Corrected Upper Bounds for Free-Cut Elimination

Arnold Beckmann*
Department of Computer Science
Swansea University
Swansea SA2 8PP, UK
a.beckmann@swansea.ac.uk

Samuel R. Buss†
Department of Mathematics
University of California, San Diego
La Jolla, CA 92093-0112, USA
sbuss@math.ucsd.edu

December 20, 2010

Abstract

Free-cut elimination allows cut elimination to be carried out in the presence of non-logical axioms. Formulas in a proof are anchored provided they originate in a non-logical axiom or non-logical inference. This paper corrects and strengthens earlier upper bounds on the size of free-cut elimination. The correction requires that the notion of a free-cut be modified so that a cut formula is anchored provided that all of its introductions are anchored, instead of only requiring that one of its introductions is anchored. With the correction, the originally proved size upper bounds remain unchanged. These results also apply to partial cut elimination. We also apply these bounds to elimination of cuts in propositional logic.

If the non-logical inferences are closed under cut and infer only atomic formulas, then all cuts can be eliminated. This extends earlier results of Takeuti and of Negri and von Plato.

1 Introduction

The notion of *free-cut elimination* was introduced by G. Takeuti [12] as an extension of cut elimination that can be used in the presence of induction inference rules. In short, the free-cut elimination theorem states that any provable sequent can be proved using only cuts in which at least one of the cut formulas was introduced as a principal formula of an induction axiom. Takeuti did not provide a detailed proof of the free-cut elimination, however.

*Supported in part by EPSRC grant EP/D03809X/1.
†Supported in part by NSF grant DMS-0700533.

Free-cut elimination has been important for results about computational complexity or constructivity in proof theory. For instance, the second author used free-cut elimination for witnessing theorems in bounded arithmetic [3], and many other researchers have used it for similar purposes.

A different version of free-cut elimination was later used by the second author in the expository article [5]. In this variant, a set $\mathfrak{S}$ of non-logical axioms is allowed, and any formula that occurs in a non-logical axiom is called *anchored*. Cuts in which neither cut formula is anchored are called *free*, and the modified free-cut elimination theorem states that any provable sequent is provable by a proof in which no cuts are free.

However, as William Scott [private communication] first pointed out, there is an error in the proof of the free-cut elimination theorem in [5]. As a result, although the free-cut elimination theorem is indeed correct, the upper bounds on the size of free-cut free proofs that are obtained in [5] are not correct as stated.

Part of goal of the present paper is to correct this. The fix does not involve changing the upper bounds themselves, rather it involves changing the definition of anchored and free formulas, as well as the definition of a *depth* of a cut formula. In fact, the revised theorem proved in the present paper is stronger than the result proved in [5], since the new definition of anchored is stricter than the original definition. The basic difference in the two notions of anchored is that the original definition specified that a formula is anchored if at least one of the places it is introduced is an anchor, whereas the revised definition requires that every place the formula is introduced be an anchor.

At the same time, we will prove the free-cut elimination theorem in a somewhat more general setting, by allowing a more general notion of non-logical initial sequents and non-logical rules. This unifies the two notions of free-cut elimination from [12] and [5].

For propositional logic, this gives a proof that non-atomic free cuts can be eliminated with only an exponential blowup in the size of proofs. This generalizes results of Zhang [14] and Gerhardy [6] by showing that these bounds apply even in the presence of non-logical axioms when eliminating free cuts.

Section 6 proves theorems about when cuts can be completely eliminated even in the presence of non-logical axioms and inferences. This generalizes work of Takeuti on generalized equality axioms, as well as the non-logical rules of inference used by Negri and von Plato to simulate arbitrary quantifier-free (i.e., purely universal) axioms.

It should be stressed that the free-cut elimination theorems stated in prior works [12, 5] are correct as stated, with the sole exception of the upper bounds in [5]. Fortunately, it seems that the applications of anchored cuts and formulas depth as defined in [5] have been used only in ways that have not generated further errors. This is because those upper bounds have been used only for common systems, not for contrived systems. Indeed, the results and upper bounds as stated in [5] *are* correct for all commonly used systems such as $I\Sigma_k$, S_2^k, T_2^k, etc., because of the special nature of the induction axioms. Section 5 proves results about partial cut elimination, and these results seemingly cover all existing applications of free-cut elimination to-date.

Our proofs will all use "global" transformations of proofs in the style of the proof of cut elimination in [5]. It would also be possible to prove the theorems using induction on the height of proofs, by using reductions that act on the final inferences of proofs as was done by Gentzen in the original proofs of cut elimination. Indeed, induction on the height of proofs is the most common way to carry out the proofs and is used by many authors, see for instance in the proofs by [10, 14, 6, 13] who obtain bounds very similar to those of the present paper. An advantage to our global proof method is that it makes more explicit how proofs are transformed for cut elimination.

A rather different approach to cut elimination is given by Baaz and Leitsch [1, 2], who reduce cut elimination to resolution. In some special cases they obtain super-elementarily better upper bounds on the size of cut free proofs than can be obtained by Gentzen reduction methods, but they do not give the same kind of tight bounds for general cut elimination as the present paper.

2 The sequent calculus and free-cuts

We presume the reader has basic familiarity with the sequent calculus and cut elimination, but begin by reviewing the necessary definitions for the systems used later in the paper. We work with a sequent calculus for classical logic over the connectives $\forall$, $\exists$, $\land$, $\lor$, $\supset$, and $\neg$. Lines in a sequent calculus proof are called *sequents* and have the form $\Gamma{\longrightarrow}\Delta$, where the *cedents* Γ and Δ are finite sequences of formulas. The *logical initial sequents* are $A{\longrightarrow}A$, with A required to be an atomic formula. The valid *logical inferences* are as shown below.

$$\text{Exchange:left}\ \frac{\Gamma, A, B, \Lambda \longrightarrow \Delta}{\Gamma, B, A, \Lambda \longrightarrow \Delta} \qquad \text{Exchange:right}\ \frac{\Gamma \longrightarrow \Delta, A, B, \Lambda}{\Gamma \longrightarrow \Delta, B, A, \Lambda}$$

$$\text{Contraction:left}\ \frac{A, A, \Gamma \longrightarrow \Delta}{A, \Gamma \longrightarrow \Delta} \qquad \text{Contraction:right}\ \frac{\Gamma \longrightarrow \Delta, A, A}{\Gamma \longrightarrow \Delta, A}$$

$$\text{Weakening:left}\ \frac{\Gamma \longrightarrow \Delta}{A, \Gamma \longrightarrow \Delta} \qquad \text{Weakening:right}\ \frac{\Gamma \longrightarrow \Delta}{\Gamma \longrightarrow \Delta, A}$$

$$\neg\text{:left}\ \frac{\Gamma \longrightarrow \Delta, A}{\neg A, \Gamma \longrightarrow \Delta} \qquad \neg\text{:right}\ \frac{A, \Gamma \longrightarrow \Delta}{\Gamma \longrightarrow \Delta, \neg A}$$

$$\wedge\text{:left}\ \frac{A, B, \Gamma \longrightarrow \Delta}{A \wedge B, \Gamma \longrightarrow \Delta} \qquad \wedge\text{:right}\ \frac{\Gamma \longrightarrow \Delta, A \qquad \Gamma \longrightarrow \Delta, B}{\Gamma \longrightarrow \Delta, A \wedge B}$$

$$\vee\text{:left}\ \frac{A, \Gamma \longrightarrow \Delta \qquad B, \Gamma \longrightarrow \Delta}{A \vee B, \Gamma \longrightarrow \Delta} \qquad \vee\text{:right}\ \frac{\Gamma \longrightarrow \Delta, A, B}{\Gamma \longrightarrow \Delta, A \vee B}$$

$$\supset\text{:left}\ \frac{\Gamma \longrightarrow \Delta, A \qquad B, \Gamma \longrightarrow \Delta}{A \supset B, \Gamma \longrightarrow \Delta} \qquad \supset\text{:right}\ \frac{A, \Gamma \longrightarrow \Delta, B}{\Gamma \longrightarrow \Delta, A \supset B}$$

$$\forall\text{:left}\ \frac{A(t), \Gamma \longrightarrow \Delta}{(\forall x)A(x), \Gamma \longrightarrow \Delta} \qquad \forall\text{:right}\ \frac{\Gamma \longrightarrow \Delta, A(b)}{\Gamma \longrightarrow \Delta, (\forall x)A(x)}$$

$$\exists\text{:left}\ \frac{A(b), \Gamma \longrightarrow \Delta}{(\exists x)A(x), \Gamma \longrightarrow \Delta} \qquad \exists\text{:right}\ \frac{\Gamma \longrightarrow \Delta, A(t)}{\Gamma \longrightarrow \Delta, (\exists x)A(x)}$$

$$\text{Cut}\ \frac{\Gamma \longrightarrow \Delta, A \qquad A, \Gamma \longrightarrow \Delta}{\Gamma \longrightarrow \Delta}$$

The first six rules (exchange, contraction, and weakening) are *weak* inferences. Weak inferences can be viewed as unimportant "bookkeeping" inferences which are needed since we treat cedents as sequences of formulas, rather than as sets or multisets of formulas. The size of a proof will be defined by ignoring the weak inferences. The free variable b of the $\forall$:right and $\exists$:left inferences is the *eigenvariable* of the inference and must not appear in the lower sequent. A complex formula introduced in the lower line of an inference I is called the *principal formula* of I; the corresponding formula(s) in the upper sequent are the *auxiliary formulas*. For example, in the $\exists$:right inference, $(\exists x)A(x)$ and $A(t)$ are the principal and auxiliary formulas, respectively.

