

# Cut-elimination for Provability Logic GL

Rajeev Goré and Revantha Ramanayake

Computer Sciences Laboratory  
The Australian National University  
{ Rajeev.Gore , revantha }@rsise.anu.edu.au

**Abstract.** In 1983, Valentini presented a syntactic proof of cut-elimination for a sequent calculus *GLS* for the provability logic *GL*. The sequents in *GLS* were built from sets, as opposed to multisets, thus permitting implicit contractions. Recently it has been claimed that Valentini’s transformations to eliminate cut do not terminate when applied to a multiset formulation of *GLS* with a rule of contraction. However it appears that the algorithm used to show non-termination is *not* a faithful representation of Valentini’s arguments. Here we show how Valentini’s transformations can in fact be applied in a multiset setting. As usual, it is the case of contractions above cut that requires special care, especially when the cut-formula is boxed. We deal with this instance using a transformation based on Valentini’s original arguments.

## 1 Introduction

The provability logic *GL* is obtained by adding Löb’s axiom  $\Box(\Box A \supset A) \supset \Box A$  to the standard Hilbert Calculus for propositional normal modal logic *K* [9]. Interpreting the modal operator  $\Box A$  as the provability predicate “*A* is provable in Peano arithmetic”, it can be shown that *GL* is complete with respect to the formal provability interpretation in Peano arithmetic (see [11]). Thus the provability logic *GL* is often regarded with special interest.

In 1981, Leivant [3] proposed a syntactic proof of cut-elimination for *GL*. Valentini [13] soon described a counter-example to this proof, proposing a more complicated proof for the sequent calculus *GLS* for *GL* shown in Table 1. Borga [1] presented another solution, while Sasaki [10] described a proof of cut-elimination for a sequent calculus very similar to *GLS* but relying on cut-elimination for *K4*. Note that only Leivant and Valentini use Gentzen-style methods. All four authors used sequents  $X \Rightarrow Y$  built from the *sets* *X* and *Y*. The significance of this is that these calculi did not require a rule of contraction as there is no notion of a set containing an element multiple times (unlike a multiset where the number of occurrences is important). Thus the following instance of the  $L\wedge$  rule is legal in *GLS* even though it ‘hides’ a contraction on  $P \wedge Q$ :

$$\frac{P \wedge Q, P \Rightarrow R}{P \wedge Q \Rightarrow R} L\wedge$$

From a syntactic point of view, it is more satisfying and formal to explicitly identify the contractions that are ‘hidden’ in these set-based proofs of cut-elimination. The appropriate formalization to understand the reliance on contraction is to use multisets. This may well be expected to cause problems – a

Initial sequents:  $\perp \Rightarrow \perp$   $A \Rightarrow A$  for each formula  $A$

Logical rules:

$$\begin{array}{c}
\frac{X \Rightarrow Y, A}{X, \neg A \Rightarrow Y} L\neg \\
\frac{A_i, X \Rightarrow Y}{A_1 \wedge A_2, X \Rightarrow Y} L\wedge \\
\frac{A_1, X \Rightarrow Y \quad A_2, X \Rightarrow Y}{A_1 \vee A_2, X \Rightarrow Y} L\vee \\
\frac{X \Rightarrow Y, A \quad B, U \Rightarrow W}{A \supset B, X, U \Rightarrow Y, W} L\supset
\end{array}
\qquad
\begin{array}{c}
\frac{A, X \Rightarrow Y}{X \Rightarrow Y, \neg A} R\neg \\
\frac{X \Rightarrow Y, A_1 \quad X \Rightarrow Y, A_2}{X \Rightarrow Y, A_1 \wedge A_2} R\wedge \\
\frac{X \Rightarrow Y, A_i}{X \Rightarrow Y, A_1 \vee A_2} R\vee \\
\frac{A, X \Rightarrow Y, B}{X \Rightarrow Y, A \supset B} R\supset
\end{array}$$

Modal rule:

$$\frac{\Box X, X, \Box A \Rightarrow A}{\Box X \Rightarrow \Box A} GLR_v$$

Structural rules:

$$\begin{array}{c}
\frac{X \Rightarrow Y}{A, X \Rightarrow Y} LW \\
\frac{X \Rightarrow Y}{X \Rightarrow Y, A} RW \\
\frac{X \Rightarrow Y, D \quad D, U \Rightarrow W}{X, U \Rightarrow Y, W} cut
\end{array}$$

**Table 1.** Valentini's sequent calculus  $GLS$  for  $GL$

good example is how Gentzen [2] was forced to introduce a ‘multicut’ rule to deal with contractions above cut in his original proof of the *Hauptsatz*.

Moen [5] has recently attempted to transform Valentini's set-based arguments to the multiset formulation of  $GLS$  with an explicit rule of contraction. However, Moen claims that Valentini's algorithm does not always terminate in this new setting. More specifically, using the multiset formulation of  $GLS$ , he gives a concrete derivation  $\epsilon$  containing cut, and claims that the multiset formulation of Valentini's cut-elimination algorithm does not terminate when applied to  $\epsilon$ . This claim has resulted in much confusion over whether syntactic cut-elimination for  $GL$  holds in a Gentzen-style multiset setting.

**There is no *a priori* reason why Valentini's arguments should succeed using sequents built from sets, but fail when using sequents built from multisets, so Moen's claim and Valentini's cut-elimination proof cannot both be correct.**

In response, Negri [6] and Mints [4] proposed two different solutions. Negri uses a new non-standard multiset sequent calculus in which sequents are built from multisets of labelled formulae of the form  $x : A$ , where  $A$  is a traditional formula and  $x$  is an explicit name for a Kripke world. She shows that contraction is height-preserving admissible in this calculus and uses this to handle contractions above cut in her cut-elimination argument. In our view, the use of semantic information in the calculus deviates from a purely proof theoretic approach. Mints [4] solves the problem using a sequent calculus similar to the multiset-formulation of  $GLS$ , but does not state how to handle contractions above cut.

Our contribution is as follows: We have successfully transformed Valentini's set-based arguments for cut-elimination to a multiset formulation of  $GLS$ . We thus demonstrate that no new difficulties arise from Valentini's algorithm *per*

se. However, to handle contractions above cut, care must be taken when the cut-formula in either premise is principal by a contraction rule. We use von Plato's arguments [8] for formulae that are not boxed, but a new argument is required when the cut-formula is boxed and is presented in Lemma 4. We also believe that we have identified the mistake in Moen's claim that Valentini's arguments (in a multiset setting) do not terminate. It appears that Moen has not used a faithful representation of Valentini's arguments for the inductive case, but instead a transformation he titles Val-II(core). This transformation is not mentioned in [13], but is already known to be insufficient [9]. We discuss this further in Section 5 where we eliminate cut from Moen's problematic example  $\epsilon$  using our multiset reductions.

