EPiC Series in Computing
Volume 100, 2024, Pages 232–245
Proceedings of 25th Conference on Logic for Programming, Artificial Intelligence and Reasoning

On Translations of Epsilon Proofs to **LK**

Matthias Baaz[1]*and Anela Lolić[2]†

[1] Institute of Discrete Mathematics and Geometry,
TU Wien, Vienna, Austria
baaz@logic.at
[2] Kurt Gödel Society,
Institute of Logic and Computation,
TU Wien, Vienna, Austria
anela@logic.at

Abstract

In this paper we present the proof that there is no elementary translation from cut-free derivations in the sequent calculus variant of the epsilon calculus to **LK**-proofs with bounded cut-complexity. This is a partial answer to a question by Toshiyasu Arai. Furthermore, we show that the intuitionistic format of the sequent calculus variant of the epsilon calculus is not sound for intuitionistic logic, due to the presence of all classical quantifier-shift rules.

1 Introduction

The epsilon calculus gives the impression to provide shorter proofs than other proof mechanisms. Consequently, Toshiyasu Arai asked:

"What is the expense to translate a derivation with cuts in the sequent formulation of epsilon calculus to **LK** derivations with cuts?"[1].

We provide as a partial answer to the question a proof that there is no elementary translation from cut-free derivations in the sequent calculus variant of the epsilon calculus to **LK**-proofs with bounded cut-complexity. The main property of the epsilon calculus used is its ability to overbind bound variables.

The intuitionistic format of the sequent calculus variant of the epsilon calculus is not sound for intuitionistic logic, due to the presence of all classical quantifier-shift rules. The quantifier shift rules that are classically valid but not valid in intuitionistic logic are

$$\forall x(A(x) \lor B) \to \forall x A(x) \lor B,$$

$$(\forall x A(x) \to B) \to \exists x(A(x) \to B),$$

$$(A \to \exists x B(x)) \to \exists x(A \to B(x)).$$

*This work is partially supported by FWF project P 36571.
†This work is partially supported by FWF project I 5848.
[1]From personal communication.

2 Epsilon Calculus

Hilbert's epsilon formalism, which is also the oldest framework for proof theory [4], uses epsilon terms to represent

$$\exists x A(x) \quad \text{by} \quad A(\varepsilon_x A(x)).$$

Consequently, the representation of

$$\forall x A(x) \quad \text{is} \quad A(\varepsilon_x \neg A(x)).$$

The language of epsilon calculus is based on the term language of epsilon expressions and other function symbols and on propositional language otherwise.

Definition 1. *Let L be a first-order language. The set of terms, epsilon terms and the set of formulas of L are defined as follows:*

- *each constant of L is a term,*
- *each variable is a term,*
- *if $s_1, \ldots, s_k$ are terms and f is a k-ary function symbol of L, then $f(s_1, \ldots, s_k)$ is a term,*
- *if $s_1, \ldots, s_k$ are terms and R is a k-ary relation symbol of L, then $R(s_1, \ldots, s_k)$ is a formula,*
- *if A and B are formulas, then $\neg A$ and $A \circ B$, for $\circ \in \{\wedge, \vee, \rightarrow\}$ are formulas,*
- *if A is a formula and x is a variable, then $\varepsilon_x A(x)$ is an epsilon term.*

Expressions in first-order language can be easily translated to expressions in the epsilon calculus language.

Definition 2. *Let A be a formula in first-order language. Its epsilon translation is denoted as $[A]^\varepsilon$ and inductively defined as*

- *A is an atom, then $[A]^\varepsilon = A$.*
- *$A = \neg B$, then $[A]^\varepsilon = \neg [B]^\varepsilon$.*
- *$A = B \circ C$, where $\circ \in \{\wedge, \vee, \rightarrow\}$ and B and C formulas, then $[A]^\varepsilon = [B]^\varepsilon \circ [C]^\varepsilon$.*
- *$A = \exists x A'(x)$, then $[A]^\varepsilon = [A'(\varepsilon_x A'(x))]^\varepsilon$.*
- *$A = \forall x A'(x)$, then $[A]^\varepsilon = [A'(\varepsilon_x \neg A'(x))]^\varepsilon$.*

For sequents $S\colon A_1, \ldots, A_n \vdash B_1, \ldots, B_m$ the epsilon translation is $[S]^\varepsilon : [A_1]^\varepsilon, \ldots, [A_n]^\varepsilon \vdash [B_1]^\varepsilon, \ldots, [B_m]^\varepsilon$.

The intended understanding of the term $\varepsilon_x A(x)$ is that it denotes some x satisfying A, if there is one. The epsilon calculus is only based on the representation by critical formulas

$$A(t) \rightarrow A(\varepsilon_x A(x)) \quad \text{for} \quad A(t) \rightarrow \exists x A(x)$$

and propositional axioms and rules, the unrestricted deduction theorem of propositional calculus transfers to this formalization of first-order logic:

Definition 3. *The epsilon proof in its classical form is a propositional tautology*

$$(\bigwedge_{i=1}^{n} A_i(t_i) \to A_i(\varepsilon_x A_i(x))) \to E.$$

Note that E is $[E']^\varepsilon$, E' with weak quantifiers only, if E corresponds to a first-order formula.