In addition to the logical inferences listed above, we allow an additional set $\mathfrak{S}$ of nonlogical axioms and inferences. The set $\mathfrak{S}$ is a set of axioms or inferences whose principal formulas serve to anchor cuts. The intent is that formulas introduced by $\mathfrak{S}$-inferences may be used as cut formulas in a free-cut free proof. The main criteria for the nonlogical axioms in $\mathfrak{S}$ are that they admit substitution by terms, and that each inference rule in $\mathfrak{S}$ must have a consistent policy about admitting side formulas. To formalize this, we define the notion of an inference skeleton.

Definition An *inference skeleton*, $\mathcal{I}$, consists of the following:

(a) A k-hypothesis inference form

$$\frac{\{\Psi_i, \mathcal{C}_i \longrightarrow \mathcal{D}_i, \Xi_i\}_{i=1}^{k}}{\Psi, \mathcal{C} \longrightarrow \mathcal{D}, \Xi} \tag{1}$$

where $k \geq 0$ and Ψ, Ξ, Ψ_i, Ξ_i are cedents, and where $\mathcal{C}, \mathcal{D}, \mathcal{C}_i, \mathcal{D}_i$ are meta-variables for cedents. The value $k = 0$ is allowed, so the inference form may have no hypotheses; in this case, (1) is an inference form for non-logical initial sequents. The formulas in Ψ and Ξ are the *principal formulas* of the inference, the formulas in the Ψ_i's and Ξ_i's are the *auxiliary formulas*, and the $\mathcal{C}$'s and $\mathcal{D}$'s contain the *side formulas*.

(b) A list of *side formula indicators*, $s_1, \ldots, s_k \in \{0, 1\}$. These indicate which of the hypotheses are permitted to have side formulas.

(c) A (possibly empty) list of free variables $a_1, \ldots, a_\ell$ called *eigenvariables*, where each a_j must appear in exactly one (sub)sequent $\Psi_i \longrightarrow \Xi_i$ and must not appear in $\Psi \longrightarrow \Xi$.

The inference skeleton $\mathcal{I}$ specifies a set $\mathfrak{S} = \mathit{Instances}(\mathcal{I})$ of inferences. The sequents in $\mathit{Instances}(\mathcal{I})$ are obtained as follows: Let Γ and Δ be any cedents that do not contain any eigenvariables, and let $\mathcal{C} = \Gamma$ and $\mathcal{D} = \Delta$. Further, for each i such that $s_i = 1$, let $\mathcal{C}_i = \Gamma$ and $\mathcal{D}_i = \Delta$; and for each i with $s_i = 0$, let $\mathcal{C}_i$ and $\mathcal{D}_i$ be empty. The resulting form of (1) is an inference in $\mathit{Instances}(\mathcal{I})$, and every member of $\mathit{Instances}(\mathcal{I})$ is obtained in this way.

Let $\mathcal{I}$ be an inference skeleton, and suppose σ is a substitution that maps free variables to terms, such that no eigenvariable of $\mathcal{I}$ occurs in either the domain or range of σ. (As usual, σ acts as the identity on any variable not in its domain.) Then $\mathcal{I}\sigma$ is called a *substitution instance* of $\mathcal{I}$, and is obtained by applying the substitution σ to every formula in $\mathcal{I}$, that is to say, by replacing every free variable a in (1) with the term $\sigma(a)$.[1]

[1]By convention, inference rules are closed under renaming of variables, and so this is not stated explicitly as part of the definition. For example, the $\forall$:right and $\exists$:left inferences implicitly already use this convention as the eigenvariable b may be any variable.

Definition A set $\mathfrak{S}$ of inferences is called *acceptable* provided there is a set $\mathbb{I}$ of inference skeletons such that $\mathbb{I}$ is closed under substitutions, and $\mathfrak{S}$ equals the union of the sets $Instances(\mathcal{I})$ for $\mathcal{I} \in \mathbb{I}$.

Henceforth, all sets $\mathfrak{S}$ of inferences are implicitly assumed to be acceptable.

It is useful to consider a few simple examples of acceptable sets $\mathfrak{S}$. First, consider the situation where $\mathfrak{S}$ is a set of initial sequents and is closed under substitution. An *initial sequent* is a zero hypothesis inference, and thus $\mathfrak{S}$ corresponds to a set of non-logical axioms. Since there are zero hypotheses, there are no eigenvariables. One example of this is the set of equality axioms.

A second example is the induction inferences. Frequently one wishes to restrict induction by specifying a set X of formulas that admit induction. For each formula $A(x)$ which admits induction, and each term t, there is an inference skeleton of the form

$$\frac{A(b), \mathcal{C}_1 \longrightarrow \mathcal{D}_1, A(S(b))}{A(0), \mathcal{C} \longrightarrow \mathcal{D}, A(t)}$$

Here, b is the only eigenvariable. The induction inferences are equivalent to the usual induction axioms for formulas from X provided side formulas are permitted, and for this we take $s_1 = 1$. It is necessary that the set X of formulas that admit induction be closed under substitution.

For a third example, consider the Π^0_2-induction rule of Parsons [11]. For each $A(x) \in \Pi^0_2$ and each term t, let $\mathcal{I}$ be the inference skeleton with form

$$\frac{\mathcal{C}_1 \longrightarrow \mathcal{D}_1, A(0) \qquad A(b), \mathcal{C}_2 \longrightarrow \mathcal{D}_2, A(b+1)}{\mathcal{C} \longrightarrow \mathcal{D}, A(t)}$$

where b is the eigenvariable and where $s_1 = 1$ and $s_2 = 0$, so side formulas are allowed in the left hypothesis but not in the right hypothesis (following the convention of [11]).

A fourth example is the collection rule of [4]:

$$\frac{\mathcal{C}_1 \longrightarrow \mathcal{D}_1, (\forall x \le t)(\exists y)A(x,y)}{\mathcal{C} \longrightarrow \mathcal{D}, (\exists z)(\forall x \le t)(\exists y \le z)A(x,y)}$$

where $s_1 = 1$ so side formulas are allowed. In [4], the quantifier complexity of A was restricted, and free-cut elimination allowed the quantifier complexity of formulas in the proof to be similarly restricted.

A fifth example is the non-logical inference rules used by Negri and von Plato [8, 9] to simulate quantifier-free axioms. These rules are inferences with $k \geq 0$ hypotheses of the form

$$\frac{Q_1, \mathcal{C}_1 {\longrightarrow} \mathcal{D}_1 \qquad \cdots \qquad Q_k, \mathcal{C}_k {\longrightarrow} \mathcal{D}_k}{P_1, \ldots, P_m, \mathcal{C} {\longrightarrow} \mathcal{D}} \tag{2}$$

where the formulas Q_i and P_j are all atomic. All hypotheses may have side formulas so $s_i = 1$ for $1 \leq i \leq k$.[2] Negri and von Plato proved that these rules admit elimination of all cuts in the **G3ipm** sequent calculus, and we will reprove this for the sequent calculus defined above.

It is also interesting to note that the logical inferences can also be viewed as a set $\mathfrak{S}$ of inferences. For example, the $\wedge$:right inference can be expressed as the union of the sets *Instances*($\mathcal{I}$) where $\mathcal{I}$ ranges over inferences skeletons with the form

$$\frac{\mathcal{C}_1 {\longrightarrow} \mathcal{D}_1, A \qquad \mathcal{C}_2 {\longrightarrow} \mathcal{D}_2, B}{\mathcal{C} {\longrightarrow} \mathcal{D}, A \wedge B}$$

with $s_1 = s_2 = 1$ so that side formulas are permitted. Although it would be unusual to include the $\wedge$:right inferences in $\mathfrak{S}$, the effect would be to allow cuts on formulas with outermost connective $\wedge$ to count as anchored (non-free) cuts. In Section 5, we use this idea to prove theorems about partial cut elimination, namely elimination of cut on formulas above a given logical complexity.

The *size*, $|P|$, of a proof P is defined to equal the total number of non-weak inferences with one or more hypotheses. Note that size does not count any initial sequents either in $\mathfrak{S}$ or of the form $A{\longrightarrow}A$. The *height*, $h(P)$, of P is equal to the maximum number of non-weak inferences with one or more hypotheses on any branch in P.

