A LOGICAL APPROACH TO ABSTRACT ALGEBRA

THIERRY COQUAND AND HENRI LOMBARDI

Abstract. Recent work in constructive mathematics show 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 a large number of abstract non effective proofs of "concrete" statements into elementary proofs. Surprisingly the arguments we get 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.

Introduction

The purpose of this paper is to survey some of our recent works in constructive algebra [5, 6, 7, 8, 9, 11] 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, being geometric and being 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 which has a particularly simple tree form [2, 8, 11].

We present first some basic examples in algebra which are directly formulated with the required logical complexity: the first one is an implication between equational statements, and the second one is coherent, that is geometric and first-order. We present then a more elaborate example, that was a mathematical conjecture and where a first-order formulation is not obvious. We can transform further it to a coherent formulation. Knowing a priori that we had to look for an "analytical" proof involving only simple algebraic manipulations helps then in finding a proof. We show then 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 the work [12] 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.

1. Logical complexity

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

$$x + (-x) = 0, \;\; x + (y + z) = (x + y) + z, \;\; x + y = y + x, \;\; x + 0 = x$$

$$x1 = x,\ xy = yx,\ x(yz) = (xy)z,\ x(y + z) = xy + xz$$

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

$$xy = 0 \rightarrow (x = 0 \vee y = 0)$$

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. If it is furthermore 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.

If we take however a basic book in abstract algebra such as Atiyah-Macdonald or Matsumura [1, 23] we discover that even basic theorems are not formulated in a first-order way because of the introduction of abstract notions. Such abstract notions are

(1) arbitrary ideals of the rings, that are defined as subsets, and thus not expressed in a first-order way,
(2) *prime* or *maximal* ideals, whose existence relies usually on Zorn's lemma,
(3) Noetherian hypotheses.

These notions have different levels of non effectivity. To be Noetherian can be captured by a generalised inductive definition [19], 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 [35] points out the relevance of the notion of *geometric formula* 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 formula (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 the true formula $\top$. We allow also existential quantification and infinite disjunction indexed over natural numbers[1]. A *geometric* formula is an implication between two positive formulae. A *coherent* formula is a formula which is both geometric and first-order. Notice that, as special cases, any positive formula is geometric, and the negation of a positive formula is geometric. As a special case of coherent formula, we have the notion of *Horn* formula, which is an implication $C \rightarrow A$ where C is a conjunction of atomic formulae, and A an atomic formula. Horn theories correspond to the notion of *atomic systems* in [26]. For instance, equational theories are Horn theories.

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

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

[1] Usually, the notion of "arbitrary" infinite disjunction is allowed, but we shall only need this generality here in the last section.

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, "to be 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 [35] of notion expressed geometrically is the notion of *flat* module M over a ring R. It says that if we have a relation $PX = 0$ where P is a row vector with coefficient in R and X 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 don't say anything about the size of Q this statement involves implicitely an infinite disjunction over natural numbers. Thus the notion of flat module is not first-order but geometric.

As stressed by G. Wraith the importance of geometric formula comes from *Barr's theorem.*

Theorem 1.1. *If a geometric sentence is deducible from a geometric theory in classical logic, with the axiom of choice, then it is also deducible from it intuitionistically.*

Furthermore in this case there is always a proof with a simple branching tree form, of a *dynamical* proof [8, 2, 11]. In general, this tree may be infinitely branching, but, if the theory is *coherent*, that is geometric *and* first-order, then the proof is a finitely branching tree [8, 2, 11].

In order to describe these proofs, it is convenient first to notice 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

$$C ::= \ \top \mid C \wedge A \qquad\qquad D ::= \bot \ \mid D \vee E \qquad\qquad E ::= \ (\exists \overrightarrow{v})C$$

We may write D for $\top \rightarrow D$, A for $\top \wedge A$ and so on, economizing on empty conjunctions, disjunctions, existential quantifications and brackets as much as possible. Let us call a closed atomic formula to be a *fact*. In most algebraic theories, the only facts are equalities. We can thus consider that a coherent theory is a collection of formulae of the form $C \rightarrow D$. 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 we start with. Each node consists of a set of facts, representing a state of information. The sets increase monotonically along the way from the root to the leaves. The successors of a node are determined by the dynamical rules that 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 under investigation F. If all leaves contain a contradiction then the given set of atomic formulas is contradictory.