Finally, we remark that it is trivial to show that the cut-rule can be eliminated in  $GLS$  (for both set and multiset formulations) by proving that  $GLS$  without the cut-rule is sound and complete for the Kripke Semantics. Our purpose here however is to resolve the claim about the failure of Valentini's arguments in a multiset formulation of  $GLS$ .

## 2 Preliminaries

We use the letters  $P, Q, \dots$  to denote propositional variables. Formulae are defined in the usual way in terms of propositional variables, the logical constant  $\perp$  and the logical connectives  $\wedge$  (conjunction),  $\vee$  (disjunction),  $\supset$  (implication) and  $\Box$  (necessity, or in this context, provability). Formulae are denoted by  $A, B, \dots, E$ . Multisets of formulae are denoted by  $X, Y, U, V, W, G$ . A *sequent* is a multiset tuple  $(X, Y)$  and is written  $X \Rightarrow Y$ . A formula  $A$  is said to be *boxed* if it is of the form  $\Box B$  for some formula  $B$  and is *not boxed* otherwise. The relation ' $\equiv$ ' is used to denote syntactic equality. The notation  $A \in X$  denotes that at least one occurrence of the formula  $A$  occurs in the multiset  $X$ . Suppose the multiset  $X = [A_1, \dots, A_n]$ , then we define the multiset  $\Box X = [\Box A_1, \dots, \Box A_n]$ .

The sequent calculus we use here is called  $GLS^m$  (the superscript  $m$  is for multiset). The initial sequents, logical rules, structural rules and modal rule are identical in form to  $GLS$  (Table 1) except the  $X, Y, U, W$  now represent multisets. In addition,  $GLS^m$  also contains the following structural rules of contraction:

$$\frac{A, A, X \Rightarrow Y}{A, X \Rightarrow Y} LC \qquad \frac{X \Rightarrow Y, A, A}{X \Rightarrow Y, A} RC$$

A *pre-derivation* (in  $GLS^m$ ) is inductively defined as an arbitrary sequent or an application of a logical, modal or structural rule to pre-derivations concluding its premises. Viewing a pre-derivation as a tree, we call the root of the tree the *end-sequent* of the derivation. The leaves of the tree are called the *top-sequents*. A *derivation* (in  $GLS^m$ ) is a pre-derivation where every top-sequent is an initial sequent. This conforms to the standard definition.

For the initial sequents and logical rules in Table 1, and the structural rules  $LC$  and  $RC$ , the multisets  $X, Y$  (and also  $U, W$  for  $L\supset$ ) are called the *context*. In

the conclusion of each of these rules, the formula occurrence not in the context is called the *principal formula*. This follows standard practice (see [12]). Any formula in the antecedent or succedent of the conclusion of a  $GLR_v$  rule application is considered to be a principal formula. The  $\Box B$  in the succedent of the conclusion of  $GLR_v$  rule is called the *diagonal formula* (and is of course boxed).

A binary rule where the context in both premises is required to be identical is called an *additive* binary rule (examples:  $L\vee, R\wedge$ ). A binary rule where the context in each premise need not be identical is called a *multiplicative* binary rule (examples:  $L\supset, cut$ ). The term context-sharing (context-independent) is also used to refer to an additive (multiplicative) rule (see [12]).

Let  $\rho$  be a rule in  $GLS^m$ . We write  $\rho(A)$  to identify the formula in the conclusion of  $\rho$  that is a principal formula by the rule. The notation  $\rho^n$  denotes  $n$  applications of the rule  $\rho$ , and  $\rho^*$  denotes some positive number of applications of the rule  $\rho$ . By a slight abuse of notation,  $\rho^*(X)$  will be used to denote repeated applications of  $\rho$  that make each formula occurrence (including multiple formula occurrences) in the multiset  $X$  a principal formula. We will occasionally write  $\rho_i$  (where  $i = \alpha, \beta$  or some natural number), instead of  $\rho$ , to denote a *particular application* of the rule  $\rho$  in some derivation in this paper, especially when many applications of the rule  $\rho$  occur in the displayed derivation. This is purely to enable us to conveniently refer to that particular  $\rho$  application in the text.

The notation  $(A)^m$  is used to denote the multiset  $\overbrace{[A, \dots, A]}^{m \text{ times}}$ . We will sometimes omit the parenthesis for the sake of clarity. We write  $\tau'/X \Rightarrow Y$  ( $\tau' \quad \tau''/X \Rightarrow Y$ ) to denote the derivation  $\frac{\tau'}{X \Rightarrow Y} \left( \frac{\tau' \quad \tau''}{X \Rightarrow Y} \right)$ .

**Definition 1.** The antecedent operator  $\oplus_a : \text{pre-derivation} \times \text{multiset} \mapsto \text{pre-derivation}$  is defined recursively as follows:

1.  $(X \Rightarrow Y) \oplus_a G = X, G \Rightarrow Y$
2.  $(\tau'/X \Rightarrow Y) \oplus_a G = \tau' \oplus_a G/X, G \Rightarrow Y$
3.  $(\tau' \quad \tau''/X \Rightarrow Y) \oplus_a G = \tau' \oplus_a G \quad \tau'' \oplus_a G'/X, G \Rightarrow Y$  where  $\tau', \tau''$  represent the premise derivations of a binary rule and  $G' = G$  ( $G' = \emptyset$ ) if the rule is additive (multiplicative).

Thus, in case 3., if the rule is multiplicative binary then the right premise derivation  $\tau'' \oplus_a G'$  is the original  $\tau''$  because  $G'$  is the empty multiset. Also, in general  $\mathcal{T} \oplus_a G$  is not a derivation.

The height of a derivation is defined in the standard manner as below.

**Definition 2.** The height  $s(\tau)$  of a derivation  $\tau$  is defined recursively as follows:

1.  $s(X \Rightarrow Y) = 1$  where  $X \Rightarrow Y$  is an initial sequent or  $\perp \Rightarrow \perp$ ,
2.  $s(\tau'/X \Rightarrow Y) = s(\tau') + 1$  if the end-sequent is obtained by a unary rule,
3.  $s(\tau' \quad \tau''/X \Rightarrow Y) = \max(s(\tau'), s(\tau'')) + 1$  if the end-sequent is obtained by a binary rule.

