effect, there is a series of local transformations which directly relates the steps in T_{BTT} to the steps in T_{LJ} such that each local step in T_{BTT} grows at most linearly to become a local step in T_{LJ}. It is therefore easy to see that the size of T_{LJ} is linear in the size of T_{BTT}. This completes the p-simulation.

□

In effect, the obvious brute-force BTT proof translates into a brute-force cut-free, tree-like LJ proof.

It is interesting to note that the Disjunction Property of IPL (See section 6) is what allows for this p-simulation to work. However, it is also the Disjunction Property which allows for the proof of our main theorem, which ultimately shows that the translation is not feasible. The Disjunction Property therefore acts as both an aid as well as a hindrance to Statman's reduction.

3.4 The System $\mathsf{LJ}[\vec{E_S}]$

Given that we have the definition of $\mathsf{LJ}[\vec{E}]$ as well as the details of Statman's translation (the full version, with extension variables), we may now define the system $\mathsf{LJ}[\vec{E_S}]$. This is the system which we will use for our main theorem in Section 7.

Definition 3.4. *We define* $\mathsf{LJ}[\vec{E_S}]$ *to be* LJ *augmented with Statman's extension axioms for a formula* $F_{QBF} = Q_n x_n, ..., Q_1 x_1 B_0$*, where* B_0 *is a quantifier-free prenex formula, and each* $Q_i = \forall$ *or* $\exists$*. More precisely,* $\vec{E_S} = E_0, ..., E_n$*, where*

- $E_0 = \{\neg\neg B_0 \mapsto y_0, y_0 \mapsto \neg\neg B_0\}$
- $E_k = \{(x_k \vee \neg x_k) \supset y_{k-1} \mapsto y_k, y_k \mapsto (x_k \vee \neg x_k) \supset y_{k-1}\}$ *if* $Q_k = \forall$*, and*
- $E_k = \{(x_k \supset y_{k-1}) \vee (\neg x_k \supset y_{k-1}) \mapsto y_k, y_k \mapsto (x_k \supset y_{k-1}) \vee (\neg x_k \supset y_{k-1})\}$ *if* $Q_k = \exists$

3.5 Manipulating the Result of Statman's Translation

For the purpose of our main theorem, instead of working with F_{IPL}, the result of Statman's translation, we will need to access its innermost extension variable. We therefore need to show that this extension variable can be efficiently extracted.

Lemma 3.5. *Let* P *be a size-*N $\mathsf{LJ}[\vec{E_S}]$*-proof of* $\mapsto F_{IPL}$ *ie. of* $\mapsto B_0^{\surd} \supset (B_1^{\surd} \supset (...(B_{n-1}^{\surd} \supset (B_n^{\surd} \supset y_n))...))$*, where* N *is the number of bits required to encode* P*. We can produce a size-*$O(N^3)$ $\mathsf{LJ}[\vec{E_S}]$ *proof* P' *of* $\mapsto y_n$ *which contains every possible extension axiom as a sequent.*

Proof: First note that although $\mathsf{LJ}[\vec{E_S}]$ does not have the rule of modus ponens, we can simulate it using cut with nothing more than a polynomial increase in proof-size: Suppose we have proofs of $\Gamma \mapsto A$, $\Gamma \mapsto A \supset B$, $A \mapsto A$, and $B \mapsto B$. From these we can produce a proof of $\Gamma \mapsto B$ as follows:

$$\dfrac{\dfrac{\dfrac{\vdots\ P_1}{\Gamma \mapsto A} \qquad \dfrac{\dfrac{\vdots\ P_2}{\Gamma \mapsto A \supset B} \qquad \dfrac{\dfrac{\vdots\ P_3}{A \mapsto A} \qquad \dfrac{\vdots\ P_4}{B \mapsto B}}{A, A \supset B \mapsto B}\supset\text{-}Left}{\Gamma, A \mapsto B}Cut}{\Gamma \cup \Gamma \mapsto B}Cut}{\vdots}\ Weaken$$

Therefore, our size-N $\mathsf{LJ}[\vec{E_S}]$ proof P of $\mapsto B_0^{\surd} \supset (B_1^{\surd} \supset (...(B_{n-1}^{\surd} \supset (B_n^{\surd} \supset y_n))...))$ corresponds to the proof P_2 of $\Gamma \mapsto A \supset B$, so to simulate modus ponens, we need only show that we can create proofs of the analogs of $\Gamma \mapsto A$, $A \mapsto A$, and $B \mapsto B$ that are short relative to N, the length of P_2.

In our case, $\Gamma \mapsto A$ is of the form $\mapsto B_i^{\surd}$ for some $B_i^{\surd}$. Each $B_i^{\surd}$ is associated with an E_k, which contains two extension axioms, one of the form $Formula \mapsto Variable$ and the other of the form $Variable \mapsto Formula$. In other words, $B_i^{\surd} = (F \supset v) \wedge (v \supset F)$, and can be proved as follows:

$$\dfrac{\dfrac{\dfrac{}{F \mapsto v}}{\mapsto F \supset v}\supset\text{-}Right \qquad \dfrac{\dfrac{}{v \mapsto F}}{\mapsto v \supset F}\supset\text{-}Right}{\mapsto (F \supset v) \wedge (v \supset F)}\wedge\text{-}Right$$

Therefore we can construct P_1 in our modus ponens simulation using a constant-sized proof. Note that this proof includes both extension axioms associated with $B_i^{\surd}$.

In our case, $A \mapsto A$ is of the form $B_i^{\surd} \mapsto B_i^{\surd}$ for some $B_i^{\surd}$. This can be proved by first proving $\mapsto B_i^{\surd}$, and then using Weakening-L to prove $B_i^{\surd} \mapsto B_i^{\surd}$. Therefore we can construct P_3 in our modus ponens simulation using a constant-sized proof.

The case of $B \mapsto B$ is a little more complicated, and is of the form $B_i^{\checkmark} \supset (B_{i+1}^{\checkmark} \supset (...(B_{n-1}^{\checkmark} \supset (B_n^{\checkmark} \supset y_n))...)) \mapsto B_i^{\checkmark} \supset (B_{i+1}^{\checkmark} \supset (...(B_{n-1}^{\checkmark} \supset (B_n^{\checkmark} \supset y_n))...))$. As shorthand, we will also refer to this sequent as $F_{IPL} \mapsto F_{IPL}$. The proof P_4 of this sequent is constructed as follows:

