Algebraically Closed Fields in Isabelle/HOL

Paulo Emílio de Vilhena[1] and Lawrence C. Paulson[2][0000−0003−0288−4279]

[1] Inria, France, `paulo-emilio.de-vilhena@inria.fr` **
[2] Computer Laboratory, University of Cambridge, UK, `lp15@cam.ac.uk`

Abstract. A fundamental theorem states that every field admits an algebraically closed extension. Despite its central importance, this theorem has never before been formalised in a proof assistant. We fill this gap by documenting its formalisation in Isabelle/HOL, describing the difficulties that impeded this development and their solutions.

1 Introduction

The *fundamental theorem of algebra* states that the field of complex numbers is algebraically closed: every nonconstant polynomial with complex coefficients has at least one complex root. By extending the field of real numbers with a single root of the polynomial $X^2 + 1$, we obtain a field (the complex numbers) where not only has $X^2 + 1$ a root but also every other polynomial.

At the beginning of the 20th century, this theorem about the reals raised the question of which other fields could similarly be extended to be algebraically closed. Curiously, the same mathematician to introduce the concept of a field, Ernst Steinitz, was the one to answer the question: *every* field admits an algebraically closed extension.

This result has important consequences. It guarantees that every polynomial has a splitting field; it links algebra and geometry. Kevin Buzzard comments:

> The theorems of local and global class field theory are one of the highlights of early 20th century mathematics. ...The Langlands Philosophy, one of the central questions in modern number theory, is a vast conjectural generalisation of these theorems, and one cannot even state the fundamental conjectures in this theory without mentioning algebraic closures. Wiles and Taylor proved an extremely small fragment of these conjectures in 1994 and deduced Fermat's Last Theorem. [3]

Despite its importance, the existence of algebraic closures has never been formalised in a proof assistant. The gap suggests that the formal proof is challenging and that there might be ill-understood technical difficulties. Here we propose to settle this:

– We formally prove that every field has an algebraic closure: an equivalent way to state that every field admits an algebraically closed extension (Section 5).

** The author was affiliated to École Polytechnique during the realisation of this work.

- We discuss, from a mathematical perspective, how our proof relates to existing ones and how we planned the formalisation effort (Section 4).
- We describe a limitation of Isabelle's type system along with a general solution, where *identity* is replaced by *isomorphism* (Section 3).

Attempting such proof in Isabelle/HOL is aligned with our intent to investigate the hurdles of formalising algebraic reasoning in a simply-typed setting.

2   Background

Here we recall some key elements of algebra. We suppose familiarity with the definitions of rings, fields, homomorphisms, ideals and the quotient of a ring by an ideal. We start with the definition of *canonical surjections*, which show that every element of a quotient ring is accessible in a structure-preserving way.

Definition 1 (Canonical surjection). *Let R be a ring and I an ideal of R. The* canonical surjection $\pi_{[R,I]} : R \to R/I$ *is a surjective homomorphism from the ring R to the quotient R/I that associates an element r of R to its equivalence class in R/I, that is, $r \mapsto \{r + i \mid i \in I\}$.*

The most common application of the canonical surjection in this paper is when considering the quotient between the ring of polynomials with coefficients in some field K, denoted by $K[X]$, and the *ideal generated by* some polynomial $P \in K[X]$, defined as

$$(P) \triangleq \{\, P\,Q \mid Q \in K[X] \,\}.$$

This gives rise to the canonical surjection $\pi_{[K[X],(P)]}$, which can sometimes be seen as a homomorphism from K to $K[X]/(P)$ and not from $K[X]$ to $K[X]/(P)$. Justification comes from the usual abuse of notation of identifying elements of K with constant polynomials in $K[X]$.

The next proposition elucidates why it is interesting to see $\pi_{[K[X],(P)]}$ as such: if P only has trivial factors, that is, if P is an *irreducible polynomial*, then the restriction of $\pi_{[K[X],(P)]}$ to K is in fact a homomorphism of *fields*.

