Available online at www.sciencedirect.com
ScienceDirect
Theoretical Computer Science 403 (2008) 160–175
Theoretical Computer Science
www.elsevier.com/locate/tcs

CERES: An analysis of Fürstenberg's proof of the infinity of primes☆

Matthias Baaz[a], Stefan Hetzl[b], Alexander Leitsch[b,*], Clemens Richter[b], Hendrik Spohr[b]

[a] *Institute of Discrete Mathematics and Geometry (E104), Vienna University of Technology, Wiedner Hauptstraße 8-10, 1040 Vienna, Austria*
[b] *Institute of Computer Languages (E185), Vienna University of Technology, Favoritenstraße 9, 1040 Vienna, Austria*

Received 12 June 2007; received in revised form 8 February 2008; accepted 20 February 2008

Communicated by A. Avron

Abstract

The distinction between analytic and synthetic proofs is a very old and important one: An analytic proof uses only notions occurring in the proved statement while a synthetic proof uses additional ones. This distinction has been made precise by Gentzen's famous cut-elimination theorem stating that synthetic proofs can be transformed into analytic ones.

CERES (cut-elimination by resolution) is a cut-elimination method that has the advantage of considering the original proof in its full generality which allows the extraction of different analytic arguments from it. In this paper we will use an implementation of CERES to analyze Fürstenberg's topological proof of the infinity of primes. We will show that Euclid's original proof can be obtained as one of the analytic arguments from Fürstenberg's proof. This constitutes a proof-of-concept example for a semi-automated analysis of realistic mathematical proofs providing new information about them.

Keywords: Cut-elimination; Proof analysis; Proof theory; Resolution

1. Introduction

Leibniz was the first to distinguish analytic proofs from synthetic proofs in the sense that analytic proofs use only the notions of the proved statement, synthetic proofs use additional notions. One main contribution of logic to mathematics is to make such vague classifications workable via (unfortunately sometimes premature) precision: "Analytic" is nowadays associated with a (sequent calculus) proof, where all formulas are subformulas of the end sequent, i.e. with an (almost) cut-free proof. The cut-elimination theorem states, that a synthetic proof can be converted into an analytic one (at least in first order logic, which is the focus of this paper).

The logical benefits gained from such a transformation are obvious (immediate arguments for consistency, construction of interpolants for derived implications and, last but not least, the calculation of a most general term minimal proof of the same structure). The mathematical benefits are not so obvious, especially if we follow the

☆ Supported by the Austrian Science Fund (project no. P17995-N12).
* Corresponding author. Tel.: +43 1 58801 18540; fax: +43 1 58801 18597.
E-mail addresses: baaz@logic.at (M. Baaz), hetzl@logic.at (S. Hetzl), leitsch@logic.at (A. Leitsch), richter@logic.at (C. Richter), spohr@logic.at (H. Spohr).

0304-3975/$ - see front matter
doi:10.1016/j.tcs.2008.02.043

opinion of Kant: "Die philosophische Erkenntnis ist die Vernunfterkenntnis aus Begriffen, die mathematische aus der Konstruktion der Begriffe" (Philosophical cognition is the cognition of reason by means of conceptions; mathematical cognition is cognition by means of the construction of conceptions.).[1] This means that the non-trivial concepts of the proof are contained in the cuts, and mathematical activity means introduction, not elimination of cuts.

The situation changes, if we consider the synthetic proof as an abstract representation of (one or several) analytic proofs whose explicit description might lead to the strengthening of the result (e.g. via Herbrand-Disjunctions [16]) or which might provide insight in the operational behavior of the abstract notions on the level of objects [14, annex 4.A].

The usual approaches of proof theory focus on cut-elimination mechanisms as confluent as possible to obtain a strong bond between the synthetic and the analytic proof [10,17]. The advantage is, that a difference in the analytic results of cut-elimination induces a difference on the synthetic preimages, which are otherwise in general hard to compare. The price to pay is however, that the full strength of the synthetic proof is lost and its mathematical flexibility is curtailed by a priori chosen preference rules of logical, not mathematical nature.

In this paper, we choose an opposite approach: Cut-elimination by resolution (CERES) is a method for cut-elimination in first-order logic. It is based on the following bipartition: 1. the parts of the proof leading to cuts and 2. the parts of the proof leading to the result, which are cut-free by definition. CERES uses the fact, that the cuts derive contradictions from the information completing the proof parts in 1. The derivation of the cut is replaced by any atomic refutation of the information leading to an essentially cut-free proof. The (sometimes infinite) set of these atomic refutations can be considered as a solution space representing analytic proofs for the synthetic proof considered as functional by CERES.

Second-order logic is a natural basis for formalizing large parts of mathematics [19]. In order to apply a first-order cut-elimination method, we face the problem of projecting a second-order formalization to a first-order one. The most appropriate solution to this problem is to introduce a many-sorted language. Theory axioms pose no problems, but valid logical expressions of second-order logic might turn into theory axioms not logically valid in first-order logic. The best way to handle this is to axiomatize relevant first-order consequences of second-order reasoning in the proof not obtainable by first-order reasoning using suitable theory axioms.

We apply CERES to Fürstenberg's proof of the existence of infinitely many primes. The arguments of this proof are of topological nature, which form the synthetic notions of this synthetic proof. A natural formalization of this argument in second-order arithmetic is constructed and then translated to many-sorted first-order logic. In order to avoid induction axioms, the proof is eventually formalized as a scheme representing an infinite sequence of ordinary first-order proofs, demonstrating the existence of more and more primes. We show that the analytic sequence corresponding to Euclid's proof belongs to the solution space of sequences.

2. Fürstenberg's proof

In 1955 the renowned mathematician H. Fürstenberg published a proof of the infinity of primes by topological means [12]: He proved the infinity of primes using a topology induced by arithmetic progressions over the integers.

We give a proof with a topology over the natural numbers in order to have a simpler formulation of the proof later on. We start with the definition of a topological space:

Definition 2.1 (*Topological Space*). A *topological space* is a set X together with a collection T of subsets of X satisfying the following axioms:

(1) The empty set and X are in T.
(2) The union of any collection of sets in T is also in T.
(3) The intersection of any pair of sets in T is also in T.

The collection T is called a *topology* on X. The sets in T are the *open sets*, and their complements in X are the *closed sets*.

[1] I. Kant: The Critique of Pure Reason, Transcendental Doctrine of Method, Chapter I, Section I

The arithmetic progressions can be used as a basis for a topology over the natural numbers. We will denote an arithmetic progression by

$$\nu(a, b) = \{a + bn \mid n \in \mathbb{N}\}$$

for $a \in \mathbb{N}$ and $b \in \mathbb{N} \setminus \{0\}$.

Proposition 1. *By defining a set $A \subseteq \mathbb{N}$ as open, when A is either empty or for each $x \in A$ exists an $a \in \mathbb{N} \setminus \{0\}$ such that $\nu(x, a) \subseteq A$, one obtains a topology over $\mathbb{N}$.*

Proof. We check Definition 2.1:

(1) The empty set and $\mathbb{N}$ are open. Trivial.
(2) The union of a collection of open sets is also open. Trivial.
(3) The intersection of two open sets is also open.
Let A and B two open sets. If $x \in A \cap B$, then there exist $a, b > 0$ such that $\nu(x, a) \subseteq A$ and $\nu(x, b) \subseteq B$ holds. Let c be the least common multiple of a and b, then $\nu(x, c) \subseteq \nu(x, a)$ and $\nu(x, c) \subseteq \nu(x, b)$, and hence $\nu(x, c) \subseteq A \cap B$. □

A nice property of this topology is that every arithmetic progression starting at 0 is not only open but closed as well. Indeed this holds for every progression $\nu(a, b)$ where $a < b$, but this is not needed for the theorem.