The notation  $X \stackrel{k}{\Rightarrow} Y$  denotes a derivation of height  $k$  with end-sequent  $X \Rightarrow Y$ . Where the height is not required, with a slight abuse of notation,  $X \Rightarrow Y$  will denote a derivation with end-sequent  $X \Rightarrow Y$ . In the context it will be clear if  $X \Rightarrow Y$  refers to a sequent or a derivation.

**Definition 3.** The degree  $d(A)$  of a formula  $A$  is defined recursively as follows:

1.  $d(P) = 1$  for a propositional variable  $P$ ,
2.  $d(\perp) = 1$ ,
3.  $d(A \wedge B) = d(A \vee B) = d(A \supset B) = d(A) + d(B) + 1$ ,
4.  $d(\neg A) = d(\Box A) = d(A) + 1$ .

**Definition 4.** The cut-height of an instance of the cut-rule with premise derivations  $\tau_1$  and  $\tau_2$  is the sum of the heights  $s(\tau_1)$  and  $s(\tau_2)$  of the premise derivations

Let  $\tau$  be a derivation containing the sequent  $X \Rightarrow Y$  and let  $J$  be an application of the  $GLR_v$  rule above  $X \Rightarrow Y$ . Following Valentini [13] we say that  $J$  is  $n$ -ary  $GLR_v$  over  $X \Rightarrow Y$  if the branch from  $X \Rightarrow Y$  up to the conclusion of  $J$  contains exactly  $n - 1$  applications of  $GLR_v$ . Let  $G_2(D)$  be the set of all the applications  $J$  of the  $GLR_v$  rule such that:

- (Con1): Application  $J$  is 2-ary  $GLR_v$  over the sequent occurrence  $X \Rightarrow Y, D$
- (Con2): Formula  $D$  is the diagonal formula of the 1-ary  $GLR_v$  rule over  $X \Rightarrow Y, D$  below  $J$
- (Con3): Formula  $D$  is not introduced via weakening in any antecedent below the application  $J$ .

**Definition 5.** Let  $X \Rightarrow Y, D$  be a sequent occurrence in some derivation  $\tau$ . The width  $n(D; X \Rightarrow Y, D; \tau)$  of the formula  $D$  in that sequent occurrence is  $|G_2(D)|$ .

When it is clear which derivation and sequent occurrence we refer to, we will simply write  $n(D)$ .

$$\begin{array}{c}
 \text{Example 1.} \quad \frac{\frac{\frac{\Box C, C, \Box \Box B, \Box B, \Box B \Rightarrow B}{\Box C, \Box \Box B \Rightarrow \Box B} GLR_{v1} \quad \frac{\frac{\Box D, D, \Box B \Rightarrow B}{\Box D \Rightarrow \Box B} GLR_{v2}}{\Box D, \Box \Box B \Rightarrow \Box B} LW_{\alpha}(\Box \Box B)}{\Box C \vee \Box D, \Box \Box B \Rightarrow \Box B} LV \\
 \frac{\frac{\Box(\Box C \vee \Box D), \Box C \vee \Box D, \Box \Box B \Rightarrow \Box B}{\Box(\Box C \vee \Box D) \Rightarrow \Box \Box B} LW(\Box(\Box C \vee \Box D))}{\Box(\Box C \vee \Box D) \Rightarrow \Box \Box B} GLR_{v0}
 \end{array}$$

Both  $GLR_{v1}$  and  $GLR_{v2}$  are 2-ary over the root  $\Box(\Box C \vee \Box D) \Rightarrow \Box \Box B$ . Thus they both satisfy (Con1). There is only one 1-ary  $GLR_v$  rule application over the root, namely  $GLR_{v0}$ . The diagonal formula is  $\Box \Box B$  so (Con2) is satisfied for both  $GLR_{v1}$  and  $GLR_{v2}$ . Observe that  $\Box \Box B$  is introduced *below*  $GLR_{v2}$  by  $LW_{\alpha}(\Box \Box B)$ . So  $GLR_{v1}$  satisfies (Con3) but  $GLR_{v2}$  does not satisfy (Con3). The width  $n(\Box \Box B) = |G_2(\Box \Box B)| = |\{GLR_{v1}\}| = 1$ .

Let  $cut$  and  $cut'$  be two instances of the cut-rule, with degree, width and cut-height given by  $(d, n, h)$  and  $(d', n', h')$  respectively. Then we say that  $cut'$  is *well-behaved* compared to  $cut$  if  $d'\omega^2 + n'\omega + h' \leq d\omega^2 + n\omega + h$ . That is,

$$\begin{aligned}
 h' > h &\text{ implies } (n' < n \text{ or } d' < d); \text{ and} \\
 n' > n &\text{ implies } d' < d; \text{ and} \\
 d' &\leq d.
 \end{aligned}$$

### 3 Properties of $GLS^m$

Unlike in [8] the logical rules that we use here are *not* invertible. Nevertheless Lemma 1 corresponds precisely to Lemma 3 of von Plato [8].

**Lemma 1.** *For all  $m \geq 0$ ,*

- (i) *If  $(\neg A)^{m+1}, X \Rightarrow Y$  is derivable, then  $X \Rightarrow Y, A^{m+1}$  is derivable.*
- (ii) *If  $X \Rightarrow Y, (\neg A)^{m+1}$  is derivable, then  $A^{m+1}, X \Rightarrow Y$  is derivable.*
- (iii) *If  $(A \wedge B)^{m+1}, X \Rightarrow Y$  is derivable, then  $A^{m+1}, B^{m+1}, X \Rightarrow Y$  is derivable.*
- (iv) *If  $X \Rightarrow Y, (A \wedge B)^{m+1}$  is derivable, then  $X \Rightarrow Y, A^{m+1}$  and  $X \Rightarrow Y, B^{m+1}$  are derivable.*
- (v) *If  $(A \vee B)^{m+1}, X \Rightarrow Y$  is derivable, then  $A^{m+1}, X \Rightarrow Y$  and  $B^{m+1}, X \Rightarrow Y$  are derivable.*
- (vi) *If  $X \Rightarrow Y, (A \vee B)^{m+1}$  is derivable, then  $X \Rightarrow Y, A^{m+1}, B^{m+1}$  is derivable.*
- (vii) *If  $(A \supset B)^{m+1}, X \Rightarrow Y$  is derivable, then  $B^{m+1}, X \Rightarrow Y$  is derivable.*
- (viii) *If  $X \Rightarrow Y, (A \supset B)^{m+1}$  is derivable, then  $A^{m+1}, X \Rightarrow Y, B^{m+1}$  is derivable.*

*Proof.* Induction on the height of derivation. ⊥

We will only require Lemma 1 for the instance when  $m = 0$ . The general statement is necessary in order to obtain a sufficiently strong induction hypothesis.