Definition 2 (Irreducible polynomial). *Let K be a field and P a nonconstant polynomial with coefficients in K. The polynomial P is* irreducible *if for every $Q \in K[X]$, if Q divides P then $Q = k$ or $Q = k\,P$ for some $k \in K$.*

Proposition 1. *If K is a field and P is a polynomial with coefficients in K, then P is irreducible iff the quotient of $K[X]$ by the ideal (P) is a field.*

Given a homomorphism ϕ between two rings A and B, and a polynomial $Q \in A[X]$, we can build the polynomial $Q^\phi \in B[X]$ by applying ϕ to each of the coefficients of Q:

$$Q^\phi = \left(\sum_i a_i X^i\right)^\phi \triangleq \sum_i \phi(a_i)X^i \text{ where } a_i \in A$$

It follows from this definition that ϕ maps the evaluation of Q at an element $a \in A$ to the evaluation of Q^ϕ at $\phi(a)$:

$$\phi(Q(a)) = Q^\phi(\phi(a)). \tag{1}$$

Now, consider a field K and a polynomial $P \in K[X]$. From our previous discussion, the coefficients of P can be seen as constant polynomials in $K[X]$ so that P can be seen as lying in $(K[X])[X]$. Therefore, the evaluation of P at elements in $K[X]$ is meaningful. For instance, what happens if we evaluate P at the monomial $X \in K[X]$? We recover P itself: $P = P(X)$. Together with equation (1), we obtain the following result:

$$\pi(P) = \pi(P(X)) = P^\pi(\pi(X))$$

where π stands for the canonical surjection $\pi_{[K[X],(P)]}$. Since the equivalence class of P in $K[X]/(P)$ is the same as the zero polynomial, we have proved that $\pi(X)$ is a root of P^π. The next proposition rephrases this result using the terminology of a *field extension* when P is an irreducible polynomial.

Definition 3 (Field extension). *Let K and L be fields and $\phi : K \to L$ a homomorphism. Since K is a field, ϕ is either the trivial map $k \mapsto 0$ or an injective map. When it is injective, L is called a* field extension of K under ϕ.

Proposition 2. *Let K be a field and P an irreducible polynomial in $K[X]$. The quotient of $K[X]$ by the ideal (P) is a field extension of K under the homomorphism $\pi_{[K[X],(P)]}$. Moreover, the polynomial P^π admits $\pi(X)$ as a root in the field $K[X]/(P)$.*

Now let's see how to build a field isomorphic to the field of complex numbers. Consider the real polynomial $X^2 + 1$. It's irreducible, therefore by Proposition 2 the ring $\mathbb{R}[X]/(X^2 + 1)$ is actually a field extension of $\mathbb{R}$ under the homomorphism π. Furthermore, $\pi(X)$ is a root of the polynomial $(X^2+1)^\pi$. Thus, we have built a field where a square root of -1 exists. This fact can be formally stated by considering the map $\pi(a + b\,X) \mapsto a + bi$, which establishes an isomorphism from $\mathbb{R}[X]/(X^2 + 1)$ to the field $\mathbb{C}$ of complex numbers.

Related to the notion of a field extension is the notion of a subfield:

Definition 4 (Subfield). *Let L be a field. A subset $K \subseteq L$ is called a* subfield *of L if it verifies the axioms of a field when equipped with the same laws as L.*

This relation is actually bidirectional:

1. If L is a field extension of K under the homomorphism ϕ, then the image $\phi(K)$ is a subfield of L.
2. If K is a subfield of L, then L is a field extension of K under the identity homomorphism $k \mapsto k$.

It is important to notice the nuances of the definition of an irreducible polynomial in the context of subfields. If K is a subfield of L, then a polynomial P in $K[X]$ has its coefficients in both K and L. When we say that P is an irreducible polynomial we have to specify in which field: while P may have no nontrivial factor with coefficients in K, it may have one with coefficients in L. To make the distinction unambiguous, we say that P is irreducible *in* K expressing that it only has trivial factors in the ring $K[X]$.

The next natural notion is that of an algebraic element:

Definition 5 (Algebraic). *Let L be a field and K a subfield of L. Then an element $l \in L$ is* algebraic over K *if there exists a polynomial P with coefficients in K such that $P(l) = 0$, that is, l is a root of P.*

The subset of elements of L which are algebraic over K is a subfield of L. Moreover, since the polynomial $X - k$ belongs to the ring $K[X]$ if $k \in K$, every element of the field K is algebraic over K. In addition, the field K is a subfield of the subset of algebraics of L, which gives the following inclusions:

$$K \subseteq \{l \in L \mid l \text{ is algebraic over } K\} \subseteq L$$

Finally, we introduce the formal definition of an algebraically closed field and the closely related notion of the algebraic closure.

