Math. Struct. in Comp. Science (2006), *vol.* 16, *pp.* 885–900. © 2006 Cambridge University Press
doi:10.1017/S0960129506005627 Printed in the United Kingdom

A logical approach to abstract algebra

THIERRY COQUAND and HENRI LOMBARDI

Computer Science and Engineering, Chalmers University of Technology and Göteborg University, 412 96 Göteborg, Sweden
Email: coquand@cs.chalmers.se

Received 15 September 2005; revised 28 September 2005, 21 October 2005 and 4 March 2006

Recent work in constructive mathematics shows that Hilbert's program works for a large part of abstract algebra. Using in an essential way the ideas contained in the classical arguments, we can transform most of the highly abstract proofs of 'concrete' statements into elementary proofs. Surprisingly, the arguments we produce are not only elementary but also mathematically clearer, and not necessarily longer. We present an example where the simplification was significant enough to suggest an improved version of a classical theorem. For this we use a general method to transform some logically complex first-order formulae into a geometrical form, which may be interesting in itself.

1. Introduction

The purpose of this paper is to survey some of our recent work in constructive algebra (Coquand and Lombardi 2002; Coquand *et al.* 2005; Coquand 2004; Coquand 2005; Coquand *et al.* 2004; Coste *et al.* 2001) from the point of view of mathematical logic. We illustrate the relevance of simple logical considerations in the development of constructive algebra.

We analyse the logical complexity of statements and proofs in abstract algebra. Two notions of formulae, one geometric the other first-order, will play an important role. The two notions are in general incomparable. Both notions have a fundamental 'analytical' property: if a statement is formulated in first-order logic and has a proof, then we know that it can be proved in a first-order way. Similarly, if a geometric statement holds, it has a constructive proof with a particularly simple tree form (Bezem and Coquand 2003; Coquand 2005; Coste *et al.* 2001).

We first present some basic examples in algebra, which are directly formulated with the required logical complexity: the first is an implication between equational statements, and the second is coherent, that is geometric and first-order. We then present a more elaborate example, which was a mathematical conjecture and for which a first-order formulation is not obvious. We can further transform it to a coherent formulation. Knowing *a priori* that we had to look for an 'analytical' proof involving simple algebraic manipulations only then helps in finding a proof. We then show on one concrete example, due to Kronecker, that in this way we can get non-trivial algorithms on polynomials. One main theme, which is also present in Ducos *et al.* (2004) is the elimination of Noetherian hypotheses to get a proof of simple first-order statements. In some complex examples one needs a concrete interpretation of the notion of minimal prime ideals, and we present such an interpretation.

2. Logical complexity

The theory of commutative rings is a first-order theory, and is actually even equational. We need 3 symbols of functions $+, \times, -$ (we will often write ab for $a \times b$), two constants $0, 1$ and the axioms

$$x + (-x) = 0, \qquad x + (y + z) = (x + y) + z, \qquad x + y = y + x, \qquad x + 0 = x$$
$$x1 = x, \qquad xy = yx, \qquad x(yz) = (xy)z, \qquad x(y + z) = xy + xz\,.$$

Some elementary concepts and theorems of commutative abstract algebra can be formulated in this language. For instance, the notion of *integral* ring is not equational but can be represented by the universally quantified first-order formula

$$xy = 0 \to (x = 0 \vee y = 0)\,.$$

By the completness theorem of first-order logic, we know that if a theorem can be formulated in a first-order way, it has a proof in first-order logic. Furthermore, if it is formulated equationally, we even know, by Birkhoff's completness theorem, that there is a purely equational proof. As we shall explain below, this can be seen as a partial realisation of Hilbert's program.

However, if we take a basic book in abstract algebra, such as Atiyah and Macdonald (1969) or Matsumura (1986), we discover that even basic theorems are not formulated in a first-order way because of the introduction of abstract notions. These abstract notions include:

1 arbitrary ideals of the rings, which are defined as subsets and are not a first-order notion;
2 *prime* or *maximal* ideals, whose existence usually relies on Zorn's lemma;
3 Noetherian hypotheses.

These notions have different levels of non-effectivity. The property of being Noetherian can be captured by a generalised inductive definition (Jacobson and Lofwall 1991), but then we leave first-order logic. The notion of prime ideals seems even more ineffective, the existence of prime ideals being usually justified by the use of Zorn's lemma.

Furthermore, a notion such as 'being nilpotent' cannot be expressed in a first-order way since it involves an infinite countable disjunction.

G. Wraith (Wraith 1980) points out the relevance of the notion of *geometric formulae* for constructive algebra. One defines first the notion of *positive formulae*: a positive formula is one formula of the language of rings built using positive atomic formulae (equality between two terms) and the connectives $\vee, \wedge$. Special cases are the empty disjunction, which is the false formula $\bot$, and the empty conjunction, which is a true formula. We also allow existential quantification and infinite disjunction indexed over natural numbers[†]. A

[†] Sometimes, the notion of an 'arbitrary' infinite disjunction is allowed, but in this paper we shall only need this generality in the final section.

https://doi.org/10.1017/S0960129506005627 Published online by Cambridge University Press

geometric formula is an implication between two positive formulae. A *coherent* formula is a formula that is both geometric and first-order. Note that, as special cases, any positive formula is geometric, and the negation of a positive formula is geometric. As a special case of a coherent formula, we have the notion of a *Horn* formula, which is an implication $C \rightarrow A$ where C is a conjunction of atomic formulae and A is an atomic formula. Horn theories correspond to the notion of an *atomic system* in Prawitz (1971). For example, equational theories are Horn theories.