### 4 Cut-elimination for multiset sequent calculus $GLS^m$

The proof of cut-elimination is very similar to the classical case when the cut-formula is not boxed. The main case for  $GLS^m$  is when the cut-formula is principal by the  $GLR_v$  rule in both the left and right premises of cut. Such a derivation is said to be in Sambin Normal Form (SNF) — see Fig. 1. From now on, we shall only consider derivations in SNF with cut-free premises. In Section 2, we defined

$$\frac{\frac{\Pi}{\frac{\square X, X, \square B \stackrel{k}{\Rightarrow} B}{\square X \stackrel{k \pm 1}{\Rightarrow} \square B} GLR_{v1}}{\square X, \square U \Rightarrow \square D} \quad \frac{\frac{\Omega}{\frac{\square B, B, \square U, U, \square D \stackrel{l}{\Rightarrow} D}{\square B, \square U \stackrel{l \pm 1}{\Rightarrow} \square D} GLR_{v2}}{\square X, \square U \Rightarrow \square D} cut(\square B)}$$

**Fig. 1.** The Sambin Normal Form (SNF) for  $GLS^m$

width in similar terms to Valentini’s original definition in order to establish conformity with that paper. However, it will be helpful to formulate the conditions (Con1), (Con2) and (Con3) in terms of the  $\square B$ -trace defined below.

**Definition 6.** ( *$\square B$ -trace*) Let  $\tau$  be a derivation with end-sequent  $\square B, U \Rightarrow W$ . Trace up the sequents from the end-sequent  $\square B, U \Rightarrow W$  the formula occurrence  $\square B$ . If a contraction rule  $LC(\square B)$  is encountered trace up from the premise of the contraction rule both formula occurrences. In the case of a binary rule, trace up from both the premises the formula occurrence  $\square B$ .

The trace is terminated at a termination sequent where

- (type 1) an initial sequent  $\Box B \Rightarrow \Box B$  is encountered, or
- (type 2) a sequent that is the conclusion of the weakening rule  $LW(\Box B)$  is encountered, or
- (type 3) a sequent that is the conclusion of the  $GLR_v$  rule is encountered.

Such a trace is called a  $\Box B$ -trace (of  $\Box B, U \Rightarrow W$ ).

If a  $\Box B$ -trace terminates in (type 1), we refer to that initial sequent  $\Box B \Rightarrow \Box B$  as a  $\Box B$ -traced initial sequent. If a  $\Box B$ -trace terminates in (type 2) then we call the application of the weakening rule a  $\Box B$ -traced weakening rule. Finally, if the  $\Box B$ -trace terminates in (type 3) we call the application of the  $GLR_v$  rule a  $\Box B$ -traced  $GLR_v$  rule.

By inspection we note that a  $\Box B$ -trace will always terminate. Because of the contraction rule and the binary logical rules, a  $\Box B$ -trace may have several termination sequents, where each of the termination sequents is of type 1,2 or 3.

*Remark 1.* The antecedent of every sequent in the  $\Box B$ -trace contains  $\Box B$ .

*Remark 2.* For a derivation in SNF with cut-formula  $\Box B$  (Fig. 1),  $n(\Box B) = 0$  if and only if the  $\Box B$ -trace of the premise of  $GLR_{v1}$  has termination sequents of type 1 or type 2 only.

The formula occurrence  $\Box B$  in the antecedent of each termination sequent is called a *parametric ancestor* of the  $\Box B$  in the antecedent of  $\Box B, U \Rightarrow W$ .

We now define the  $\Box B$ -trace\* which is identical to the  $\Box B$ -trace except that if a multiplicative binary rule is encountered in the trace up, only one premise containing  $\Box B$  in the antecedent is traced.

**Definition 7.** ( $\Box B$ -trace\*) Suppose a derivation  $\tau$  contains an application of the binary multiplicative rule  $L\supset$  where the antecedent of the conclusion of the rule contains two or more occurrences of  $\Box B$ :

$$\frac{E, U_1' \Rightarrow W_1' \quad U_2' \Rightarrow W_2', D}{D \supset E, \Box B, \Box B, U_1, U_2 \Rightarrow W_1, W_2} L\supset$$

If  $\Box B \in U_1'$  the  $\Box B$ -trace\* traces up from the left premise of  $L\supset$  a single occurrence of  $\Box B$ . If  $\Box B \notin U_1'$  then it must be the case that  $\Box B \in U_2'$ . In this case  $\Box B$ -trace\* traces up from the right premise of  $L\supset$  a single occurrence of  $\Box B$ . With respect to the other rules,  $\Box B$ -trace\* is identical to  $\Box B$ -trace. The termination sequents are defined in the same manner as for  $\Box B$ -trace.

The terms  $\Box B$ -trace\* initial sequent,  $\Box B$ -trace\* weakening rule and  $\Box B$ -trace\*  $GLR_v$  rule are defined analogously to the definitions for  $\Box B$ -trace.

Note that Remark 1 is true for  $\Box B$ -trace\* if we replace the term ' $\Box B$ -trace' with ' $\Box B$ -trace\*'. However Remark 2 is no longer true if we replace the term ' $\Box B$ -trace' with ' $\Box B$ -trace\*', as there exist derivations such that the  $\Box B$ -trace has a termination sequent of type 3 (so  $n(\Box B) > 0$ ) and yet every termination sequent of a  $\Box B$ -trace\* is of type 1 or type 2.

The proof that follows is essentially Valentini's [13], adapted to  $GLS^m$ .

**Lemma 2.** *Let  $\tau$  be a derivation in  $GLS^m$  in SNF (Fig. 1), where the cut is of cut-height  $h$ , the cut-formula of degree  $d$  and width  $n$  and the premises are cut-free. Then  $\tau$  can be transformed to a derivation  $\tau'$  with identical end-sequent containing several cuts, where each introduced cut has either cut-height  $< h$  or degree  $< d$  or width  $< n$ . Also, each introduced cut is well-behaved compared to the original cut.*

*Proof.* Consider the width of the cut-formula in the conclusion of  $GLR_{v1}$ .

First suppose that  $n(\Box B) = 0$ . The idea is to locate the parametric ancestors of the  $\Box B$  in the antecedent of  $\Box X, X, \Box B \Rightarrow B$  (the left premise of  $GLR_{v1}$ ), and replace each parametric ancestor with  $\Box X$ . By Remark 2 we know that the parametric ancestors must be introduced by an initial sequent  $\Box B \Rightarrow \Box B$  or the weakening rule  $LW(\Box B)$ .