Definition 6 (Algebraically closed). *A field K is* algebraically closed *if every nonconstant polynomial with coefficients in K has at least one root in K.*

Definition 7 (Algebraic closure). *Let K and L be two fields. Then L is an* algebraic closure *of K if there exists a homomorphism $\phi : K \to L$ such that*

1. *Every polynomial P of degree n with coefficients in the subfield $\phi(K)$ has n roots in L: that is, P* splits *in L.*
2. *Every element of L is algebraic over the subfield $\phi(K)$.*

It is usual to refer to L as *the* algebraic closure: they are unique up to isomorphism. An additional remark is that if P splits in L, then P only has trivial irreducible factors. That is, for every irreducible polynomial $Q \in L[X]$, if Q divides P, then Q must have degree 1. The converse also holds.

Our last claim in this section formally connects these notions:

Proposition 3. *Let K be a field. The following three statements are equivalent:*

1. *There exists an algebraically closed field extension of K.*
2. *There exists an algebraic closure of K.*
3. *There exists a field extension L of K under a homomorphism ϕ such that every polynomial with coefficients in the subfield $\phi(K)$ splits in L.*

The way we establish that every field admits an algebraically closed extension in our formal development is by proving that (3) implies (2), that (2) implies (1) and that assertion (3) holds. The first two proofs are straightforward and our focus will be in describing how we prove (3).

3   Formalisation

The formalisation of algebra in the absence of dependent types can sometimes be challenging. Below we describe how we addressed these situations. We begin by presenting a technique for changing the underlying type of an algebraic structure. The application of this procedure is named the *induction of a structure*. In addition, we discuss our formalisation of multivariate polynomials. Both of these two features play an important role in the formal proof of the existence of the algebraic closure, and they surely have other applications.

3.1   Induced Structures

The type of monoids is formalised in the HOL-Algebra library as a record with three fields: the carrier, which has the polymorphic type $'a\ set$, the composition operator and finally the unit.

$$\mathtt{record}\ 'a\ monoid\ =\ carrier :: 'a\ set\quad mult :: 'a \Rightarrow 'a \Rightarrow 'a\quad one :: 'a$$

The binary or *direct product* of two monoids is defined as follows:

```
definition DirProd (infix ××) where
```
$$G \times\!\!\times H\ =\ (\!|\ carrier = carrier\ G \times carrier\ H;\ one = (one\ G, one\ H);$$
$$mult = \lambda(g_1,\ h_1)\,(g_2,\ h_2).\,(mult\ G\ g_1\ g_2, mult\ H\ h_1\ h_2)\,|\!)$$

This takes two monoids as arguments, G and H. It returns a monoid whose carrier is the Cartesian product of the carriers of G and H, with composition defined element-wise. Isabelle assigns *DirProd* the type

$$'a\ monoid \Rightarrow 'b\ monoid \Rightarrow ('a * 'b)\ monoid.$$

Now, let's say we want to define the direct product of a list of monoids Gs, that is, the n-ary product of monoids where n is the length of Gs. Consider the following attempt to define this by recursion on the list:

```
fun DirProd_list where
```
$$DirProd\_list\,(G \mathbin{\#} Gs) = G \times\!\!\times DirProd\_list\ Gs$$
$$|\ DirProd\_list\,[] = (\!|\ carrier = \{[]\};\ mult = \lambda as\ bs.\ [];\ one = []\ |\!)$$

Such an attempt to define an n-ary product of monoids from a list of monoids must fail, as the result type would depend on the length of the list.

To solve this problem, we introduce the concept of *induced structures*. The idea is that, given the pair of an algebraic structure such as a monoid G of type $'a\ monoid$ and an injective function f of type $'a \Rightarrow 'b$, we can *induce* a monoid H of type $'b\ monoid$ such that f is an isomorphism between the monoids G and H. Thus, H has the same algebraic properties as G but with the type we want.