The *direct ancestor* relation on occurrences of formulas in a proof P is defined in the usual way so as to keep track of the identity of formulas from line to line. Let C and C' be two occurrences of the same formula in P. We call C' an *immediate direct ancestor* of C provided C' appears in an upper sequent of an inference and C appears in the lower sequent of the same inference, and provided that either (i) the inference is any logical inference or $\mathfrak{S}$-inference, and C and C' occupy the same position in Γ, Δ or Λ in their respective sequents, or (ii) the inference is a contraction, and C and C' are both formulas denoted by "A", or (iii) the inference is an exchange, and

[2]We have slightly simplified Negri and von Plato's formulation of the inferences to take advantage of the way our system handles weak inferences.

C and C' are the formulas denoted by "A" or are the formulas denoted by "B". Note that some formulas do not have immediate direct ancestors; namely, the principal formulas of non-weak inferences, formulas introduced by a weakening rule, and formulas in the cedents Ψ and Ξ of an $\mathfrak{S}$-inference.

We next define the notion of an "anchored" cut. First, however, we must define the notion of $\mathfrak{S}$-depth.

Definition Let C be an occurrence of a formula in a proof P. The *$\mathfrak{S}$-depth of C*, denoted $\mathfrak{S}$-depth(C), is defined in terms of how it is inferred.

(1) If C is a principal formula in an $\mathfrak{S}$-inference (a formula in Ψ or Ξ), then C has $\mathfrak{S}$-depth 0.

(2) If C is in a logical initial sequent, then C has $\mathfrak{S}$-depth 1.

(3) If C is in the lower sequent of an inference I, and if either I is a weak inference or I is non-weak with C a side formula of I (in a cedent Γ or Δ), then the $\mathfrak{S}$-depth of C is equal to

$$\max\left\{\mathfrak{S}\text{-depth}(C') : C' \text{ is an immediate direct ancestor of } C\right\}.$$

The maximum of the empty set is taken to equal $-\infty$.

(4) If C is the principal formula of a non-weak, non-$\mathfrak{S}$ inference I, then the $\mathfrak{S}$-depth of C is equal to

$$1 + \max\left\{\mathfrak{S}\text{-depth}(C') : C' \text{ is an auxiliary formula of } I\right\}.$$

By convention, $1 + (-\infty) = -\infty$.

Definition The $\mathfrak{S}$-depth of a cut inference is the minimum of the $\mathfrak{S}$-depths of the two occurrences of its cut formula. The $\mathfrak{S}$-depth of a proof is the maximum $\mathfrak{S}$-depth of its cut inferences.

Definition A cut is *anchored* provided that one of the occurrences of its cut formula has $\mathfrak{S}$-depth zero. A cut is called *free* if

(i) One of the occurrences of the cut formula has $\mathfrak{S}$-depth $-\infty$, or

(ii) Its cut formula is atomic, and one of the occurrences of the cut formula has $\mathfrak{S}$-depth 1, or

(iii) It is not anchored.

A proof is *free-cut free* provided it has no free cuts.

As an immediate consequence of the definitions, we have:

Proposition 1 *A non-free cut has* $\mathfrak{S}$*-depth zero.*

Note, however, that it is possible for free cut to have $\mathfrak{S}$-depth zero; namely, a cut on an atomic cut formula with one of the occurrences of the cut formula having $\mathfrak{S}$-depth 0 and the other having $\mathfrak{S}$-depth 1.

The next section will state the free-cut elimination theorem, but first we prove Theorem 3 that allows eliminating cuts on $\mathfrak{S}$-depth $-\infty$ formulas.

Definition Then $P \preccurlyeq_{\mathfrak{S}} P'$ means that the proofs P and P' have the same endsequent $\Gamma \longrightarrow \Delta$, and that each formula A occurring in $\Gamma \longrightarrow \Delta$ has $\mathfrak{S}$-depth in P less than or equal to its $\mathfrak{S}$-depth in P'.

Proposition 2 *Suppose* P_1 *is a subproof of* P*, and* $P_2 \preccurlyeq_{\mathfrak{S}} P_1$*. Let* P' *be obtained from* P *by replacing* P_1 *with* P_2*. Then* $P' \preccurlyeq_{\mathfrak{S}} P$*.*

Proposition 2 is an immediate consequence of the monotonicity of the definition of $\mathfrak{S}$-depth.

Theorem 3 *Let* P *be a proof of* $\Gamma \longrightarrow \Delta$*. Then there is a proof* P' *of the same sequent with no* $\mathfrak{S}$*-depth* $-\infty$ *cuts, such that* $|P'| \le |P|$ *and* $h(P') \le h(P)$ *and* $\mathfrak{S}$*-depth*$(P') \le \mathfrak{S}$*-depth*(P)*. Furthermore,* $P' \preccurlyeq_{\mathfrak{S}} P$*.*

As will be evident from the proof of the theorem, P' is formed from P by discarding parts of P and possibly adding weak inferences. The idea of the proof is quite simple: namely, delete from P, whenever possible, formulas which have $\mathfrak{S}$-depth $-\infty$, and also remove cuts of $\mathfrak{S}$-depth $-\infty$. The main complication is that removing formulas of $\mathfrak{S}$-depth $-\infty$ may lower the $\mathfrak{S}$-depth of other formulas in the proof and thereby lower the $\mathfrak{S}$-depth of cut inferences. Some of these cuts may become $\mathfrak{S}$-depth $-\infty$ and thus need to be eliminated.

The next lemma is a sharpened form of Theorem 3.

Lemma 4 *Let* P *be a proof ending with the sequent* $\Gamma \longrightarrow \Delta$*. Let* $\Gamma' \longrightarrow \Delta'$ *be obtained from* $\Gamma \longrightarrow \Delta$ *by removing an arbitrary subset of the formulas that have* $\mathfrak{S}$*-depth* $-\infty$ *in the endsequent of* P*. Then there is a proof* P' *of* $\Gamma' \longrightarrow \Delta'$ *such that* P' *has no cuts of* $\mathfrak{S}$*-depth* $-\infty$ *and such that* $|P'| \le |P|$ *and* $h(P') \le h(P)$ *and* $\mathfrak{S}$*-depth*$(P') \le \mathfrak{S}$*-depth*(P)*. Furthermore, for each formula* C *appearing in* $\Gamma' \longrightarrow \Delta'$*, the* $\mathfrak{S}$*-depth of* C *in* P' *is less than or equal to the* $\mathfrak{S}$*-depth of the corresponding formula in the endsequent of* P*.*

Proof The lemma is proved by induction on $|P|$. The proof splits into cases depending on the final inference of P. The proof is trivial if the last inference of P is a weak inference.

Consider the case where the final inference of P is a $\land$:right inference:

$$\dfrac{\begin{array}{c}P_1\\ \vdots\\ \Gamma \longrightarrow \Delta, A\end{array} \qquad \begin{array}{c}P_2\\ \vdots\\ \Gamma \longrightarrow \Delta, B\end{array}}{\Gamma \longrightarrow \Delta, A \land B}$$

The goal is to find a proof of the sequent $\Gamma' \longrightarrow \Delta', (A \land B)'$ where $(A \land B)'$ indicates that either (i) the formula $A \land B$ has $\mathfrak{S}$-depth $-\infty$ and that this formulas is one of the formulas that is to be deleted, or (ii) $(A \land B)'$ is just $A \land B$. The latter case must happen if $\mathfrak{S}$-depth$(A \land B) \neq -\infty$, but can also happen with $\mathfrak{S}$-depth$(A \land B) = -\infty$ if it is not one of the deleted formulas.

In case (i), we must give a proof P' of $\Gamma' \longrightarrow \Delta'$. In this case, both the occurrence of A in the endsequent of P_1 and the occurrence of B in the endsequent of P_2 have $\mathfrak{S}$-depth $-\infty$. Thus the induction hypothesis gives two proofs P_1' and P_2' of $\Gamma' \longrightarrow \Delta'$, and either one can be taken to be P'. In case (ii), we need to give a proof P' of $\Gamma' \longrightarrow \Delta', A \land B$. The induction hypothesis gives us two proofs, P_1' and P_2' of $\Gamma' \longrightarrow \Delta', A$ and $\Gamma' \longrightarrow \Delta', B$, respectively. Combine these with a single $\land$:right inference to get the desired proof P'. In both cases (i) and (ii), the fact that the $\mathfrak{S}$-depth of formulas in the endsequent has not been increased in P' follows immediately from the definition of $\mathfrak{S}$-depth and the induction hypotheses.