Tracing upwards from the end-sequent, if we encounter a  $L\supset$  rule application before the parametric ancestor, we must replace a  $\Box B$  formula occurrence with  $\Box X$  in one premise only. For this we use the  $\Box B$ -trace\* of  $\Box X, X, \Box B \Rightarrow B$ .

Consider the  $\Box B$ -trace\* of the left premise  $\Box X, X, \Box B \Rightarrow B$  of  $GLR_{v1}$ , in Fig. 1. Let  $\Pi'$  be the derivation obtained by replacing every  $\Box B$ -traced\* initial sequent with the derivation of  $\Box X \Rightarrow \Box B$  (obtained from  $\tau$ ). We also replace any  $\Box B$ -traced\* weakening rules with  $LW^*(\Box X)$ . By the latter we mean

$$\frac{X' \Rightarrow Y'}{\Box B, X' \Rightarrow Y'} LW(\Box B) \quad \text{replaced by} \quad \frac{X' \Rightarrow Y'}{\Box X, X' \Rightarrow Y'} LW^*(\Box X)$$

Finally, it is necessary to replace any  $LC(\Box B)$  rule applications that occur along the  $\Box B$ -trace\* with the rule application  $LC^*(\Box X)$ . We can thus obtain a cut-free derivation  $\Pi' / \Box X, \Box X, X \Rightarrow B$ .

Valentini makes a mistake in his proof by *deleting* the  $LW(\Box B)$  rule instead of replacing it with  $LW^*(\Box X)$  as we do. It is necessary to weaken with  $\Box X$  because we replace initial sequents  $\Box B \Rightarrow \Box B$  with derivations  $\Box X \Rightarrow \Box B$ . Otherwise an additive binary rule with one premise derivation containing a  $\Box B$ -traced\* initial sequent and the other premise derivation containing a  $\Box B$ -traced\* weakening rule will no longer transform to a (legal) derivation.

We now obtain the required derivation as shown in Fig 2.

$$\frac{\frac{\frac{\Pi'}{\Box X, \Box X, X \Rightarrow B} LC^*(\Box X)}{\Box X, X \Rightarrow B} \quad \frac{\frac{\frac{\frac{\Pi}{X, \Box X, \Box B \stackrel{k}{\Rightarrow} B} GLR_v}{\Box X \stackrel{k+1}{\Rightarrow} \Box B} \quad \frac{\Omega}{B, \Box B, U, \Box U, \Box D \stackrel{l}{\Rightarrow} D}}{\Box X, B, \Box U, U, \Box D \Rightarrow D} cut_1(\Box B)}}{\frac{\frac{\frac{\Box X, \Box X, X, \Box U, U, \Box D \Rightarrow D}{\Box X, X, \Box U, U, \Box D \Rightarrow D} LC^*(X)}{\Box X, \Box U \Rightarrow \Box D} GLR_v} cut_2(B)}$$

**Fig. 2.** Valentini's transformations when  $n(\Box B) = 0$ .

Now  $cut_1$  has a smaller cut-height  $(k+1) + l < h = (k+1) + (l+1)$  although the degree and width of  $cut_1$  remain  $d$  and  $n$  respectively. The degree of the

cut-formula for  $cut_2$  is one less than  $d$  although the width and the cut-height may be greater than  $n$  and  $h$  respectively. We observe that  $cut_1$  and  $cut_2$  are both well-behaved compared to the original cut.

Now suppose that  $n(\Box B) > 0$ . Then  $\tau$  must be of the following form, where no  $LW(\Box B)$  rule application occurs in  $\Upsilon$  below  $\Box B, \Box G \Rightarrow \Box A$  and over  $\Box X, X, \Box B \Rightarrow B$  by (Con3):

$$\frac{\frac{\frac{\Pi}{B, \Box B, G, \Box G, \Box A \Rightarrow A}}{\Box B, \Box G \Rightarrow \Box A} GLR_{v3}}{\Upsilon} \quad \frac{\frac{\frac{\Omega}{B, \Box B, U, \Box U, \Box D \Rightarrow D}}{\Box B, \Box U \Rightarrow \Box D} GLR_{v2}}{\Box B, \Box U \Rightarrow \Box D} GLR_{v1}}{\frac{\frac{X, \Box X, \Box B \stackrel{k}{\Rightarrow} B}{\Box X \stackrel{k+1}{\Rightarrow} \Box B} GLR_{v1}}{\Box X, \Box U \stackrel{l+1}{\Rightarrow} \Box D} cut(\Box B)}$$

Formally,  $\Upsilon$  is a pre-derivation. Let  $\Lambda_1$  be the derivation

$$\frac{\frac{\frac{\frac{\Box A \Rightarrow \Box A}{A, \Box A, \Box B, \Box G \Rightarrow \Box A} LW^*(A, \Box B, \Box G)}{\Upsilon \oplus_\alpha(A, \Box A)} A, \Box A, X, \Box X, \Box B \Rightarrow B}{\Box A, \Box X \Rightarrow \Box B} GLR_v \quad \frac{\frac{\frac{\frac{\Pi}{B, \Box B, G, \Box G, \Box A \Rightarrow A}}{\Box B, \Box G \Rightarrow \Box A} GLR_v}{\Upsilon} X, \Box X, \Box B \Rightarrow B}{\Box A, \Box X, \Box X, X \Rightarrow B} cut_1(\Box B)}{\frac{\Box A, \Box X, \Box X, X \Rightarrow B}{\Box A, \Box X, X \Rightarrow B} LC^*(\Box X)}$$

Let  $\Lambda_2$  be the derivation

$$\frac{\frac{\frac{\frac{\Box A \Rightarrow \Box A}{A, \Box A, \Box B, \Box G \Rightarrow \Box A} LW^*(A, \Box B, \Box G)}{\Upsilon \oplus_\alpha(A, \Box A)} A, \Box A, X, \Box X, \Box B \Rightarrow B}{\Box A, \Box X \Rightarrow \Box B} GLR_v \quad \frac{\frac{\frac{\frac{\Pi}{B, \Box B, G, \Box G, \Box A \Rightarrow A}}{\Box B, \Box G \Rightarrow \Box A} GLR_v}{\Upsilon} X, \Box X, \Box B \Rightarrow B}{\Box A, \Box X, G, \Box G, B \Rightarrow A} cut_2(\Box B)}{\frac{\Box A, \Box A, \Box X, G, \Box G, B \Rightarrow A}{\Box A, \Box X, G, \Box G, B \Rightarrow A} LC(\Box A)}$$

Let  $\Lambda_3$  be the derivation

$$\frac{\frac{\frac{\frac{\Pi}{B, \Box B, G, \Box G, \Box A \Rightarrow A}}{\Box B, \Box G \Rightarrow \Box A} GLR_v}{\Upsilon} \quad \frac{\frac{\frac{\Omega}{B, \Box B, U, \Box U, \Box D \Rightarrow D}}{\Box B, \Box U \Rightarrow \Box D} GLR_{v2}}{\Box B, \Box U \Rightarrow \Box D} GLR_{v1}}{\frac{\frac{X, \Box X, \Box B \stackrel{k}{\Rightarrow} B}{\Box X \stackrel{k+1}{\Rightarrow} \Box B} GLR_{v1}}{\Box X, B, U, \Box U, \Box D \Rightarrow D} cut_5(\Box B)}$$

Then we transform  $\tau$  to the derivation  $\tau'$  shown in Fig. 3. For the remainder of this proof,  $cut_i (1 \leq i \leq 6)$  and  $GLR_{vj} (1 \leq j \leq 4)$  will be with respect to the derivations  $\tau$  and  $\tau'$  in the case  $n(\Box B) > 0$ .