A coherent way to express the fact that a ring is a field is

$$\forall x.x = 0 \vee \exists y.xy = 1\,.$$

On the other hand, the following formula, classically equivalent, is *not* geometric

$$\forall x.(\neg x = 0) \rightarrow \exists y.xy = 1\,.$$

The notion for $a \in R$ to be nilpotent is not first-order, but it can be expressed as a positive formula: a is nilpotent if and only if $a^n = 0$ for some $n \in \mathbb{N}$. On the other hand, 'being reduced', that is to have only 0 as a nilpotent element, can be expressed by the following Horn formula

$$\forall x.x^2 = 0 \rightarrow x = 0\,.$$

Another typical example (Wraith 1980) of a notion expressed geometrically is the notion of a *flat* module M over a ring R. This says that if we have a relation $PX = 0$ where P is a row vector with coefficient in R and X is a column vector with elements in M, then we can find a rectangular matrix Q and a vector Y such that $QY = X$ and $PQ = 0$. Since we do not say anything about the size of Q, this statement implicitly involves an infinite disjunction over natural numbers. Thus, the notion of a flat module is not first-order but geometric.

As stressed by G. Wraith, the importance of geometric formulae comes from *Barr's Theorem.*

Theorem 2.1. If a geometric sentence is deducible from a geometric theory in classical logic with the axiom of choice, it is also deducible from it intuitionistically.

Furthermore, in this case there is always a proof with a simple branching tree form, a *dynamical* proof (Coquand 2005; Bezem and Coquand 2003; Coste *et al.* 2001). In general, this tree may be infinitely branching, but if the theory is *coherent*, that is, geometric *and* first-order, the proof is a finitely branching tree.

In order to describe these proofs, it is convenient first to note that any coherent formula is equivalent to a conjunction of formulae of the form $C \rightarrow D$ where C and D are given by the following grammar:

$$C ::= 1 \mid C \wedge A \qquad D ::= 0 \mid D \vee E \qquad E ::= (\exists \overrightarrow{v})C\,.$$

Here 0 and 1 represent the empty disjunction and conjuntion, respectively. We may write D for $1 \rightarrow D$, A for $1 \wedge A$, and so on, to economies on empty conjunctions, disjunctions, existential quantifications and brackets as much as possible. We will call a closed atomic formula a *fact*. In most algebraic theories, the only facts are equalities. We can thus

https://doi.org/10.1017/S0960129506005627 Published online by Cambridge University Press

consider a coherent theory to be a collection of formulae of the form $C \to D$. We look at the formulae of the theory T as a collection of *rules*. The purpose of a dynamical proof is to establish the correctness of a fact with reference to some given set of facts X and the dynamical rules belonging to T, starting from a given set of facts. A dynamical proof shows when a given fact F is a consequence of the given set of facts X. Formally, a dynamical proof is a rooted tree. At the root of the tree is the set of facts X that we start with. Each node consists of a set of facts, representing a state of information. The sets increase monotonically as we progress from the root to the leaves. The successors of a node are determined by the dynamical rules, which add new information to the set of already available atomic formulas. The different immediate successors of a node correspond to case distinctions. Every leaf of a dynamical proof contains either a contradiction or the fact F under investigation. If all leaves contain a contradiction, the given set of atomic formulas is contradictory.

In the special case in which all formulae are of the form $C \to A$, the tree has no branching. We get something equivalent to the notion of the *atomic systems* introduced by Prawitz (Prawitz 1971). In particular, equational theories are of this form. The crucial point is that this notion of a dynamical proof is *complete* for deducibility in a coherent theory, and that a dynamical proof uses only intuitionistically valid inference steps. Barr's theorem, which we have cited above, follows from this: if a geometric sentence is deducible from a geometric theory in classical logic, even with the axiom of choice, it is a semantical consequence of the theory, and so, by completeness, it can be derived by a dynamic proof.

In the more general case of a geometric theory, where we also allow countable disjunctions in positive formulae, we have to generalise the notion of a dynamical proof with countable branching, but it can be proved that completeness still holds.

We can now explain the sense in which these completeness theorems can be seen as a realisation of Hilbert's program. We consider the facts, or atomic sentences, as *concrete statements*. A dynamic proof can be seen as a 'logic-free' and elementary way of deriving new concrete statements from a given collection of concrete statements. By completeness, we know that if we can derive a concrete statement from this theory with the use of ideal methods (typically using Zorn's lemma), there is also an elementary derivation. Prawitz has a similar analysis for the case of Horn theories.

It is suggestive to interpret the construction of such a dynamical proof in computational terms: each geometric axiom can be interpreted as the specification of a subprogram. The actual computation of a witness using these subprograms can then be seen as a branch in the dynamical proof. For instance, the coherent axiom for fields

$$x = 0 \vee \exists y.1 = xy$$

can be seen as the specification of a program that, given an element a, tests whether $a = 0$ or not, and in the later case, gives an element b such that $ab = 1$.

However, both the completeness theorem and Barr's theorem are purely heuristic results from a constructive point of view. Indeed, they are both proved using non-constructive methods, and do not give algorithms to transform a non-effective proof to an effective one. In practice, however, in all the examples analysed so far, it has been possible to extract effective arguments from the ideas present in the non-effective proofs.

https://doi.org/10.1017/S0960129506005627 Published online by Cambridge University Press

3. Some basic examples

In this section we provide two elementary examples where Barr's theorem can be invoked. They are directly expressed with the appropriate logical complexity. In the next section, we present more elaborate examples where some work has to be done in order to get the right logical complexity. For the first example of this section, Birkhoff's completeness theorem for equational logic is enough. Both examples appear at the beginning of Matsumura (1986).

3.1. *Dimension over rings*

The following result is usually proved using maximal ideals (Matsumura 1986).

Theorem 3.1. If $n < m$ and $f : R^n \to R^m$ is a surjective linear map, then $1 = 0 \in R$.

What is the logical complexity of this statement? If we fix n and m, say $n = 2$ and $m = 3$, the statement becomes an implication from a conjunction of equalities to $1 = 0$. More precisely, the hypothesis is that we have a 2×3 matrix P and a 3×2 matrix Q such that $PQ = I$. That is, we have 9 equations of the form

$$p_{i1}q_{1j} + p_{i2}q_{2j} = \delta_{ij}$$

with $i, j = 1, 2, 3$.

A typical classical proof uses the existence of maximal ideals: if R is non-trivial, it has a maximal ideal $\mathfrak{m}$. If $k = R/\mathfrak{m}$, we have a surjective map from k^n to k^m, and this is a contradiction.