Note that strong quantifier inferences[2] are deleted and replaced by substitutions of $\varepsilon_x \neg A(x)$ for $\forall x A(x)$ positive and $\varepsilon_x A(x)$ for $\exists x A(x)$ negative for eigenvariables. Valid propositional formulas do not influence an epsilon proof.

First-order logic can be embedded in the epsilon calculus. However, the converse is not true, as not every formula in the epsilon calculus can be translated to first-order logic. Let A be a formula in the epsilon calculus language, then we denote with $[A]^{\forall\exists}$ the translation form epsilon calculus language to first-order language, if $A = [B]^\varepsilon$ for some first-order expression B, and undefined otherwise.

Example 1. *Note that $[A]^{\forall\exists}$ for an epsilon calculus expression A does not always exist: Consider for example the axiom $[\forall x.x = x]^\varepsilon = \varepsilon_x \neg(x = x) = \varepsilon_x \neg(x = x)$. With help of the critical formula*

$$\varepsilon_x \neg(x = x) = \varepsilon_x \neg(x = x) \to \varepsilon_v(v = \varepsilon_x \neg(x = x)) = \varepsilon_x \neg x = x$$

we obtain

$$\varepsilon_v(v = \varepsilon_x \neg x = x) = \varepsilon_x \neg x = x$$

which has no meaning in first-order logic.

Because of the untractability of almost all nonclassical logics by any adaptation of the epsilon formalism and the clumsiness of the epsilon formalism itself, the epsilon calculus has never become popular in computational proof theory as it deserves.

Example 2. *Consider the first-order formula $\exists x \exists y \exists z\ A(x, y, z)$. Its epsilon translation is $[\exists x \exists y \exists z\ A(x, y, z)]^\varepsilon =$*

$$\begin{aligned}
&A(\varepsilon_x A(x, \varepsilon_y A(x, y, \varepsilon_z A(x, y, z)), \varepsilon_z A(x, \varepsilon_y A(x, y, \varepsilon_z A(x, y, z)), z)),\\
&\varepsilon_y A(\varepsilon_x A(x, \varepsilon_y A(x, y, \varepsilon_z A(x, y, z)), \varepsilon_z A(x, \varepsilon_y A(x, y, \varepsilon_z A(x, y, z)), z)), y,\\
&\varepsilon_z A(\varepsilon_x A(x, \varepsilon_y A(x, y, \varepsilon_z A(x, y, z)), \varepsilon_z A(x, \varepsilon_y A(x, y, \varepsilon_z A(x, y, z)), z)), y, z)),\\
&\varepsilon_z A(\varepsilon_x A(x, \varepsilon_y A(x, y, \varepsilon_z A(x, y, z)), \varepsilon_z A(x, \varepsilon_y A(x, y, \varepsilon_z A(x, y, z)), z)),\\
&\varepsilon_y A(\varepsilon_x A(x, \varepsilon_y A(x, y, \varepsilon_z A(x, y, z)), \varepsilon_z A(x, \varepsilon_y A(x, y, \varepsilon_z A(x, y, z)), z)), y,\\
&\varepsilon_z A(\varepsilon_x A(x, \varepsilon_y A(x, y, \varepsilon_z A(x, y, z)), \varepsilon_z A(x, \varepsilon_y A(x, y, \varepsilon_z A(x, y, z)), z)), y, z)), z))
\end{aligned}$$

The example above demonstrates that the language has to be adapted for concrete applications. What is lost is the generality of epsilon terms.

The historically first results on the epsilon calculus were the first and second epsilon theorems [4]. The first epsilon theorem states that if a formula without quantifiers or epsilon terms is provable in the epsilon calculus, then it is already provable in the quantifier-free calculus. The second epsilon theorem removes Skolem functions from Herbrand disjunctions of prenex formulas.

[2]Positive occurrences of universal and negative occurrences of existential quantifiers are called strong, all other quantifiers are called weak.

In the proofs of the above theorems the rank and degree of epsilon expressions are used. These are two measures of complexity of epsilon expressions, where the degree applies to epsilon terms only and measures the depth of nesting of epsilon terms, and the rank measures the complexity of cross-bindings of epsilon expressions.

Definition 4 (degree)**.** *The degree of an epsilon term is inductively defined as*

1. *If $A(x)$ contains no epsilon subterms, then $d(\varepsilon_x A(x)) = 1$.*
2. *If $e_1, \ldots, e_n$ are all immediate epsilon subterms of $A(x)$, then*

$$d(\varepsilon_x A(x)) = max\{d(e_1), \ldots, d(e_n)\} + 1.$$

Definition 5. *An epsilon expression e is subordinate to $\varepsilon_x A(x)$ if e is a proper sub semi-term of $A(x)$ and contains x (semi-terms are an extended notion of terms where bound variables may occur).*

Definition 6 (rank)**.** *The rank of an epsilon expression e is inductively defined as*

1. *If e contains no subordinate epsilon expressions, then $r(e) = 1$.*
2. *If $e_1, \ldots, e_n$ are all epsilon expressions subordinate to e, then*

$$r(e) = max\{r(e_1), \ldots, r(e_n)\} + 1.$$

The extended first epsilon theorem [4, 7] eliminates algorithmically the critical formulas of highest rank and within the highest rank after the highest degree obtaining a Herbrand disjunction [3]

$$\bigvee_{i=1}^{n} A(\overline{t_i}),$$

where A is the ε-translation of $\exists \overline{x} A'(\overline{x})$, A' being quantifier-free. The argument can be easily extended to formulas A' which contain only weak quantifiers.