Compared to the cut in derivation  $\tau$ ,  $cut_1$ ,  $cut_2$  and  $cut_4$  have smaller width,  $cut_3$  and  $cut_6$  have a smaller degree of cut-formula, and  $cut_5$  has a smaller cut-height by one. In particular,  $cut_1$  and  $cut_2$  have a smaller width by one because the 2-ary rule application  $GLR_{v3}$  in the original derivation has been replaced by an initial sequent in  $\Lambda_1$  and  $\Lambda_2$  (refer Con1). Also  $cut_4$  has a smaller width



*Proof.* See Appendix. ⊣

**Lemma 4.** *In the following derivation both premises of  $cut_\gamma$  are cut-free:*

$$\frac{\frac{\frac{\Box X, X, \Box B \stackrel{k}{\Rightarrow} B}{\Box X \stackrel{k+1}{\Rightarrow} \Box B} GLR_v \quad \frac{\frac{\tau}{(\Box B)^{m+2}, U \Rightarrow W} \rho}{\Box B, U \Rightarrow W} LC^{m+1}(\Box B)}{\Box X, U \Rightarrow W} cut_\gamma(\Box B)$$

*Then there is a derivation of  $\Box X, U \Rightarrow W$  where each cut in this derivation is of lesser degree, width or cut-height, and well-behaved compared to  $cut_\gamma$ .*

*Proof.* We limit ourselves to the case where a  $\Box B$  formula occurrence is principal by  $\rho$  and  $\rho = GLR_v$ . It follows that  $U \equiv \Box V$  and  $W \equiv \Box C$  for some multiset  $V$  and formula  $C$ . Thus the *right premise* is of the form

$$\frac{\frac{\tau'}{(\Box B)^{m+2}, B^{m+2}, \Box V, V, \Box C \stackrel{l}{\Rightarrow} C} \rho = GLR_v}{\frac{(\Box B)^{m+2}, \Box V \stackrel{l+1}{\Rightarrow} \Box C}{\Box B, \Box V \stackrel{l+1+m+1}{\Rightarrow} \Box C} LC^{m+1}(\Box B)}$$

Consider the derivation

$$\frac{\frac{\frac{\frac{\frac{\frac{\Box X, X \Rightarrow B}{\Box X, X, \Box X, \Box V, V, \Box C \Rightarrow C} LC^*(\Box X)}{\Box X, X, \Box V, V, \Box C \Rightarrow C} GLR_v}{\Box X, B^{m+2}, \Box V, V, \Box C \Rightarrow C} LC^{m+1}(B)}{\Box X, B^{m+2}, \Box V, V, \Box C \stackrel{l+m+1}{\Rightarrow} C} cut_\beta(B)}{\frac{\frac{\frac{\frac{\Box X, X \Rightarrow B}{\Box X, X, \Box X, \Box V, V, \Box C \Rightarrow C} LC^*(\Box X)}{\Box X, X, \Box V, V, \Box C \Rightarrow C} GLR_v}{\Box X, B^{m+2}, \Box V, V, \Box C \stackrel{l+m+1}{\Rightarrow} C} cut_\alpha(\Box B)}{\Box X, X \Rightarrow B} LC^{m+1}(\Box B)}$$

The derivation concluding  $\Box X, X \Rightarrow B$  is obtained by Corollary 1. Compared to  $cut_\gamma$ , all cuts in this derivation are seen to be well-behaved and have cut-formulae with lesser degree or width. Also  $cut_\alpha$  has lesser cut-height  $((k+1) + (l+m+1)) < (k+1) + (l+1+m+1)$ , and the degree of cut-formula for  $cut_\beta$  is smaller by 1.

See the Appendix for further details. ⊣

**Theorem 1.** *Cut-elimination holds for  $GLS^m$*

*Proof.* Without loss of generality we will eliminate the topmost cut. Let  $D$  be the cut-formula and  $h$  the cut-height. Induction on  $d(D)\omega^2 + n(n)\omega + h$ . Any introduced cut with lesser degree, width or cut-height that is well-behaved compared to the original cut can be eliminated by the induction hypothesis.

When the cut-formula  $D$  is not boxed, it suffices to use classical arguments and von Plato's [8] argument for avoiding Gentzen's multicut. The Appendix contains the full details. Now suppose that the cut-formula  $D$  is boxed. We can permute the cut with any logical rules appearing above the left and right premises in the standard manner. This is because the cut-formula  $D$  cannot be principal by a logical rule as it is boxed. If the cut-formula has been introduced by weakening in either premise, the end-sequent  $X, U \Rightarrow Y, W$  can be obtained by replacing the  $LW(D)$  or  $RW(D)$  rule with appropriate weakening rules.

The remaining cases to consider are contractions on the cut-formula  $D$  above the 1. left (2. right) premise, and when 3. the derivation is in SNF. Case 1 can be handled by Lemma 3. Case 2 can be handled by Lemma 4. If the derivation is in SNF, then we can use Lemma 2 to obtain the required result.  $\dashv$

## 5 Moen’s Val-II(core) is not Valentini’s reduction

We have carefully examined Moen’s slides titled “The proposed algorithms for eliminating cuts in the provability calculus  $GLS$  do not terminate” [5].

Moen sets out to reduce a cut in SNF (Fig. 1) using the transformation he titles Val-II(core). Moen claims that Val-II(core) is the “. . . core of Valentini’s reduction” [5]. Yet Val-II(core) does not appear at any point in [13]. However it appears in [9, page 322] with the comment “this reduction is not sufficient”.