$$\dfrac{\mapsto B_i^{\checkmark} \qquad \dfrac{\dfrac{\dfrac{\ddots \; \cdot^{\cdot^{\cdot}}}{B_{i+1}^{\checkmark} \supset (...(B_{n-1}^{\checkmark} \supset (B_n^{\checkmark} \supset y_n))...) \mapsto B_{i+1}^{\checkmark} \supset (...(B_{n-1}^{\checkmark} \supset (B_n^{\checkmark} \supset y_n))...)}}{B_{i+1}^{\checkmark} \supset (...(B_{n-1}^{\checkmark} \supset (B_n^{\checkmark} \supset y_n))...), B_i^{\checkmark} \mapsto B_{i+1}^{\checkmark} \supset (...(B_{n-1}^{\checkmark} \supset (B_n^{\checkmark} \supset y_n))...)} \, Weaken}{B_{i+1}^{\checkmark} \supset (...(B_{n-1}^{\checkmark} \supset (B_n^{\checkmark} \supset y_n))...) \mapsto B_i^{\checkmark} \supset (B_{i+1}^{\checkmark} \supset (...(B_{n-1}^{\checkmark} \supset (B_n^{\checkmark} \supset y_n))...))} \supset\text{-}Right}{B_i^{\checkmark} \supset (B_{i+1}^{\checkmark} \supset (...(B_{n-1}^{\checkmark} \supset (B_n^{\checkmark} \supset y_n))...)) \mapsto B_i^{\checkmark} \supset (B_{i+1}^{\checkmark} \supset (...(B_{n-1}^{\checkmark} \supset (B_n^{\checkmark} \supset y_n))...))} \supset\text{-}Left$$

Repeat this process at most n times, each time stripping off one $B_i^{\checkmark}$ from each side. This will yield the first line of P_4 to be $y_n \mapsto y_n$, an axiom. We already showed that $\mapsto B_i^{\checkmark}$ has a constant-sized proof, so the total length of our proof of P_4 will be linear in n, but since F_{IPL} occurs in P, n is $O(N)$, so each P_4 adds at most $O(N^2)$ to the size of our modus ponens simulation.

Since P_1 and P_3 have constant size, P_2 has size N, and P_4 has size-$O(N^2)$, every application of modus ponens adds one P_4 and therefore $O(N^2)$ to the size of our proof. In order to get at y_n, we must apply modus ponens n times, but as we already said, n is $O(N)$, so our entire proof of $\mapsto y_n$ requires size-$O(N^3)$.

Therefore, if given a size-N proof of $\mapsto B_0^{\checkmark} \supset (B_1^{\checkmark} \supset (...(B_{n-1}^{\checkmark} \supset (B_n^{\checkmark} \supset y_n))...))$, we can repeatedly apply modus ponens with the $\mapsto B_i^{\checkmark}$ sequents to produce a size-$O(N^3)$ of $\mapsto y_n$. In addition, since each subproof of $\mapsto B_i^{\checkmark}$ contains both extension axioms, and since all $\mapsto B_i^{\checkmark}$s are present, our overall proof of $\mapsto y_n$ contains each possible extension axiom, as required.

□

4 Cut-Elimination

In this section we extend the cut-elimination technique developed by Buss and Pudlák in [BP01] (which itself was adapted from [BM99]) so that it holds for any system $\mathsf{LJ}[\vec{E}]$.

4.1 Definitions

We shall make use of the following definitions:

Definition 4.1. *The closure of a proof P, denoted $cl(P)$, is the smallest set of sequents which includes the sequents of P and is closed under both weakening and cut.*

Definition 4.2. *In a proof P, the direct ancestors of a formula A are all instances of A comprising an unbroken path towards the leaves of P from A to the first instance where A was introduced.*

Definition 4.3. *A principal cut is a cut in which at least one of its two input sequents is an extension axiom.*

4.2 Cut-Elimination Theorem

Theorem 4.4 (Cut-Elimination). *For any proof P in any system $\mathsf{LJ}[\vec{E}]$, it is possible to eliminate all non-principal cuts from P to produce a pseudo-cut-free proof P' such that $cl(P') \subseteq cl(P)$.*

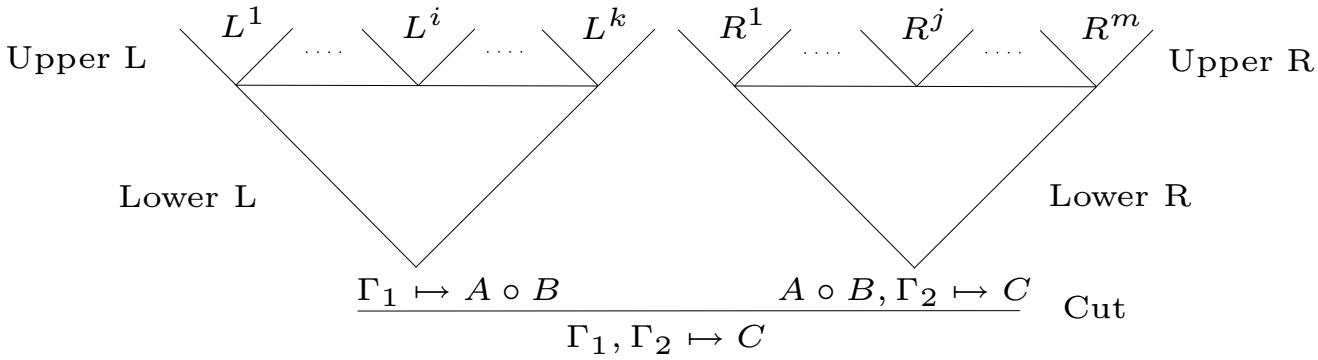


Figure 1: The Template For Cut-Elimination

Proof: We will first show the general technique for eliminating a cut on an arbitrary binary logical connective $\circ$, and will then provide the details specific to the connectives $\supset, \vee, \wedge$. Finally, we will show how to remove cuts on atoms. While reading this proof, please refer to Figure 1 below.

We wish to eliminate a cut on the formula $A \circ B$. Let L be the proof of the left-hand input to the cut, and let R be the proof of the right-hand input to the cut. Let 'lower L' be the portion of L which contains sequents with direct ancestors of $A \circ B$ appearing in the succedent. In other words, each sequent in lower L is of the form $\Theta \mapsto A \circ B$. At the upper border of lower L, there are k subproofs labeled L^i, each of which introduces $A \circ B$ for the first time along its branch of the proof. More specifically, each L^i proves the sequent $\Pi_i \mapsto A \circ B$.

Similarly, 'lower R' is the portion of R in which the sequents contain the direct ancestors of $A \circ B$ in the antecedent. In other words, each sequent in lower R is of the form $A \circ B, \Theta \mapsto X$. Along the upper border of lower R, there are m subproofs labeled R^j, each of which introduces $A \circ B$ for the first time along its branch of the proof. More specifically, each R^j proves the sequent $A \circ B, \Delta_j \mapsto D_j$.