Theorem 1 (extended first epsilon theorem)**.** *The critical formulas can be eliminated from every epsilon proof to obtain*

$$\bigvee_{i=1}^{n} A(\overline{t_i}).$$

Corollary 1. *Let n be the number of critical formulas in the proof described in the theorem above, then the Herbrand disjunction*

$$\bigvee_{i} A(t_{i1}, \ldots, t_{in})$$

is bounded by $4^{4^{\cdot^{\cdot^{\cdot}}}}\}^{n}$. (For cut-elimination the bound is $2^{2^{\cdot^{\cdot^{\cdot}}}}\}^{n}$. This is caused by the stronger expressivity of epsilon calculus.)

Note that the corollary above remains true even if equality is added to the sequent calculus [6].

Remark 1. *Recall that a function on the natural numbers is elementary if it can be defined by a quantifier-free formula from* $+$, $\times$, *and the function* $x \to 2^x$. *By independent results of R. Statman [9] and of V. P. Orevkov [8], the sizes of the smallest cut-free* **LK***-proofs of sequents of size* n *are not bounded by any elementary function on* n.

As the extended first epsilon theorem is faithful only for translations of formulas with weak quantifiers, strong quantifiers have to be replaced by Skolem terms. The second epsilon theorem allows to recover the original formulation.

The optimal calculation of Herbrand disjunctions from unformalized or formalized mathematical proofs is one of the most prominent problems in proof theory of first-order logic. Herbrand disjunctions can also be calculated via Gentzen's midsequent theorem, based on the cut-elimination theorem [2]. However, the extended first epsilon theorem has been the first complete and correct proof of Herbrand's theorem.

Moreover, the epsilon calculus is not sensitive w.r.t. addition of (arbitrary) tautologies. This implies that bounds depend on first-order features only. More precisely, the complexity of the Herbrand disjunction does not depend on the complexity of the derivation, but only on the number of critical formulas.

Furthermore, the extended first epsilon theorem develops the Herbrand disjunctions as contraposition of all case distinctions in the proof: the Herbrand disjunction may be understood as the once and for all case distinction. This explains why Herbrand disjunctions may be short even for high level mathematical proofs (cf. the complexity of the Herbrand disjunction in [5]). Another advantage is that a proof might be formalized disregarding propositional features: represent all substitutions in the proof; if the obtained epsilon proof is not a tautology, a substitution has been overlooked.

3 Lε and LK

To compare cut-free derivations we consider a sequent calculus format of the epsilon calculus.

Definition 7 (**L**ε)**.** **L**ε *is in the language of epsilon calculus.*

Axiom schema: $A \vdash A$, A *atomic.*

The inference rules are:

- *the structural rules:*

  for weakening

$$\frac{\Gamma \vdash \Delta}{A, \Gamma \vdash \Delta}\, w_l \qquad\qquad \frac{\Gamma \vdash \Delta}{\Gamma \vdash \Delta, A}\, w_r$$

  for contraction

$$\frac{A, A, \Gamma \vdash \Delta}{A, \Gamma \vdash \Delta}\, c_l \qquad\qquad \frac{\Gamma \vdash \Delta, A, A}{\Gamma \vdash \Delta, A}\, c_r$$

  for cut

$$\frac{\Gamma_1 \vdash \Delta_1, A \qquad A, \Gamma_2 \vdash \Delta_2}{\Gamma_1, \Gamma_2 \vdash \Delta_1, \Delta_2}\, cut$$

- *the propositional logical rules*

for conjunction

$$\frac{A, B, \Gamma \vdash \Delta}{A \wedge B, \Gamma \vdash \Delta} \wedge_l \qquad \frac{\Gamma_1 \vdash \Delta_1, A \qquad \Gamma_2 \vdash \Delta_2, B}{\Gamma_1, \Gamma_2 \vdash \Delta_1, \Delta_2, A \wedge B} \wedge_r$$

for disjunction

$$\frac{A, \Gamma_1 \vdash \Delta_1 \qquad B, \Gamma_2 \vdash \Delta_2}{A \vee B, \Gamma_1, \Gamma_2 \vdash \Delta_1, \Delta_2} \vee_l \qquad \frac{\Gamma \vdash \Delta, A, B}{\Gamma \vdash \Delta, A \vee B} \vee_r$$

for implication

$$\frac{\Gamma_1 \vdash \Delta_1, A \qquad B, \Gamma_2 \vdash \Delta_2}{A \to B, \Gamma_1, \Gamma_2 \vdash \Delta_1, \Delta_2} \to_l \qquad \frac{A, \Gamma \vdash \Delta, B}{\Gamma \vdash \Delta, A \to B} \to_r$$

for negation

$$\frac{\Gamma \vdash \Delta, A}{\neg A, \Gamma \vdash \Delta} \neg_l \qquad \frac{A, \Gamma \vdash \Delta}{\Gamma \vdash \Delta, \neg A} \neg_r$$