Lemma 2.1. *Every arithmetic progression starting at 0 is closed.*

Proof. Let be $A = \nu(0, b)$ an arithmetic progression. Then the complement of A is a union of arithmetic progressions:

$$\bar{A} = \bigcup_{i=1}^{b-1} \nu(i, b).$$

The sets $\nu(i, b)$ are open, and the union of any collection of open sets is open; therefore $\bar{A}$ is open, hence A is closed. □

Theorem 2.1. *There are infinitely many primes.*

Proof. Denote with P the set of all primes and assume P is finite. Let $X = \bigcup\{\nu(0, p) \mid p \in P\}$. By Lemma 2.1 every $\nu(0, p)$ for $p \in P$ is closed, so X is a finite union of closed sets and therefore closed as well. As every number different from 1 has a prime divisor we get $\bar{X} = \{1\}$. Being a complement of a closed set, $\bar{X}$ is open. But $\{1\}$ is neither empty nor does it contain an arithmetic progression, and so $\{1\}$ is not open. Contradiction! We conclude that P must be infinite. □

3. Cut-elimination by resolution

3.1. The sequent calculus LKDe

Basically we use Gentzen's version of **LK** [13] for the specification of mathematical proofs. **LKDe** is an extension of **LK** by definition- and equality rules (see [5] for the complete definition). As axioms, arbitrary atomic sequents are admitted.

- The *definition rules* directly correspond to the *extension principle* (see [11]) in predicate logic. It simply consists in introducing new predicate- and function symbols as abbreviations for formulas and terms. Mathematically this corresponds to the introduction of new concepts in theories. Let A be a first-order formula with the free variables $x_1, \ldots, x_k$ (denoted by $A(x_1, \ldots, x_k)$) and P be a *unique* k-ary predicate symbol corresponding to the formula A. Then the rules are:

$$\frac{A(t_1, \ldots, t_k), \Gamma \vdash \Delta}{P(t_1, \ldots, t_k), \Gamma \vdash \Delta}\ \mathrm{def}_P{:}l \qquad \frac{\Gamma \vdash \Delta, A(t_1, \ldots, t_k)}{\Gamma \vdash \Delta, P(t_1, \ldots, t_k)}\ \mathrm{def}_P{:}r$$

for arbitrary sequences of terms $t_1, \ldots, t_k$. There are also definition introduction rules for new function symbols which are of similar type.

- The *equality rules* are:

$$\frac{\Gamma \vdash \Delta, s = t \quad A[s]_\Sigma, \Pi \vdash \Lambda}{A[t]_\Sigma, \Gamma, \Pi \vdash \Delta, \Lambda} =: l \qquad \frac{\Gamma \vdash \Delta, s = t \quad \Pi \vdash \Lambda, A[s]_\Sigma}{\Gamma, \Pi \vdash \Delta, \Lambda, A[t]_\Sigma} =: r$$

where Σ denotes a set of positions of subterms where replacement of s by t has to be performed. We call $s = t$ the *active equation* of the rules. For complete treatment of equality we have to include the reflexivity axiom $\vdash s = s$, symmetry and transitivity can be derived.

3.2. The resolution calculus **PR**

The main part of the proof-transformation method CERES consists in the refutation of clause sets by resolution and paramodulation. We use the following resolution calculus **PR** based on atomic sequents as clauses.

Definition 3.1. The rules of **PR** are:

(1) Factoring:

$$\frac{A_1, \ldots, A_n, \Gamma \vdash \Delta}{(A_1, \Gamma \vdash \Delta)\sigma}\ factor(\sigma): l \qquad \frac{\Gamma \vdash \Delta, A_1, \ldots, A_n}{(\Gamma \vdash \Delta, A_1)\sigma}\ factor(\sigma): r$$

where σ is an m.g.u. of the set $\{ A_i \}_{1 \leq i \leq n}$. The conclusion of a factoring rule is called factor and additionally nontrivial in case $1 < n$.

(2) Binary resolution:

$$\frac{\Gamma \vdash \Delta, A \quad A', \Pi \vdash \Lambda}{(\Gamma, \Pi \vdash \Delta, \Lambda)\sigma}\ \mathrm{res}(\sigma)$$

where σ is an m.g.u. of $\{ A, A' \}$ and the premises are variable disjoint clauses. The conclusion of a binary resolution inference is also called resolvent.

(3) Paramodulation (for some position set Σ):

$$\frac{\Gamma \vdash \Delta, s = t \quad A[s']_\Sigma, \Pi \vdash \Lambda}{(A[t]_\Sigma, \Gamma, \Pi \vdash \Delta, \Lambda)\sigma}\ \mathrm{p}(\sigma): l \quad \frac{\Gamma \vdash \Delta, s = t \quad \Pi \vdash \Lambda, A[s']_\Sigma}{(\Gamma, \Pi \vdash \Delta, \Lambda, A[t]_\Sigma)\sigma}\ \mathrm{p}(\sigma): r$$

where s, t are arbitrary terms containing only free variables, σ is an m.g.u. of $\{ s, s' \}$ and the premises are variable disjoint clauses.

(4) Permutation:

$$\frac{A_1, \ldots, A_n \vdash \Delta}{A_{\tau(1)}, \ldots, A_{\tau(n)} \vdash \Delta}\ \pi(\tau): l \qquad \frac{\Gamma \vdash A_1, \ldots, A_n}{\Gamma \vdash A_{\tau(1)}, \ldots, A_{\tau(n)}}\ \pi(\tau): r$$

where τ is a permutation of $\{ 1, \ldots, n \}$.

A deduction of the empty clause based on a set of clauses using the rules of the resolution calculus **PR** is called a **PR**-proof resp. **PR**-refutation.

Note that, after application of the global m.g.u. of the resolution proof, all of the **PR**-rules defined above become rules of **LKDe**.

3.3. Skolemization of proofs

Definition 3.2. Let B be a formula. If $(\forall x)$ occurs positively (negatively) in B then $(\forall x)$ is called a strong (weak) quantifier. If $(\exists x)$ occurs positively (negatively) in B then $(\exists x)$ is called a weak (strong) quantifier.

Skolemization is a transformation on closed formulas which removes all strong quantifiers. There are different types of skolemizations which may strongly differ in the proof complexity of the transformed formula (see [6]). Below we define the structural skolemization operator *sk* introduced in [7].

Definition 3.3. *sk* is a function which maps closed formulas into closed formulas; it is defined in the following way: $sk(F) = F$ if F does not contain strong quantifiers, and

$$sk(F) = sk(F_{(Qy)}\{y \leftarrow f(x_1, \ldots, x_n)\})$$

where (Qy) is a strong quantifier appearing in the scope of the weak quantifiers $(Q_1x_1), \ldots, (Q_nx_n)$ (in this order) and $F_{(Qy)}$ denotes F after omission of (Qy); f is a function symbol which does not occur in F (if $n = 0$ then f is a constant symbol).

In model theory and automated deduction the definition of skolemization mostly is dual to Definition 3.3, i.e. in case of prenex forms the existential quantifiers are eliminated instead of the universal ones. The skolemization of sequents, defined below, represents a more general framework covering both concepts.