In order to eliminate the cut on $A \circ B$, we perform the following steps:

1. For $i = 1$ to k create a proof R^{i*} by taking a copy of R and doing the following: modify each R^j along the entire border between upper and lower R so that instead of proving $A \circ B, \Delta_j \mapsto D_j$, each R^j now proves $\Pi_i, \Delta_j \mapsto D_j$. In the case where $A \circ B$ was introduced by weakening, simply introduce Π_i via weakening instead. In all other cases, take a copy of L^i (which is where Π_i comes from) and splice it into each R^j along the entire border in order to replace $A \circ B$ with Π_i (the details of this splicing depend on which connective is being eliminated, see below). To be clear, we do this k times; for each of the k separate R^{i*}s that we are building, splice each of the k L^is in once with each of the m R^js appearing in upper R.

2. Next, to complete the construction of each R^{i*}, replace each sequent $A \circ B, \Theta \mapsto X$ in lower R with the sequent $\Pi_i, \Theta \mapsto X$. Each R^{i*} therefore proves $\Pi_i, \Gamma_2 \mapsto C$ instead of $A \circ B, \Gamma_2 \mapsto C$.

3. Modify L: First, replace each L^i in upper L with R^{i*}. Now every sequent along the border between upper and lower L is no longer $\Pi_i \mapsto A \circ B$, it is $\Pi_i, \Gamma_2 \mapsto C$. Next, replace each sequent $\Theta \mapsto A \circ B$ in lower L with $\Theta, \Gamma_2 \mapsto C$. Therefore, instead of proving $\Gamma_1 \mapsto A \circ B$, L now proves $\Gamma_1, \Gamma_2 \mapsto C$, which was precisely the result of our cut, so it has been eliminated.

It is not hard to see that this process of eliminating cuts does not add any new sequents to the closure of the proof; our cut-elimination produced new sequents in only three ways: Firstly, we created new sequents while constructing the $R^{i*}s$ when we replaced each $A \circ B, \Theta \mapsto X$ with the new sequent

$\Pi_i, \Theta \mapsto X$. However, each L^i proves the sequent $\Pi_i \mapsto A \circ B$, and when we cut this with $A \circ B, \Theta \mapsto X$, we get $\Pi_i, \Theta \mapsto X$, showing that it was in the original closure of P.

Secondly, we created new sequents while modifying lower L when we replaced each sequent $\Theta \mapsto A \circ B$ in lower L with the new sequent $\Theta, \Gamma_2 \mapsto C$. However, when we cut $\Theta \mapsto A \circ B$ with $A \circ B, \Gamma_2 \mapsto C$ (which was the original result of R), we get $\Theta, \Gamma_2 \mapsto C$, showing that it too was in the original closure of P.

Finally, we created new sequents during the splicing of the L^is into the R^js. Below, alongside the details of how to actually perform the splicing, we will show that each of these new sequents is in the closure of the original proof.

These were the only types of sequents which we introduced during our cut-elimination of binary connectives; all others came from the original proof P. Therefore, the cut elimination technique does not introduce any new sequents to the closure.

All that remains to be shown is how to splice L^i into each R^j to turn $A \circ B, \Gamma_2 \mapsto C$ into $\Pi_i, \Delta_j \mapsto D_j$ for each binary connective, and that the closure condition is satisfied. Each L^i proves the sequent $\Pi_i \mapsto A \circ B$, which is itself an extension axiom, or was introduced by weakening or $\circ$-$Right$. Similarly, each R^j proves the sequent $A \circ B, \Delta_j \mapsto D_j$, which is itself an extension axiom, or was introduced by weakening or $\circ$-$Left$. We must show how to combine each of these potential L^is with each of the R^js to get $\Pi_i, \Delta_j \mapsto D_j$:

Usually replacing $A \circ B$ by Π_i requires splicing, but in the case where R^j introduced $A \circ B$ by weakening, no splicing is required; simply weaken to get Π_i rather than $A \circ B$. Each of the remaining combinations is spliced as follows:

Case 1

If both $\Pi_i \mapsto A \circ B$ and $A \circ B, \Delta_j \mapsto D_j$ are extension axioms of the form $y_{A \circ B} \mapsto A \circ B$ and $A \circ B \mapsto y_{A \circ B}$, simply get $y_{A \circ B} \mapsto y_{A \circ B}$ by using a single principal cut.

Case 2

If $\Pi_i \mapsto A \circ B$ is an extension axiom of the form $y_{A \circ B} \mapsto A \circ B$ and $A \circ B, \Delta_j \mapsto D_j$ was introduced by $\circ$-$Left$, again get $y_{A \circ B}, \Delta_j \mapsto D_j$ by using a single principal cut.

Case 3

If $\Pi_i \mapsto A \circ B$ was introduced by $\circ$-$Right$ and $A \circ B, \Delta_j \mapsto D_j$ is an extension axiom of the form $A \circ B \mapsto y_{A \circ B}$, once again get $\Pi_i \mapsto y_{A \circ B}$ by using a single principal cut.

Case 4

However, if both $\Pi_i \mapsto A \circ B$ and $A \circ B, \Delta_j \mapsto D_j$ came from introduction rules, then splicing to get $\Pi_i, \Delta_j^\alpha, \Delta_j^\beta \mapsto D_j$ is somewhat more complicated and depends on what connective $\circ$ represents.

Case 4a If $A \circ B$ is the formula $A \supset B$, then L^i is of the form

$$\dfrac{\begin{array}{c}\overset{L^i}{\ddots \vdots \iddots} \\ \Pi_i, A \mapsto B\end{array}}{\Pi_i \mapsto A \supset B} \supset\text{-}Right$$

and R^j is of the form

$$\dfrac{\begin{array}{cc}\overset{R^{j\alpha}}{\vdots} & \overset{R^{j\beta}}{\vdots} \\ \Delta_j^\alpha \mapsto A & B, \Delta_j^\beta \mapsto D_j\end{array}}{A \supset B, \Delta_j^\alpha, \Delta_j^\beta \mapsto D_j} \supset\text{-}Left$$

we may therefore splice them together as follows:

$$\dfrac{\dfrac{\begin{array}{cc}\overset{R^{j\alpha}}{\vdots} & \overset{L^{i}}{\vdots} \\ \Delta_j^\alpha \mapsto A & \Pi_i, A \mapsto B\end{array}}{\Pi_i, \Delta_j^\alpha, \mapsto B}\, Cut \qquad \begin{array}{c}\overset{R^{j\beta}}{\vdots} \\ B, \Delta_j^\beta \mapsto D_j\end{array}}{\Pi_i, \Delta_j^\alpha, \Delta_j^\beta \mapsto D_j}\, Cut$$

to produce $\Pi_i, \Delta_j \mapsto D_j$, as required. Note that $\Pi_i, \Delta_j^\alpha, \mapsto B$ is a new sequent that potentially did not occur in our original proof, but it was produced via cut from sequents from the original proof, so no new sequents are added to the closure.