In the special case where all formulae are of the form $C \to A$, the tree has no branching. We get something equivalent to the notion of *atomic systems* introduced by Prawitz [26]. In particular, equational theories are of this form. The crucial point is that this notion of dynamical proof is *complete* for deducibility in a coherent theory [8, 2, 11], and that a dynamical proof uses only intuitionistically valid inference steps. Barr's theorem that we have cited above is a simple consequence: if a coherent sentence is deducible from a coherent theory in classical logic, even with the axiom of choice, it is a *semantical* consequence of the theory, and so, by completness, it can be derived by a dynamic proof, which is intuitionistically valid.

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

We can now explain in what sense these completeness theorems are related to 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 to derive new concrete statements from given 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 [26] has a similar analysis in 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 from 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 which, given an element a, tests if $a = 0$ or not, and in the later case, gives an element b such that $ab = 1$.

Both the completeness theorem and Barr's theorem are purely heuristic results from a constructive point of view however. Indeed, they are both proved using non constructive means, and do not give algorithms to transform a non effective proof to an effective one. In practice however, in all examples analysed so far, it has been possible to extract effective arguments from the ideas present in the non effective proofs. We think that our work, complementary to the work done in constructive mathematics [28, 14] or in Computable Algebra [31], provides a partial realisation of Hilbert's program in abstract commutative algebra.

2. 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 completness theorem for equational logic is enough. Both examples appear at the beginning of [23].

2.1. **Dimension over rings.** The following result is usually proved using maximal ideals [23].

Theorem 2.1. *If $n < m$ and $f : R^n \to R^m$ is surjective linear map then R is a trivial ring, that is $1 = 0$ in R.*

What is the logical complexity of this statement? If we fix n and m, let 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 existence of maximal ideals: if R is not 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 remark 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.

2.2. **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, that R is local is expressed by the positive formula

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

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

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

Since $Inv(xy) \leftrightarrow (Inv(x) \wedge Inv(y))$, we have, for all x

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

Classically it is possible to derive from this

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

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. Classically the Jacobson radical can be also defined as the intersection of all maximal ideals of R and 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, classically we have shown that in a local ring, an element is invertible or in the Jacobson radical.

We analyse the following theorem.

Theorem 2.2. *If M is a finitely generated projective module over a local ring R then M is free.*

The concrete formulation of this theorem [22] is the following.

Theorem 2.3. *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, that we can transform by proof-theoretic methods to a constructive first-order derivation.

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

It is interesting that the next constructive argument we give, and 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 $f'_1, \ldots, f'_n$ so that $f'_i = f_i$ or $e_i - f_i$ and that for each m the top $m \times m$ minor of the matrix $f'_1, \ldots, f'_m$ is invertible. This is possible since the sum of the minor for $f'_1, \ldots, f'_{m-1}, f_m$ and the minor for $f'_1, \ldots, f'_{m-1}, e_m - f_m$ is the minor for $f'_1, \ldots, f'_{m-1}, e_m$ which is invertible by induction.

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. □

Notice that this last proof can be read as an algorithm: given the matrix F and the "subprogram" which 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 2.3 has an interesting history in intuitionistic algebra. It was noticed in [24] 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 $C(M)$ [34]. The result in [24] is formulated in higher-order intuitionistic logic. In [27] it is noticed that one can formulate the theorem in first-order logic. The formulation there, attributed to A. Kock, is a priori weaker than the formulation of Theorem 2.3[2].

Theorem 2.4. *If F is a $n \times n$ projection matrix over a local ring R then we can find a $n \times r$ matrix X and a $r \times n$ matrix Y such that $XY = F$ and $YX = I_r$.*

This is essentially what is proved in [24]. Notice however that the proof there uses, a priori, that an element is invertible or not, and is not, as it stands, intuitionistically valid. We present here 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 $X = U_1, \dots, U_m$ that generate $Im(F)$ (we start with $m = n$ and $X = F$.) We can then find a $m \times n$ matrix Y such that $XY = F$ (at the beginning, we can take $X = F$ and $Y = I_n$ or $Y = F$.) Then $YX = G$ is a $m \times m$ projection matrix since $G^2 = YXYX = YX = G$. We also have $XG = XYX = FX = X$. If we write $G = (c_{ij})$, we have thus $U_j = \Sigma c_{ij} U_i$ for each j. Since R is local, c_{jj} invertible or $1 - c_{jj}$ 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 term 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. □