- *the quantifier rules:*

  the weak quantifier inferences $\exists_r$

$$\frac{\Pi \vdash \Delta, A(t)}{\Pi \vdash \Delta, A(\varepsilon_x A(x))} \exists_r$$

  and $\forall_l$

$$\frac{A(t), \Pi \vdash \Delta}{A(\varepsilon_x \neg A(x)), \Pi \vdash \Delta} \forall_l$$

  the strong quantifier inferences:

  $\exists_l$*: substitution of* $\varepsilon_x A'(x)$ *for the eigenvariable*

  $\forall_r$*: substitution of* $\varepsilon_x \neg A'(x)$ *for the eigenvariable.*

Definition 8. *An epsilon term* $\varepsilon_x A(x)$ *overbinds if it is generated by the following inferences*

$$A(f(e_1, \ldots, e_n), g(e_1, \ldots, e_n)) \to A(\varepsilon_x A(x, g(e_1, \ldots, e_n), g(e_1, \ldots e_n)),$$

$$\frac{\Pi \vdash \Delta, A(t)}{\Pi \vdash \Delta, A(\varepsilon_x A(x))} \exists_r$$

$$\frac{A(t), \Pi \vdash \Delta}{A(\varepsilon_x \neg A(x)), \Pi \vdash \Delta} \forall_l$$

Definition 9 (**LK**)**.** **LK** *is in the language of first-order logic. The rules are the same as* $\mathbf{L}\varepsilon$*, except for the quantifier rules:*

$\forall$*-introduction*

$$\frac{A(t), \Gamma \vdash \Delta}{\forall x A(x), \Gamma \vdash \Delta} \forall_l$$

$$\frac{\Gamma \vdash \Delta, A(a)}{\Gamma \vdash \Delta, \forall x A(x)} \forall_r$$

where t *is an arbitrary term and* a *is a free variable which may not occur in* Γ, Δ, A. a *is called an eigenvariable.*

$\exists$*-introduction*

$$\frac{A(a), \Gamma \vdash \Delta}{\exists x A(x), \Gamma \vdash \Delta} \exists_l$$

$$\frac{\Gamma \vdash \Delta, A(t)}{\Gamma \vdash \Delta, \exists x A(x)} \exists_r$$

where the variable conditions for $\exists_l$ *are the same as those for* $\forall_r$ *and similarly for* $\exists_r$ *and* $\forall_l$*. The quantifier-rules* $\forall_l, \exists_r$ *are called weak, the rules* $\exists_l, \forall_r$ *strong.*

Remark 2. *As the main result of this paper concerns non-elementary transformations the version of the sequent calculus* **LK** *considered is not essential.*

In this work we will use the notion of proof size and cut complexity.

Definition 10 (size)**.** *The size of a formula is the number of symbols occurring in it. The size of a sequent is the sum of the sizes of the formula occurrences in it. The size of an* **L**ε *(***LK***) derivation is the sum of the sizes of the sequents occurring in it.*

Definition 11 (cut complexity)**.** *The cut complexity is the number of symbols in the cut.*

Proposition 1. *Every* **LK**-*derivation possibly with cuts can be translated into an* **L**ε-*derivation of double exponential size.*

Proof. All inference steps are replaced by corresponding inference steps with exception of strong quantifier rules, which are replaced by substitution. This means, that every bound variable is replaced by an epsilon expression at most of the size of the existing derivation. □

Remark 3. *Note that the usual form of epsilon proofs can be obtained by deleting the quantifier inferences of* **L**ε*, and replacing them by*

$$\frac{\overset{(\psi')}{\Pi' \vdash \Delta', A'(t)} \qquad A'(\varepsilon_x A(x)) \vdash A'(\varepsilon_x A(x))}{A'(t) \to A'(\varepsilon_x A(x)), \Pi' \vdash \Delta'} \exists_r$$

and

$$\cfrac{\cfrac{\overset{(\psi')}{A'(t), \Pi' \vdash \Delta'} \qquad A'(\varepsilon_x \neg A(x)) \vdash A'(\varepsilon_x \neg A(x))}{A'(\varepsilon_x \neg A(x)) \to A'(t), \Pi' \vdash \Delta'} \forall_l}{\neg A'(t) \to \neg A'(\varepsilon_x \neg A(x)), \Pi' \vdash \Delta'}$$