Definition 8 (Induced monoid). *Let G be a monoid, H a set and $f : G \to H$ an injection from G to H. Now the image $f(G)$ is a subset of H, which can be equipped with a monoid structure as defined below:*

$$\mathbf{1}_{f(G)} \triangleq f(\mathbf{1}_G), \qquad f(g_1) \otimes_{f(G)} f(g_2) \triangleq f(g_1 \otimes_G g_2).$$

We obtain a monoid $f(G)$, called the monoid induced by G and f. *Furthermore, f is an isomorphism between the monoids G and $f(G)$.*

Here is the formal Isabelle definition (where ' denotes the image operator):

> **definition** *image_monoid* **where**
>   *image_monoid f G =* ⦇ *carrier = f ' carrier G* ; *one = f (one G)* ;
>   *mult = $\lambda h_1 h_2$. f (mult G (inv_into (carrier G) f h_1)*
>   *(inv_into (carrier G) f h_2))* ⦈

Note that composition is defined as $f(f^{-1}h_1 \otimes_G f^{-1}h_2)$, using the inverse of the function f. We use the *inv_into* function (from Isabelle's standard library), which denotes the inverse of $f \upharpoonright A$ (the function f restricted to domain A).

Let's return to the definition of the function *DirProd_list*. With this new tool, we are able to handle the inductive case:

$$DirProd\_list(G \# Gs) = image\_monoid \; (\lambda(a, as).\; a \# as) \; (G \times\times DirProd\_list \, Gs)$$

Now, the definition is accepted by Isabelle, yielding the isomorphism

$$DirProd\_list(G \# Gs) \cong G \times\times DirProd\_list \; Gs.$$

Algebraically speaking, an isomorphism is enough: we do not need true equality.

More generally, given a monoid G of type $'a\ monoid$ and an injective function f of type $'a \Rightarrow 'b$, we prove that the function f is an isomorphism between the monoids G and *image_monoid f G*.

$$f \in iso \; G \; (image\_monoid \; f \; G)$$

Other algebraic structures such as groups, rings and fields can also benefit from this construction. The difference between groups and monoids in HOL-Algebra is only logical: they satisfy different axioms but both have the same type, so *image_monoid* can be applied to groups. The same idea works for other abstract mathematical structures, wherever isomorphism (as opposed to equality) is good enough.

The ability to choose the type of an algebraic structure while preserving its abstract properties can also be useful in the formalisation of certain proofs of existence. We illustrate this use case through the following theorem.

Theorem 1. *Let K be a field and P a polynomial with coefficients in K. Then there exists a field extension L under a homomorphism $\phi : K \to L$ such that the polynomial P^ϕ splits in L.*

Proof. By induction on the degree of P. If $\deg P = 0$, then P already splits in K. Thus, K itself (with the identity homomorphism) is the required field extension.

If $\deg P = n+1$ for some n then there exists an irreducible polynomial Q with coefficients in K such that Q divides P. Since Q is irreducible, by Proposition 2 we obtain the field extension $K[X]/(Q)$ where Q^π has $\pi(X)$ as a root. Since $R \mapsto R^\pi$ is a homomorphism, Q^π divides P^π. Consequently, $\pi(X)$ is also a root of P^π and hence $X - \pi(X)$ divides P^π.

Let R denote the division of P^π by the polynomial $X - \pi(X)$, that is, R is a polynomial with coefficients in $K[X]/(Q)$ such that $P^\pi = (X - \pi(X))\, R$. Then $\deg(R) = n$ and by induction hypothesis we obtain a field L and a homomorphism $\phi : K[X]/(Q) \to L$ such that R^ϕ splits in L. Clearly L under the homomorphism $\phi \circ \pi$ is the required field extension. $\square$

There are a number of obstacles to the formalisation of this proof in Isabelle, starting with the statement itself. The theorem asserts the existence of a field, which in Isabelle has type $'a\,ring$, where $'a$ is the type of the elements of its carrier. We need to find a type for the field whose existence is being claimed.

In a dependent type setting, the problem could be avoided. It would suffice to quantify the type existentially: to announce the existence of both the type and the field. Then, we would be able to build the precise type during the proof.