3. Serre's splitting-off theorem

3.1. **Classical formulation.** The example we are going to present has its origin in a paper of Serre [30] from 1958. It is a purely algebraic theorem, but it has a geometrical intuition. The geometrical statement is roughly that if we have a vector fibre bundle over a space of finite dimension, and each fiber has a large enough dimension, then we can find a non vanishing section. We give first the classical formulation, where both hypotheses and conclusions have a non elementary form, and then a version where the conclusion is first-order.

We assume R to be a Noetherian ring, and we let $\mathsf{Max}(R)$ to 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 chains 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 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,

[2] In our formulation, we express that both the image and the kernel of F are free. In the formulation of [27], we express 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.

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 fiber 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 3.1. *(Serre, 1958) If M is a finitely generated projective module over R such that $n \leq r_x$ for all maximal ideals x of R then 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 [9, 22]. 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 3.2. *(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 which is unimodular.*

Interestingly, in this form, the theorem can then be seen as a special case of Swan's theorem [32], a theorem conjectured by Serre. We give first the abstract form of the theorem.

Theorem 3.3. *(Swan 1967) If M is a finitely generated module over R such that for each $x \in \mathsf{Max}(R)$ the fiber $M(x)$ can be generated by p elements then M can be generated by $p + n - 1$ elements.*

Theorem 3.4. *(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 which is unimodular.*

Only the concrete formulation of these two results reveals their similarities. The generalisation of these theorems to the non Noetherian case has been first established in [9], by analysing the paper [17] using the techniques that are presented in this note.

Notice that the conclusion of this theorem is expressed in first-order logic, and even in a positive way. The hypothesis however is non elementary: we suppose both that R is Noetherian and we have an 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 that we want to analyse. It is left to express the hypothesis of the theorem $\mathsf{dim}\ (\mathsf{Max}(R)) < n$ in a first-order way.

3.2. **Geometric formulation of Krull dimension.** The first step is to give an elementary formulation of the notion of Krull dimension. It is not so easy a priori since the usual definition is in term of 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 [6]. We introduce first the notion of *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 define then 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 $\mathsf{Kdim}\ R < n$ which 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 (that motivated Swan's theorem in the Noetherian case).

Theorem 3.5. *(Heitmann, 1984, concrete version) If* $\mathsf{Kdim}\ R < n$ *and if* F *is a rectangular matrix over* R *such that* $\Delta_n(F) = 1$ *then there exists a linear combination of the columns of* F *which 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, of a very simple nature furthermore. In this case, it is enough to extract this direct proof from the argument in [17]. This is carried out in [9].

3.3. **A new notion of dimension.** We present now a notion of dimension, introduced in [9] and which appears implicitly in [17]. This notion is finer than the notion of Krull dimension: we always have $\mathsf{Hdim}\ R \leq \mathsf{Kdim}\ R$. Interestingly $\mathsf{Hdim}\ R \leq n$ can be expressed by a first-order formula, but the logical complexity of this formula increases with n, contrary to $\mathsf{Kdim}\ R \leq n$ which stays a positive formula for all n.

We get this definition by changing the nilradical in the definition of Krull dimension by the Jacobson radical J which is classically the intersection of all maximal ideals, but, as we have seen, 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 introduce then a new notion of *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 define then inductively $\mathsf{Hdim}\ R < n$: for $n = 0$ it means that $1 = 0 \in R$ and for $n > 0$ it means that we have $\mathsf{Hdim}\ (R/N_a) < n - 1$ 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 raw 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, by using general proof-theoretic arguments that, if we have a first-order classical proof, then we also have an intuitionistic proof. From proof theory, one can use Gentzen sharpened Hauptsatz [16], 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 which illustrates well the strength of this notion. We illustrate the idea only for $n = 1$. 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 has then a particular simple form, as the fact that in this equational theory, extended with the equation $\Delta_1(F) = 1$ we can build a raw vector X and a column vector Y such that $1 = XFY$.