The other non-weak logical inferences are handled similarly to the $\land$:right inference, except for cut inferences. Suppose the final inference of P is

$$\dfrac{\begin{array}{c}P_1\\ \vdots\\ \Gamma \longrightarrow \Delta, A\end{array} \qquad \begin{array}{c}P_2\\ \vdots\\ A, \Gamma \longrightarrow \Delta\end{array}}{\Gamma \longrightarrow \Delta}$$

The induction hypotheses give us proofs P_1' and P_2' of the sequents $\Gamma' \longrightarrow \Delta', A$ and $A, \Gamma' \longrightarrow \Delta'$, respectively. If the $\mathfrak{S}$-depth of A in the final sequent of P_1' is equal to $-\infty$, then we apply the induction hypothesis again to P_1' to get a proof P' of $\Gamma' \longrightarrow \Delta'$: this P' is immediately seen to satisfy the desired conditions. Likewise, if the $\mathfrak{S}$-depth of A in P_2' is $-\infty$, we can apply the induction hypothesis to P_2' to obtain the desired P'. Finally, if neither case holds, form P' by combining P_1' and P_2' with a cut inference. By definition, this cut inference has $\mathfrak{S}$-depth $> -\infty$.

The case where the final inference of P is a $\mathfrak{S}$-inference is very simple to handle with the induction hypothesis since only side formulas can have $\mathfrak{S}$-depth equal to $-\infty$. $\square$

3 The free-cut elimination theorems

Theorem 5 and the proof of Lemma 6 contain our basic results on upper bounds on free-cut elimination in the presence of $\mathfrak{S}$-inferences.

Definition For $i, k \in \mathbb{N}$, the *superexponential* function 2^i_k is defined inductively by $2^i_0 = i$ and $2^i_{k+1} = 2^{2^i_k}$.

Theorem 5 *Suppose P is a proof of $\mathfrak{S}$-depth $\leq d$, where $d \geq 0$. Then there is a proof P' of the same endsequent which contains no free cuts and has height $h(P') < 2^{h(P)+1}_{d+1}$. Therefore, P' has size $|P'| < c^{2^{h(P)+1}_{d+1}} \leq c^{2^{|P|+1}_{d+1}}$, where c is the maximum of 2 and the maximum arity of the $\mathfrak{S}$-inferences that appear in P.*

The next lemma is the main tool for the proof of the theorem.

Lemma 6 *Suppose P ends with a free cut inference of $\mathfrak{S}$-depth $d \geq 0$ and all other free cuts in P have $\mathfrak{S}$-depth $< d$. Then there is a proof P' of the same endsequent, such that all free cuts in P' have $\mathfrak{S}$-depth $< d$, and $h(P') \leq 2 \cdot h(P)$ and $P' \preccurlyeq_{\mathfrak{S}} P$. If the cut formula is not atomic, then $|P'| \leq |P|^2$. Otherwise, the cut formula is atomic and $d \leq 1$, and $|P'| \leq (c-1)|P|^2$, where c is as in Theorem 5.*

To prove the lemma, we will let P_1 and P_2 be the two immediate subproofs of P as displayed in (3) below, and prove that

(a) If A has outermost connective $\neg$, $\vee$, $\wedge$, or $\supset$, then

$$h(P') \leq h(P) + 2.$$

(b) If A has outermost connective $\forall$ or $\exists$, then

$$h(P') \leq 2 \cdot h(P).$$

(c) If A is atomic, then

$$h(P') \leq h(P_1) + h(P_2) + 1.$$

We conclude in all cases that $h(P') \leq 2h(P)$.

The bounds (b) and (c) can be compared to the similar results of Orevkov and others [10, 14, 6, 13] who all give a bound of $h(P') \leq h(P_1) + h(P_2)$. Their upper bound is slightly better than ours because their proof systems are for pure first-order logic and do not admit $\mathfrak{S}$-inferences. Our upper bounds are slightly larger because of the need to add cuts on $\mathfrak{S}$-depth zero formulas.

The bound (a) should similarly be compared to results of [14, 6] that prove bounds of $h(P') \leq h(P) + 1$. Again, our bound is higher by 1 because of cuts on $\mathfrak{S}$-depth zero formulas.[3]

We can always assume w.l.o.g. that any proof P is in *free variable normal form*: this implies that no variable is used more than once as eigenvariable in P and furthermore that if a variable c is used as an eigenvariable then c appears in the proof only above the inference where it is used as an eigenvariable. In particular, if c appears in the endsequent of P, then c is not used as an eigenvariable in P. In this case, we write $P(t/c)$ to denote the result of replacing every occurrence of c in P with the term t. If no eigenvariable of P occurs in t, then $P(t/c)$ is a valid proof.

Proof (of Lemma 6) The proof is by induction on the size of the proof P. Suppose P ends with a free cut inference

$$\dfrac{\begin{array}{c} P_1 \\ \vdots \\ \Gamma \longrightarrow \Delta, A \end{array} \qquad \begin{array}{c} P_2 \\ \vdots \\ A, \Gamma \longrightarrow \Delta \end{array}}{\Gamma \longrightarrow \Delta} \tag{3}$$

with one of the two occurrences of A having $\mathfrak{S}$-depth d and the other having $\mathfrak{S}$-depth $\geq d$. We begin by assuming that the cut formula A is not atomic. The proof splits into cases depending on the outermost connective of A. We'll consider the cases of $\neg$, $\vee$, and $\forall$; the remaining cases $\wedge$, $\supset$, and $\exists$ are essentially the same.

Since A is not atomic and the cut is free, we have $d \geq 1$ and the $\mathfrak{S}$-depths of both occurrences of A are ≥ 1. Since A is not atomic, each subproof P_i contains at least one non-$\mathfrak{S}$ inference with a direct ancestor of A as its principal formula. Thus, $|P_i| \geq 1$ and $h(P_i) \geq 1$ holds for $i = 1, 2$.

Suppose the cut formula A is of the form $\neg B$. (This case is rather simple, but we cover it in detail since rest of the cases use similar techniques.) Examining the subproof P_1, find all the occurrences of direct ancestors of

[3] For more on this, see the discussion at the end of Section 5.

$\neg B$ which do not have an immediate direct ancestor. These occurrences are where $\neg B$ *originates* in P_1. They can be principal formulas of weakenings, $\mathfrak{S}$-inferences, or $\neg$:right inferences

$$\frac{B, \Pi \longrightarrow \Lambda}{\Pi \longrightarrow \Lambda, \neg B}\ \neg\text{:right} \tag{4}$$

If the occurrence of $\neg B$ as a cut formula in the endsequent of P_1 has $\mathfrak{S}$-depth d, then, in (4), $\neg B$ has $\mathfrak{S}$-depth $\leq d$ and hence $\mathfrak{S}$-depth$(B) < d$. We modify P_1 to construct a proof P_1' of $\Gamma, B \longrightarrow \Delta, \neg B$, by replacing each $\neg$:right inference (4) with

$$\frac{B, \Pi \longrightarrow \Lambda}{\Pi, B \longrightarrow \Lambda, \neg B}\ \text{Weakening and exchanges}$$

where the $\neg B$ formula is introduced by weakening, and thus has $\mathfrak{S}$-depth $-\infty$. We repeat this construction for every $\neg$:right inference where $\neg B$ originates. The new occurrence of the formula B in the antecedent is propagated down the proof to all descendents of these $\neg$:right inferences. And, by adding weak inferences, a new occurrence of B is added to all side formulas of inferences whose lower sequents contain a new occurrence of B; with the exception that, for $\mathfrak{S}$-inferences, only the upper sequents with side indicators equal to 1 are given a new occurrence of B. Note that $\neg B$ cannot originate from an inference above any $\mathfrak{S}$-inferences hypothesis which has side indicator 0, as no direct ancestor $\neg B$ can lie above any sequent with side indicator 0.

This gives a proof P_1' of $\Gamma, B \longrightarrow \Delta, \neg B$. We have $\mathfrak{S}$-depth$(\neg B) \leq 0$ because the direct ancestors of $\neg B$ can originate only from $\mathfrak{S}$-inferences or weakenings (since only atomic formulas are allowed in logical initial sequents). If $\mathfrak{S}$-depth$(A) = d$ in the endsequent of P_1, then $\mathfrak{S}$-depth$(B) < d$ in P_1'. And, it is clear from the construction that any formula in Γ, Δ has $\mathfrak{S}$-depth in the endsequent of P_1' less than or equal to its $\mathfrak{S}$-depth in the endsequent of P_1.

If $\mathfrak{S}$-depth$(\neg B)$ equals $-\infty$ in the endsequent of P_1', then Lemma 4 gives a proof P_1'' of $\Gamma, B \longrightarrow \Delta$. If the $\mathfrak{S}$-depth of $\neg B$ equals 0, we instead form P_1'' as

$$\dfrac{\begin{matrix} P_1' \\ \vdots \\ \Gamma, B \longrightarrow \Delta, \neg B \end{matrix} \qquad \dfrac{\begin{matrix} P_2 \\ \vdots \\ \neg B, \Gamma \longrightarrow \Delta \end{matrix}}{\neg B, \Gamma, B \longrightarrow \Delta}\ \text{Weak inferences}}{\Gamma, B \longrightarrow \Delta}\ \text{Cut}$$

The $\neg B$ in the endsequent of P_2 has $\mathfrak{S}$-depth ≥ 1; thus the cut has $\mathfrak{S}$-depth zero and is anchored and not free. A similar construction lets us form proofs P_2' and P_2'', with P_2'' a proof of $\Delta {\longrightarrow} B, \Delta$. If $\mathfrak{S}$-depth$(A) = d$ in the endsequent of P_2, then $\mathfrak{S}$-depth$(B) < d$ in the endsequent of P_2''.