Case 4b If $A \circ B$ is the formula $A \vee B$, then L^i is of the form

$$\dfrac{\begin{array}{c}\overset{L^{i}}{\vdots} \\ \Pi_i, \mapsto A\end{array}}{\Pi_i \mapsto A \vee B} \vee\text{-}Right$$

and R^j is of the form

$$\dfrac{\begin{array}{cc}\overset{R^{j\alpha}}{\vdots} & \overset{R^{j\beta}}{\vdots} \\ A, \Delta_j^\alpha \mapsto D_j & B, \Delta_j^\beta \mapsto D_j\end{array}}{A \vee B, \Delta_j^\alpha, \Delta_j^\beta \mapsto D_j} \vee\text{-}Left$$

we may therefore splice them together as follows:

$$\dfrac{\dfrac{\begin{array}{cc}\overset{L^{i}}{\vdots} & \overset{R^{j\alpha}}{\vdots} \\ \Pi_i, \mapsto A & A, \Delta_j^\alpha \mapsto D_j\end{array}}{\Pi_i, \Delta_j^\alpha, \mapsto D_j}\, Cut}{\Pi_i, \Delta_j^\alpha, \Delta_j^\beta \mapsto D_j}\, Weaken$$

to produce $\Pi_i, \Delta_j \mapsto D_j$, as required (the case involving $\Pi_i, \mapsto B$ is entirely similar; just use $R^{j\beta}$). Note that $\Pi_i, \Delta_j^\alpha, \mapsto D_j$ is a new sequent that potentially did not occur in our original proof, but it was produced via cut from sequents from the original proof, so no new sequents are added to the closure.

Case 4c If $A \circ B$ is the formula $A \wedge B$, then L^i is of the form

$$\dfrac{\begin{array}{cc}\overset{L^{i\alpha}}{\ddots} & \overset{L^{i\beta}}{\ddots} \\ \Pi_i^\alpha, \mapsto A & \Pi_i^\beta, \mapsto B\end{array}}{\Pi_i^\alpha, \Pi_i^\beta \mapsto A \wedge B} \wedge\text{-}Right$$

and R^j is of the form

$$\dfrac{\begin{array}{c}\overset{R^{j}}{\ddots} \\ A, \Delta_j \mapsto D_j\end{array}}{A \wedge B, \Delta_j \mapsto D_j} \wedge\text{-}Left$$

we may therefore splice them together as follows:

$$\dfrac{\dfrac{\begin{array}{cc}\overset{L^{i\alpha}}{\ddots} & \overset{R^{j}}{\ddots} \\ \Pi_i^\alpha, \mapsto A & A, \Delta_j \mapsto D_j\end{array}}{\Pi_i^\alpha, \Delta_j, \mapsto D_j}\, Cut}{\Pi_i^\alpha, \Pi_i^\beta, \Delta_j \mapsto D_j}\, Weaken$$

to produce $\Pi_i, \Delta_j \mapsto D_j$, as required (the case involving $B, \Delta_j \mapsto D_j$ is entirely similar; just use $L^{i\beta}$). Note that $\Pi_i^\alpha, \Delta_j, \mapsto D_j$ is a new sequent that potentially did not occur in our original proof, but it was produced via cut from sequents from the original proof, so no new sequents are added to the closure.

Cases $1-3$ above introduce only principal cuts, which do not have to be eliminated. Note that each of the constructions in Case 4 is working to eliminate a cut on a binary connective, but in the process introduces at least one new cut. However, this is all right, since the only cuts being introduced are on simpler formulas, so this process when iterated will inevitably result in a proof in which the only cuts remaining are on atoms. As our last case, we must therefore show how cuts on atoms may be removed without introducing any new cuts at all:

The Atomic Case

The atomic case is fairly simple, and does not involve notions such as upper and lower L or R. Suppose we wish to eliminate a cut on an atomic formula x such as in

$$\dfrac{\begin{array}{cc}\overset{L}{\ddots} & \overset{R}{\ddots} \\ \Gamma_1 \mapsto x & x, \Gamma_2 \mapsto C\end{array}}{\Gamma_1, \Gamma_2, \mapsto C}\, Cut$$

Note that x could be a normal variable, or an extension variable $y_{A\circ B}$. All axioms in L are either of the form $x \mapsto x$ or $A \circ B \mapsto y_{A\circ B}$, and every sequent in L is of the form $\Theta \mapsto x$. The atomic cut-elimination proceeds as follows: First, take L, and replace each sequent $\Theta \mapsto x$ by $\Theta, \Gamma_2 \mapsto C$. The result is a proof of $\Gamma_1, \Gamma_2 \mapsto C$ in which all of the leaves are either $x, \Gamma_2 \mapsto C$ or $A \circ B, \Gamma_2 \mapsto C$.

For each leaf of the form $x, \Gamma_2 \mapsto C$, simply place a copy of R, which proves $x, \Gamma_2 \mapsto C$, on top of it. It is not hard to see that the closure of the proof is not affected; we introduced the sequents $\Theta, \Gamma_2 \mapsto C$

by modifying $\Theta \mapsto x$, but when we cut $\Theta \mapsto x$ with $x, \Gamma_2 \mapsto C$ (which was the original result of R), we get $\Theta, \Gamma_2 \mapsto C$, showing that it too was in the original closure of P.

For each leaf of the form $A \circ B, \Gamma_2 \mapsto C$, rather than adding R on top, add the following subproof on top instead:

$$\dfrac{\overline{A \circ B, \mapsto y_{A\circ B}} \qquad \overset{R}{\overset{\vdots}{y_{A\circ B}, \Gamma_2 \mapsto C}}}{A \circ B, \Gamma_2 \mapsto C}\ P.Cut$$

The cut is a principal cut, so it will not have to be eliminated. In addition, the axiom $A \circ B, \mapsto y_{A\circ B}$ which we introduced is not new; it was the original sequent which we modified in order to require this subproof in the first place, and therefore is not a new addition to the closure.

In effect, we can eliminate cuts on atoms such that no new non-principal cuts are introduced, thereby concluding the proof that cuts can be eliminated even in the presence of extension axioms. □

5 The Proof Closure Property

In this section we formally prove the Proof Closure Property which is stated but never proved in [BM99] and [BP01]. Informally, the Proof Closure Property guarantees that any sequent in the closure of an $\mathsf{LJ}[\vec{E_S}]$ proof P can be derived via a polynomial number of cuts and weakenings from the sequents in P. This property is strongly related to resolution involving Horn clauses. For more information on these topics, please refer to [Sch89].