This is not possible in Isabelle, but let's say that we have a type $'b$ that could satisfy the requirements of Theorem 1. Then, another problem appears: the application of the inductive hypothesis to the polynomial R fails with a type unification issue, because its coefficients belong to $K[X]/(Q)$ while those of P belong to K. In Isabelle, these fields have different types: if K has type $'a\,ring$ and $'a\,poly$ is the type of polynomials with coefficients of type $'a$, then the quotient field $K[X]/(Q)$ would have the type $(('a\,poly)\,set)\,ring$. So even if we generalised the induction hypothesis with respect to the field, we would not be able to instantiate it with $K[X]/(Q)$. And again, dependent types are a solution: one could generalise the induction hypothesis with respect to the *type* of the field.

The solution we propose goes in another direction. The idea is to prove an intermediate result where we fix a well chosen type $'b$ for the elements of both the fields K and L. Then, after we build the field $K[X]/(Q)$ of type $(('b\,poly)\,set)\,ring$ during the proof, we use our type-switching mechanism to induce a field of type $'b\,ring$ with the same properties of $K[X]/(Q)$. At this point, it will be possible to use the induction hypothesis and to conclude the proof.

The tricky part is choosing the type $'b$. In the definition of the function *DirProd_list*, we faced a similar problem. We had to come up with a type $'b$ such that an injective function of type $'a * 'b \Rightarrow 'b$ existed. Clearly, in that case, the type $'a\,list$ was sufficient. Now, the situation is less clear: we need to come up with a type $'b$ such that an injective function of type $('b\,poly)\,set \Rightarrow 'b$ exists.

Observe that the function only needs to be injective on the carrier of the structure we are planning to use as a model for the induction of a new one. Therefore, the definition of an injective function of type $('b\,poly)\,set \Rightarrow 'b$ does not constitute a violation of Cantor's theorem, since the injectivity only needs to

hold in the subset of the elements of type $('b\,poly)\,set$ composed by those which belong to the carrier of the field $K[X]/(Q)$.

The type we conceived to satisfy these conditions is the type of *multivariate polynomials*, or, polynomials with indexed variables. Below we discuss how these are formalised and how they solve the problem for our development.

3.2   Multivariate Polynomials

Polynomials with coefficients in a field K are usually treated as linear combinations of successive powers of a formal letter X. Multivariate polynomials follow the same idea, but, instead of dealing with the powers of a fixed formal letter X^n, we manipulate the linear combination of arbitrary expressions of the form $\prod_j \mathcal{X}_j^{n_j}$, where j runs over a finite subset of a fixed indexing set J.

The formalisation of multivariate polynomials that we are about to present relies on the notion of finite maps. A finite map from the set A to the set B is a partial function from elements of A to elements of B whose support is finite. The set of finite maps from A to B is written $A \xrightarrow{\text{fin}} B$.

First, we formally define monomials over indexed variables: a monomial is uniquely identified by a finite map from elements of the set J to positive integers $\mathbb{Z}_{>0}$. Imagine it as the choice of exponent for each indexed letter. If Mon_J denotes the set of monomials over variables indexed by J, then we have

$$Mon_J \triangleq J \xrightarrow{\text{fin}} \mathbb{Z}_{>0}.$$

We define multivariate polynomials similarly, as finite maps from monomials Mon_J to K^*, the set of nonzero elements of K. The elements in the support of the finite map are the monomials involved in the linear combination. The value in K^* associated to each monomial is the choice of coefficient. Accordingly, if $K[J]$ denotes the set of polynomials over variables indexed by J and coefficients in K, then

$$K[J] \triangleq Mon_J \xrightarrow{\text{fin}} K^*.$$

Isabelle's predefined type of *multisets* comprises the finite maps into $\mathbb{Z}_{>0}$. A multiset is a function denoting the number of occurrences of each element, and multisets are finite. In our development, monomials have the same type as multisets. So, if the indexing set J has type $'c\,set$, then a monomial would have the type $'c\,multiset$.

For polynomials, there was no shortcut. They are modelled as functions from monomials to K and we require that the image of this function is zero save for a finite set of monomials. Therefore, a multivariate polynomial has the type $'c\,multiset \Rightarrow 'a$, where, as usual, a is the type of the elements in the field K.

Now, we substantiate our claim that if the type $'b$ is instantiated with the type of multivariate polynomials, then there exists an injective function of type $('b\,poly)\,set \Rightarrow 'b$.