Thus Moen is incorrect in claiming that he has demonstrated that Valentini’s algorithm does not terminate — he is using the wrong algorithm. In fact, for his concrete derivation  $\epsilon$ , the width of the cut-formula is 1 so a single application of Valentini’s reduction (Fig. 3) is sufficient to reduce it to the base case. Applying the base case transformations, and then the classical transformations, we obtained a cut-free derivation of the end-sequent of  $\epsilon$ .

## References

1. M. Borga. On Some Proof Theoretical Properties of the Modal Logic  $GL$  *Studia Logica*, 42:453–459, 1983.
2. G. Gentzen. The Collected Papers of Gerhard Gentzen, ed. M. Szabo.
3. On the Proof Theory of the Modal Logic for Arithmetic Provability *Journal of Symbolic Logic*, 46:531–538, 1981.
4. G. Mints. Cut elimination for provability logic. *Collegium Logicum* 2005.
5. A. Moen. The proposed algorithms for eliminating cuts in the provability calculus  $GLS$  do not terminate *NWPT 2001*, Norwegian Computing Center, 2001-12-10. <http://publ.nr.no/3411>
6. S. Negri. Proof Analysis in Modal Logic *Journal of Philosophical Logic*, 34:507–544, 2005.
7. S. Negri and J. von Plato. *Structural Proof Theory*. CUP, 2001.
8. J. von Plato. A proof of Gentzen’s *Hauptsatz* without multicut *Archive of Mathematical Logic*, 40:9–18, 2001.
9. G. Sambin and S. Valentini. The Modal Logic of Provability. The Sequential Approach *Journal of Philosophical Logic*, 11:311–342, 1982.
10. K. Sasaki. Löb’s Axiom and Cut-elimination Theorem *Journal of the Nanzan Academic Society Math. Sci. and Information Engineering*, 1:91–98, 2001.
11. R.M. Solovay. Provability Interpretations of Modal Logic *Israel Journal of Mathematics*, 25:287–304, 1976.
12. A.S. Troelsltra and H. Schwichtenberg. *Basic Proof Theory*. CUP, 2000.
13. S. Valentini. The Modal Logic of Provability: Cut-elimination *Journal of Philosophical Logic*, 12:471–476, 1983.

## 6 Appendix

**Lemma 3.** *In the following derivation both premises of the  $cut_\gamma$  are cut-free:*

$$\frac{\frac{\tau}{X \Rightarrow Y, (\Box B)^{m+2}} \rho}{X \Rightarrow Y, \Box B} RC^{m+1}(\Box B) \quad \frac{\Box B, U \stackrel{l}{\Rightarrow} W}{X, U \Rightarrow Y, W} cut_\gamma(\Box B)$$

*Then there is a derivation of  $X, U \Rightarrow Y, W$  where each cut in this derivation is of strictly less cut-height, and well-behaved compared to  $cut_\gamma$ .*

*Proof.* Without loss of generality,  $\rho \neq RC(\Box B)$ . Also it is clear that  $X \Rightarrow Y, (\Box B)^{m+2}$  cannot be an initial sequent or the conclusion of a  $GLR_v$  rule application. Thus, if a  $\Box B$  formula occurrence is principal by  $\rho$ , it must be that case that  $\rho = RW(\Box B)$ . In this case  $\tau$  is of the form  $\tau'/X \stackrel{k}{\Rightarrow} Y, (\Box B)^{m+1}$ . Applying  $RC^m$  and then the cut-rule we obtain a derivation with end-sequent  $X, U \Rightarrow Y, W$  and lesser cut-height ( $m+l < (m+1)+l$ ).

Now suppose that no  $\Box B$  formula occurrence is principal by  $\rho$ .

1. If  $\rho$  is a unary rule, then  $\tau$  is of the form  $\tau'/X' \stackrel{k}{\Rightarrow} Y', (\Box B)^{m+2}$ . Applying  $RC^{m+1}(\Box B)$  to  $\tau$  and then the cut-rule with  $\Box B, U \Rightarrow W$  as the right premise derivation, followed by  $\rho$  we obtain the required derivation.
2. If  $\rho$  is an additive binary rule, then  $\tau$  is of the form

$$\frac{\tau_1 \quad \tau_2}{X \stackrel{\max(k_1, k_2)+1}{\Rightarrow} Y, (\Box B)^{m+2}}$$

where  $\tau_1 := \tau'/X' \stackrel{k_1}{\Rightarrow} Y', (\Box B)^{m+2}$  and  $\tau_2 := \tau''/X'' \stackrel{k_2}{\Rightarrow} Y'', (\Box B)^{m+2}$ . For  $i \in \{1, 2\}$ , apply  $RC^{m+1}(\Box B)$  to  $\tau_i$  and then the cut-rule with  $\Box B, U \stackrel{l}{\Rightarrow} W$  as the right premise derivation. Using these derivations as the premises for an application of rule  $\rho$  we obtain the required derivation.

3. If  $\rho$  is the multiplicative binary rule  $L\supset$ , then  $\tau$  is of the form

$$\frac{\frac{\tau'}{X' \stackrel{k_1}{\Rightarrow} Y', C, (\Box B)^s} \quad \frac{\tau''}{D, X'' \stackrel{k_2}{\Rightarrow} Y'', (\Box B)^t}}{C \supset D, X', X'' \stackrel{\max(k_1, k_2)+1}{\Rightarrow} Y', Y'', (\Box B)^{m+2}} L\supset$$

where  $s+t = m+2$ . The transformation is similar to case 2.

By inspection, each introduced cut is well-behaved compared to  $cut_\gamma$ . ⊣

**Lemma 4.** *In the following derivation both premises of  $cut_\gamma$  are cut-free:*

$$\frac{\frac{\Box X, X, \Box B \stackrel{k}{\Rightarrow} B}{\Box X \stackrel{k+1}{\Rightarrow} \Box B} GLR_v \quad \frac{\frac{\tau}{(\Box B)^{m+2}, U \Rightarrow W} \rho}{\Box B, U \Rightarrow W} LC^{m+1}(\Box B)}{\Box X, U \Rightarrow W} cut_\gamma(\Box B)$$

*Then there is a derivation of  $\Box X, U \Rightarrow W$  where each cut in this derivation is of lesser degree, width or cut-height, and well-behaved compared to  $cut_\gamma$ .*

*Proof.* Without loss of generality  $\rho \neq LC(\Box B)$ . Also it is clear that  $(\Box B)^{m+2}, U \Rightarrow W$  cannot be an initial sequent (however, unlike Lemma 3, the sequent may be the conclusion of a  $GLR_v$  rule application).