Definition 5.1. *A Horn clause is a clause in which at most one literal is positive. A Horn formula is one in which every clause is a Horn clause.*

Lemma 5.2. *Every unsatisfiable Horn formula F has a size-$O(n)$ regular input resolution refutation, where n is the number of distinct variables in F.*

Proof: An important fact to note is that every unsatisfiable Horn formula F has at least one clause containing just a single positive literal. To see this, consider a Horn formula in which every clause contains one or more negative literals; setting every variable to false therefore satisfies it, since it sets every clause to true. The fact that F contains a positive unit clause will allow us to build a short DLL refutation tree. Let $\{x_1\}$ be F's unit clause. Build a DLL tree based on x_1; the branch where x_1 is set to false terminates immediately, and the other branch gives us the formula $F \restriction_{x_1=T}$, which is clearly still an unsatisfiable Horn formula since restrictions cannot add positive literals. Therefore $F \restriction_{x_1=T}$ contains a positive unit clause, call it $\{x_2\}$. Simply repeat this process of building the DLL tree, each time branching on the positive unit variable. Since there are only n variables, and since one branch always terminates immediately, this process yields a size-$O(n)$ DLL tree. This tree can easily be turned into a resolution refutation by turning it upside down. It is not hard to see that since DLL trees are regular, this refutation is also regular, and since every resolution step involves an input clause, it is also an input resolution, as required.

□

Corollary 5.3. *Given any set of Horn clauses Σ and any Horn clause H such that $\Sigma \vdash_{RES} H$, there exists a size-$O(N)$ regular input resolution derivation of $D \subseteq H$ from Σ, where N is the number of bits required to encode Σ.*

Proof: Let ϕ be any minimal truth assignment which sets H to false. Since $\Sigma \models_{RES} H$, we know that $\Sigma \restriction_\phi$ is unsatisfiable. Therefore, by Lemma 5.2, $\Sigma \restriction_\phi$ has a size-$O(n)$ regular input resolution refutation,

call it R, where n is the number of distinct sentence letters in $\Sigma \restriction_\phi$. All of the literals in H were eliminated, so this resolution refutation never resolves on H's literals. It is therefore easy to see that if we create R' by replacing every clause in R that came from $\Sigma \restriction_\phi$ with its corresponding clause in Σ, the literals from H are present and will simply be carried down the proof so that instead of proving $\varnothing$ like R did, R' proves $D \subseteq H$, and its size is clearly bounded by $O(n)$. Since n is bounded above by N, the size of R' is bounded by $O(N)$, as required.

□

Lemma 5.4 (Proof Closure Property). *Let P be a size-N $\mathsf{LJ}[\vec{E_S}]$ proof, where N is the number of bits required to encode P. If $\Gamma \mapsto A \in cl(P)$ then there exists a size-$O(N^2)$ tree-like and a size-$O(N)$ DAG-like $\mathsf{LJ}[\vec{E_S}]$-proof of $\Gamma \mapsto A$.*

Proof: It is easy to see that sequents are very similar to Horn clauses. A singular right-side sequent $\Gamma \mapsto A$ is interpreted as meaning that a conjunction of all the formulas in Γ implies A. A Horn clause $(\neg x_1 \vee \neg x_2 \vee ... \vee \neg x_i \vee x_j)$ is equivalent to $(x_1 \wedge x_2 \wedge ... \wedge x_i \supset x_j)$. This gives rise to the following translation between sequents and Horn clauses: if given a sequent $A_1, A_2, ..., A_i \mapsto A_j$ where the As are formulas or negated formulas, we convert this sequent to the Horn clause by replacing each A_i with a variable x_i to give $(\neg x_1 \vee \neg x_2 \vee ... \vee \neg x_i \vee x_j)$. It is easy to see that a sequent with no formula on the right-hand side corresponds to a Horn clause containing no positive literal. Furthermore, a cut on two sequents corresponds to the resolution of two Horn clauses.

Therefore, suppose that we have an $\mathsf{LJ}[\vec{E_S}]$ proof P of size N and we know that $\Gamma \mapsto A \in cl(P)$. As shown above, translate the sequents $S_1, S_2, ..., S_k$ in P to Horn clauses. These clauses comprise our initial set of clauses Σ upon which we will resolve. Let H be the Horn clause corresponding to the sequent $\Gamma \mapsto A$. Since $\Gamma \mapsto A \in cl(P)$, we know by Corollary 5.3 that there is a size-$O(N)$ linear input resolution derivation R of $D \subseteq H$ from Σ. Translate the clauses in R back into sequents. This corresponds to a size-$O(N)$ derivation of $\Gamma \mapsto A$ from the sequents $S_1, S_2, ..., S_k$ in P using cut. To complete the proof, simply weaken D to get H. If our $\mathsf{LJ}[\vec{E_S}]$ proof is required to be tree-like, then add a proof of each $S_1, S_2, ..., S_k$ immediately above it. Each of these proofs is a sub-proof of P, and therefore has size at most $O(N)$. Since there are at most $O(N)$ S_is, this gives an overall size-$O(N^2)$ $\mathsf{LJ}[\vec{E_S}]$-proof of $\Gamma \mapsto A$ for tree-like proofs. If our $\mathsf{LJ}[\vec{E_S}]$ proof can be DAG-like, then these proofs are not necessary, and we have an overall size-$O(N)$ $\mathsf{LJ}[\vec{E_S}]$-proof of $\Gamma \mapsto A$.

□

6 The Disjunction & Implication Properties

Normally the Disjunction Property in intuitionistic propositional logic states that if Γ contains no formula containing $\vee$, then $\Gamma \vdash_{IPL} A \vee B$ implies that $\Gamma \vdash_{IPL} A$ or $\Gamma \vdash_{IPL} B$. However, this form of the Disjunction Property fails when applied to certain extension axioms (for example, just take an extension axiom with $\Gamma = y$ on the left where y is an extension variable, and any formula containing $\vee$ as the major logical particle on the right). We must therefore prove a weaker form of the Disjunction Property that is still strong enough to help us prove our main result:

Lemma 6.1 (Modified Disjunction Property). *If P is a $\mathsf{LJ}[\vec{E_S}]$-proof of the sequent $l_1, ..., l_k \mapsto A \vee B$ such that the only cuts in P are principal cuts, $l_1, ..., l_k$ are literals which do not include any y extension variables, then there either exists a proof P' of $l_1, ..., l_k \mapsto A$ or of $l_1, ..., l_k \mapsto B$ in which all cuts are principal cuts such that $cl(P') \subseteq cl(P)$.*

Proof: Suppose that P is a $\mathsf{LJ}[\vec{E_S}]$-proof of the sequent $l_1, ..., l_k \mapsto A \vee B$ such that the only cuts in P are principal cuts and $l_1, ..., l_k$ are literals which do not include any y extension variables as above. Since P does not contain normal cuts, $l_1, ..., l_k \mapsto A \vee B$ could only have come from one of the following rules:

1. $\vee$-*Right*

2. $Weaken\text{-}Right$
3. $Weaken\text{-}Left$
4. $Contraction\text{-}Left$
5. $APrincipalCut$

In the first case, the penultimate line in P was either $l_1, ..., l_k \mapsto A$ or $l_1, ..., l_k \mapsto B$, so we are done. In the second case, the penultimate line in P was $l_1, ..., l_k \mapsto$, which can be weakened to get what we want, and again we are done.