First, remember that the *set* constructor was introduced because it is the type of the elements of the quotient field $K[X]/(Q)$: equivalence classes are encoded

as cosets of the form $\pi(P)$ for some $P \in K[X]$. However, these equivalence classes could also be uniquely identified by polynomials:

Definition 9. *Let M be a field and Q a polynomial with coefficients in M. We let $mod_{[M,Q]}$ denote the function which assigns the remainder of the euclidean division of P by Q to each equivalence class $\pi(P)$ for $P \in M[X]$. It is a well-defined injective function from $M[X]/(Q)$ to $M[X]$.*

We use the letter M for an arbitrary field instead of the usual letter K, because we instantiate this definition with a field whose elements have type $'b$, that is, a field included in $K[J]$. Thus, back in Isabelle, we are able to define an injective function of type $('b\,poly)\,set \Rightarrow 'b\,poly$. The problem becomes simpler, since now it suffices to define an injective function of type $'b\,poly \Rightarrow 'b$.

In other words, given a polynomial P with coefficients in $K[J]$, we need to define a unique way to recover an element of $K[J]$. We do this by replacing the formal letter X from the polynomial P with an indexed one, $\mathcal{X}_l$. However, in order to get injectivity, we constrain the coefficients of P to not *use* the indexed variable $\mathcal{X}_l$. For this purpose, we introduce the notion of a *free* index:

Definition 10 (Free index). *Let M be a subset of $K[J]$ and $l \in J$ be an index. Then l is* free *in M if M is a subset of $K[J \setminus \{l\}]$.*

Intuitively, l is free in M if $\mathcal{X}_l$ does not appear in the writing of any term in M. The idea of replacing the formal letter X with $\mathcal{X}_l$ is captured as follows:

Definition 11 (Eval). *Let K be a field, J be an indexing set and l an index in J. We define $Eval_l$, an injective function from polynomials with coefficients in $K[J \setminus \{l\}]$ to elements in $K[J]$:*

$$Eval_l \left(\sum_k \left(\sum_i a_{ik} \prod_{j \in J_{ik}} \mathcal{X}_j^{n_{ijk}} \right) X^k \right) = \sum_{i,k} a_{ik} \prod_{j \in J_{ik} \cup \{l\}} \mathcal{X}_j^{n'_{ijk}}$$

$$where\ n'_{ijk} = \begin{cases} k & if\ j = l \\ n_{ijk} & otherwise \end{cases}$$

Finally, we are able to define an injective function of type $('b\,poly)\,set \Rightarrow 'b$ as the composition of *Eval* and *mod*. The following technical lemma exploits this function to induce a field isomorphic to $M[X]/(Q)$.

Lemma 1. *Let K be a field, J be an indexing set, M be a field whose elements belong to $K[J]$ and Q an irreducible polynomial in $M[X]$. If $l \in J$ is a free index in M, then the composition $Eval_l \circ mod_{[M,Q]}$ is an injective map from $M[X]/(Q)$ to $K[J]$.*

$$M[X]/(Q) \xrightarrow{mod_{[M,Q]}} M[X] \subseteq (K[J \setminus \{l\}])[X] \xrightarrow{Eval_l} K[J]$$

Moreover, let L denote the structure induced by $Eval_l \circ mod_{[M,Q]}$. It is a field whose elements belong to $K[J]$. Furthermore, M is a subfield of L and the indexed variable $\mathcal{X}_l$ is a root of Q in L.

Proof. Injectivity of the map $Eval_l \circ mod_{[M,Q]}$ comes from the composition of injective maps. Now, let L denote the field induced by $Eval_l \circ mod_{[M,Q]}$. It is a field isomorphic to $M[X]/(Q)$ inheriting the same algebraic properties as such. For instance, there exists a root of Q in L: it is the element that realises $\pi(X)$, that is, $(Eval_n \circ mod_{[M,Q]})(\pi(X))$. However, remember that $\pi(X)$ was a root of the polynomial Q^π in $M[X]/(Q)$. So, the element

$$(Eval_n \circ mod_{[M,Q]})(\pi(X))$$

is actually a root of the polynomial $Q^{Eval_n \circ mod_{[M,Q]} \circ \pi}$. Fortunately,

$$(Eval_n \circ mod_{[M,Q]})(\pi(X)) = Eval_n(mod_{[M,Q]}(\pi(X)) = Eval_n(X) = \mathcal{X}_n$$

and $(Eval_n \circ mod_{[M,Q]} \circ \pi)(a) = a$ for $a \in M$.