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

Theorem 3.6. *([9], 2004) If* $\mathsf{Hdim}\ R < n$ *and if* F *is a rectangular matrix over* R *such that* $\Delta_n(F) = 1$ *then there exists a linear combination of the columns of* F *which is unimodular.*

The formulation of this theorem is now purely coherent, in a coherent theory which has a specially simple form (no branching). If it holds, it has a purely elementary proof, and knowing this helps in finding a proof [9]. We can furthermore read the proof presented in [9] as an algorithm which produces an unimodular column.

4. Kronecker's theorem

In this section, we show that, though these results may seem quite abstract, being expressed in first-order logic and 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 [17, 7]. In this case, it is possible to get from an abstract proof a concrete algorithm that could have been formulated by Kronecker [14]. We first give the abstract version, which is proved in [7].

Theorem 4.1. *If* $\mathsf{Kdim}\ R \leq n$ *and we have* $n+2$ *elements* $g_0, g_1, \ldots, g_{n+1}$ *then it is possible to find* $n + 1$ *elements* $f_0, f_1, \ldots, f_n$ *so 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 [7]. To simplify the discussion, let us take $n = 2$. As we have explained the meaning of $\mathsf{Kdim}\ R \leq 2$ 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 4.1 can thus be interpreted as follows: given such an algorithm which 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 $g_0, \ldots, g_3$.

This algorithm is furthermore simple and explicit, corresponding to the simplicity of the the proof in [7], given the algorithm corresponding to $\mathsf{Kdim}\ R \leq 2$. 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

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

where

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

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

$$1 \in \,<g_1, h_1>,\ \ g_1h_1 \in \sqrt{<g_2, h_2>},\ \ g_2h_2 \in \sqrt{<g_3, h_3>},\ \ g_3h_3 \in \sqrt{0}$$

In [6], we present a direct proof that $\mathsf{Kdim}\ \mathbb{Q}[X_1, \ldots, X_n] \leq n$. For $n = 2$ this reduces to the remark that if we take 3 elements g_1, g_2, g_3 in $\mathbb{Q}[X_1, X_2]$ then they are algebraically dependent (See [28, 14].) Such an algebraic dependence relation can always be written

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

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

We can then combine the two algorithms and we get in this way a non trivial algorithm on polynomials, which given g_0, g_1, g_2, g_3 produces f_0, f_1, f_2 so 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 4.2. *Let polynomials g_1, g_2, …, 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, …, f_{n+1} in the same indeterminates that are zero mod g_1, g_2, …, g_m and have the property that, for each $i = 1$, 2, …, m, some power of g_i is zero mod f_1, f_2, …, 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.

5. 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 4.1. The elimination of Noetherian hypotheses is also a theme in algebraic geometry [13]. However the method which is 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 one is elementary and appears in [13].

Theorem 5.1. *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 [13] consists in proving first the statement in the case where 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 $m_1, \dots, m_k$ 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 have also 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.

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 give indeed one elementary argument that gives a way also to compute the inverse of u as a polynomial in u. 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 a structure of A-module. Also, if I is the ideal of A generated by u we have $IM = M$ and so there exists $v \in 1 + I$ such that $vM = 0$ (this is Corollary 2.5 of [1] which has an elementary proof). But $vM = 0$ means $v = 0$ and so $1 \in I$, which implies that u is invertible.

The second example is more complex, and comes from the work [33]. We say that R is *seminormal* iff if $b^2 = c^3$ then there exists $a \in R$ such that $b = a^3$ and $c = a^2$. This is a remarkably simple, and first-order, condition. The work of [33] shows that 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 [33] consists in reducing the problem to the case where R is Noetherian.

In this case also, the theorem can be formulated in a geometric way. We give here only the concrete formulation.

Theorem 5.2. *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 expects then a priori a direct elementary proof. This is indeed the case, and this has been carried out in [10].

There are examples in algebra, like Krull's Principal Ideal theorem, or the Regular Element Property, which states that a regular ideal contains a regular element (see [18]), where the Noetherian hypothesis is necessary.