It is possible to transform this argument into equational reasoning. Here we simply note that the concrete statement means that 1 belongs to the ideal generated by $p_{i1}q_{1j} + p_{i2}q_{2j} - \delta_{ij}$, seeing p_{ik}, q_{kj} as indeterminates, and this can be certified with a simple algebraic identity.

3.2. *Projective modules over local rings*

We shall analyse a standard theorem on *local* rings. Classically, a local ring is defined to be a ring with only one maximal ideal. Constructively, the locality of R is expressed by the coherent formula

$$Inv(x) \vee Inv(1 - x)$$

where $Inv(a)$ means $\exists y.ay = 1$. That this condition is equivalent to the implication

$$Inv(x + y) \to (Inv(x) \vee Inv(y))$$

can be seen directly. We have

$$Inv(xy) \leftrightarrow (Inv(x) \wedge Inv(y)).$$

It then follows that for all x and y we have

$$Inv(x) \vee Inv(1 - xy).$$

https://doi.org/10.1017/S0960129506005627 Published online by Cambridge University Press

Classically, it is possible to derive

$$Inv(x) \vee \forall y. Inv(1 - xy)$$

from this, but constructively this inference is not justified. The last statement says that any element x is invertible or belongs to the *Jacobson radical* of R, which is classically the intersection of all maximal ideals of R. It is easy to see that this is the same as the set of elements x such that all $1 - xy$ are invertible, and this is a first-order characterisation of the Jacobson radical. Thus, we have shown classically that in a local ring an element is invertible or in the Jacobson radical. It follows that the Jacobson radical is the unique maximal ideal of R.

We now analyse the following theorem.

Theorem 3.2. If M is a finitely generated projective module over a local ring R, then M is free.

The concrete formulation of this theorem is as follows (Lombardi and Quitte – to appear).

Theorem 3.3. If F is an idempotent square matrix over a local ring R, then F is similar to a matrix of the form

$$\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

The statement of this theorem, for a fixed size of F, is expressed in coherent logic.

We have a first-order classical derivation, which we can transform by proof-theoretic methods to a constructive first-order derivation.

Proof (classical). Let $f_1, \dots, f_n$ be the column vectors of the matrix F, and $e_1, \dots, e_n$ be the column vectors of the identity matrix I_n, that is, the canonical basis of R^n, so that $e_1 - f_1, \dots, e_n - f_n$ are the column vectors of the matrix $I_n - F$. We have that $f_1, \dots, f_n$ generate $Im(F)$, and $e_1 - f_1, \dots, e_n - f_n$ generate $Im(I_n - F)$. Also, $R^n = Im(F) \oplus Im(I_n - F)$. Let J be the Jacobson radical of R, so that $R/J = k$ is a field, classically. We can extract from $f_1, \dots, f_n$ and $e_1 - f_1, \dots, e_n - f_n$ a basis $g_1, \dots, g_n$ of k^n such that, for each i, we have $Fg_i = g_i$ or 0, that is, each g_i is either in $Im(F)$ or in $Im(I_n - F)$. We can assume that we first group the vectors in $Im(F)$. The determinant of the matrix $P = g_1, \dots, g_n$ is not 0 modulo J, hence it is invertible in R, and $g_1, \dots, g_n$ is a basis of R^n. The matrix PFP^{-1} then has the desired form. □

It is interesting that the constructive argument we give next, which is extracted from this proof, is both simpler and more precise than the classical argument.

Proof (constructive). We build by induction a sequence of column vectors $f'_1, \dots, f'_n$ such that $f'_i = f_i$ or $e_i - f_i$, and that for each m the top $m \times m$ minor of the matrix $f'_1, \dots, f'_m$ is invertible. This is possible because the sum of the minor for $f'_1, \dots, f'_{m-1}, f_m$ and the minor for $f'_1, \dots, f'_{m-1}, e_m - f_m$ is the minor for $f'_1, \dots, f'_{m-1}, e_m$, which is invertible by induction.

https://doi.org/10.1017/S0960129506005627 Published online by Cambridge University Press

In this way, we build an invertible matrix $f'_1,\ldots,f'_n$. We also have $Ff'_i = f'_i$ or 0 for each i. For a suitable permutation $g_1,\ldots,g_n$ of these vectors, we get a matrix P such that PFP^{-1} has the required form. □

Note that the constructive proof can be read as an algorithm: given the matrix F and the 'subprogram' that for each x decides whether x or $1-x$ is invertible, it computes an invertible matrix P such that PFP^{-1} has the required form.

Theorem 3.3 has an interesting history in intuitionistic algebra. It was noted in Mulvey (1974) that an intuitionistic proof of this result could be used to give an alternative proof of Swan's theorem relating fibre bundles on a compact Hausdorff space M with finitely generated projective modules over the ring $C(M)$ (Swan 1962). The result in (Mulvey 1974) is formulated in higher-order intuitionistic logic. In Reyes (1978) it is noted that one can formulate the theorem in first-order logic. The formulation there, which is attributed to A. Kock, is *a priori* weaker than the formulation of Theorem 3.3[†].

Theorem 3.4. If F is an $n \times n$ projection matrix over a local ring R, we can find an $n \times r$ matrix X and an $r \times n$ matrix Y such that $XY = F$ and $YX = I_r$.

This is essentially what is proved in Mulvey (1974). Note, however, that the proof there uses, *a priori*, the fact that an element is invertible or not, and so is not, as it stands, intuitionistically valid. Here we present an intuitionistic version of this argument, which is very close to the classical argument.

Proof. Suppose that we have m column vectors that form a $n\times m$ matrix $X = U_1,\ldots,U_m$ that generate $Im(F)$ (we start with $m = n$ and $X = F$.) We can then find an $m \times n$ matrix Y such that $XY = F$ (to start, we can take $X = F$ and $Y = I_n$ or $Y = F$.) Then $YX = G$ is an $m \times m$ projection matrix since $G^2 = YXYX = YX = G$. We also have $XG = XYX = FX = X$. If we write $G = (c_{ij})$, we have $U_j = \Sigma c_{ij}U_i$ for each j. Since R is local, c_{jj} is invertible or $1 - c_{jj}$ is invertible.