Therefore, P_1'' and P_2'' can be combined with a cut of $\mathfrak{S}$-depth $< d$ on B to give the desired proof P':

$$\dfrac{\begin{array}{c} P_2'' \\ \vdots \\ \Gamma {\longrightarrow} B, \Delta \end{array} \qquad \begin{array}{c} P_1'' \\ \vdots \\ \Gamma, B {\longrightarrow} \Delta \end{array}}{\Gamma {\longrightarrow} \Delta} \text{ Weak inferences and Cut}$$

Note that the height, $h(P')$ is

$$\begin{aligned} h(P') &= \max\{h(P_1'') + 1,\, h(P_2'') + 1\} \\ &\leq \max\{h(P_1') + 2,\, h(P_2) + 2,\, h(P_2') + 2,\, h(P_1) + 2\} \\ &\leq \max\{h(P_1) + 2,\, h(P_2) + 2\} \\ &= h(P) + 1 \;<\; 2 \cdot h(P). \end{aligned} \tag{5}$$

It is also easy to see that

$$\begin{aligned} |P'| &\leq (|P_1'| + |P_2| + 1) + (|P_2'| + |P_1| + 1) + 1 \\ &\leq ((|P_1| - 1) + |P_2| + 1) + ((|P_2| - 1) + |P_1| + 1) + 1 \\ &\leq 2|P_1| + 2|P_2| + 1 < |P|^2, \end{aligned}$$

since $|P_i'| < |P_i|$ due to the removal of at least one $\neg$:right or $\neg$:left inference from P_i. The fact that $P' \preccurlyeq_{\mathfrak{S}} P$ follows from the construction.

Now consider the case where the cut formula A is $B \vee C$. The inferences where $B \vee C$ can originate in P_1 as a principal formula are weakenings, $\mathfrak{S}$-inferences, and $\vee$:right inferences

$$\frac{\Pi {\longrightarrow} \Lambda, B, C}{\Pi {\longrightarrow} \Lambda, B \vee C} \;\vee\text{:right} \tag{6}$$

Form a proof P_1' of $\Gamma {\longrightarrow} B, C, \Delta, B \vee C$ by replacing each such $\vee$:right inference (6) in P_1 with

$$\dfrac{\Gamma {\longrightarrow} \Delta, B, C}{\Gamma {\longrightarrow} B, C, \Delta, B \vee C} \text{ Weakening and exchanges}$$

and adding additional weak inferences to propagate the new occurrences of B, C down to the endsequent. If $\mathfrak{S}$-depth$(A) = d$ in the endsequent of P_1, then the $\mathfrak{S}$-depths of B and C occurring in the endsequent of P_1' are both $< d$. The $\mathfrak{S}$-depth of the $B \vee C$ in the endsequent of P' is either $-\infty$ or 0. If the depth is $-\infty$, use Lemma 4 to form a proof P_1'' of $\Gamma \longrightarrow B, C, \Delta$. If it is zero, form P_1'' with an $\mathfrak{S}$-depth zero cut as:

$$\dfrac{\begin{array}{c} P_1' \\ \vdots \\ \Gamma \longrightarrow B, C, \Delta, B \vee C \end{array} \qquad \begin{array}{c} P_2 \\ \vdots \\ B \vee C, \Gamma \longrightarrow \Delta \\ \hline\hline B \vee C, \Gamma \longrightarrow B, C, \Delta \end{array}\ \text{Weak inferences}}{\Gamma \longrightarrow B, C, \Delta}\ \text{Cut}$$

The inferences in P_2 where $B \vee C$ originates can be weak inferences, $\mathfrak{S}$-inferences, and $\vee$:left inferences

$$\frac{B, \Gamma \longrightarrow \Delta \qquad C, \Gamma \longrightarrow \Delta}{B \vee C, \Gamma \longrightarrow \Delta}\ \vee\text{:left} \tag{7}$$

Letting X denote either B or C, form a proof P_2^X of the sequent $B \vee C, \Gamma, X \longrightarrow \Delta$, by replacing each inference (7) with

$$\begin{array}{c} X, \Gamma \longrightarrow \Delta \\ \hline\hline B \vee C, \Gamma, X \longrightarrow \Delta \end{array}\ \text{Weakening and exchanges}$$

and propagate the new occurrence of X down to the endsequent, adding weak inferences as necessary to form a valid proof. If the occurrence of $B \vee C$ in the endsequent of P_2^X has $\mathfrak{S}$-depth $-\infty$, we can form a proof $P_2'^X$ of $\Gamma, X \longrightarrow \Delta$ using Lemma 4. Otherwise, it has $\mathfrak{S}$-depth zero, and we form $P_2'^X$ using a cut of $\mathfrak{S}$-depth zero against P_1. If $\mathfrak{S}$-depth$(A) = d$ in the endsequent of P_2, then $\mathfrak{S}$-depth$(X) < d$ in the endsequent of $P_2'^X$.

The desired proof P' is formed as

$$\begin{array}{c} \begin{array}{c} \begin{array}{c} P_1'' \\ \vdots \\ \Gamma \longrightarrow B, C, \Delta \end{array} \quad \begin{array}{c} P_2'^B \\ \vdots \\ \Gamma, B \longrightarrow \Delta \end{array} \\ \hline\hline \Gamma \longrightarrow C, \Delta \end{array} \quad \begin{array}{c} P_2'^C \\ \vdots \\ \Gamma, C \longrightarrow \Delta \end{array} \\ \hline\hline \Gamma \longrightarrow \Delta \end{array}$$

using weak inferences and two cuts of $\mathfrak{S}$-depth $< d$.

Note that the height of P' can be bounded by

$$\begin{aligned} h(P') &\leq \max\{h(P_1')+3,\, h(P_2)+3,\, h(P_2^B)+3,\, h(P_1)+3,\, h(P_2^C)+2\} \\ &\leq \max\{h(P_1)+3,\, h(P_2)+3\} \\ &= h(P)+2 \;\leq\; 2\cdot h(P) \end{aligned} \tag{8}$$

since $h(P) \geq 2$. Also,

$$\begin{aligned} |P'| &\leq |P_1'| + |P_2| + |P_2^B| + |P_2^C| + 2|P_1| + 5 \\ &\leq (|P_1|-1) + |P_2| + 2(|P_2|-1) + 2|P_1| + 5 \\ &\leq 3|P_1| + 3|P_2| + 2 \\ &< (1+|P_1|+|P_2|)^2 \;=\; |P|^2. \end{aligned}$$

since $|P_1|, |P_2| \geq 1$.

Now consider the case where the cut formula A is a universal formula $(\forall x)B(x)$. The inferences where $(\forall x)B(x)$ can originate in P_1 are weakenings, $\mathfrak{S}$-inferences, and $\forall$:right inferences

$$\frac{\Pi \longrightarrow \Lambda, B(c_i)}{\Pi \longrightarrow \Lambda, (\forall x)B(x)}\ \forall\text{:right} \tag{9}$$

where c_i is an eigenvariable. (Of course, Π, Λ, and c_i are different for each inference (9).) Letting c be a new variable, we form a proof P_1' of $\Pi \longrightarrow B(c), \Lambda, (\forall x)B(x)$, by replacing each inference (9) with

$$\frac{\Pi \longrightarrow \Lambda, B(c)}{\Pi \longrightarrow B(c), \Lambda, (\forall x)B(x)}\ \text{Weakening and exchanges}$$

and replacing all occurrences of all eigenvariables c_i with c, adding the formula $B(c)$ to every sequent below each inference (9), and adding weak inferences as needed to form a valid proof. From P_1', we form a proof P_1'' of $\Gamma \longrightarrow B(c), \Delta$. Namely, if the formula $(\forall x)B(x)$ has $\mathfrak{S}$-depth $-\infty$ in the endsequent of P_1', then use Lemma 4, and if it has $\mathfrak{S}$-depth zero then combine P_1' and P_2 with an $\mathfrak{S}$-depth zero cut to form P_1''. If $\mathfrak{S}\text{-depth}(A) = d$ in the endsequent of P_1, then $\mathfrak{S}\text{-depth}(B(c)) < d$ in the endsequent of P_1''.

For a term t not containing any eigenvariable from P_1, we write $P_1''(t)$ to denote the result of replacing c everywhere in P_1'' with t. $P_1''(t)$ is still a valid proof, and the $\mathfrak{S}$-depths of formulas in $P_1''(t)$ are unchanged from their $\mathfrak{S}$-depths in P''.