Definition 3.4. Let S be the sequent $A_1, \ldots, A_n \vdash B_1, \ldots, B_m$ consisting of closed formulas only and $(A'_1 \wedge \cdots \wedge A'_n) \rightarrow (B'_1 \vee \cdots \vee B'_m)$ be the structural skolemization of $(A_1 \wedge \cdots \wedge A_n) \rightarrow (B_1 \vee \cdots \vee B_m)$. Then the sequent $S' : A'_1, \ldots, A'_n \vdash B'_1, \ldots, B'_m$ is called the *skolemization* of S.

Example 3.1. Let S be the sequent $(\forall x)(\exists y)P(x, y) \vdash (\forall x)(\exists y)P(x, y)$. Then the skolemization of S is $S' : (\forall x)P(x, f(x)) \vdash (\exists y)P(c, y)$ for a one-place function symbol f and a constant symbol c. Note that the skolemization of the left-hand-side of the sequent corresponds to the refutational skolemization concept for formulas mentioned above.

Not only formulas and sequents, but also whole proofs can be skolemized. By a skolemized proof we mean a proof of a skolemized end-sequent. Also proofs with cuts can be skolemized, but the cut formulas themselves cannot. Only the strong quantifiers which "go" into the end sequent are eliminated. Skolemization does not increase the length of proofs. We do not give a formal definition of proof skolemization *sk* but refer to [7]. The example below illustrates the transformation.

Example 3.2. Let $\varphi =$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{P(c,\alpha) \vdash P(c,\alpha) \quad Q(\alpha) \vdash Q(\alpha)}{P(c,\alpha), P(c,\alpha) \rightarrow Q(\alpha) \vdash Q(\alpha)} \rightarrow : l}{P(c,\alpha) \rightarrow Q(\alpha), (\forall x)P(c,x) \vdash Q(\alpha)} \forall : l}{P(c,\alpha) \rightarrow Q(\alpha), (\forall x)P(c,x) \vdash (\exists y)Q(y)} \exists : r}{(\exists y)(P(c,y) \rightarrow Q(y)), (\forall x)P(c,x) \vdash (\exists y)Q(y)} \exists : l}{(\forall x)(\exists y)(P(x,y) \rightarrow Q(y)), (\forall x)P(c,x) \vdash (\exists y)Q(y)} \forall : l$$

Then $sk(\varphi) =$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{P(c,f(c)) \vdash P(c,f(c)) \quad Q(f(c)) \vdash Q(f(c))}{P(c,f(c)), P(c,f(c)) \rightarrow Q(f(c)) \vdash Q(f(c))} \rightarrow : l}{P(c,f(c)) \rightarrow Q(f(c)), (\forall x)P(c,x) \vdash Q(f(c))} \forall : l}{P(c,f(c)) \rightarrow Q(f(c)), (\forall x)P(c,x) \vdash (\exists y)Q(y)} \exists : r}{(\forall x)(P(x,f(x)) \rightarrow Q(f(x))), (\forall x)P(c,x) \vdash (\exists y)Q(y)} \forall : l$$

3.4. The characteristic clause set

The *characteristic clause set* encodes the structure of a proof and the essence of the cut formulas used in the proof. Its construction is based on an analysis, which inferences go into a cut and which go into the end-sequent.

Definition 3.5. The characteristic clause set $\mathrm{CL}(\varphi)$ of a skolemized **LKDe**-proof φ is defined inductively. For every node ν of φ we define:

- If ν is an occurrence of an axiom S and S' is the subsequent of S consisting of the ancestors of cut formulas then $\mathcal{C}_\nu = \{ S' \}$.
- Let ν_1, ν_2 be the predecessors of ν in a binary inference then we distinguish:

. The auxiliary formulas of ν_1, ν_2 are ancestors of cut formulas then
$$\mathcal{C}_\nu = \mathcal{C}_{\nu_1} \cup \mathcal{C}_{\nu_2}.$$
. Otherwise
$$\mathcal{C}_\nu = \mathcal{C}_{\nu_1} \times \mathcal{C}_{\nu_2},$$
for $\mathcal{C} \times \mathcal{D} = \{C \circ D \mid C \in \mathcal{C}, D \in \mathcal{D}\}$, where $C \circ D$ is the merge of the clauses C and D, i.e. $(\Gamma \vdash \Delta) \circ (\Pi \vdash \Lambda) = \Gamma, \Pi \vdash \Delta, \Lambda$.

- Let ν' be the predecessor of ν in a unary inference then $\mathcal{C}_\nu = \mathcal{C}_{\nu'}$.

The characteristic clause set $\mathrm{CL}(\varphi)$ is defined as $\mathcal{C}_\nu$ where ν is the root node of φ. Note that for a cut-free proof φ we have $\mathrm{CL}(\varphi) = \{\vdash\}$.

Theorem 3.1. *Let φ be a skolemized proof. Then* $\mathrm{CL}(\varphi)$ *is unsatisfiable.*

Proof. In [8]. □

In case of equational clause sets the concept of unsatisfiability has to be replaced by E-unsatisfiability. For details see [5]. Note that, in practice, $\mathrm{CL}(\varphi)$ will be reduced under subsumption and tautology-deletion.

3.5. Projection and refutation

Every clause C in the characteristic clause set $\mathrm{CL}(\varphi)$ of a skolemized proof φ defines a *proof projection* $\varphi[C]$. $\varphi[C]$ is a cut-free proof of the end-sequent augmented by C and is defined from φ by omitting all inferences on ancestors of cuts. For a formal definition of projection we refer to [8] and [18].

Let φ be a skolemized proof of S and let γ be a ground resolution refutation of $\mathrm{CL}(\varphi)$. Then γ can be transformed into a proof $\varphi(\gamma)$ of S with at most atomic cuts. $\varphi(\gamma)$ is constructed from γ simply by replacing the clauses by the corresponding proof projections. The construction of the refutation γ is the essential part of the method CERES for two reasons: 1. The computational complexity of cut-elimination is defined by the size of the resolution tree [8]. 2. The resolution proof contains all substitutions defined in the cut-elimination procedure (and thus the mathematical content). The final elimination of atomic cuts – provided the atomic axiom sequents are closed under cuts, i.e. it is possible at all – is mathematically inessential. For proofs $\varphi(\gamma)$ as above, the complete cut-elimination is at most linear in size.

Finally we give the general definition of CERES (cut-elimination by resolution) as a whole. As an input we take a skolemized proof φ.

- construct $\mathrm{CL}(\varphi)$,
- construct a **PR**-refutation γ of $\mathrm{CL}(\varphi)$,
- compute the output $\varphi(\gamma')$ for a ground projection γ' of γ.

Theorem 3.2. *CERES is a cut-elimination method, i.e., for every skolemized proof φ of a sequent S, CERES produces a proof ψ of S s.t. ψ contains at most atomic cuts.*

Proof. In [8] and [5]. □

4. Formalization of Fürstenberg's proof

The automated processing of Fürstenberg's proof requires a nontrivial logical preprocessing by humans. The first important step consists in the right choice of the logical language. As Fürstenberg's proof contains a topology defined over natural numbers and topological lemmas, an adequate candidate is *second-order arithmetic*. Below we will give a formalization of the main concepts and lemmas in second-order arithmetic. In a second step we will translate this specification into a scheme of sorted first-order definitions and proofs.