If $1 - c_{jj}$ is invertible for some j, we can express U_j in term of U_i, $i \neq j$, and reduce m by one.

Otherwise, c_{jj} is invertible for all j. The determinant of G is of the form $r + \Pi c_{jj}$ with r in the ideal generated by c_{ij}, $i \neq j$. Since R is local, and Πc_{jj} is invertible, either this determinant is invertible or there exists $i \neq j$ such that c_{ij} is invertible. In the later case, since $U_j = \Sigma c_{ij}U_i$, we can express U_i in terms of U_l, $l \neq i$, and reduce m by one. In the former case, we have that G is invertible. Since $G(I_m - G) = 0$, this implies $G = I_m$, and we have finished. □

[†] In our formulation, we assume that both the image and the kernel of F are free. The formulation of Reyes (1978) assumes only that the image of F is free. However, since the kernel of F is the image of $I_n - F$, and the theorem holds for *all* projection matrix, the two formulations turn out to be equivalent.

https://doi.org/10.1017/S0960129506005627 Published online by Cambridge University Press

4. Serre's splitting-off theorem

4.1. *Classical formulation*

The example we are going to present has its origin in a paper of Serre from 1958 (Serre 1957/1958). It is a purely algebraic theorem, but with a geometrical intuition. The geometrical statement is roughly that if we have a vector fibre bundle over a space of finite dimension, and each fibre has a large enough dimension, we can find a non-vanishing section. We first give the classical formulation, in which both hypotheses and conclusions have a non-elementary form, and then a version in which the conclusion is first-order.

We assume R to be a Noetherian ring, and let $\mathsf{Max}(R)$ be the space of maximal ideals with the topology induced from the Zariski topology. We assume that the dimension of $\mathsf{Max}(R)$ is finite and $< n$ (that is there is no proper chain of irreducible closed sets of length n). For instance, if R is a local ring, then $\mathsf{Max}(R)$ is a singleton and we can take $n = 1$.

If M is a finitely generated module over R and x is a maximal ideal of R, then M/xM is a finite-dimensional vector space over R/x and we let $r_x(M)$ be its dimension. Intuitively, M represents the module of global section of a vector bundle over the space $\mathsf{Max}(R)$ and $r_x(M)$ is the dimension of the fibre at the point x. If $s \in M$, it is suggestive to write $s(x)$ the equivalence class of s in $M(x) = M/xM$. Intuitively, $s(x)$ is a continuous family of sections.

Theorem 4.1 (Serre 1958). If M is a finitely generated projective module over R such that $n \leqslant r_x$ for all maximal ideals x of R, there exists $s \in M$ such that $s(x) \neq 0$ for all $x \in \mathsf{Max}(R)$.

The first step is to give a more concrete formulation of this result. We give only the end result (Coquand *et al.* 2004; Lombardi and Quitte – to appear). If F is a matrix over R, we let $\Delta_k(F)$ be the ideal generated by all minors of F of order k. We say that a vector of elements of R is *unimodular* if and only if 1 belongs to the ideal generated by these elements. With the same hypothesis as before, that the dimension of $\mathsf{Max}(R)$ is $< n$, we can state the following result.

Theorem 4.2 (Serre 1958 – concrete version). If F is an idempotent matrix over R and $\Delta_n(F) = 1$, then there exists a linear combination of the columns of F that is unimodular.

Interestingly, in this form, the theorem can then be seen as a special case of Swan's Theorem (Swan 1967), which is a theorem that was conjectured by Serre. We give the abstract form of the theorem first.

Theorem 4.3 (Swan 1967). If M is a finitely generated module over R such that for each $x \in \mathsf{Max}(R)$ the fibre $M(x)$ can be generated by p elements, then M can be generated by $p + n - 1$ elements.

Theorem 4.4 (Swan 1967 – concrete version). If F is a rectangular matrix over R and $\Delta_n(F) = 1$, then there exists a linear combination of the columns of F that is unimodular.

https://doi.org/10.1017/S0960129506005627 Published online by Cambridge University Press

Only the concrete formulation of these two results reveals their similarities. The generalisation of these theorems to the non-Noetherian case was first established in Coquand *et al.* (2004), by analysing the paper Heitmann (1984) using the techniques that are presented in this paper.

Note that the conclusion of this theorem is expressed in first-order logic, and even positively. The hypothesis, however, is non-elementary: we suppose both that R is Noetherian and that we have a hypothesis on the dimension of $\mathsf{Max}(R)$. It was conjectured that the theorem holds without the hypothesis that R is Noetherian, and this is the statement we want to analyse. We will leave expressing the hypothesis of the theorem $\mathsf{dim}\ (\mathsf{Max}(R)) < n$ in a first-order way.

4.2. *Geometric formulation of the Krull dimension*

The first step is to give an elementary formulation of the notion of the Krull dimension. This is not so easy *a priori* since the usual definition is in terms of a chain of prime ideals: a ring R is of Krull dimension $< n$ if and only if there is no proper chain of prime ideals of length n. An elementary definition is presented in Coquand *et al.* (2005). We first introduce the notion of the *boundary* of an element of a ring: the boundary N_a of a is the ideal generated by a and the elements x such that ax is nilpotent. We then define inductively $\mathsf{Kdim}\ R < n$: for $n = 0$ it means that $1 = 0 \in R$, and for $n > 0$ it means that we have $\mathsf{Kdim}\ (R/N_a) < n - 1$ for all $a \in R$.

For each n, we get a formulation of $\mathsf{Kdim}\ R < n$ that is positive, but *not* first-order. For instance, $\mathsf{Kdim}\ R < 1$ is expressed by the formula

$$\forall x.\exists a. \bigvee_{k \in \mathbb{N}} x^k(1 - ax) = 0,$$

while $\mathsf{Kdim}\ R < 2$ is expressed by

$$\forall x, y.\exists a, b. \bigvee_{k,l \in \mathbb{N}} y^k(x^l(1 - ax) - by)) = 0.$$

We can now express the concrete form of the non-Noetherian version of Forster's theorem (which motivated Swan's theorem in the Noetherian case).