Example 3. *Consider the sequent*

$$\exists y(A(y) \to \forall x A(x)).$$

Its shortest cut-free **LK**-*derivation is*

$$
\cfrac{\cfrac{\cfrac{\cfrac{\cfrac{\cfrac{\cfrac{\cfrac{\cfrac{A(a) \vdash A(a)}{A(a) \vdash A(a), \forall x A(x)}\, w_r}{\vdash A(a), A(a) \to \forall x A(x)}\, \to_r}{\vdash A(a), \exists y(A(y) \to \forall x A(x))}\, \exists_r}{\vdash \forall x A(x), \exists y(A(y) \to \forall x A(x))}\, \forall_r}{A(b) \vdash \forall x A(x), \exists y(A(y) \to \forall x A(x))}\, w_l}{\vdash A(b) \to \forall x A(x), \exists y(A(y) \to \forall x A(x))}\, \to_r}{\vdash \exists y(A(y) \to \forall x A(x)), \exists y(A(y) \to \forall x A(x))}\, \exists_r}{\vdash \exists y(A(y) \to \forall x A(x))}\, c_r}{}
$$

When we translate the sequent to epsilon calculus we obtain

$$\vdash A(\varepsilon_y(A(y) \to A(\varepsilon_x \neg A(x)))) \to A(\varepsilon_x \neg A(x)),$$

and the translation of the above proof to **L**ε *is*

$$
\cfrac{\cfrac{\cfrac{\cfrac{\cfrac{\cfrac{\cfrac{A(\varepsilon_x \neg A(x)) \vdash A(\varepsilon_x \neg A(x))}{A(\varepsilon_x \neg A(x)) \vdash A(\varepsilon_x \neg A(x)), A(\varepsilon_x \neg A(x))}\, w_r}{\vdash A(\varepsilon_x \neg A(x)), A(\varepsilon_x \neg A(x)) \to A(\varepsilon_x \neg A(x))}\, \to_r}{\vdash A(\varepsilon_x \neg A(x)), A(\varepsilon_y(A(y) \to A(\varepsilon_x \neg A(x)))) \to A(\varepsilon_x \neg A(x))}\, \exists_r}{A(b) \vdash A(\varepsilon_x \neg A(x)), A(\varepsilon_y(A(y) \to A(\varepsilon_x \neg A(x)))) \to A(\varepsilon_x \neg A(x))}\, (*) \;\; + w_l}{\vdash A(b) \to A(\varepsilon_x \neg A(x)), A(\varepsilon_y(A(y) \to A(\varepsilon_x \neg A(x)))) \to A(\varepsilon_x \neg A(x))}\, \to_r}{\vdash A(\varepsilon_y(A(y) \to A(\varepsilon_x \neg A(x)))) \to A(\varepsilon_x \neg A(x)), A(\varepsilon_y(A(y) \to A(\varepsilon_x \neg A(x)))) \to A(\varepsilon_x \neg A(x))}\, \exists_r}{\vdash A(\varepsilon_y(A(y) \to A(\varepsilon_x \neg A(x)))) \to A(\varepsilon_x \neg A(x))}\, c_r
$$

Note that the formula in the end-sequent above is $[\exists y(A(y) \to \forall x A(x))]^\varepsilon$*, and for* $(*)$*:* $\forall_r$ *has been replaced by the substitution of* $\varepsilon_x \neg A(x)$ *for* a*.*

The shortest cut-free derivation of the sequent

$$\vdash A(\varepsilon_y(A(y) \to A(\varepsilon_x \neg A(x)))) \to A(\varepsilon_x \neg A(x))$$

in **L**ε *is however*

$$
\cfrac{\cfrac{A(\varepsilon_x \neg A(x)) \vdash A(\varepsilon_x \neg A(x))}{\vdash A(\varepsilon_x \neg A(x)) \to A(\varepsilon_x \neg A(x))}\, \to_l}{\vdash A(\varepsilon_y(A(y) \to A(\varepsilon_x \neg A(x)))) \to A(\varepsilon_x \neg A(x))}\, \exists_r
$$

Example 4. *The extended first epsilon theorem applies only to sequents with weak quantifiers. Consequently, we have to consider a Skolemization of the given proof of*

$$\exists y(A(y) \to \forall x A(x))$$

$$\begin{array}{cl}
A(f(b)) \vdash A(f(b)) & \\ \hline
A(f(b)) \vdash A(f(b)), A(f(f(b))) & w_r \\ \hline
\vdash A(f(b)), A(f(b)) \to A(f(f(b))) & \to_r \\ \hline
\vdash A(f(b)), \exists x(A(x) \to A(f(x))) & \exists_r \\ \hline
A(b) \vdash A(f(b)), \exists x(A(x) \to A(f(x))) & w_l \\ \hline
\vdash A(b) \to A(f(b)), \exists x(A(x) \to A(f(x))) & \to_r \\ \hline
\vdash \exists x(A(x) \to A(f(x)), \exists x(A(x) \to A(f(x))) & \exists_r \\ \hline
\vdash \exists x(A(x) \to A(f(x))) & c_r
\end{array}$$

The translation to epsilon calculus is

$$\begin{array}{cl}
A(f(b)) \vdash A(f(b)) & \\ \hline
A(f(b)) \vdash A(f(b))), A(f(f(b))) & w_r \\ \hline
\vdash A(f(b)), A(f(b))) \to A(f(f(b))) & \to_r \\ \hline
\vdash A(f(b)), A(e) \to A(f(e))) & \exists_r \\ \hline
A(b) \vdash A(f(b)), A(e) \to A(f(e))) & w_l \\ \hline
\vdash A(b) \to A(f(b)), A(e) \to A(f(e))) & \to_r \\ \hline
\vdash A(e) \to A(f(e))), A(e) \to A(f(e))) & \exists_r \\ \hline
\vdash A(e) \to A(f(e))) & c_r
\end{array}$$

where $e = \epsilon_x(A(x) \to A(f(x))))$.