Therefore, the field M is a subfield of L, the polynomial Q has its coefficients in the field L and $\mathcal{X}_n$ is a root of Q in L.   □

We can finally proceed to the proof of the intermediate result that enjoys a direct analogue in Isabelle as suggested in the previous subsection.

Lemma 2. *Let $M \subseteq K[\mathbb{N}]$ be a field and let P be a polynomial with coefficients in M. Suppose that for every $j \in \{0, \ldots, \deg P - 1\}$, the index j is free in M. Under these hypotheses, there exists a field $L \subseteq K[\mathbb{N}]$, such that M is a subfield of L and P splits in L.*

Proof. By induction on the degree of P. If $\deg P = 0$, then P splits in M and we are done. If $\deg P > 0$, then there exists an irreducible polynomial $Q \in M[X]$ such that Q divides P. We can suppose that $\deg Q > 1$, otherwise P have only trivial irreducible factors and would split in M.

With the polynomial Q and the index n, we are in position to apply Lemma 1; let S be the field induced by the injective map $Eval_n \circ mod_{[M,Q]}$ such that M is a subfield of S and $\mathcal{X}_n$ is a root of Q in S.

Since the only free index in M that becomes nonfree in S is the index n, we are allowed to instantiate the induction hypothesis with the field S and the polynomial $P/(X - \mathcal{X}_n)$, whose degree is equal to n. We obtain a field $L \subseteq K[\mathbb{N}]$, such that S is a subfield of L and the polynomial $P/(X - \mathcal{X}_n)$ splits in L. It is easy to see that P splits in L as well.   □

To recover the proper statement of Theorem 1, we only need to find a way to embed K into the set $K[\mathbb{N}]$. If we have a homomorphism $\phi : K \to M$ such that M is a subset of $K[\mathbb{N}]$, then the field we get from Lemma 2 together with ϕ will be the field extension satisfying the conditions assured by Theorem 1. The required embedding simply maps each $a \in K$ to the constant polynomial a. Using our representation of multivariate polynomials as finite maps, we can make this intuition precise:

Definition 12. *Let K be a field and J an indexing set. We define $const^\#$ a function from K to the set $K[J]$:*

$$const^\#(a) = \begin{cases} \{\} & if\, a = 0 \\ \{\} \mapsto a & otherwise \end{cases}$$

where $\_ \mapsto \_$ denotes the singleton map and $\{\}$, the empty one.

It is an injective map, therefore we can promote the function $const^\#$ to an isomorphism between K and the induced field $const^\#(K)$, written $K^\#$ for brevity.

Corollary 1 (Same statement as Theorem 1).

Proof. Consider the field $K^\# \subseteq K[\mathbb{N}]$. Observe that every index $n \in \mathbb{N}$ is free in $K^\#$. Indeed, an element of $K^\#$ is the image $const^\#(a)$ for some a in K and no indexed variable is involved in such terms. Lemma 2 applies. We obtain a field $L \subseteq K[\mathbb{N}]$ such that $K^\#$ is a subfield of L and $P^{const^\#}$ splits in L. Rephrase it with $\phi = const^\#$: the field L is a field extension of K under the homomorphism ϕ such that the polynomial P^ϕ splits in L.   □

The proof of Lemma 2 follows the same structure as the one from Theorem 1: obtain an irreducible factor, build the quotient field and then apply induction. The major difference is that now we have instrumented the proof with technical arguments concerning indexes. These might seem artificial at first, but they do have their mathematical value. After all, we've built a splitting field of a polynomial in the same set from which we started. In other words, we had greater control over the field that was being built during the proof. Notice that all field extensions explicitly mentioned in the proof are those under the identity homomorphism.

A final remark: we do not define how multivariate polynomials compose with each other; we just show how they can be represented. We do not define the ring operations as we do not need their algebraic properties. We use only their set properties, which ease the definition of an injective function as described and allow the construction of an induced structure. Moreover, this procedure of inducing structures engenders the composition laws for free.