The basic language of second-order arithmetic contains the constant symbols 0, 1, the function symbols $+$ and $*$ and the predicate symbols $=, <$. There are two types of variables: individual variables (which we denote by $k, l, m, k', \ldots$) and set variables (denoted by $X, Y, \ldots$). Atoms of the form $X(t)$ are written as $t \in X$. Apart from some basic first order axioms we have the second-order axiom of induction

$$\forall X((0 \in X \wedge \forall n(n \in X \to n + 1 \in X)) \to \forall n.n \in X).$$

Moreover we have the comprehension scheme

$$\exists X \forall n (n \in X \leftrightarrow F(n))$$

where $F(n)$ is a formula in which X does not occur freely (for further details see [19]).

Below we give formal definitions of the main concepts used in Fürstenberg's proof:

(a) $m \in \nu(k, l) \equiv \exists n (m = k + n * l)$.
(b) $\mathrm{DIV}(l, k) \equiv \exists m . l * m = k$.
(c) $\mathrm{PRIME}(k) \equiv 1 < k \wedge \forall l (\mathrm{DIV}(l, k) \rightarrow (l = 1 \vee l = k))$.
(d) $X \subseteq Y \equiv \forall n (n \in X \rightarrow n \in Y)$, and $X = Y \equiv X \subseteq Y \wedge Y \subseteq X$.
(e) $n \in \overline{X} \equiv n \notin X$.
(f) A function $p: \mathbb{N} \rightarrow \mathbb{N}$ which enumerates primes is one that fulfills the property:

$$\forall i \forall k (p(i) = k \rightarrow \mathrm{PRIME}(k)).$$

For the definition of p the comprehension principle is needed; for information about function definitions in second-order arithmetic see [19].

(g) $n \in \mathrm{S}[l] \equiv \exists m (m \leq l \wedge n \in \nu(0, p(m)))$.
S$[l]$ describes the set of all elements n which occur in some $\nu(0, k)$, where k is one of the first $l + 1$ primes enumerated by p. In mathematical notation we get

$$\mathrm{S}[l] = \bigcup_{m=0}^{l} \nu(0, p(m)).$$

(h) $\mathrm{F}[l] \equiv \forall k (\mathrm{PRIME}(k) \leftrightarrow \exists m (m \leq l \wedge k = p(m)))$.
F$[l]$ is a formula which asserts that there are only $l + 1$ primes, namely $\{p(0), \ldots, p(l)\}$.
(i) $\mathrm{O}(X) \equiv \forall m (m \in X \rightarrow \exists l \; \nu(m, l + 1) \subseteq X)$.
(j) $\mathrm{C}(X) \equiv \mathrm{O}(\overline{X})$.
(k) $\infty(X) \equiv \forall k \exists l \; k + l + 1 \in X$.

Let (*) be the assumption that all primes occur in the set M: $\{p(0), \ldots, p(l)\}$. The first lemma in Fürstenberg's proof states that, under the assumption (*), every natural number different from 1 occurs in some $\nu(0, m)$ for $m \in M$. The corresponding formula is

$$\text{(I)} \; \forall l (\mathrm{F}[l] \; \rightarrow \; \mathrm{S}[l] = \overline{\{1\}}).$$

The second lemma states that, under the assumption (*), the set S$[l]$ is closed. The formula expressing this lemma is

$$\text{(II)} \; \forall l (\mathrm{F}[l] \rightarrow \mathrm{C}(\mathrm{S}[l])).$$

Proofs of (I) and (II) can easily combined to a proof of

$$\text{(III)} \; \forall l (\mathrm{F}[l] \rightarrow \mathrm{C}(\overline{\{1\}})).$$

The proofs of (II) and (III) in second order arithmetic require induction. By (j) it is straightforward to prove

$$\text{(IV)} \; \forall l (\mathrm{F}[l] \rightarrow \mathrm{O}(\{1\})).$$

Another main lemma of the proof states that nonempty open sets are infinite:

$$\text{(V)} \; \forall X (\mathrm{O}(X) \wedge X \neq \emptyset \rightarrow \infty(X)).$$

While (I), (II), (III) and (IV) can be directly translated to first order logic (via the definitions), (V) is genuinely second order. Using (V) we show that $\infty(\{1\})$ holds giving a contradiction to $\neg\infty(\{1\})$, which is easily derivable in second order arithmetic.

To formulate Fürstenberg's proof in **LKDe** it is necessary to schematize it in order to avoid induction. In particular, induction is needed to prove the lemmas (II) and (III) above. The tool `hlk` [2] allows to define an infinite sequence of

LKDe-proofs by specifying a proof scheme. The k-th proof can then be generated automatically from the scheme for any k.

The k-th proof shows that there cannot be $\leq k+1$ prime numbers.

To compile the second-order formulation to first order we work in a two-sorted logic containing sorts for 1. the natural numbers (denoted by $k, l, m, n, \ldots$ as before) and 2. sets of natural numbers (denoted by $x, y, \ldots$). Addition ($+$), multiplication ($*$) and the less-than relation ($<$) in the natural numbers are axiomatized. The background theory is purely universal and thus can be expressed as a set of clauses AX; It contains 34 clauses, among them associativity, commutativity and distributivity laws plus some derived laws like e.g. the cancellation law ($k+l=m+l \vdash k=m$). For the full list of axioms see the documentation on the web.[2] All of these axiom clauses are valid axiom sequents for the **LKDe**-proof.

Some of the definitions (a)–(k) given above can be taken over without change. This holds for (a), (b) and (c). For the others we get:

(d′) $x \subseteq y \equiv \forall n(n \in x \to n \in y)$, and $x = y \equiv x \subseteq y \wedge y \subseteq x$. Here we only replaced the set variables by variables of the sort "set of natural numbers".

(e′) $n \in \overline{x} \equiv n \notin x$.

(f′) Instead of p we introduce a finite set P[k] defined by

$$\mathrm{P}[k] \equiv \{p_0\} \cup \cdots \cup \{p_k\}.$$

where the p_i are constant symbols denoting primes. Note that the k appearing in the definition is a metavariable, not an object variable as l in the definition of F[l] and S[l].

(g′) $\mathrm{S}[k] \equiv \nu(0, p_0) \cup \cdots \cup \nu(0, p_k)$. Note that, in place of the object variable l in the definition (g), we have the metavariable k of the scheme.

(h′) $\mathrm{F}[k] \equiv \forall m(\mathrm{PRIME}(m) \leftrightarrow m \in \mathrm{P}[k])$.

(i′) $\mathrm{O}(x) \equiv \forall m(m \in x \to \exists l\ \nu(m, l+1) \subseteq x)$.

(j′) $\mathrm{C}(x) \equiv \mathrm{O}(\overline{x})$.

(k′) $\infty(x) \equiv \forall k \exists l\ k + l + 1 \in x$.

In order to avoid induction we also introduce three axioms (which can be proven in Peano arithmetic): (1) Every number greater than 0 has a predecessor, (2) every number is in a remainder class modulo l and (3) every number has a prime divisor. These axioms will be carried down to the antecedent of the end sequent of the **LKDe**-proof.

(1) $\mathrm{PRE} \equiv \forall k(0 < k \to \exists m\ k = m + 1)$
(2) $\mathrm{REM} \equiv \forall l(0 < l \to \forall m \exists k(k < l \wedge m \in \nu(k, l)))$
(3) $\text{PRIME-DIV} \equiv \forall m(m \neq 1 \to \exists l(\mathrm{PRIME}(l) \wedge \mathrm{DIV}(l, m)))$.