The next three cases are more complicated. In the weakening-left case, the only difference is that the left-hand side of the previous sequent contained one fewer literal. In the contraction case, the only difference is that the left-hand side of the previous sequent contained one more duplicated literal. In our final case, if $l_1, ..., l_k \mapsto A \vee B$ came from a principal cut, then this cut must have been on the sequent $l_1, ..., l_k \mapsto y$ and the extension axiom $y \mapsto A \vee B$. The sequent $l_1, ..., l_k \mapsto y$ could only have come from one of the following rules:

1. $Weaken\text{-}Right$
2. $Weaken\text{-}Left$
3. $Contraction\text{-}Left$
4. $APrincipalCut$

In the first case, the previous line was $l_1, ..., l_k \mapsto$, which can be weakened to $l_1, ..., l_k \mapsto A$, and we are done. The weakening and contraction cases are similar to the ones before; each just affects the number of literals on the left by one. Finally, if $l_1, ..., l_k \mapsto y$ came from a principal cut, then this cut must have been on the sequent $l_1, ..., l_k \mapsto A \vee B$ and the extension axiom $A \vee B \mapsto y$, which brings us back to what we started with.

Therefore, in order to avoid the cases in which we are done, let us assume that P contains no $\vee$-$Right$ or $Weaken\text{-}Right$ rules. Hence, P has $l_1, ..., l_k \mapsto A \vee B$ as its last line, preceeded by 0 or more weakenings and contractions on the left, preceeded by a principal cut on the sequent $l_1, ..., l_j \mapsto y$, which itself was preceeded by 0 or more weakenings and contractions on the left, and a principal cut on $l_1, ..., l_i \mapsto A \vee B$, as shown here:

$$\dfrac{\dfrac{\dfrac{\dfrac{\vdots}{\vdots}}{l_1, ..., l_i \mapsto A \vee B} \qquad \overline{A \vee B \mapsto y}}{l_1, ..., l_i \mapsto y}\, P.Cut}{\vdots \text{ 0 or more Weaken-L or Contraction-L} \vdots}$$

$$\dfrac{l_1, ..., l_j \mapsto y \qquad \overline{y \mapsto A \vee B}}{l_1, ..., l_j \mapsto A \vee B}\, P.Cut$$

$$\vdots \text{ 0 or more Weaken-L or Contraction-L} \vdots$$

$$l_1, ..., l_k \mapsto A \vee B$$

However, this pattern cannot go on indefinitely, since proofs are only finitely long. Note that every sequent P must therefore have either y or $A \vee B$ on the right-hand side. P cannot begin at a sequent $l \mapsto A \vee B$, since we said that none of the l literals are y extension variables, so $l \mapsto A \vee B$ cannot be an axiom. Similarly, the proof cannot begin at a sequent $l \mapsto y$. Alternatively, P cannot contain the sequents $\mapsto A \vee B$ or $\mapsto y$ as axioms, since these are not axioms. Therefore P cannot contain any axioms, a contradiction. In other words, P must contain an application of $\vee$-*Right* or *Weaken-Right*, so the line preceeding this application allows us to easily show that $l_1, ..., l_k \mapsto A \in cl(P)$ or $l_1, ..., l_k \mapsto B \in cl(P)$, as required.

□

Lastly, we need one final lemma that is very similar to the Modified Disjunction Property:

Lemma 6.2 (Implication Property). *If P is a $\mathsf{LJ}[\vec{E_S}]$-proof of the sequent $l_1, ..., l_k, (x_{k+1} \vee \neg x_{k+1}), ..., (x_j \vee \neg x_j) \mapsto A \supset B$ such that the only cuts in P are principal cuts, and $l_1, ..., l_k$ are literals which do not include any y extension variables, then there exists a proof P' in which all cuts are principal cuts of $l_1, ..., l_k, (x_{k+1} \vee \neg x_{k+1}), ..., (x_j \vee \neg x_j), A \mapsto B$ such that $cl(P') \subseteq cl(P)$.*

Proof: The proof is almost identical to that of Lemma 6.1 with $A \supset B$ replacing all instances of $A \vee B$, and the rule $\supset$-*Right* replacing all instances of $\vee$-*Right*.

□

7 Main Result

We are now ready to prove our main result.

Theorem 7.1 (Main Theorem). *Let $F_{Prop} = A(x_1, ...x_n)$ be any arbitrary classical propositional tautology containing n distinct variables, and consider the formula $F'_{Prop} = A(x_1, ...x_n) \vee \neg A(x_1, ...x_n)$. Let F_{QBF} be the prenex QBF translation of F'_{Prop} where each quantifier is $\forall$, and let F_{IPL} be Statman's translation of F_{QBF}. If there exists a size-N DAG-like $\mathsf{LJ}[\vec{E_S}]$ proof of F_{IPL}, where N is the number of bits required to encode F_{IPL}, then F_{Prop} has a DAG-like classical $\mathsf{LK}[\vec{E_S}]$ proof of size-$O(N^4)$.*

Proof: Since $F'_{Prop} = A(x_1, ...x_n) \vee \neg A(x_1, ...x_n)$, the quantified form is $\forall x_n, ..., \forall x_1 A(x_1, ...x_n) \vee \neg \forall x_n, ..., \forall x_1 A(x_1, ...x_n)$. In order to turn this formula into prenex form, we have to rename the variables in $\neg A(x_1, ...x_n)$. Therefore, $F_{QBF} = \exists x_{2n}, ..., \exists x_{n+1}, \forall x_n, ..., \forall x_1 B_0$, where $B_0 = F'_{Prop} = A(x_1, ...x_n) \vee \neg A(x_{n+1}, ...x_{2n})$ Applying Statman's translation yields

$$F_{IPL} = B_0^{\surd} \supset (B_1^{\surd} \supset (...(B_{2n-1}^{\surd} \supset (B_{2n}^{\surd} \supset y_{2n}))...)).$$

The extension axioms associated with F_{IPL} that we will need are:

- $y_0 \mapsto \neg\neg B_0$
- $y_k \mapsto (x_k \vee \neg x_k) \supset y_{k-1}$ for $k \leq n$ (these are the $\forall$ axioms).
- $y_k \mapsto (x_k \supset y_{k-1}) \vee (\neg x_k \supset y_{k-1})$ for $n+1 \leq k \leq 2n$ (these are the $\exists$ axioms).

Suppose that there exists a size-N $\mathsf{LJ}[\vec{E_S}]$ proof P of $\mapsto F_{IPL}$. We will now show how to build a DAG-like classical $\mathsf{LK}[\vec{E_S}]$ proof of F_{Prop}. By Lemma 3.5, there exists a size-$O(N^3)$ proof of $\mapsto y_{2n}$. Note that this proof contains every extension axiom. Cut this sequent with the extension axiom $y_{2n} \mapsto (x_{2n} \supset y_{2n-1}) \vee (\neg x_{2n} \supset y_{2n-1})$ to get $\mapsto (x_{2n} \supset y_{2n-1}) \vee (\neg x_{2n} \supset y_{2n-1})$, and call this entire proof P_1.