The epsilon proof in its traditional form is

$$[(A(f(b))) \to A(f(f(b)))) \to (A(e) \to A(f(e))) \wedge$$

$$(A(b) \to A(f(b)) \to A(e) \to A(f(e)))] \to (A(e) \to A(f(e))).$$

The extended first epsilon theorem yields

$$A(f(b)) \to A(f(f(b)) \vee A(b) \to A(f(b)) \vee A(e) \to A(f(e)))$$

which can be contracted after replacing e *by* b *to*

$$A(f(b)) \to A(f(f(b)) \vee A(b) \to A(f(b))).$$

e *can be substituted by* b *as* e *does not occur in* $f(b), f(f(b))$.

Note that an **LK***-proof with or without cuts can be skolemized relative to the end-sequent with at most double exponential increase of size. Without adding new cuts,* **L**ε*-proofs cannot be skolemized at all.*

The main argument of this paper is based on the existence of suitable sequences of proofs with cuts.

Theorem 2 ([8,9]). *There is a specific family of sequents* $\{S_i\}_{i<\omega}$ *described in [1] and due to Statman or Orevkov, and specific* **LK***-proofs thereof, such that they have the following properties:*

1. *the size of* S_i *is polynomial in* i*;*

2. *there is no bound on the size of their smallest cut-free* **LK***-proofs that is elementary in* i*;*

3. *the size of these proofs (with cuts), however, is polynomially bounded in i.*

In the following we will consider the sequence of sequents $\{S_i\}_{i<\omega}$ from Theorem 2 above.

Corollary 2. *Each worst-case sequence as formulated in Theorem 2 generates a worst-case sequence, where the end-sequents contain weak quantifiers only.*

Proof. Strong quantifiers in a cut-free **LK** proof can be replaced by Skolem functions with at most double exponential increase of size. □

Bound variables are replaced by Skolem expressions that are not larger in size than the existing proof.

In the theorem below we will consider matrices of first-order formulas.

Definition 12. *The matrix A^M of a first-order formula A is A, after deletion of all quantifiers and after replacement of bound variables by free variables.*

Example 5. *Consider the formula*

$$\exists x(\forall y A(x, y) \vee B(x)).$$

Its matrix is

$$[\exists x(\forall y A(x, y) \vee B(x))]^M = A(a, b) \vee B(a).$$

Lemma 1. *There is a specific family of sequents $\{S_i\}_{i<\omega}$ such that they have the following properties:*

1. *the size of S_i is polynomial in i;*
2. *there is no bound on the size of their smallest cut-free **LK**-proofs that is elementary in i;*
3. *the size of these proofs (with cuts), however, is polynomially bounded in i;*
4. *they contain only weak quantifiers;*
5. *on the left-side of the conclusion for every cut A, $\forall\overline{x}(A^M \to A^M)$ is added ($\forall\overline{x}(A^M \to A^M)$ is the closure of $A^M \to A^M$).*

Proof. For the proofs with cut the addition of

$$\forall\overline{x}(A^M \to A^M)$$

leads to at most double exponentially larger proofs.

For the cut-free proofs the proofs may be double exponentially shorter if the newly added universal formulas are eliminated in the following way: In the moment where the corresponding implication left is inferred, replace this inference by a cut. In consequence, there is a proof with propositional cuts only, which can be eliminated in at most double exponential expense [10].

The idea is to reduce the cuts to their sub-formulas without moving them. This keeps the depth of the proof, consequently the size is exponential. Another exponent is connected to the elimination of the atomic cuts. □

Corollary 3. *There is a specific family of sequents $\{S_i\}_{i<\omega}$ containing only weak quantifiers such that they have the following properties:*

1. *the size of S_i is polynomial in i;*

2. *there is no elementary bound of their smallest* **LK***-proofs with bounded cut-complexity;*

3. *the size of these proofs with unrestricted cuts, however, is polynomially bounded in i;*

4. *on the left-side of the conclusion for every cut A, $\forall\overline{x}(A^M \to A^M)$ is contained.*

Proof. Note that the elimination of cuts with fixed complexity is elementary. Consequently, Lemma 1 applies. □

Theorem 3. *There is a sequence of cut-free* **L**ε*-proofs of S_i such that*

1. *the size of S_i is elementary in i;*

2. *the end-sequents S_i are translations of first-order sequents S'_i with weak quantifiers only, i.e. $S_i = [S'_i]^\varepsilon$;*