We now formulate a proof $\varphi_1(k)$ which proves the translation of (IV) above:

$\varphi_1(k) :=$

$$\dfrac{\dfrac{\dfrac{\begin{matrix}\psi_1(k)\\ \vdots\end{matrix}}{\mathrm{F}[k], \text{PRIME-DIV} \vdash \mathrm{S}[k] = \overline{\{1\}}} \quad \dfrac{\begin{matrix}\psi_2(k)\\ \vdots\end{matrix}}{\mathrm{F}[k], \mathrm{PRE}, \mathrm{REM} \vdash \mathrm{C}(\mathrm{S}[k])}}{\mathrm{F}[k], \Gamma \vdash \mathrm{C}(\overline{\{1\}})} =: r \quad \dfrac{\begin{matrix}\vdots\\ \mathrm{C}(\overline{\{1\}}) \vdash \mathrm{O}(\{1\})\end{matrix}}{\mathrm{F}[k], \Gamma \vdash \mathrm{O}(\{1\})}\ cut.$$

The proof $\psi_1(k)$ shows that if there are $\leq k+1$ primes, then by the prime divisor axiom PRIME-DIV, the complement of all multiples of these primes is $\{1\}$, and the proof $\psi_2(k)$ demonstrates (under the assumption of $\leq k+1$ primes and the remainder axiom REM) that the set of these multiples is closed. With the help of these lemmas we can show that the set $\{1\}$ is open — if there are $\leq k+1$ primes.

[2] http://www.logic.at/ceres/primeproof/.

The proof φ_2 (which does not depend on k) shows that every (non-empty) open set is infinite. This lemma yields that, under the assumption of the set of primes being finite, the set $\{1\}$ must be either empty or infinite — of course neither is the case, which is easily shown. Hence we get our end-sequent, stating that there cannot be $\leq k+1$ primes:

$\varphi(k) :=$

$$\dfrac{\dfrac{\vdots}{\vdash \{1\} \neq \emptyset} \quad \dfrac{\dfrac{\dfrac{\varphi_1(k)}{\vdots}}{\mathrm{F}[k], \Gamma \vdash \mathrm{O}(\{1\})} \quad \dfrac{\dfrac{\dfrac{\varphi_2}{\vdots}}{\vdash \forall x((\mathrm{O}(x) \wedge x \neq \emptyset) \rightarrow \infty(x))} \quad \dfrac{\vdots}{\mathrm{O}(\{1\}), \{1\} \neq \emptyset \vdash \infty(\{1\})}\, cut}{\{1\} \neq \emptyset, \mathrm{F}[k], \Gamma \vdash \infty(\{1\})}\, cut}{\mathrm{F}[k], \Gamma \vdash \infty(\{1\})} \quad \dfrac{\vdots}{\infty(\{1\}) \vdash}\, cut}{\underbrace{\mathrm{F}[k], \Gamma \vdash}_{\Gamma}}$$

$$\dfrac{}{\underbrace{\mathrm{PRIME\text{-}DIV}, \mathrm{PRE}, \mathrm{REM}}_{\Gamma} \vdash \neg \mathrm{F}[k]}\, \neg : r$$

The above proof scheme has been formalized with the help of the `hlk`-tool [2]. The design goal of `hlk` is to provide a comfortable input format for the formalization of large sequent calculus proofs. It acts as a compiler transforming a proof specified in a higher-level language into a fully formalized proof in the calculus **LKDe** (described in Section 3.1). The output proof is coded as an xml-file. `hlk` generates the structural rules as well as purely propositional parts of proofs automatically. It also allows to specify a proof scheme, each instance of which can be generated automatically. The `hlk`-script formalizing Fürstenberg's proof as a scheme has a size of approximately 1360 lines (6 kB gzipped). The instances of the scheme for $k = 0, 3, 12$ have sizes of 37, 157, 863 kB respectively (as gzipped xml files).

5. Analysis of Fürstenberg's proof

The proof, formalized as described in the previous section, has been subjected to analysis by the `CERES`-system [1]. In this section we will describe the results and important intermediate steps of this analysis. The reader interested in full formal details is referred to the documentation on the web.[3] The main result is that Euclid's proof of the infinity of primes can be constructed from Fürstenberg's proof. Euclid's proof can thus be regarded as a combinatorial kernel of Fürstenberg's proof.

5.1. Skolemization

The first step of the proof transformation consists in performing skolemization. As explained in Section 3.3 this constitutes an essential preprocessing step for the CERES-method. Besides from being necessary for the application of the resolution calculus, skolemization also has the advantage that it makes clauses easier to read because the syntax of the term level is enriched and skolem symbols often have a very natural interpretation.

In a formal proof all function symbols including the skolem symbols are purely syntactic objects and thus have – per se – no interpretation. However, the mathematical proof presented refers to a certain model: the natural numbers. Thus, mathematically, $+$ is not interpreted as any binary function symbol satisfying a certain set of axioms but rather as addition in natural numbers. In the same sense we can interpret the newly introduced skolem symbols: not as having an interpretation per se but as having a natural meaning in the standard model that makes the end-sequent true. The skolem symbols with interpretations in this sense are written down in the following table.

Skolem symbol	Interpretation
$s_1(m)$	least prime divisor of m for $m \neq 1$
$s_2(n)$	least divisor of n
$s_3(n)$	$\frac{n}{s_2(n)}$ (this division is possible without remainder)
$s_4(m)$	$\frac{m}{s_1(m)}$ (this division is possible without remainder)
$s_5(j, m)$	$\mathrm{rmd}(j, m)$ — the remainder of the division $\frac{m}{j}$
$s_6(j, m)$	$\mathrm{quot}(j, m)$ — the quotient of the division $\frac{m}{j}$
$s_7(n)$	$n - 1$ for $n > 0$

[3] http://www.logic.at/ceres/primeproof/.

When we speak of the *standard model* in the context of the skolemized proof, then the skolem symbols are understood to be interpreted as explained above.

The skolemization is performed automatically by the CERES-system. The skolem symbols need not be schematized; they are the same for all $\varphi(k)$, because they only depend on the end-sequent: The part $P[k]$ of the end-sequent that changes with k has no influence on them as it does not contain quantifiers. The above table of skolem symbols was easy to assemble manually by comparing the skolemized end-sequent to the original.

5.2. The characteristic clause set

The next step consists in the extraction of the characteristic clause set from the proof. As we are dealing with a proof scheme that gives rise to an infinite sequence of proofs, an automated extraction of the characteristic clause set working on the whole sequence is not possible. Instead, we consider a scheme of clause sets that is defined in a way that is analogous to the proof scheme and that verifiably corresponds to the characteristic clause sets of an initial segment of the infinite sequence of proofs.