Theorem 4.5 (Heitmann 1984 – concrete version). If $\mathsf{Kdim}\ R < n$ and F is a rectangular matrix over R such that $\Delta_n(F) = 1$, then there exists a linear combination of the columns of F that is unimodular.

The formulation is now geometric (but not first-order). The hypothesis is a positive statement (of the form $\forall\exists$, but the existential quantification is over natural numbers) and the conclusion is purely existential. We expect it to have a constructive proof, which, furthermore, will have a very simple nature. In this case, it is enough to extract this direct proof from the argument in Heitmann (1984). This is carried out in Coquand *et al.* (2004).

https://doi.org/10.1017/S0960129506005627 Published online by Cambridge University Press

4.3. *A new notion of dimension*

We now present a notion of dimension, which was introduced in Coquand *et al.* (2004) and appears implicitly in Heitmann (1984). This notion is finer than the notion of the Krull dimension: we always have $\mathsf{Hdim}\ R \leqslant \mathsf{Kdim}\ R$. Interestingly, $\mathsf{Hdim}\ R \leqslant n$ can be expressed by a first-order formula, but the logical complexity of this formula increases with n, contrary to $\mathsf{Kdim}\ R \leqslant n$, which stays a positive formula for all n.

We get this definition by changing the nilradical in the definition of the Krull dimension by the Jacobson radical J, which is classically the intersection of all maximal ideals, but can be defined in a first-order way as the set of elements a such that $1 - ax$ is invertible for all $x \in R$. We then introduce a new notion of the *boundary* of an element of a ring: the boundary J_a of a is the ideal generated by a and the elements x such that ax is in the Jacobson radical of R. We then define $\mathsf{Hdim}\ R < n$ inductively: for $n = 0$ it means that $1 = 0 \in R$, and for $n > 0$ it means that we have $\mathsf{Hdim}\ (R/J_a) < n - 1$ for all $a \in R$.

What is the logical complexity of $\mathsf{Hdim}\ R < n$? For $n = 1$ we get that $\mathsf{Hdim}\ R < n$ means

$$\forall x.\exists a.\forall y.\exists b.1 = b(1 - yx(1 - ax)),$$

which is a prenex formula with two alternations of quantifiers. For $n = 2$ we get an even more complex formula, and the logical complexity increases with n.

In this way we get a way to state a plausible non-Noetherian version of Swan's theorem in a purely first-order way as an implication

$$\mathsf{Hdim}\ R < n \rightarrow \Delta_n(F) = 1 \rightarrow \exists X, Y.1 = XFY\,,$$

where X is a row vector and Y a column vector. For a given n and a given size of F this is a first-order statement.

The form of the statement for $\mathsf{Hdim}\ R < n$ is particular since it is a purely *prenex* formula. It is then possible to conclude using general proof-theoretic arguments that if we have a first-order classical proof, we also have an intuitionistic proof. From proof theory, one can use the Gentzen sharpened Hauptsatz (Gentzen 1969), or a negative translation.

Yet another logical analysis can be obtained using the notion of Skolem functions, and we think that we provide an example that illustrates well the strength of this notion. We illustrate the idea for $n = 1$ only. We have seen that $\mathsf{Hdim}\ R < 1$ is equivalent to

$$\forall x.\exists a.\forall y.\exists b.1 = b(1 - yx(1 - ax)).$$

If we add two Skolem functions $f(x)$ and $g(x, y)$ to the language of rings, we can reformulate this as

$$1 = g(x, y)(1 - yx(1 - xf(x))).$$

The non-Noetherian version of Swan's theorem then takes a particularly simple form, because in this equational theory, extended with the equation $\Delta_1(F) = 1$, we can build a row vector X and a column vector Y such that $1 = XFY$.

It can be checked that if R is Noetherian, $\mathsf{Hdim}\ R < n$ if and only if $\mathsf{dim}\ (\mathsf{Max}(R)) < n$. A possible generalisation of Serre's theorem can thus be formulated as follows.

https://doi.org/10.1017/S0960129506005627 Published online by Cambridge University Press

Theorem 4.6 (Coquand *et al.* 2004). If $\mathsf{Hdim}\ R < n$ and F is a rectangular matrix over R such that $\Delta_n(F) = 1$, then there exists a linear combination of the columns of F that is unimodular.

The formulation of this theorem is now purely coherent, in a coherent theory that has a specially simple form (no branching). If it holds, it has a purely elementary proof, and knowing this helps in finding a proof (Coquand *et al.* 2004). We can, furthermore, read the proof presented in Coquand *et al.* (2004) as an algorithm that produces a unimodular column.

5. Kronecker's Theorem

In this section we show that, though these results may seem quite abstract, as they are expressed in first-order logic and are *a priori* far from actual computations, they can be used to get concrete computations on polynomials. The previous example of Serre's Theorem may involve too complicated computations, and we shall analyse a simpler statement, the abstract version of a theorem of Kronecker (Heitmann 1984; Coquand 2004). In this case, it is possible to get from an abstract proof a concrete algorithm that could have been formulated by Kronecker (Edwards 2005). We first give the abstract version, which is proved in (Coquand 2004).

Theorem 5.1. If $\mathsf{Kdim}\ R \leqslant n$ and we have $n+2$ elements $g_0, g_1, \ldots, g_{n+1}$, it is possible to find $n+1$ elements $f_0, f_1, \ldots, f_n$ such that $g_0, g_1, \ldots, g_{n+1}$ and $f_0, f_1, \ldots, f_n$ generate the same radical ideal.