First suppose that a  $\Box B$  formula occurrence is principal by  $\rho$ . If  $\rho = LW(\Box B)$ , then  $\tau$  is of the form  $\tau' / (\Box B)^{m+1}, U \stackrel{l}{\Rightarrow} W$ . Applying  $LC^m(\Box B)$  to  $\tau$ , and then the cut-rule we obtain a derivation with end-sequent  $\Box X, U \Rightarrow W$  and lesser cut-height  $((k+1) + (l+m) < (k+1) + (l+m+1))$ .

The other possibility is that  $\rho = GLR_v$ . It follows that  $U \equiv \Box V$  and  $W \equiv \Box C$  for some multiset  $V$  and formula  $C$ . Thus the *right premise* is of the form

$$\frac{\frac{\tau'}{(\Box B)^{m+2}, B^{m+2}, \Box V, V, \Box C \stackrel{l}{\Rightarrow} C}}{(\Box B)^{m+2}, \Box V \stackrel{l+1}{\Rightarrow} \Box C} \rho = GLR_v}{\Box B, \Box V \stackrel{l+1+m+1}{\Rightarrow} \Box C} LC^{m+1}(\Box B)$$

Consider the derivation

$$\frac{\frac{\frac{\frac{\frac{\frac{\frac{\Box X, X \Rightarrow B}{\Box X, X, \Box X, \Box V, V, \Box C \Rightarrow C} LC^*(\Box X)}{\Box X, X, \Box V, V, \Box C \Rightarrow C} GLR_v}{\Box X, X, \Box X, \Box V, V, \Box C \Rightarrow C} cut_\beta(B)}{\Box X, B^{m+2}, \Box V, V, \Box C \Rightarrow C} LC^{m+1}(B)}{\Box X, X, \Box X, \Box V, V, \Box C \Rightarrow C} cut_\alpha(\Box B)}{\Box X, X, \Box X, \Box V, V, \Box C \Rightarrow C} LC^{m+1}(\Box B)}{\Box X, X, \Box X, \Box V, V, \Box C \Rightarrow C} cut_\beta(B)}{\Box X, X, \Box X, \Box V, V, \Box C \Rightarrow C} LC^*(\Box X)}{\Box X, X, \Box V, V, \Box C \Rightarrow C} GLR_v}{\Box X, \Box V \Rightarrow \Box C}$$

The derivation concluding  $\Box X, X \Rightarrow B$  is obtained by Corollary 1. Compared to  $cut_\gamma$ , all cuts in this derivation are seen to be well-behaved and have cut-formulae with lesser degree or width. Also  $cut_\alpha$  has lesser cut-height  $((k+1) + (l+m+1) < (k+1) + (l+1+m+1))$ , and the degree of cut-formula for  $cut_\beta$  is smaller by 1.

If no  $\Box B$  formula occurrence is principal by  $\rho$ , we consider when  $\rho$  is a unary rule, additive binary or multiplicative binary ( $\rho = L\supset$ ) rule. We omit the proof of these cases as they are similar to the corresponding cases in Lemma 3. By inspection, each of the introduced cuts is well-behaved compared to  $cut_\gamma$ .  $\dashv$

**Theorem 1.** *Cut-elimination holds for  $GLS^m$*

*Proof.* Without loss of generality we will eliminate the topmost cut. Let  $D$  be the cut-formula and  $h$  the cut-height. Induction on  $d(D)\omega^2 + n(n)\omega + h$ . We observe that any introduced cut with lesser degree, width or cut-height that is well-behaved compared to the original cut can be eliminated by the induction hypothesis. This is because the ordinal inequality  $\leq$  in the definition of well-behaved is strict in this case.

First suppose the cut-formula  $D$  is *not boxed*. The (classical) Gentzen-style proof of cut-elimination can be carried out in this case [2]. In particular, the last rule above either premise cannot be an application of the rule  $GLR_v$  because the cut-formula would then be boxed. Also we note that the width of the cut-formula in an introduced cut is increased only if the degree of the cut-formula is smaller than before. This is because the classical transformations do not introduce any

new  $GLR_v$  rule applications. The induction hypothesis can be used to eliminate the cut-rule in this case.

For contractions above cut we use Lemma 1. We illustrate with the case when the left premise of the cut-rule is principal by  $R\supset$  and the right premise of the cut-rule is principal by a left contraction on the cut-formula. The derivation

$$\frac{\frac{A, X \xrightarrow{k} Y, B}{X \xrightarrow{k+1} Y, A \supset B} R\supset \quad \frac{\frac{(A \supset B)^{n-t}, U \xrightarrow{l_1} W, A \quad B, (A \supset B)^t, U \xrightarrow{l_2} W}{L\supset} \quad \frac{A \supset B, (A \supset B)^n, U \xrightarrow{\max(l_1, l_2)+1} W}{LC^n(A \supset B)}}{A \supset B, U \xrightarrow{\max(l_1, l_2)+1+n} W} cut(A \supset B)}{X, U \Rightarrow Y, W} cut(A \supset B)$$

is transformed to

$$\frac{\frac{\frac{(A \supset B)^t, U \xrightarrow{l_1} W, A}{LC^{t-1}(A \supset B)} \quad X \xrightarrow{k+1} Y, A \supset B}{A \supset B, U \xrightarrow{l_1+t-1} W, A} cut_1(A \supset B)}{\frac{X, U \Rightarrow Y, W, A \quad A, X \Rightarrow Y, B}{cut_2(A)} \quad \frac{X, X, U \Rightarrow Y, Y, W, B}{B, U \Rightarrow W} cut_3(B)}}{\frac{X, X, U, U \Rightarrow Y, Y, W, W}{LC^*, RC^*} \quad X, U \Rightarrow Y, W} cut_3(B)$$

where  $t - 1$  is taken as 0 if  $t = 0$ . The derivation concluding  $B, U \Rightarrow W$  is obtained by applying Lemma 1 to the right premise of the cut in the original derivation. Now  $cut_1$  has lesser cut-height; the cut-formulae for  $cut_2$  and  $cut_3$  are of lesser degree. The introduced cuts are eliminated by the induction hypotheses.

Now suppose that the cut-formula  $D$  is boxed. We can permute the cut with any logical rules appearing above the left and right premises in the standard manner. This is because the cut-formula  $D$  cannot be principal by a logical rule as it is boxed. If the cut-formula has been introduced by weakening in either premise, the end-sequent  $X, U \Rightarrow Y, W$  can be obtained by replacing the  $LW(D)$  or  $RW(D)$  rule with appropriate weakening rules.

The remaining cases to consider are contractions on the cut-formula  $D$  above the 1. left (2. right) premise, and when 3. the derivation is in SNF. Case 1 can be handled by Lemma 3. Case 2 can be handled by Lemma 4. If the derivation is in SNF, then we can use Lemma 2 to obtain the required result.  $\dashv$