To the automatically generated clause sets $\mathrm{CL}(\varphi(k))$, we add the set AX of the background theory and apply redundancy-elimination (subsumption and tautology-deletion, see e.g. [15]). This proved very useful for decreasing the size and increasing the readability of the clause sets. Let $\mathcal{C}_i$ be the subset of $\mathrm{CL}(\varphi_i)$ after redundancy-elimination w.r.t. AX. For example, for $k = 0, 3, 12$ respectively we obtained the sizes

$$|\mathrm{CL}(\varphi(0))| = 57,\ |\mathrm{CL}(\varphi(3))| = 222,\ |\mathrm{CL}(\varphi(12))| = 1041$$

for the extracted clause sets and

$$|\mathcal{C}_0| = 13,\ |\mathcal{C}_3| = 34,\ |\mathcal{C}_{12}| = 97$$

after redundancy-elimination. The characteristic clause sets after redundancy-elimination can be defined as follows:

$$\mathrm{CL}_r := \mathcal{C}_r \cup \mathrm{AX}$$

where

$$\mathcal{C}_r := A \cup \bigcup_{i=0}^{r} B_i \cup \{C_r\}$$

for

$$C_r := \ \vdash m_0 = 1, s_1(m_0) = p_0, \ldots, s_1(m_0) = p_r,$$

$B_i :=$

$$0 < p_i \vdash p_i = s_7(p_i) + 1$$
$$0 < p_i \vdash t_0 = s_5(p_i, t_0) + (s_6(p_i, t_0) * p_i)$$
$$0 < p_i, s_5(p_i, t_0) = 0 \vdash t_0 = 0 + (s_6(p_i, t_0) * p_i)$$
$$0 < p_i \vdash s_5(p_i, t_0) < p_i$$
$$t_0 = p_i, m_0 * n_0 = t_0 \vdash m_0 = 1, m_0 = t_0$$
$$t_0 = p_i \vdash 1 < t_0$$
$$t_0 = p_i, 1 = n_0 * t_0 \vdash$$

and $A :=$

$$\vdash m_0 = 1, s_1(m_0) * s_4(m_0) = m_0$$
$$\vdash m_0 + (((k * (l_0 + (1 + 1))) + (l_0 * (m_0 + 1))) + 1) = k + ((k + (m_0 + 1)) * (l_0 + 1))$$
$$m_0 = k_0 + (r_0 * ((t_0 + 1) * (t_1 + 1))) \vdash m_0 = k_0 + ((r_0 * (t_0 + 1)) * (t_1 + 1))$$

$$m_0 = k_0 + (r_0 * ((t_0 + 1) * (t_1 + 1)))$$
$$\vdash m_0 = k_0 + ((r_0 * (t_1 + 1)) * (t_0 + 1))$$
$$\vdash (((t_0 + 1) * t_1) + t_0) + 1 = (t_0 + 1) * (t_1 + 1)$$

We have verified that CL_r is the characteristic clause set of the proof $\varphi(r)$ produced by the system `ceres` after redundancy-elimination until $r \leq 10$. Given the highly regular and simple structure of the proofs in the sequence (being generated by a scheme) and of the corresponding clause sets we take that as empirical evidence that CL_r is the characteristic clause set of $\varphi(r)$ after redundancy-elimination for all r. From now on we will work with the scheme CL_r of clause sets.

Note that the question whether this is true for all r or if there is a large r violating this, is in fact of minor importance for the generation of elementary mathematical proofs. The point is that we can – in the end – give an elementary mathematical proof of the theorem under investigation which is possible as we continue to work with a scheme of clause sets giving rise to a scheme of elementary formal proofs. Nevertheless, in the long run it would be more satisfactory to integrate these kind of arguments into the method itself. Therefore one of our goals for future work is the integration of induction into the method.

5.3. The resolution refutation: Euclid's proof

The original aim in the development of CERES was a full automation of the cut-elimination method. The hard part of the whole procedure, which consists in the refutation of the characteristic clause set, was successfully mastered for simple mathematical proofs like the tape proof (see [4] and [5]). However, for the clause sets CL_r, the situation is radically different: even a long series of systematic tests with Prover9 [3], the successor of otter, gave only a refutation of CL_0, which is not very informative. It seems that, for the purpose of cut-elimination, specific theorem proving methods have to be developed. Therefore the sequence of refutations defined in this section was constructed by hand.

Below we define resolution refutations of the clause sets CL_r for arbitrary r. We begin by listing those clauses which play the major role in the proof:

$$I : \vdash m_0 = 1,\ s_1(m_0) * s_4(m_0) = m_0,$$
$$C_r : \vdash m_0 = 1,\ s_1(m_0) = p_0, \ldots, s_1(m_0) = p_r,$$
$$\mathrm{L} :\ 1 < k,\ k = 1 \vdash$$
$$\mathrm{R} :\ k < l,\ k < m,\ l < m,\ k + (i * m) = l + (j * m) \vdash$$

In clause I the function symbol s_1 is a one-place Skolem symbol which can be interpreted as: $s_1(m) =$ the least prime divisor of m. Then s_4 stands for m divided by $s_1(m)$. So clause I simply tells that for an $m_0 \neq 1$ there exists such a prime divisor $s_1(m_0)$. s_1, s_4 and s_7 (which will be used later) are the only Skolem symbols used in this proof. For the interpretation of the Skolem symbols and the definition of the standard interpretation see Section 5.1.

C_r is the central clause of the resolution refutation because it is the only one (!) in the set CL_r which is false in the standard interpretation. It says that for every number m_0, which is different from 1, the least prime divisor of m_0 is one of the primes $p_0, \ldots, p_r$.

L is trivial, but plays a technical role in the proof. It simply says that $1 < k$ and $1 = k$ cannot hold simultaneously.

R is the central number theoretic tool in the refutations. It expresses the fact that two different remainder classes modulo m (i.e. the classes congruent to k and to l modulo m for $k \neq l$) are disjoint. It is used as axiom in the proof ψ_2 in order to show that the finite union of the $\nu(0, p_i)$ progressions is closed.

For all $j \in \{1, \ldots, r\}$ the following lemmas are easily derivable:

$$II_j : \vdash 0 < p_j,$$
$$III_j : \vdash 1 < p_j.$$

Now we resolve the first atoms $m_0 = 1$ in I and in C_r with $k = 1$ in L and obtain

$$I_a :\ 1 < k \vdash s_1(k) * s_4(k) = k,$$
$$IV_r :\ 1 < k \vdash s_1(k) = p_0, \ldots, s_1(k) = p_r.$$

iterated paramodulations of I_a and IV_r give the clause

$$D_r :\ 1 < k \vdash p_0 * s_4(k) = k, \ldots, p_r * s_4(k) = k.$$

D_r (such as C_r) is false in the standard interpretation and says that for any $k > 1$ one of the primes $p_0, \ldots, p_r$ divides k.

The key steps in the proof come from the clause R. Paramodulation with $\vdash 0 + x = x$ and $\vdash x * y = y * x$ gives the clause

$$V : 0 < l,\ 0 < m,\ l < m,\ m * i = m * j + l \vdash .$$

V represents a special case of R and tells that the sets of numbers congruent 0 modulo m and of those congruent l modulo m (for $0 < l < m$) are disjoint.

Paramodulation to V with the associative law for $*$ (assoc$*$: $\vdash k * (j_1 * j_2) = (k * j_1) * j_2$ by unifying $m * j$ with $k * (j_1 * j_2)$) gives

$$V' : 0 < l,\ 0 < m,\ l < m,\ m * i = (m * j_1) * j_2 + l \vdash .$$

An iteration of the application of assoc$*$ on the innermost term of the right-hand-side of the equation (and renaming of variables) gives:

$$V^* : 0 < l,\ 0 < m,\ l < m,\ m * i = (\ldots (m * j_1) * j_2 * \ldots) * j_r + l \vdash .$$

Further applications of associativity yields

$$E_r : 0 < l,\ 0 < m,\ l < m,\ m * i = m * (\ldots (j_1 * j_2) * \ldots) * j_r) + l \vdash .$$

The clause E_r is the key to the Euclidean construction. Though E_r and V express the same number theoretic property we obtain a special replacement of the variable j by a product term which plays an important role in the refutation. Note that all substitutions appearing in the deduction above are most general unifiers produced by resolution and paramodulation.