This means that some power of f_j is zero mod $g_1, g_2, \ldots, g_{n+2}$ and some power of g_i is zero mod $f_1, f_2, \ldots, f_{n+1}$. This theorem is expressed in geometric logic, and has a simple inductive proof (Coquand 2004). To simplify the discussion, we take $n = 2$. As we have explained, the meaning of $\mathsf{Kdim}\ R \leqslant 2$ is that for all $x_1, x_2, x_3 \in R$ there exists $p_1, p_2, p_3 \in R$ and $k_1, k_2, k_3 \in \mathbb{N}$ such that

$$p_3^{k_3}(p_2^{k_2}(p_1^{k_1}(1 - p_1x_1) - p_2x_2) - p_3x_3) = 0\,.$$

Theorem 5.1 can thus be interpreted as follows: given such an algorithm that produces such an algebraic identity taking as input $x_1, x_2, x_3 \in R$, we can give another algorithm, which produces $f_0, \ldots, f_2$ as a function of $g_0, \ldots, g_3$.

Furthermore, this algorithm is simple and explicit, corresponding to the simplicity of the proof in Coquand (2004), given the algorithm corresponding to $\mathsf{Kdim}\ R \leqslant 2$. Given g_1, g_2, g_3, we find p_1, p_2, p_3 and k_1, k_2, k_3 such that

$$p_3^{k_3}(p_2^{k_2}(p_1^{k_1}(1 - p_1g_1) - p_2g_2) - p_3g_3) = 0\,,$$

and we can then take

$$f_1 = g_1 + g_0h_1, \quad f_2 = g_2 + g_0h_2, \quad f_3 = g_3 + g_0h_3$$

where

$$h_1 = 1 - p_1g_1, \quad h_2 = p_1^{k_1}(1 - p_1g_1) - p_2g_2, \quad h_3 = p_2^{k_2}(p_1^{k_1}(1 - p_1g_1) - p_2g_2) - p_3g_3\,.$$

The correction of the algorithm follows from the fact that we have

$$1 \in <g_1, h_1>, \quad g_1 h_1 \in \sqrt{<g_2, h_2>}, \quad g_2 h_2 \in \sqrt{<g_3, h_3>}, \quad g_3 h_3 \in \sqrt{0}.$$

In Coquand *et al.* (2005), we present a direct proof that Kdim $\mathbb{Q}[X_1, \ldots, X_n] \leqslant n$. For $n = 2$ this reduces to the remark that if we take 3 elements g_1, g_2, g_3 in $\mathbb{Q}[X_1, X_2]$, they are algebraically dependent (see Richman *et al.* (1988) and Edwards (2005)). Such an algebraic dependence relation can always be written

$$p_3^{k_3}(p_2^{k_2}(p_1^{k_1}(1 - p_1 g_1) - p_2 g_2) - p_3 g_3) = 0$$

for some $p_1, p_2, p_3 \in \mathbb{Q}[X_1, X_2]$. Thus, we have Kdim $\mathbb{Q}[X_1, X_2] \leqslant 2$. Since this algorithm corresponds to finding an algebraic dependence relation, complex computations are involved in general.

We can then combine the two algorithms and in this way get a non-trivial algorithm on polynomials, which, given g_0, g_1, g_2, g_3, produces f_0, f_1, f_2 such that g_0, g_1, g_2, g_3 and f_0, f_1, f_2 generate the same radical ideal. In general, we get a constructive proof for the following result, which is a formulation of Kronecker's Theorem.

Theorem 5.2. Let polynomials $g_1, g_2, \ldots, g_m$ in n indeterminates with rational coefficients be given, and let m be greater than $n+1$. Construct $n+1$ polynomials $f_1, f_2, \ldots, f_{n+1}$ in the same indeterminates that are zero mod $g_1, g_2, \ldots, g_m$ and have the property that, for each $i = 1, 2, \ldots, m$, some power of g_i is zero mod $f_1, f_2, \ldots, f_{n+1}$.

The geometrical interpretation of this statement is that any algebraic variety in $\mathbb{C}^n$ is the intersection of at $n+1$ hypersurfaces.

6. Elimination of Noetherian hypotheses

It is remarkable that the Noetherian hypothesis could be avoided in the case of Serre's Theorem or of the generalisation of Kronecker's Theorem, Theorem 5.1. The elimination of Noetherian hypotheses is also a theme in algebraic geometry (Dieudonne 1964). However, the method usually used there is to reduce the statement to the Noetherian case. This misses the fact that, given the logical simplicity of the statement without the Noetherian hypotheses, one can expect a direct simple proof.

We give two examples of this fact. The first appears in Dieudonne (1964) and is elementary.

Theorem 6.1. If M is a finitely generated module over a commutative ring R and $u : M \to M$ a surjective linear map, then u is bijective.

The proof given in Dieudonne (1964) consists of first proving the statement in the case for which the ring is Noetherian, and then reducing the general case to this case. Essentially, the argument for this reduction is as follows: if M is generated by $m_1, \ldots, m_k$, the fact that u is surjective says that we can find r_{ij} in R such that $m_i = \Sigma r_{ij} u(m_j)$. We also have s_{ij} in R such that $u(m_i) = \Sigma s_{ij} m_j$. If we let R' be the subring of R generated by the elements r_{ij} and s_{ij}, then R' is Noetherian. If the proposition is proved in the Noetherian case, we get an inverse for s_{ij} with coefficients in $R' \subseteq R$. Hence, u is bijective.

https://doi.org/10.1017/S0960129506005627 Published online by Cambridge University Press

This argument is not satisfactory from a logical point of view since it proves a first-order statement using a logically complex notion, the notion of being Noetherian. One would expect a more direct argument. In this case, one can indeed give one elementary argument that also gives a way to compute the inverse of u as a polynomial in u. For this, let A be the subring of endomorphisms of M generated by u, that is, the ring of endomorphisms that are polynomials in u. Then M has the structure of an A-module. Also, if I is the ideal of A generated by u, we have $IM = M$, so there exists $v \in A$ such that $vM = 0$ and $1 - v \in I$ (this is Corollary 2.5 of Atiyah and Macdonald (1969), which has an elementary proof). But $vM = 0$ means $v = 0$, so $1 \in I$, that is, u is invertible.