By Theorem 4.4, we can eliminate all non-principal cuts in P_1 to produce P_2 such that $cl(P_2) \subseteq cl(P_1)$. Note that P_2 may be exponentially large. Since all cuts in P_2 are principal cuts, Lemma 6.1 applies, so

there exists a proof P_3 of $\mapsto x_{2n} \supset y_{2n-1}$ or of $\neg x_{2n} \supset y_{2n-1}$. Let l_{2n} be the literal such that P_3 proves $\mapsto l_{2n} \supset y_{2n-1}$, and note that $cl(P_3) \subseteq cl(P_2)$. Furthermore, note that all cuts in P_3 are principal cuts.

By Lemma 6.2, there exists a proof P_4 of $l_{2n} \mapsto y_{2n-1}$ in which all cuts are principal cuts such that $cl(P_4) \subseteq cl(P_3)$. Cut this sequent with the extension axiom $y_{2n-1} \mapsto (x_{2n-1} \supset y_{2n-2}) \vee (\neg x_{2n-1} \supset y_{2n-2})$ to get $l_{2n} \mapsto (x_{2n-1} \supset y_{2n-2}) \vee (\neg x_{2n-1} \supset y_{2n-2})$. Call this proof P_5. Since $cl(P_4) \subseteq cl(P_3) \subseteq cl(P_2) \subseteq cl(P_1)$, we know that $cl(P_4) \subseteq cl(P_1)$. Furthermore, since P_1 contains every extension axiom, $cl(P_5) \subseteq cl(P_1)$.

Now simply repeat this process. After another iteration, we will have a proof P_6 of $l_{2n}, l_{2n-1} \mapsto y_{2n-2}$ such that $cl(P_6) \subseteq cl(P_1)$. We can repeat this n times in total for all of the $\exists$ extension axioms to get a proof P_7 of $l_{2n}, l_{2n-1}, ..., l_{n+1} \mapsto y_n$ such that $cl(P_7) \subseteq cl(P_1)$, and all cuts in P_7 are still principal cuts.

We now continue cutting with all of the $\forall$ extension axioms: Cut $l_{2n}, l_{2n-1}, ..., l_{n+1} \mapsto y_n$ with $y_n \mapsto (x_n \vee \neg x_n) \supset y_{n-1}$ to get $l_{2n}, l_{2n-1}, ..., l_{n+1} \mapsto (x_n \vee \neg x_n) \supset y_{n-1}$, and call this proof P_8. Note that all cuts in P_8 are principal cuts, and $cl(P_8) \subseteq cl(P_1)$.

By Lemma 6.2, there exists a proof P_9 of the sequent $l_{2n}, l_{2n-1}, ..., l_{n+1}, (x_n \vee \neg x_n) \mapsto y_{n-1}$ such that $cl(P_9) \subseteq cl(P_1)$.

Repeat this process. After another iteration, we will have a proof P_{10} of $l_n, l_{2n-1}, ..., l_{n+1}, (x_n \vee \neg x_n), (x_{n-1} \vee \neg x_{n-1}) \mapsto y_{n-2}$, where $cl(P_{10}) \subseteq cl(P_1)$. We can repeat this n times in total for all of the $\forall$ extension axioms to get a proof P_{11} of $l_{2n}, l_{2n-1}, ..., l_{n+1}, (x_n \vee \neg x_n), ..., (x_1 \vee \neg x_1) \mapsto y_0$ such that $cl(P_{11}) \subseteq cl(P_1)$.

Finally, cut this sequent with $y_0 \mapsto \neg\neg B_0$, where $B_0 = A(x_1, ...x_n) \vee \neg A(x_{n+1}, ...x_{2n})$. This gives us a proof P_{12} of $l_{2n}, l_{2n-1}, ..., l_{n+1}, (x_n \vee \neg x_n), ..., (x_1 \vee \neg x_1) \mapsto \neg\neg(A(x_1, ...x_n) \vee \neg A(x_{n+1}, ...x_{2n}))$ such that $cl(P_{12}) \subseteq cl(P_1)$.

Since $cl(P_{12}) \subseteq cl(P_1)$, we know that $l_{2n}, l_{2n-1}, ..., l_{n+1}, (x_n \vee \neg x_n), ..., (x_1 \vee \neg x_1) \mapsto \neg\neg(A(x_1, ...x_n) \vee \neg A(x_{n+1}, ...x_{2n})) \in cl(P_1)$. Therefore by Lemma 5.4, since P_1 is a proof of size-$O(N^3)$, there exists a size-$O(N^4)$ DAG-like proof P_{13} of the sequent $l_{2n}, l_{2n-1}, ..., l_{n+1}, (x_n \vee \neg x_n), ..., (x_1 \vee \neg x_1) \mapsto \neg\neg(A(x_1, ...x_n) \vee \neg A(x_{n+1}, ...x_{2n}))$.

Recall that every LJ proof is an LK proof. Therefore P_{13} is a size-$O(N^4)$ DAG-like classical $\mathsf{LK}[\vec{E_S}]$ proof. Restrict P_{13} by setting all of the literals $l_{2n}, l_{2n-1}, ..., l_{n+1}$ to true, and make all obvious simplifications. Since $A(x_1, ...x_n)$ is a tautology, $\neg A(x_{n+1}, ...x_{2n})$ is unsatisfiable. This leaves us with a size-$O(N^4)$ proof P_{14} of the sequent $(x_n \vee \neg x_n), ..., (x_1 \vee \neg x_1) \mapsto \neg\neg A(x_1, ...x_n)$. Since $\mapsto x_i \vee \neg x_i$ is essentially an LK axiom for each i, we can simply introduce n such axioms, one for each i, and repeatedly cut with $(x_n \vee \neg x_n), ..., (x_1 \vee \neg x_1) \mapsto \neg\neg A(x_1, ...x_n)$ to get an $O(N^4)$ proof P_{15} of the sequent $\mapsto \neg\neg A(x_1, ...x_n)$.

Finally, remove the double negation as follows:

$$\dfrac{\dfrac{\vdots\,P_{15}}{\mapsto \neg\neg A(x_1, ...x_n)} \qquad \dfrac{\dfrac{\dfrac{\vdots\,P_x}{A(x_1, ...x_n) \mapsto A(x_1, ...x_n)}}{\mapsto A(x_1, ...x_n), \neg A(x_1, ...x_n)}\,\neg\text{-}Right}{\neg\neg A(x_1, ...x_n) \mapsto A(x_1, ...x_n)}\,\neg\text{-}Left}{\mapsto A(x_1, ...x_n)}\,Cut$$

It is easy to show that any sequent of the form $A(x_1, ...x_n) \mapsto A(x_1, ...x_n)$ has a size-$O(n)$ DAG-like LK proof P_x. Since n is $O(N)$, $\mapsto A(x_1, ...x_n)$ has a classical size-$O(N^4)$ DAG-like $\mathsf{LK}[\vec{E_S}]$ proof, where $A(x_1, ...x_n) = F_{Prop}$ is any arbitrary classical tautology on n variables, as required.