By resolving the atom $m * i = m * (\ldots (j_1 * j_2) * \ldots) * j_r) + l$ in E_r and $p_0 * s_4(k) = k$ in D_r we obtain

$$VI : 1 < t,\ 0 < l,\ 0 < p_0,\ l < p_0 \vdash p_1 * s_4(t) = t, \ldots p_r * s_4(t) = t$$

For $t = p_0 * (j_1 * \cdots * j_r) + l$. The m.g.u. of the resolution is

$$\{m \leftarrow p_0,\ i \leftarrow s_4(t),\ k \leftarrow t\}.$$

resolutions of VI with III_0 and $IIII_0$ and $0 < 1$ have the effect that l becomes 1 and that only $1 < t_0$ for the term

$$t_0 = p_0 * (j_1 * \cdots * j_r) + 1$$

remains on the left hand side of the clause. The resulting clause is

$$F_1 : 1 < t_0 \vdash p_1 * s_4(t_0) = t_0, \ldots, p_r * s_4(t_0) = t_0.$$

Note that putting the m.g.u.s of both resolutions together and applying the substitutions to the resolved atoms in the resolution for VI we obtain the atom

$$p_0 * s_4(p_0 * (j_1 * \cdots * j_r) + 1) = p_0 * (j_1 * \cdots * j_r) + 1$$

which expresses that there are numbers which are congruent 0 modulo p_0 and congruent 1 modulo p_0 at the same time. But this atom is falsified by the principle R. Here we see the first step of Euclid's construction.

Now assume inductively that we have derived the clause

$$F_s : 1 < t_s \vdash p_{s+1} * s_4(t_s) = t_s, \ldots, p_r * s_4(t_s) = t_s$$

where

$$t_s = p_0 * \cdots * p_s * J_{s+1} + 1, \text{ for}$$
$$J_{s+1} = j_{s+1} * (j_{s+2} * (\ldots * j_r) \ldots),$$
$$t_r = p_0 * \cdots * p_r + 1.$$

We distinguish two cases

(a) $s = r$,
(b) $s < r$.

In case (a) the term J_{s+1} disappears and F_r is

$$1 < t_r \vdash$$

for $t_r = p_0 * \cdots * p_r + 1$.

In case (b) we transform F_s (by multiple application of commutativity and associativity of $*$) to

$$G_s : 1 < t_s \vdash p_{s+1} * s_4(t_s) = j_{s+1} * ((p_0 * \cdots * p_s) * J_{s+2}) + 1, \ldots, p_r * s_4(t_s) = t_s.$$

where the term J_{s+2} disappears if $r = s + 1$. Now G_s can be resolved with the clause:

$$V : 0 < l,\ 0 < m,\ l < m,\ m * i = m * j + l \vdash$$

on the atom

$$p_{s+1} * s_4(t_s) = j_{s+1} * ((p_0 * \cdots * p_s) * J_{s+2}) + 1$$

in G_s. The m.g.u. is

$$\sigma_t : \{m \leftarrow p_{s+1},\ j_{s+1} \leftarrow p_{s+1},\ i \leftarrow s_4(t_{s+1}),\ j \leftarrow ((p_0 * \cdots * p_s) * J_{s+2}),\ l \leftarrow 1\}.$$

where

$$t_{s+1} = p_0 * \cdots * p_{s+1} * J_{s+2} + 1.$$

The result is the clause

$$1 < t_{s+1},\ 0 < 1,\ 0 < p_{s+1},\ 1 < p_{s+1} \vdash \\ p_{s+2} * s_4(t_{s+1}) = t_{r+1}, \ldots, p_r * s_4(t_{s+1}) = t_{s+1}.$$

The resolved atom (under the m.g.u. σ_t) is

$$p_{s+1} * s_4(t_{s+1}) = p_{s+1} * ((p_0 * \cdots * p_s) * J_{s+2}) + 1.$$

Again this atom expresses, that there exists a number which is congruent 0 modulo p_{s+1} and congruent 1 modulo p_{s+1} at the same time, which is falsified by the principle V.

Now we resolve with $\vdash 0 < 1$, II_{s+1} and III_{s+1} and obtain

$$1 < t_{s+1} \vdash p_{s+2} * s_4(t_{s+1}) = t_{s+1}, \ldots, p_r * s_4(t_{s+1}) = t_{s+1}.$$

which is just F_{s+1}.

We complete the resolution proof by considering the case (a) and by deriving $\vdash$. In case (a) we have

$$F_r : 1 < t_r \vdash .$$

Paramodulations with the clauses $\vdash p_j = s_7(p_j) + 1$ for $j = 1, \ldots, r$ (obtained from the set B_j and II_j) give

$$1 < (s_7(p_0) + 1) * \cdots * (s_7(p_r) + 1) + 1 \vdash .$$

Several applications of distributivity and associativity then gives the clause

$$1 < (w + 1) + 1 \vdash$$

for a term w. Finally resolution with the clause $\vdash 1 < (k + 1) + 1$ which can be derived from the axioms yields the empty clause $\vdash$ and thus a contradiction.

To see the mathematical argument as a whole, we consider the global unifier applied to the clause D_r; then we obtain the instance

$$D'_r : 1 < t_r \vdash p_0 * s_4(t_r) = t_r, \ldots, p_r * s_4(t_r) = t_r$$

for $t_r = p_0 * \cdots * p_r + 1$. We only have to realize that $1 < t_r$ holds. Then we are left to refute that any of the primes p_i divides t_r. This is exactly Euclid's argument in the construction of infinitely many primes.

The construction above can be verified, at least for small n, by using an automated theorem prover. In the long run, however, it would be fruitful to develop resolution proofs interactively using a flexible theorem proving environment. Such a system is currently under development in the research group of the authors.

5.4. Another possible refutation

The former section might suggest that Euclid's proof is *the* combinatorial kernel of Fürstenberg's proof. This would mean that all resolution refutations of the characteristic clause sets somehow represent Euclid's construction. We show below that this is not the case, by analyzing the case of two primes, i.e. the clause set containing the clause

$$C_1\colon \vdash m_0 = 1,\ s_1(m_0) = p_0,\ s_1(m_0) = p_1.$$

Like in the resolution derivation in Section 5.3 we derive from C_1 the clause

$$D_1\colon\ 1 < k \vdash p_0 * s_4(k) = k,\ p_1 * s_4(k) = k.$$

From L we derive the special version

$$\mathrm{L}'\colon\ 1 < m, m * i = m + 1 \vdash$$

by paramodulation and resolution, corresponding to the instances $k = 0$, $l = 1$ and $j = 1$. Now we construct two resolvents from L′ and D_1. The first one selects the atom $p_0 * s_4(k) = k$ from D_1 and resolves with the atom $m * i = m + 1$ in L′. The m.g.u. is $\{m \leftarrow p_0, k \leftarrow p_0 + 1, i \leftarrow s_4(p_0 + 1)\}$ and the resolvent

$$R_0\colon\ 1 < p_0, 1 < p_0 + 1 \vdash p_1 * s_4(p_0 + 1) = p_0 + 1.$$

In a completely symmetric way we obtain the resolvent

$$R_1\colon\ 1 < p_1, 1 < p_1 + 1 \vdash p_0 * s_4(p_1 + 1) = p_1 + 1.$$

As the unit clauses $\vdash 1 < p_i, \vdash 1 < p_i + 1$ for $i = 0, 1$ are easily derivable (see Section 5.3) we resolve them with R_0 and R_1 and obtain

$$R_{00}\colon \vdash p_1 * s_4(p_0 + 1) = p_0 + 1,$$
$$R_{11}\colon \vdash p_0 * s_4(p_1 + 1) = p_1 + 1.$$

Now we continue at little more informally, but would like to point out that a fully formal derivation can be obtained on the basis of the axioms (which, however, is quite long). We distinguish three cases:

(a) $p_0 = p_1$.
In this case we obtain from R_{00} the clause

$$R'_{00}\colon \vdash p_0 * s_4(p_0 + 1) = p_0 + 1$$

which gives $1 < p_0 \vdash$ under resolution with L′. A further resolvent with $\vdash 1 < p_0$ then gives contradiction $\vdash$. Note that an analogous refutation can be obtained via R_{11}.