6. Interpretation of minimal prime ideals

Besides Noetherian hypotheses, proofs in algebra use abstract objects such as prime ideals, and even minimal prime ideals, i.e. prime ideals that are minimal for inclusion. This is used for instance in the classical proof of Theorem 5.2, and in Peskine's proof of the Main Theorem of Zariski [25]. 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 consider only the case where the commutative ring R is *reduced*, that is we assume

$$x^2 = 0 \rightarrow x = 0$$

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

We recall first the elementary description of the Zariski spectrum of R, following Joyal [5, 20]. We consider the following coherent 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

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

holds if, and only if, the monoid generated by $g_1, \dots, g_n$ meets the ideal generated by $f_1, \dots, f_m$ [5]. 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

$$D(f) \vee \bigvee_{gf=0} D(g) \tag{$*$}$$

Indeed the axiom expresses that $\{f \in R \mid D(f)\}$ is a maximal filter, and so that its complement is a minimal prime ideal. The axiom $(*)$ is a geometric infinitary axiom. Together 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 orthogality 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 [3, 29] 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 6.1. *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 complement of minimal prime ideals.*

Proof. Notice first that if $X \in L$ and $a \in X$ then $au \in X$ for all $u \in R$. Indeed if $b \in X^{\perp}$ then $ab = 0$ and 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 $g \perp f$ then we have $a \perp f$ and so $a \perp a$. This implies $a^2 = 0$ and so $a = 0$ since R is reduced. □

Corollary 6.2. *$D(f) = 0$ is derivable in the theory M iff $f = 0$. More generally, we can derive $D(f_1) \wedge \cdots \wedge D(f_n) \rightarrow D(g_1) \vee \cdots \vee D(g_m)$ in the theory M iff $hg_1 = \cdots = hg_m = 0$ implies $hf_1 \ldots f_n = 0$.*

Proof. If $D(f_1) \wedge \cdots \wedge D(f_n) \rightarrow D(g_1) \vee \cdots \vee D(g_m)$ is derivable then we have by the previous Theorem

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

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

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

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.

Notice that a consequence of the theory M is

$$D(f) \vee \neg D(f) \qquad (**)$$

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

Acknowledgement

We thank the referee for his comments.

References

[1] M.F. Atiyah and I.G. Macdonald. *Introduction to Commutative Algebra.* Addison Wesley, 1969. 2, 12

[2] M. Bezem and Th. Coquand. Newman's lemma—a case study in proof automation and geometric logic. Bull. Eur. Assoc. Theor. Comput. Sci. EATCS No. 79 (2003), 86–100. 1, 3, 4

[3] G. Birkhoff. *Lattice theory.* Third edition. American Mathematical Society Colloquium Publications, Vol. XXV American Mathematical Society, Providence, R.I. 1967. 13

[4] A. Blass. Topoi and computation, *Bulletin of the EATCS* **36**, October 1988, pp. 57–65.

[5] Th. Coquand and H. Lombardi. Hidden constructions in abstract algebra (3) Krull dimension. in: Commutative ring theory and applicatoins. Eds: Fontana M., Kabbaj S.-E., Wiegand S. M.Dekker LNPAM 131. (2002) 477-499. 1, 13

[6] Th. Coquand, H. Lombardi and M.F. Roy. Une caractérisation élémentaire de la dimension de Krull. In *From Sets and Types to Topology and Analysis Towards practicable foundations for constructive mathematics*, Edited by Laura Crosilla and Peter Schuster, 2005. 1, 8, 11

[7] Th. Coquand. Sur un théorème de Kronecker concernant les variétés algébriques C. R. Acad. Sci. Paris, Ser. I 338 (2004), Pages 291-294 1, 10, 11

[8] Th. Coquand. A Completness Proof for Geometrical Logic in *Logic, Methodology and Philosophy of Sciences*, Hajek, Valdes-Villuaneva, Westertahl, editors, 79-90, 2005. 1, 3, 4

[9] Th. Coquand, H. Lombardi, C. Quitte. Generating non-Noetherian modules constructively. Manuscripta mathematica, 115, 513-520 (2004) 1, 8, 9, 10

[10] Th. Coquand. On Seminormality. to appear in the Journal of Algebra, 2006. 12

[11] M. Coste, H. Lombardi, and M.F. Roy. Dynamical methods in algebra: effective Nullstellensätze, *Annals of Pure and Applied Logic* **111**(3):203–256, 2001. 1, 3, 4

[12] L. Ducos, H. Lombardi, C. Quitté and M. Salou. Théorie algorithmique des anneaux arithmétiques, de Prüfer et de Dedekind. Journal of Algebra 281, (2004), 604-650. 1

[13] J. Dieudonné. *Fondements de la géométrie algébrique moderne.* Les Presses de l'Université de Montréal, 1964. 12

[14] H. Edwards. *Essays in constructive mathematics.* Springer-Verlag, New York, 2005. 4, 10, 11

[15] O. Forster. Über die Anzahl der Erzeugenden eines Ideals in einem Noetherschen Ring. Math. Z. 84 1964 80–87.