The second example is more complex, and comes from Swan (1980). We say that R is *seminormal* if and only if, if $b^2 = c^3$, there exists $a \in R$ such that $b = a^3$ and $c = a^2$. This is a remarkably simple, and first-order, condition. Swan (1980) shows that for reduced rings this is a necessary and sufficient condition for the canonical map $\mathsf{Pic}\ R \to \mathsf{Pic}\ R[X]$ to be an isomorphism. The proof in Swan (1980) consists of reducing the problem to the case in which R is Noetherian.

In this case also the theorem can be formulated in a geometric way. Here, we just give the concrete formulation.

Theorem 6.2. If R is seminormal and M is an idempotent matrix of rank 1 over $R[X]$ such that there is a unimodular combination of the columns of $M(0)$ over R, then there is a unimodular combination of the columns of M over $R[X]$.

The hypotheses are coherent without branching for a fixed size of the matrix. One then expects *a priori* a direct elementary proof. This is indeed the case, and was carried out in Coquand (2006).

There are examples in algebra, such as Krull's Principal Ideal theorem or the Regular Element Property, which states that a regular ideal contains a regular element (Kaplansky 1974), where the Noetherian hypothesis is necessary.

7. Interpretation of minimal prime ideals

Besides Noetherian hypotheses, proofs in algebra use abstract objects such as prime ideals, and even minimal prime ideals, that is, prime ideals that are minimal for inclusion. This is used, for instance, in the classical proof of Theorem 6.2, and in Peskine's proof of Zariski's Main Theorem (Peskine 1966). Classically, the existence of such prime ideals rely on Zorn's lemma. Contrary to the use of Noetherian hypotheses, it can be shown generally that the use of minimal prime can always be eliminated. To simplify, we just consider the case in which the commutative ring R is *reduced*, that is, we add the first-order axiom

$$x^2 = 0 \to x = 0\,,$$

and show in this case how to interpret the existence of a minimal prime ideal of R.

We first recall the elementary description of the Zariski spectrum of R, following Joyal (Coquand and Lombardi 2002; Joyal 1975). We consider the following coherent

https://doi.org/10.1017/S0960129506005627 Published online by Cambridge University Press

proposition theory, with axioms

$$\neg D(0) = 0, \qquad D(1), \qquad D(fg) \leftrightarrow D(f) \wedge D(g), \qquad D(f+g) \rightarrow D(f) \vee D(g)\,.$$

It can be shown directly that

$$D(g_1) \wedge \ldots \wedge D(g_n) \rightarrow D(f_1) \vee \ldots \vee D(f_m)$$

holds if, and only if, the monoid generated by $g_1, \ldots, g_n$ meets the ideal generated by $f_1, \ldots, f_m$ (Coquand and Lombardi 2002). Since R is reduced, $\neg D(f)$ is derivable in this theory if and only if $f = 0$ in R. This is a constructive interpretation of the fact that the intersection of all prime ideals of R is $\{0\}$.

A 'model' of the propositional theory $D(f)$ corresponds classically to a complement of a prime ideal. In order to get a complement of a minimal prime ideal, it is enough to add the axiom

$$D(f) \vee \bigvee_{gf=0} D(g)\,. \qquad (*)$$

Indeed, the axiom says that $\{f \in R \mid D(f)\}$ is a maximal filter, and thus that its complement is a minimal prime ideal. The axiom $(*)$ is a geometric infinitary axiom. Taken with the previous coherent axioms, this defines a geometric theory M whose models are classically the complement of minimal prime ideals. We are going to show the formal consistency of this theory M by building constructively a topological model. For this we introduce the orthogonality relation: $f \perp g$ if and only if $fg = 0$. If $X \subseteq R$, we define the orthogonal of X to be

$$X^{\perp} = \{y \in R \mid \forall x \in X. y \perp x\}\,.$$

It is standard (Birkhoff 1967) that the lattice of sets equal to their biorthogonal is a complete lattice L. In L we have $\vee X_i = (\cup X_i)^{\perp\perp}$ and $\wedge X_i = \cap X_i$.

Theorem 7.1. The lattice L is a complete Heyting algebra. Furthermore, if we take $D(f) = f^{\perp\perp} \in L$, we get a model of the theory M of the complement of minimal prime ideals.

Proof. Note first that if $X \in L$ and $a \in X$, then $au \in X$ for all $u \in R$. Indeed, if $b \in X^{\perp}$, we have $ab = 0$, so $aub = 0$. This implies $au \in X^{\perp\perp} = X$. From this fact, it follows by elementary reasoning that we have $X \wedge (\vee Y_i) = \vee(X \wedge Y_i)$ in L, that is, L is a complete Heyting algebra. The axiom $(*)$ is satisfied since, if $a \in f^{\perp}$ and $a \in g^{\perp}$ for all $g \perp f$, we have $a \perp f$ and thus $a \perp a$. This implies $a^2 = 0$, so $a = 0$ since R is reduced. □

Corollary 7.2. $D(f) = 0$ is derivable in the theory M iff $f = 0$. More generally, we can derive $D(f_1) \wedge \ldots \wedge D(f_n) \rightarrow D(g_1) \vee \ldots \vee D(g_m)$ in the theory M iff $hg_1 = \ldots = hg_m = 0$ implies $hf_1 \ldots f_n = 0$.

Proof. If $D(f_1) \wedge \ldots \wedge D(f_n) \rightarrow D(g_1) \vee \ldots \vee D(g_m)$ is derivable, we have, by the previous theorem,

$$f_1^{\perp\perp} \cap \ldots \cap f_m^{\perp\perp} \subseteq (g_1^{\perp} \cap \ldots \cap g_m^{\perp})^{\perp}\,,$$

which is equivalent to $g_1^{\perp} \cap \ldots \cap g_m^{\perp} \subseteq (f_1 \ldots f_n)^{\perp}$. Conversely, if $hg_1 = \ldots = hg_m = 0$ implies $hf_1 \ldots f_n = 0$ and $D(f_1 \ldots f_n)$ holds, it follows from $(*)$ that we have $D(g_1) \vee \ldots \vee D(g_m)$.

https://doi.org/10.1017/S0960129506005627 Published online by Cambridge University Press

In particular, $D(f) = 0$ is derivable so we get $f^{\perp} = R$, and thus $f = 0$. □

One interpretation of this corollary is that the intersection of all minimal prime ideals of R is $\{0\}$. This gives an effective interpretation of the existence of minimal prime ideals.