The inferences in P_2 where direct ancestors of $(\forall x)B(x)$ originate can be weakenings, $\mathfrak{S}$-inferences, and $\forall$:left inferences

$$\frac{B(t_j), \Pi \longrightarrow \Lambda}{(\forall x)B(x), \Pi \longrightarrow \Lambda}\ \forall\text{:left} \tag{10}$$

where $\mathfrak{S}$-depth$(B(t_j)) < d$ if $\mathfrak{S}$-depth$(A) = d$ in the endsequent of P_2.

We form a proof P_2' with the same endsequent $(\forall x)B(x), \Gamma \longrightarrow \Delta$ as P_2, but with the $\mathfrak{S}$-depth of $(\forall x)B(x) \leq 0$ in P_2', and additionally with $P_2' \preccurlyeq_{\mathfrak{S}} P_2$. To construct P_2', replace each inference (10) with

$$\dfrac{\dfrac{\begin{array}{c} P_1''(t_j) \\ \vdots \\ \Gamma \longrightarrow B(t_j), \Delta \end{array} \qquad B(t_j), \Pi \longrightarrow \Lambda}{\Pi, \Gamma \longrightarrow \Delta, \Lambda}\ \text{Weak inferences and a cut}}{(\forall x)B(x), \Pi, \Gamma \longrightarrow \Delta, \Lambda}\ \text{Weakening}$$

The cut on $B(t_j)$ has $\mathfrak{S}$-depth $< d$; the formula $(\forall x)B(x)$ is now introduced by weakening. The newly appearing formulas Γ and Δ are propagated down to the endsequent, adding weak inferences as necessary to make P_2'' a valid proof.

The formula $(\forall x)B(x)$ in the endsequent of P_2' has $\mathfrak{S}$-depth equal to either $-\infty$ or zero. If it is $-\infty$, Lemma 4 gives the desired proof P' of $\Gamma \longrightarrow \Delta$. Otherwise, form P' by combining P_1 and P_2' with a cut of $\mathfrak{S}$-depth zero. The height of P' is

$$\begin{aligned} h(P') &\leq \max\{h(P_1'') + h(P_2) + 1,\ h(P_1) + 1\} \\ &\leq \max\{\max\{h(P_1){+}1, h(P_2){+}1\} + h(P_2) + 1,\ h(P_1) + 1\} \\ &\leq \max\{h(P_1) + h(P_2) + 2,\ 2h(P_2) + 2)\} \\ &\leq 2 \cdot h(P). \end{aligned}$$

The size of P' is

$$\begin{aligned} |P'| &\leq |P_1''| \cdot |P_2| + |P_2| + |P_1| + 1 \\ &\leq (|P_1| + |P_2|) \cdot |P_2| + |P_2| + |P_1| + 1 \\ &< (|P_1| + |P_2| + 1)^2 \ = \ |P|^2. \end{aligned}$$

Finally, suppose the cut formula A is atomic so $d = 0$ or 1. The inferences in P_1 and P_2 where direct ancestors of the cut formula originate can be weakenings, $\mathfrak{S}$-inferences, and initial sequents $A \longrightarrow A$. We form a proof P_1' of

$\Gamma{\longrightarrow}\Delta, A$ as follows. Each logical initial sequent $A{\longrightarrow}A$ in P_1 which contains a direct ancestor of the cut formula is replaced by a copy of P_2 plus a weakening:

$$\dfrac{\begin{array}{c} P_2 \\ \ddots\vdots\,\cdot^{\cdot^{\cdot}} \\ A, \Gamma{\longrightarrow}\Delta \end{array}}{A, \Gamma{\longrightarrow}\Delta, A}$$

so that the A in the succedent has $\mathfrak{S}$-depth $-\infty$. The new occurrences of Γ and Δ are propagated to the endsequent, using weak inferences as necessary to keep it a valid proof. In the end, we have a proof P_1' of $\Gamma{\longrightarrow}\Delta, A$, with the $\mathfrak{S}$-depth of A equal to either $-\infty$ or zero, and with $P_1' \preccurlyeq_{\mathfrak{S}} P_1$. A proof P_2' of $A, \Gamma{\longrightarrow}\Delta$ is formed similarly, again with $\mathfrak{S}$-depth$(A) \leq 0$ and $P_2' \preccurlyeq_{\mathfrak{S}} P_2$.

If $\mathfrak{S}$-depth$(A) = -\infty$ in the endsequent of either P_1' or P_2', then Lemma 4 gives us the desired proof of $\Gamma{\longrightarrow}\Delta$. Otherwise, we combine P_1' and P_2' with a cut on A to form the proof P'. Note this cut is not free, since both cut formulas have $\mathfrak{S}$-depth zero.

The height of P' is bounded by

$$h(P') \;\leq\; h(P_1) + h(P_2) + 1 \;<\; 2 \cdot h(P). \tag{11}$$

The size of P' can be bounded by

$$|P'| \;\leq\; |P_2| \cdot ((c-1)|P_1| + 1) + |P_1| \cdot ((c-1)|P_2| + 1) + 1 \;\leq\; (c-1) \cdot |P|^2,$$

since $(c-1)|P_i| + 1$ is an upper bound on the number of initial sequents in in $|P_i|$. Q.E.D. Lemma 6

The next lemma uses Lemma 6 iteratively to remove all free cuts of $\mathfrak{S}$-depth d.

Lemma 7 *Suppose $\mathfrak{S}$-depth$(P) \leq d$, where $d \geq 0$. Then there is a proof P', of the same endsequent, in which all free cuts have $\mathfrak{S}$-depth $< d$ and $h(P') < 2^{h(P)+1}$ and $P' \preccurlyeq_{\mathfrak{S}} P$.*

Proof The proof is by induction on the height of P, using the bounds obtained during the proof of Lemma 6. Let $f(i)$ be the least integer such that, for all P, if $h(P) \leq i$, then $h(P') \leq f(i)$. We have $f(0) = 0$ since in this case there are no cuts in P at all. Next, suppose P has height $i = 1$. If P does not contain a free cut, take $P' = P$. Otherwise, P's only non-weak, non-initial inference is a free cut: the cut formula must be atomic, and, by (11), Lemma 6 gives P' such that $h(P') < 2$. Thus, $f(1) = 1$.

Now suppose $i \geq 2$. W.l.o.g., P ends with a non-weak inference. Apply the induction hypothesis to the immediate subproof(s) of P to transform each immediate subproof P_j into a proof $P'_j \preccurlyeq_{\mathfrak{S}} P_j$ of height $\leq f(i-1)$ in which all free cuts have $\mathfrak{S}$-depth $< d$. Form P'' from P by replacing each P_j with P'_j. If P'' does not end with a free cut, take $P' = P''$ and this directly gives P' of height $\leq f(i-1)+1$. Suppose instead that P'' ends with a free cut. By the conditions $P'_j \preccurlyeq_{\mathfrak{S}} P_j$, the cut must have $\mathfrak{S}$-depth $\leq d$. If it has $\mathfrak{S}$-depth strictly less than d, just set $P' = P''$. Otherwise, apply Lemma 6 to P'' to form the desired proof P'; the proof P' has height $\leq 2f(i-1)+2$.

We have proved that $f(0) = 0$ and, for all $i > 0$, $f(i) \leq 2f(i-1)+2$. Thus, by induction, $f(i) < 2^{i+1} - 1$ for all i. In particular, $f(i) < 2^{i+1}$. □

To prove Theorem 5, Lemma 7 used $d+1$ times gives a proof P'' of height $< 2_{d+1}^{h(P)+1}$. Every free cut in P'' has $\mathfrak{S}$-depth < 0, i.e., $\mathfrak{S}$-depth $-\infty$. Lemma 4 gives the desired proof P' with no free cuts. Q.E.D. Theorem 5

4 Eliminating propositional cuts

The bounds in Theorem 5 apply to first-order logic. The proof, however, gives somewhat better bounds for cut elimination in propositional logic. The definition of $\mathfrak{S}$ being a set of non-logical axioms still makes sense for propositional logic, but the notion of being closed under term substitution does not apply and this requirement is dropped. In most applications, $\mathfrak{S}$ is a set of 0-ary inferences, namely an arbitrary set of non-logical initial sequents.

Theorem 8 *Let P be a proof in propositional logic over a set $\mathfrak{S}$ of non-logical inferences. Let $d = \mathfrak{S}$-depth(P) and assume $d > 0$. Then there is a proof P' of the same endsequent as P such that $h(P') < 3^d \cdot h(P)$ and such that every cut in P' either (a) has $\mathfrak{S}$-depth zero or (b) has $\mathfrak{S}$-depth one and has an atomic formula as cut formula. Furthermore $P' \preccurlyeq_{\mathfrak{S}} P$.*

Note that P' can still contain free cuts as Theorem 8 does not remove all cuts on atomic formulas. The proof uses the next lemma.