(b) $p_0 < p_1$.
Then $p_0 + 1 \leq p_1$ for $x \leq y =_{def} \vdash x < y, x = y$. So R_{00} yields

$$p_1 * s_4(p_0 + 1) \leq p_1 \quad \text{and} \quad p_1 * s_4(p_0 + 1) = p_0 + 1$$

admitting the only solution $s_4(p_0 + 1) = 1$ and $p_1 = p_0 + 1$. Substituting this new equation into R_{11} gives

(i) $p_0 * s_4(p_1 + 1) = (p_0 + 1) + 1$.

Elementary arithmetic based on case analysis yields the only solution $p_0 = 2,\ s_4(p_1 + 1) = 2$. So for p_1 we get

$$p_1 = p_0 + 1 = 3.$$

The interesting point here is the derivation of concrete witnesses 2 and 3 from the clause set. Using the clauses $\vdash p_0 = 2, \vdash p_1 = 3$ in paramodulation with D_1 gives the clause

(ii) $1 < k \vdash 2 * s_4(k) = k,\ 3 * s_4(k) = k$.

By evaluating k to 5, we obtain

(iii) $1 < 5 \vdash 2 * s_4(5) = 5,\ 3 * s_4(5) = 5$.

(iii) can be refuted by elementary "school" arithmetic.

(c) $p_1 < p_0$.
This case is completely symmetric to (b) and we obtain $p_0 = 3,\ p_1 = 2$.

The derivation above illustrates that mathematical arguments are derivable from the clause set which essentially differ from Euclid's one. The above argument provides the concrete witnesses 2 and 3, while Euclid's argument extends an arbitrary set of primes by a new one. Therefore we must consider the characteristic clause set as a reservoir of combinatorial proofs rather than the representation of a single one.

6. Conclusion

We have applied the CERES-method to Fürstenberg's topological proof of the infinity of primes in order to demonstrate that Euclid's original proof is a combinatorial kernel of Fürstenberg's argument. But it is not the only one: Other combinatorial proofs can be obtained from it. This work constitutes a proof-of-concept example for a semi-automated analysis of realistic and interesting mathematical proofs providing new information about them.

The analysis has shown that the characteristic clause set of a proof is not only a formal tool but contains essential mathematical information about the proof. Thus the extraction of the characteristic clause set alone (without going through the whole cut-elimination) is of considerable mathematical interest. The characteristic clause set provides structural information about the proof which – in principle – cannot be obtained by a Gentzen-like method. Note that all Gentzen normal forms can be obtained by the CERES-method, see [9]. Moreover, there are CERES normal forms which are not Gentzen normal forms; refutations of the characteristic clause set need not respect the ancestral relations in the proof.

The skolemization of the input proof as preprocessing is not only necessary for the resolution part, but actually makes the clauses easier to read as there are natural interpretations of the skolem symbols as number-theoretic functions. We could also see that the formal language becomes more understandable to the human reader by adding more symbols (e.g. also $<$) and by allowing more axioms (in contrast to the reductionist approach usually taken in the mathematical analysis of formal theories).

This application has also demonstrated the limits of the automation of the current method: While the characteristic clause set can be generated fully automatically, these clause sets are hard problems for current automated theorem provers. However, there exists serious potential for improvement: We plan to develop and implement resolution refinements that are targeted specifically towards characteristic clause sets extracted from proofs with cuts, for example semantic resolution or variants of indexed resolution taking the structure of the original proof into account.

The proof under investigation has been formalized as an infinite sequence of pure first-order proofs. For further applications to mathematical proofs it is more satisfactory to extend the method to cover also induction which would allow the formalization of such a proof as a formal derivation without schemes. We plan to extend CERES to the theory ACA_0 (which covers also induction). This will further contribute to the long-range goal of this work: To establish the use of automated and semi-automated methods for the logical analysis of mathematical arguments.

Acknowledgements

We would like to thank one of the referees for constructive criticism and many suggestions for improvements; his (her) comments had a strong impact on the final version of this paper.

References

[1] CERES, http://www.logic.at/ceres/.
[2] Handy LK, http://www.logic.at/hlk/.
[3] Prover9, http://www.prover9.org/.
[4] M. Baaz, S. Hetzl, A. Leitsch, C. Richter, H. Spohr, Cut-elimination: Experiments with CERES, in: F. Baader, A. Voronkov (Eds.), Logic for Programming, Artificial Intelligence, and Reasoning, LPAR 2004, in: Lecture Notes in Computer Science, vol. 3452, Springer, 2005.
[5] M. Baaz, S. Hetzl, A. Leitsch, C. Richter, H. Spohr, Proof transformation by CERES, in: J.M. Borwein, W.M. Farmer (Eds.), Mathematical Knowledge Management, MKM 2006, in: Lecture Notes in Artificial Intelligence, vol. 4108, Springer, 2006.
[6] M. Baaz, A. Leitsch, On skolemization and proof complexity, Fundamenta Informaticae 20 (4) (1994) 353–379.
[7] M. Baaz, A. Leitsch, Cut normal forms and proof complexity, Annals of Pure and Applied Logic 97 (1999) 127–177.
[8] M. Baaz, A. Leitsch, Cut-elimination and redundancy-elimination by resolution, Journal of Symbolic Computation 29 (2) (2000) 149–176.
[9] M. Baaz, A. Leitsch, Towards a clausal analysis of cut-elimination, Journal of Symbolic Computation 41 (3–4) (2006) 381–410.
[10] V. Danos, J.-B. Joinet, H. Schellinx, A new deconstructive logic: Linear logic, Journal of Symbolic Logic 62 (3) (1997) 755–807.
[11] E. Eder, Relative Complexities of First-order Calculi, Vieweg, 1992.
[12] H. Fürstenberg, On the infinitude of primes, American Mathematical Monthly 62 (1955) 353.
[13] G. Gentzen, Untersuchungen über das logische Schließen, Mathematische Zeitschrift 39 (1934–1935) 176–210; 405–431.

[14] J.-Y. Girard, Proof Theory and Logical Complexity, Elsevier, 1987.
[15] A. Leitsch, The resolution calculus, in: Texts in Theoretical Computer Science, Springer, 1997.
[16] H. Luckhardt, Herbrand-Analysen zweier Beweise des Satzes von Roth: Polynomiale Anzahlschranken, Journal of Symbolic Logic 54 (1) (1989) 234–263.
[17] M. Parigot, $\lambda\mu$-calculus: An algorithmic interpretation of classical natural deduction, in: A. Voronkov (Ed.), Logic Programming and Automated Reasoning, International Conference LPAR'92, Proceedings, in: Lecture Notes in Computer Science, vol. 624, Springer, 1992.
[18] C. Richter, Proof transformations by resolution — computational methods of cut-elimination, Ph.D. Thesis, Vienna University of Technology, 2006.
[19] S.G. Simpson, Subsystems of second-order arithmetic, in: Perspectives in Mathematical Logic, Springer, 1999.