□

8 Immediate Implications

Our main result implies that if Statman's translation maps trivial instances of the law of excluded middle in QBF to IPL instances which have polynomially-bounded proof complexity, then $\mathsf{LK}[\vec{E_S}]$ is a super proof system. This leads to some immediate corollaries:

8.1 Conditional Separation Between Classical & Intuitionistic Logic

The most important implication of our main result is that under the assumption that $\mathsf{LK}[\vec{E_S}]$ is not super, there is a superpolynomial separation between $\mathsf{LK}[\vec{E_S}]$ and $\mathsf{LJ}[\vec{E_S}]$:

Corollary 8.1. *The F_{IPL} formulas have size-$O(n)$ $\mathsf{LK}[\vec{E_S}]$ proofs, where n is the number of distinct variables in F_{IPL}, but unless $\mathsf{LK}[\vec{E_S}]$ is super, there are no $\mathsf{LJ}[\vec{E_S}]$ proofs of F_{IPL} with polynomial upper bounds.*

Proof: Our main result shows that unless $\mathsf{LK}[\vec{E_S}]$ is super, there are no $\mathsf{LJ}[\vec{E_S}]$ proofs of F_{IPL} with polynomial upper bounds. All that remains to be shown is the polynomial $\mathsf{LK}[\vec{E_S}]$ upper bounds. The proof is straightforward, so we shall omit the details; all that is required is to take the size-$O(n)$ LK proof P_x from Theorem 7.1. This gives us $A(x_1, ...x_n) \mapsto A(x_1, ...x_n)$. A few simple steps gives us $\mapsto \neg\neg B_0$, and from there we can easily build F_{IPL} in another $O(n)$ steps by doing little more than repeatedly cutting with extension axioms to produce $\mapsto y_{2n}$, and then repeatedly applying $Weaken\text{-}Left$ and $\supset\text{-}Right$. Since F_{IPL} is valid both Classically and Intuitionistically, we have a conditional separation between $\mathsf{LK}[\vec{E_S}]$ and $\mathsf{LJ}[\vec{E_S}]$.

□

An interesting point of note is how powerful $\mathsf{LJ}[\vec{E_S}]$ really is. LJ and LK are very similar, and $\mathsf{LJ}[\vec{E_S}]$ is strictly stronger than LJ. But LK with cut is p-equivalent to any Frege system, and $\mathsf{LJ}[\vec{E_S}]$ has cut. All of these things together suggest that it is a fairly powerful system.

8.2 Dangerous Reductions

The second implication of our main result is that Statman's reduction is 'dangerous' in the sense that unless $\mathsf{LK}[\vec{E_S}]$ is super, it translates some trivial formulas from QBF to very difficult formulas in IPL. In order to formalize this notion, we take the formal definitions of what constitute dangerous and safe proof complexity reductions from [HH06]. Note that $\mathsf{LJ}[\vec{E_S}]$ is a strictly stronger proof system than LJ.

Definition 8.2. *Let α be a proof system for a language L_1, let β be a proof system for a language L_2, and let $R : L_1 \to L_2$ be a reduction from L_1 to L_2. If there exists some family of strings $X = \{x_1, x_2, ...\}$, $X \subseteq L_1$ such that for all k and for all $x_i \in X$ there exists an α-proof P_1 of x_i, but there exists no β-proof P_2 of $R(x_i)$ such that $|P_2| < |P_1|^k$, then we say that the reduction R is (α, β)-Explosive.*

Applying this definition to our main result, we get the following Corollary:

Corollary 8.3. *Unless $\mathsf{LK}[\vec{E_S}]$ is super, Statman's reduction is (α,$\mathsf{LJ}[\vec{E_S}]$)-Explosive for every proof system α for QBF which has polynomially-bounded proofs for every prenex instance of the law of excluded middle.*

The complement to the above concept of explosive reductions is that of a stable reduction:

Definition 8.4. *Let α, β, L_1, L_2, and R be as in Definition 8.2. If there exists a k such that for all $x \in L_1$ and any α-proof P_1 of x there exists a β-proof P_2 of $R(x)$ where $|P_2| < |P_1|^k$, then we say that R is (α, β)-Safe.*

Corollary 8.5. *If Statman's reduction is* $(\alpha,\mathsf{LJ}[\vec{E_S}])$*-Safe for some QBF proof system* α *which has polynomially-bounded proofs for every prenex instance of the law of excluded middle, then* $\mathsf{LK}[\vec{E_S}]$ *is a super proof system.*

9 Open Problems

This work has highlighted a few interesting open problems:

The first open problem is to prove superpolynomial LJ lower bounds which do not depend on any assumptions.

This may or may not solve the second open problem, namely showing a superpolynomial or even exponential separation between LK and LJ which does not depend on any assumptions. Another way of phrasing this is as follows: Statman's translation is probably explosive, but perhaps some other safe translation is possible. Does a safe translation exist, or are all reductions from QBF to IPL explosive? Since IPL is believed to be weaker than QBF, it is reasonable to conjecture that no safe translation exists. However, if a safe translation does exist, then it almost certainly isn't as natural or intuitive as Statman's.

The third open problem has to do with the reduction in the reverse direction: Does a safe translation from IPL to QBF exist? Finding such a translation appears to be difficult, but it is reasonable to conjecture that the answer is yes.

References

[BM99] S. R. Buss and G. Mints. The Complexity of the Disjunction and Existential Properties in Intuitionistic Logic. *Annals of Pure and Applied Logic*, 99:93 – 104, 1999.

[BP01] S. R. Buss and P. Pudlák. On The Computational Content of Intuitionistic Propositional Proofs. *Annals of Pure and Applied Logic*, 109:49 – 64, 2001.

[HH06] A. Hertel and P. Hertel. Formalizing Dangerous Reductions. 2006.

[Pud99] P. Pudlák. On The Complexity of Propositional Calculus, Sets and Proofs. In *Logic Colloquium '97*, pages 197 – 218. Cambridge University Press, 1999.

[Sch89] U. Schöning. *Logic For Computer Scientists.* Birkhäuser, Berlin, 1989.

[Sta79] R. Statman. Intuitionistic Propositional Logic is Polynomial-Space Complete. *Theoretical Computer Science*, 9:67 – 72, 1979.

[Sza69] M.E. Szabo. *The Collected Papers of Gerhard Gentzen.* North-Holland Publishing Company, Amsterdam, 1969.