Lemma 9 *Let P and d be as above. There is a proof P' with the same endsequent as P such that $h(P') < 3 \cdot h(P)$, and $P' \preccurlyeq_{\mathfrak{S}} P$, and every cut in P' either (a) has $\mathfrak{S}$-depth $< d$, or (b) has an atomic formula as its cut formula.*

Proof Let $f(i)$ be the least value such that if $h(P) \leq i$, then $h(P') \leq f(i)$. It is immediate that $f(1) = 1$. For $i \geq 2$, we have $f(i) \leq (f(i-1)+1)+2 = f(i-1)+3$ by the bound (a) discussed after Lemma 6. Thus $f(i) < 3i$ for all $i \geq 1$. Since $d > 0$, P contains at least one cut and has height ≥ 1, so this proves the lemma. □

To prove Theorem 8, use induction on d and Lemma 9 to prove that there exists a proof P'' with the $h(P'') < 3^d \cdot h(P)$ and such that every cut in P'' either has $\mathfrak{S}$-depth ≤ 0 or satisfies condition (b). Then obtain the desired P' from P'' by using Lemma 4.

It is interesting to note that, in first-order logic, the above construction also allows eliminating cuts on formulas which have outermost connective a propositional connective (while allowing cuts on formulas that have outermost connective a quantifier). The bound 3^d still applies, where now d is the maximum nesting of propositional connectives outside of quantifiers. In the setting of pure first-order logic, with no non-logical axioms, the factor 3^d can be replaced by 2^d. This is used already by Zhang [14, Corollary 2.16] and Gerhardy [7] to get improved bounds on the size of cut free proofs. Namely, they show that, for first order logic, if n bounds the nesting depth of quantifiers in cut formulas, then cuts can be eliminated with an increase in proof height bounded by $2_{n+2}^{\alpha \cdot h(P)}$ where α is slightly bigger than 1.[4] Similar bounds hold for free-cut elimination, but we omit formalizing this here.

5 Partial cut elimination

Partial cut elimination refers to the property of being able to restrict cut formulas to lie in a given complexity class. This section shows that Theorem 5 can be used to prove partial cut elimination.

Let Φ be a set of formulas, and assume that Φ is closed under the operations of taking subformulas and replacing terms with other terms. Examples of Φ include the sets Σ_i^0 and Π_i^0 in Peano arithmetic, or Σ_i^{b} and Π_i^{b} in bounded arithmetic.

Theorem 10 *Suppose every $\mathfrak{S}$-inference has only Φ formulas as principal formulas. Let P be a proof. Then there is a proof P' of the same endsequent such that every cut in P' has cut formula in Φ.*

[4]This slightly generalizes the bounds of Zhang and Gerhardy, but follows immediately from their construction as $h(P)$ bounds the nesting of propositional connectives in any cut formula.

Proof Define $\mathfrak{S}'$ to be the inferences of $\mathfrak{S}$ plus all non-weak, non-$\mathfrak{S}$ logical inferences which have principal formula in Φ. In addition, for each atomic formula A in Φ, add the initial sequent $A{\longrightarrow}A$ to $\mathfrak{S}'$. In P, interpret every logical inference or initial sequent with principal formula in Φ as an $\mathfrak{S}'$-inference. Theorem 5 gives the desired proof P' over $\mathfrak{S}'$ with no free cuts.

The cuts in P' have $\mathfrak{S}'$-depth zero and therefore must have cut formulas in Φ. □

Theorem 5 also gives a bound on the height and size of P'. Define the Φ-*depth* of a formula by letting every formula in Φ have Φ-depth zero and, for $A \notin \Phi$, letting the Φ-depth of A equal one plus the maximum Φ-depth of proper subformulas of A. Clearly, any Φ-depth d formula appearing in P has $\mathfrak{S}'$-depth equal to either $-\infty$ or d. Let d be the maximum Φ-depth of any cut formula in P. Using Lemma 6 and arguing by induction as in the proofs of Lemma 7 and Theorem 5, one can show that the proof P' of Theorem 10 has height bounded by $h(P') \leq 2_d^{h(P)+1}$.

The proof of Theorem 10 used a special set $\mathfrak{S}'$ of non-logical inferences, based on a set of formulas Φ which is closed under subformulas. For this set $\mathfrak{S}'$, Lemma 6 can be strengthened by replacing the bounds (a)-(c) after Lemma 6 with the respective bounds

(a') $h(P') \leq h(P) + 1$.

(b') $h(P') \leq 2 \cdot h(P) - 2$.

(c') $h(P') \leq h(P_1) + h(P_2)$.

To prove this recall that, in the proof of Lemma 6, there were various places where it was sometimes possible to use Lemma 4 to form new proofs instead of adding a cut of $\mathfrak{S}$-depth zero. In particular, this arises when forming the proofs of P_1'' and P_2'' in the case $\neg$, the proofs of P_1'' and $P_2'^X$ in the case of $\vee$, the proofs of P_1'' and P' in the case of $\forall x$, and the proof of P' in the case of atomic formulas. Note that a formula A not in Φ can never have $\mathfrak{S}'$ depth equal to zero. Therefore, in each of the cases listed, Lemma 4 is used instead of adding an $\mathfrak{S}$-depth zero cut.

The bounds (a')-(c') for the case of pure first-order logic match results of Zhang [14] and Gerhardy [6].

6 Eliminating all cuts

For general sets $\mathfrak{S}$ of non-logical inferences, one cannot expect to eliminate *all* cuts, since it may be unavoidable to have some anchored cuts. There are, however, some special cases where all cuts can be eliminated. As an example of this, it is a consequence of Theorem 5 that the inferences (2) of Negri and von Plato admit elimination of all cuts. In fact, letting $\mathfrak{S}$ be any (acceptable) set of inferences of the form (2), we claim that a free-cut free proof P cannot contain *any* cuts. This is proved using the fact that every principal formula of an $\mathfrak{S}$-inference is atomic, and is in the antecedent of the conclusion of the inference. Any cut in P must have $\mathfrak{S}$-depth zero and its cut formula A must be atomic. Since the cut is not free, both occurrences of the cut formula A must have $\mathfrak{S}$-depth zero. But it is impossible for the occurrence of A in the succedent of the upper left hypothesis to have $\mathfrak{S}$-depth zero, since the $\mathfrak{S}$-inferences (2) have no principal formula in the succedent. Therefore, P cannot contain any cuts.

Negri and von Plato's methods allowed arbitrary quantifier-free initial sequents to be transformed into inference rules that admit complete cut elimination; for example, they used this to formulate a sequent calculus proof system for first-order logic with equality that admits complete cut elimination. A different approach was taken by Takeuti [12], who used generalized equality axioms to form a proof system for first-order logic with equality that admits complete cut elimination. The *generalized equality axioms* are the sequents that can be derived from the (ordinary) equality axioms, expressed as sequents, using only exchanges, contractions and cuts. Takeuti showed that when the generalized equality axioms are allowed as initial sequents, then all cuts can be eliminated. It is also well-known that this holds for any set of non-logical initial sequents which contain only atomic formulas provided the set of initial sequents is closed under cuts.

We generalize Takeuti's construction by considering arbitrary sets $\mathfrak{S}$ of inferences which are closed under cut, and have only atomic formulas as principal formulas. For technical reasons, it is slightly easier to deal with being closed under "mixes" rather than cuts.

Definition A *Mix* inference is an inference of the form

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

such that there is a *mix* formula A so that Π', respectively Δ', is obtained from Π, respectively Δ, by removing one or more occurrences of the formula A.

Note that a cut inference is a special case of a mix. Conversely, a mix inference can be simulated using weak inferences and cut.

Definition A set $\mathfrak{S}$ of inferences is *closed under mix*, provided that the following two properties hold. First, if $\mathcal{I}$ is an $\mathfrak{S}$-inference skeleton and σ is a term substitution that respects the eigenvariable conditions of $\mathcal{I}$, then $\mathcal{I}\sigma$ is also an $\mathfrak{S}$-inference skeleton. Second, suppose that $\mathcal{I}$ and $\mathcal{I}'$ are $\mathfrak{S}$-inference skeletons, which are k-ary and k'-ary, respectively, and they have conclusions

$$\Psi, \mathcal{C} \longrightarrow \mathcal{D}, \Xi \qquad \text{and} \qquad \Psi', \mathcal{C} \longrightarrow \mathcal{D}, \Xi',$$

respectively. Let $\Psi'' \longrightarrow \Xi''$ be obtained by a mix from $\Psi \longrightarrow \Xi$ and $\Psi' \longrightarrow \Xi'$. Then there is an inference skeleton in $\mathfrak{S}$ of arity $\leq k+k'$ which has as hypotheses a subset of the hypotheses of $\mathcal{I}$ and $\mathcal{I}'$, and which has the conclusion $\Psi'', \mathcal{C} \longrightarrow \mathcal{D}, \Xi''$.