3. *the size of these* **L**ε*-proofs is elementarily bounded in i;*

4. *let the cut-complexity be bounded, then there is no elementary bound on the size of the smallest* **LK***-proofs of S'_i.*

Proof. We choose a sequence of **LK**-proofs from Corollary 3 which has non-elementarily growing cut-free proofs. We translate the **LK**-proofs with cut into epsilon calculus. This lengthens the proof at most double exponentially according to Proposition 1. In the **L**ε-proof we replace all cuts on A by inferences of $A \to A$ on the left side. Then we replace the epsilon terms in $A \to A$ by epsilon terms representing the universal quantifiers to obtain

$$[\forall\overline{x}(A^M \to A^M)]^\varepsilon$$

on the left side. Finally, we contract with the existing universal formulas in the end-sequent. □

Note that the extended first epsilon theorem provides an upper bound for cut-free **L**ε-derivations: The number of weak quantifier inferences determines the number of critical formulas. From the Herbrand disjunction according to Theorem 1 we obtain easily a cut-free **LK** derivation.

The question remains however, whether **L**ε-derivations with cuts can be translated into **LK**-derivations with arbitrary cuts in an elementary way.

Example 6. *This example shows how* **L**ε *reduces cuts of arbitrary complexity to universal cuts. Consider the following proof in* **LK**

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{P(a,b)\vdash P(a,b)}{\forall y P(a,y)\vdash P(a,b)}\,\forall_l}{\forall x\forall y P(x,y)\vdash P(a,b)}\,\forall_l}{\forall x\forall y P(x,y)\vdash \forall y P(a,y)}\,\forall_r}{\forall x\forall y P(x,y)\vdash \exists x\forall y P(x,y)}\,\exists_r \qquad \dfrac{\dfrac{\dfrac{\dfrac{P(c,d)\vdash P(c,d)}{P(c,d)\vdash \exists y P(c,y)}\,\exists_r}{P(c,d)\vdash \exists x\exists y P(x,y)}\,\exists_r}{\forall y P(c,y)\vdash \exists x\exists y P(x,y)}\,\forall_l}{\exists x\forall y P(x,y)\vdash \exists x\exists y P(x,y)}\,\exists_l}{\forall x\forall y P(x,y)\vdash \exists x\exists y P(x,y)}\,cut$$

The cut is replaced by an application of $\to_l$

$$
\dfrac{
\dfrac{\dfrac{\dfrac{\dfrac{P(a,b) \vdash P(a,b)}{\forall y P(a,y) \vdash P(a,b)}\,\forall_l}{\forall x \forall y P(x,y) \vdash P(a,b)}\,\forall_l}{\forall x \forall y P(x,y) \vdash \forall y P(a,y)}\,\forall_r}{\forall x \forall y P(x,y) \vdash \exists x \forall y P(x,y)}\,\exists_r
\qquad
\dfrac{\dfrac{\dfrac{\dfrac{P(c,d) \vdash P(c,d)}{P(c,d) \vdash \exists y P(c,y)}\,\exists_r}{P(c,d) \vdash \exists x \exists y P(x,y)}\,\exists_r}{\forall y P(c,y) \vdash \exists x \exists y P(x,y)}\,\forall_l}{\exists x \forall y P(x,y) \vdash \exists x \exists y P(x,y)}\,\exists_l
}{\forall x \forall y P(x,y), \exists x \forall y P(x,y) \rightarrow \exists x \forall y P(x,y) \vdash \exists x \exists y P(x,y)}\,\rightarrow_l
$$

The translation of this derivation to $\mathbf{L}\varepsilon$ *is*

$$
\dfrac{\dfrac{\dfrac{
\dfrac{\dfrac{P(a, x_1(a)) \vdash P(a, x_1(a))}{P(x_2, x_1(x_2)) \vdash P(a, x_1(a))}}{P(x_2, x_1(x_2)) \vdash P(x_3, x_1(x_3))}
\qquad
\dfrac{\dfrac{\dfrac{P(x_3, d) \vdash P(x_3, d)}{P(x_3, d) \vdash P(x_3, x_4(x_3))}}{P(x_3, d) \vdash P(x_5, x_4(x_5))}}{P(x_3, x_1(x_3)) \vdash P(x_5, x_4(x_5))}
}{P(x_2, x_1(x_2)), P(x_3, x_1(x_3)) \rightarrow P(x_3, x_1(x_3)) \vdash P(x_5, x_4(x_5))}}{P(x_2, x_1(x_2)), P(x_3, x_6(x_3)) \rightarrow P(x_3, x_6(x_3)) \vdash P(x_5, x_4(x_5))}}{P(x_2, x_1(x_2)), P(x_7, x_6(x_7)) \rightarrow P(x_7, x_6(x_7)) \vdash P(x_5, x_4(x_5))}
$$