Note that a consequence of the theory M is

$$D(f) \vee \neg D(f), \tag{$**$}$$

and this gives a direct explanation of why the Krull dimension decreases by at least one when we quotient R by the boundary ideal N_f of f: the prime ideals of R/N_f corresponds exactly to the prime ideals containing N_f, and $(**)$ implies that no minimal prime ideals of R contain N_f.

References

Atiyah, M. F. and Macdonald, I. G. (1969) *Introduction to Commutative Algebra,* Addison Wesley.

Bezem, M. and Coquand, Th. (2003) Newman's lemma – a case study in proof automation and geometric logic. *Bulletin of the EATCS* **79** 86–100.

Birkhoff, G. (1967) *Lattice theory,* Third edition, American Mathematical Society Colloquium Publications, Vol. XXV.

Blass, A. (1988) Topoi and computation. *Bulletin of the EATCS* **36** 57–65.

Coquand, Th. (2004) Sur un théorème de Kronecker concernant les variétés algébriques. *C. R. Acad. Sci. Paris* (Ser. I) **338** 291–294.

Coquand, Th. (2005) A Completeness Proof for Geometrical Logic (to appear).

Coquand, Th. (2006) On Seminormality. *Journal of Algebra* (to appear).

Coquand, Th. and Lombardi, H. (2002) Hidden constructions in abstract algebra (3). Krull Dimension of distributive lattices and commutative rings. In: Fontana M., Kabbaj, S.-E. and Wiegand S. (eds.) Commutative ring theory and applications. *Lecture notes in pure and applied mathematics* **231**, M. Dekker 477–499.

Coquand, Th., Lombardi, H. and Quitté, C. (2004) Generating non-Noetherian modules constructively. *Manuscripta mathematica* **115** 513–520.

Coquand, Th., Lombardi, H. and Roy, M. F. (2005) Une caractérisation élémentaire de la dimension de Krull. In: Crosilla, L. and Schuster, P. (eds.) *From Sets and Types to Topology and Analysis Towards practicable foundations for constructive mathematics.*

Coste, M., Lombardi, H. and Roy, M. F. (2001) Dynamical methods in algebra: effective Nullstellensätze. *Annals of Pure and Applied Logic* **111** (3) 203–256.

Ducos, L., Lombardi, H., Quitté, C. and Salou, M. (2004) Théorie algorithmique des anneaux arithmétiques, de Prüfer et de Dedekind. *Journal of Algebra* **281** 604–650.

Dieudonné, J. (1964) *Fondements de la géométrie algébrique moderne*, Les Presses de l'Université de Montréal.

Edwards, H. (2005) *Essays in constructive mathematics,* Springer-Verlag.

Forster, O. (1964) Über die Anzahl der Erzeugenden eines Ideals in einem Noetherschen Ring. *Math. Z.* **84** 80–87.

Gentzen, G. (1969) *Collected Works* (edited by Szabo), North-Holland.

Heitmann, R. (1984) Generating non-Noetherian modules efficiently. *Michigan Math. J.* **31** (2) 167–180.

Kaplansky, I. (1974) *Commutative Rings,* University of Chicago Press.

https://doi.org/10.1017/S0960129506005627 Published online by Cambridge University Press

Jacobsson, C. and Löfwall, C. (1991) Standard bases for general coefficient rings and a new constructive proof of Hilbert's basis theorem. *J. Symbolic Comput.* **12** (3) 337–371.

Joyal, A. (1975) Le théorème de Chevalley-Tarski. *Cahiers de Topologie et Géométrie Différentielle* **16** 256–258.

Kronecker, L. (1882) Grundzüge einer arithmetischen Theorie der algebraischen Grössen. *J. reine angew. Math.* **92** 1–123 (reprinted in *Leopold Kronecker's Werke*, II, 237–387.)

Lombardi, H. and Quitté, C. (to appear) Modules projectifs de type fini.

Matsumura, H. (1986) *Commutative ring theory,* (translated from the Japanese by M. Reid), Cambridge Studies in Advanced Mathematics **8**, Cambridge University Press.

Mines, R., Richman, F. and Ruitenburg, W. (1988) *A course in constructive algebra,* Springer-Verlag.

Mulvey, C. (1974) Intuitionistic algebra and representations of rings. In: Recent advances in the representation theory of rings and C^*-algebras by continuous sections. *Mem. Amer. Math. Soc.* **148** 3–57.

Peskine, C. (1966) Une généralisation du "main theorem" de Zariski. *Bull. Sci. Math.* **90** (2) 119–127.

Prawitz, D. (1971) Ideas and results in proof theory. *Proceedings of the Second Scandinavian Logic Symposium,* Studies in Logic and the Foundations of Mathematics **63**, North-Holland 235–307.

Reyes, G. (1978) Théorie des modèles et faisceaux. *Adv. in Math.* **30** (2) 156–170.

Serre, J. P. (1957/1958) Modules projectifs et espaces fibrés à fibre vectorielle, Séminaire P. Dubreil.

Swan, R. G. (1962) Vector bundles and projective modules. *Trans. Amer. Math. Soc.* **105** 264–277.

Swan, R. G. (1967) The Number of Generators of a Module. *Math. Z.* **102** 318–322.

Swan, R. G. (1980) On Seminormality. *Journal of Algebra* **67** 210–229.

Wraith, G. (1980) Intuitionistic algebra: some recent developments in topos theory. *Proceedings of the International Congress of Mathematicians (Helsinki, 1978)*, Acad. Sci. Fennica, Helsinki 331–337.

https://doi.org/10.1017/S0960129506005627 Published online by Cambridge University Press