Theorem 11 *Suppose $\mathfrak{S}$ is closed under mix, and all principal formulas of $\mathfrak{S}$-inferences are atomic. If there is a proof P of $\Gamma \longrightarrow \Delta$ of $\mathfrak{S}$-depth d, then there is a cut free proof P' of the same endsequent with $h(P') < 2_{d+1}^{h(P)+1}$.*

Theorem 11 will be proved by a construction similar to Lemma 7. The construction is however more complicated in the present setting due to the symmetric nature of $\mathfrak{S}$-inferences with atomic principal formulas.

Lemma 12 *Let $\mathfrak{S}$ satisfy the hypotheses of Theorem 11. Suppose P contains a single cut, as its final inference, and that the cut formula is atomic. Then there is a proof P' of the same endsequent containing no cuts, with $|P'| < |P|^2$ and $h(P') \leq h(P_1) + h(P_2)$, where P_1 and P_2 are the two immediate subproofs of P.*

Proof We use the conventions and notations of the proof of Lemma 6. In P_1, locate all inferences where a direct ancestor of the cut formula A originates, and let I_{src} be the set containing their lower sequents. Thus, a sequent in I_{src} is either an initial sequent $A \longrightarrow A$ or the lower sequent of an $\mathfrak{S}$-inference

$$\frac{\{\Psi_\ell, \Pi \longrightarrow \Lambda, \Xi_\ell\}_{\ell=1}^{k}}{\Psi, \Pi \longrightarrow \Lambda, \Xi} \tag{12}$$

where at least one direct ancestor of the cut formula A is present in Ξ. The cedents Π and Λ contain the side formulas of the $\mathfrak{S}$-inference and may not be present in all upper sequents. Let S_0 be the endsequent of P_1, namely $\Gamma \longrightarrow \Delta, A$.

Similarly define J_{src} to be the set of sequents where the cut formula A originates in P_2. These sequents are initial sequents $A \longrightarrow A$ or are inferred by $\mathfrak{S}$-inferences (12), but now with the direct ancestor of the cut formula present in Ψ. Likewise, let T_0 be the endsequent, $A, \Gamma \longrightarrow \Delta$, of P_2.

For S a sequent in P, let P_S be the subproof of P with endsequent S. If S and T are sequents in P_1 and P_2, respectively, define the sequent $mix_A(S,T)$ as follows. Suppose

$$S \text{ is } \Pi^S \longrightarrow \Lambda^S \qquad \text{and} \qquad T \text{ is } \Pi^T \longrightarrow \Lambda^T.$$

Then $mix_A(S,T)$ is

$$\Pi^S, \Pi^{T*} \longrightarrow \Lambda^{S*}, \Lambda^T,$$

where Π^{T*} and Λ^{S*} are Π^T and Λ^S, respectively, with all direct ancestors (if any) of the cut formula A removed.

The idea for proving Lemma 12 is that, for each sequent S from P_1 and sequent T from P_2, we construct a cut free proof $P_{S,T}$ of $mix_A(S,T)$, of height $\le h(P_S) + h(P_T)$. Then P_{S_0,T_0} will be the desired proof P'. The size and height bounds of Lemma 12 will be immediate from the construction.

We do not actually need to form $P_{S,T}$ for all pairs of sequents S and T; instead, we only define $P_{S,T}$ when at least one of S and T contain a direct ancestor of the cut formula A. (In fact, not even all of these are needed.) If S does not contain a direct ancestor of A, then Λ^{S*} is the same as Λ^S and we can define $P_{S,T}$ to be the proof obtained by adding weak inferences to the end of P_S to introduce the formulas in Π^{T*} and Λ^T. $P_{S,T}$ is defined similarly from P_T if T does not contain a direct ancestor of A.

Now suppose both S and T contain a direct ancestor of A. If S, respectively T, is an initial sequent $A \longrightarrow A$, then $P_{S,T}$ is defined to be just P_T, respectively P_S (plus weak inferences to reorder the formulas in the sequent).

If $S \notin I_{\rm src}$, then S is inferred in P_1 by a logical inference:

$$\frac{S_1}{S} \qquad \text{or} \qquad \frac{S_1 \qquad S_2}{S}$$

where the principal formula of the inference is not a direct ancestor of the cut formula A. In this case, $P_{S,T}$ is formed by using the same kind of inference to infer $mix_A(S,T)$ from the proof(s) $P_{S_i,T}$:

$$\frac{mix_A(S_1,T)}{\overline{mix_A(S,T)}} \qquad \text{or} \qquad \frac{mix_A(S_1,T) \qquad mix_A(S_2,T)}{\overline{mix_A(S,T)}}$$

where the double line means that weak inferences may be needed to reorder the formulas in the sequents.

$P_{S,T}$ is formed dually if $T \notin J_{\rm src}$. If neither $S \in I_{\rm src}$ nor $T \in J_{\rm src}$, there are two possible ways to form $P_{S,T}$: either way may be used.

Finally, consider the case where both S and T are inferred by $\mathfrak{S}$-inferences:

$$\frac{S_1 \quad S_2 \quad \cdots \quad S_k}{S} \qquad \text{and} \qquad \frac{T_1 \quad T_2 \quad \cdots \quad T_{k'}}{T}$$

where $k, k' \geq 0$, and S and T both contain a direct ancestor of the cut formula. Let $s_1, \ldots, s_k$ and $t_1, \ldots, t_{k'}$ be the side formula indicators for the two $\mathfrak{S}$-inferences. Let the notation $m_A^{s_i}(S_i, T)$ denote $mix_A(S_i, T)$ if $s_i = 1$ and denote just S_i if $s_i = 0$. Define the notation $m_A^{t_j}(S, T_j)$ similarly. Then, by the closure of $\mathfrak{S}$ under mix, we can form a proof of $mix_A(S,T)$ by using a single $\mathfrak{S}$-inference and weak inferences:

$$\frac{m_A^{s_1}(S_1, T) \quad \cdots \quad m_A^{s_k}(S_k, T) \quad m_A^{t_1}(S, T_1) \quad \cdots \quad m_A^{t_{k'}}(S, T_{k'})}{mix_A(S,T)}$$

where possibly some of the upper sequents are omitted.

That completes the proof of Lemma 12. □

We can now prove Theorem 11. By applying Lemma 7 d times, there is a proof P'' with the same endsequent as P such that all cuts in P' are on atomic formulas and such that $h(P'') < 2_d^{h(P)+1}$. Now using Lemma 12, and using induction on the height of P'' as in the proof of Lemma 7, we obtain the desired cut free proof P' with $h(P') < 2^{h(P'')+1} \leq 2_{d+1}^{h(P)+1}$.

Acknowledgements. We thank William Scott for extensive email correspondence, and for corrections to the first draft of this paper. We also thank Sara Negri, and two anonymous referees, for comments and corrections.

References

[1] M. Baaz and A. Leitsch, *Cut-elimination and redundancy-elimination by resolution*, Journal of Symbolic Computation, 29 (2000), pp. 149–176.

[2] ——, *Fast cut-elimination by CERES*. To appear in *Proofs, Categories and Computations*, College Publications, 2010.

[3] S. R. Buss, *Bounded Arithmetic*, Bibliopolis, 1986. Revision of 1985 Princeton University Ph.D. thesis.

[4] ——, *A conservation result concerning bounded theories and the collection axiom*, Proceedings of the American Mathematical Society, 100 (1987), pp. 709–716.

[5] ——, *An introduction to proof theory*, in Handbook of Proof Theory, S. R. Buss, ed., North-Holland, 1998, pp. 1–78.

[6] P. Gerhardy, *Refined complexity analysis of cut elimination*, in Proc. 17th Workshop on Computer Science Logic (CSL), Lecture Notes in Computer Science #2803, Springer Verlag, 2003, pp. 212–225.

[7] ——, *The role of quantifier alternations in cut elimination*, Notre Dame Journal of Formal Logic, 46 (2005), pp. 165–171.

[8] S. Negri and J. von Plato, *Cut elimination in the presence of axioms*, Bulletin of Symbolic Logic, 4 (1998), pp. 418–435.

[9] ——, *Structural Proof Theory*, Cambridge University Press, 2001. With an appendix by Aarna Ranta.

[10] V. P. Orevkov, *Upper bound on the lengthening of proofs by cut elimination*, Journal of Soviet Mathematics, 34 (1986), pp. 1810–1819. Original Russian version in Zap. Nauchn. Sem. L.O.M.I. Steklov **137** (1984) 87-98.

[11] C. Parsons, *On n-quantifier induction*, Journal of Symbolic Logic, 37 (1972), pp. 466–482.

[12] G. Takeuti, *Proof Theory*, North-Holland, Amsterdam, 2nd ed., 1987.

[13] A. S. Troelstra and H. Schwichtenberg, *Basic Proof Theory*, Tracts in Theoretical Computer Science #43, Cambridge University Press, Cambridge, 2nd ed., 2000.

[14] W. Zhang, *Cut elimination and automatic proof procedures*, Theoretical Computer Science, 91 (1991), pp. 265–284.