[16] G. Gentzen. *Collected Works.* Edited by Szabo, North-Holland, 1969. 10

[17] R. Heitmann. Generating non-Noetherian modules efficiently. Michigan Math. J. 31 (1984), no. 2, 167–180. 8, 9, 10

[18] I. Kaplansky. *Commutative Rings.* The University of Chicago Press, Chicago, 1974. 13

[19] C. Jacobsson and C. Löfwall. Standard bases for general coefficient rings and a new constructive proof of Hilbert's basis theorem. *J. Symbolic Comput.* 12 (1991), no. 3, 337–371. 2

[20] A. Joyal. Le théorème de Chevalley-Tarski. *Cahiers de Topologie et Géométrie Différentielle* 16, 256–258 (1975). 13

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

[22] H. Lombardi and C. Quitté. *Modules projectifs de type fini.* To appear. 6, 8

[23] H. Matsumura. *Commutative ring theory.* Translated from the Japanese by M. Reid. Cambridge Studies in Advanced Mathematics, 8. Cambridge University Press, Cambridge, 1986. 2, 5

[24] C. Mulvey. Intuitionistic algebra and representations of rings. In *Recent advances in the representation theory of rings and C^*-algebras by continuous sections*, pp. 3–57. Mem. Amer. Math. Soc., No. 148, Amer. Math. Soc., Providence, R. I., 1974. 6, 7

[25] C. Peskine. Une gnralisation du "main theorem" de Zariski. *Bull. Sci. Math.* (2) 90 1966 119–127. 13

[26] D. Prawitz. Ideas and results in proof theory. Proceedings of the Second Scandinavian Logic Symposium, pp. 235–307. Studies in Logic and the Foundations of Mathematics, Vol. 63, North-Holland, Amsterdam, 1971 2, 4

[27] G. Reyes. Théorie des modèles et faisceaux. *Adv. in Math.* 30 (1978), no. 2, 156–170. 6, 7

[28] R. Mines, F. Richman and W. Ruitenburg. *A course in constructive algebra.* Springer-Verlag, 1988 4, 11

[29] G. Sambin. A new and elementary method to represent every complete Boolean algebra. in *Logic and Algebra*, eds. A. Ursini and P. Agliano, New-York, Dekker, 1996, 655-665. 13

[30] J.P. Serre. Modules projectifs et espaces fibrés à fibre vectorielle. Séminaire P. Dubreil, Année 1957/1958. 7

[31] V Stoltenberg-Hansen and J V Tucker. Computable rings and fields. in E Griffor (ed.), *Handbook of Computability Theory*, Elsevier, 1999, pp.363-447. 4

[32] R.G. Swan. The Number of Generators of a Module. *Math. Z.* 102 (1967), 318-322. 8

[33] R.G. Swan. On Seminormality. *Journal of Algebra*, 67, 210-229 (1980) 12

[34] R.G. Swan. Vector bundles and projective modules. *Trans. Amer. Math. Soc.* 105 1962 264–277. 6

[35] G. Wraith. Intuitionistic algebra: some recent developments in topos theory. Proceedings of the International Congress of Mathematicians (Helsinki, 1978), pp. 331–337, Acad. Sci. Fennica, Helsinki, 1980. 2, 3

COMPUTER SCIENCE, CHALMERS UNIVERSITY, SE-412 96 GÖTEBORG, SWEDEN, WWW.CS.CHALMERS.SE/ COQUAND.
DEPARTMENT OF MATHEMATICS, UNIVERSITY OF FRANCHE-COMTÉ, 25030 BESANÇON, FRANCE, HTTP://HLOMBARDI.FREE.FR.