Note that

$$[\forall x \forall y (P(x,y) \rightarrow P(x,y))]^{\varepsilon}) = P(x_7, x_6(x_7)) \rightarrow P(x_7, x_6(x_7)),$$

and

$$
\begin{array}{rcl}
x_1(a) & = & \varepsilon_y \neg P(a,y), \\
x_2 & = & \varepsilon_x \neg P(x, x_1(x)), \\
x_3 & = & \varepsilon_x P(x, x_1(x)), \\
x_4(c) & = & \varepsilon_x P(c, x), \\
x_5 & = & \varepsilon_x P(x, x_4(x)), \\
x_6(d) & = & \varepsilon_y \neg P(d, y), \\
x_7 & = & \varepsilon_x \neg (P(x, x_6(x)) \rightarrow P(x, x_6(x)).
\end{array}
$$

4 Intuitionistic $\mathbf{L}\varepsilon$

Intuitionistic $\mathbf{L}\varepsilon$, i.e. with right-side restriction to at most one formula, is not sound for intuitionistic logic.

Theorem 4. *Intuitionistic* $\mathbf{L}\varepsilon$ *is not sound.*

Proof. Consider the following derivation of

$$[\exists x (A(x) \rightarrow A(f(x)))]^{\varepsilon}$$

$$
\dfrac{\dfrac{\dfrac{A(f(x)) \vdash A(f(x))}{A(x) \vdash A(f(x))}}{\vdash A(x) \rightarrow A(f(x))}}{\vdash A(y) \rightarrow A(f(y))}
$$

where

$$\begin{aligned} x &= \varepsilon_u \neg A(u), \\ y &= \varepsilon_w (A(w) \to A(f(w))) \end{aligned}$$

Note that
$$[\exists x(A(x) \to A(f(x)))]^\varepsilon = A(y) \to A(f(y)).$$

The result is not valid in intuitionistic logic. The reason of this phenomenon is that intuitionistic $\mathbf{L}\varepsilon$ proves all classical quantifier shifts. Also, the three shifts which are not valid in intuitionistic logic:

$$\forall x(A(x) \vee B) \to \forall x A(x) \vee B$$

$$(\forall x A(x) \to B) \to \exists x(A(x) \to B)$$

$$(A \to \exists x B(x)) \to \exists x(A \to B(x)$$

The derivations are

$$\frac{A(x) \vee B \vdash A(x) \vee B}{A(y) \vee B \vdash A(x) \vee B}$$

where

$$\begin{aligned} x &= \varepsilon_x \neg A(x), \\ y &= \varepsilon_y \neg (A(y) \vee B) \end{aligned}$$

$$\frac{(A(x) \to B) \vdash A(x) \to B}{A(x) \to B \vdash A(y) \to B}$$

where

$$\begin{aligned} x &= \varepsilon_x \neg A(x), \\ y &= \varepsilon_y (A(y) \to B) \end{aligned}$$

and

$$\cfrac{\cfrac{A \to B(x) \vdash A \to B(x)}{A \to B(x) \vdash A \to B(y)}}{\vdash (A \to B(x)) \to A \to B(y)}$$

where

$$\begin{aligned} x &= \varepsilon_x B(x), \\ y &= \varepsilon_y (A \to B(y)) \end{aligned}$$

□

5 Conclusion

The results of this paper depend on the effect that bound variables in epsilon calculus are terms, and that the epsilon calculus is able to overbind bound variables. This is a property which should be exploited for automated deduction systems.

The best approach to answer Toshiyasu Arai's question in full is to find a suitable translation of the language of the epsilon calculus to first-order language. The straightforward translation of this and other papers does not work, as not all epsilon expressions have a first-order meaning (see Example 1).

References

[1] Matthias Baaz and Alexander Leitsch. On skolemization and proof complexity. *Fundam. Informaticae*, 20(4):353–379, 1994.

[2] Gerhard Gentzen. Untersuchungen über das logische Schließen. *Mathematische Zeitschrift*, 39:176–210,405–431, 1934-35.

[3] Jacques Herbrand. *Recherches sur la théorie de la démonstration.* PhD thesis, Université de Paris, 1930.

[4] David Hilbert and Paul Bernays. Grundlagen der Mathematik. 2, 1939.

[5] Horst Luckhardt. Herbrand-analysen zweier Beweise des Satzes von Roth: Polynomiale Anzahlschranken. *Journal of Symbolic Logic*, pages 234–263, 1989.

[6] Kenji Miyamoto and Georg Moser. Herbrand complexity and the epsilon calculus with equality. *Arch. Math. Log.*, 63(1):89–118, 2024.

[7] Georg Moser and Richard Zach. The epsilon calculus and Herbrand complexity. *Studia Logica*, 82(1):133–155, 2006.

[8] Vladimir P Orevkov. Lower bounds for increasing complexity of derivations after cut elimination. *Journal of Soviet Mathematics*, 20:2337–2350, 1982.

[9] Richard Statman. Lower bounds on herbrand's theorem. *Proceedings of the American Mathematical Society*, 75(1):104–107, 1979.

[10] Daniel Weller. On the elimination of quantifier-free cuts. *Theor. Comput. Sci.*, 412(49):6843–6854, 2011.