4 Conception

In this section we outline the main ideas behind our formal proof of the existence of an algebraic closure. We start with a short survey of existing proofs. Technical details are left for the next section.

We mentioned in the introduction that the existence of an algebraic closure was proved by Steinitz in 1910. However, because set theory was not yet well understood, constructions relying on infinite collections were unnecessarily involved and his proof was 20 pages long. The proof preferred by modern authors [2,9], due to Emil Artin, is much shorter and simpler:

Proof. Let F be a field. Consider the ring $F[F[X]]$ of multivariate polynomials over variables indexed by polynomials in $F[X]$ and coefficients in F. Then the ring $F[F[X]]$ admits an ideal I with two key properties: (1) the quotient $F[F[X]]/I$ is a field; (2) every nonconstant polynomial in $F[X]$ has a root in $F[F[X]]/I$. This yields a field extension of F where every nonconstant polynomial with coefficients in F has at least one root. Furthermore, the construction is parametric in F. Now, consider the chain of field extensions

$$K = E_0 \to E_1 \to E_2 \to \cdots \to E_n \to \cdots$$

where K is the field for which we intend to find an algebraically closed extension and E_{n+1} is the field obtained by instantiating F with the field E_n. The sequence admits a limit. It is a field E left invariant by the construction described above, that is, a fixed point: every nonconstant polynomial in $E[X]$ has a root in E. Thus, E is an algebraically closed field.   □

There are two main components in Artin's proof: the construction of a field and its iteration. We might wonder which field we obtain if we change the first component. Would we still have an algebraically closed field? What would be the impact on the iteration part? Hernandez and Laszlo [6] present a variant of the proof where the construction is modified to build a field with stronger properties. They consider the larger ring of multivariate polynomials $K[K[X] \times \mathbb{N}]$ with the product $K[X] \times \mathbb{N}$ as the choice of indexing set and prove that a slightly different construction engenders an algebraically closed field extension of K directly, without iteration.

Our proof goes in the opposite direction. We weaken the construction and strengthen the iteration method. If F is the current field, we find an irreducible polynomial Q with coefficients in F and define the usual quotient $F[X]/(Q)$ of the ring $F[X]$ by the ideal (Q) to be the next field in the sequence.

The increment now is small. Proposition 2 states that we add only one root at each step, while the previous construction found one root for every polynomial with coefficients in the current field. We compensate for this deficit through the use of Zorn's lemma, which guarantees the existence of a maximal element of a partially ordered set. Thus we abstract away from the process of iteration and leap immediately to the required limit.

In our proof, we need this ability. With the construction from Artin's proof, we were content to iterate "vertically" over the degree of each polynomial at the same time. Now, we have to iterate over each polynomial individually. A step of iteration considers one polynomial at a time. So, we iterate both over each polynomial and its degree, both "horizontally" and "vertically".

More precisely, given both a field E_n in the sequence of Artin's proof and a polynomial P with coefficients in the field E_n, we can find a k such that P splits in the field E_{n+k}. It suffices to take k larger than the degree of P. With our new construction however, the distance between the field where a polynomial has its coefficients and the field in which the same polynomial splits might be larger than every natural number. Since our method iterates over polynomials, the subset of steps that interleave the two fields might have the same cardinality

as the set of polynomials, which could be uncountable, depending on the the cardinality of the field of coefficients.

This simplification eases the formalisation in a number of ways. The field $F[X]/(Q)$ depends on the ring of polynomials with coefficients in F and the ideal generated by Q, two straightforward structures. The field $F[F[X]]/I$, on the other hand, depends on the *ring* of multivariate polynomials and the ideal I. While we don't see any issues with the definition of the composition laws for the set $F[F[X]]$, the proof that they satisfy the axioms of a commutative ring seems laborious. Moreover, the proof that I enjoys the two key properties discussed in the proof sketch relies on the intermediate result that every field admits an extension that splits a finite set of polynomials (a corollary of Theorem 1). We would have to prove a lemma that would immediately be subsumed by our intended theorem: the existence of an algebraically closed extension.

The idea of coupling a simpler construction with Zorn's lemma is also seen in Jelonek [8]. But while he relies on set-theoretical arguments to obtain a set sufficiently large to host an algebraic closure, we exhibit this set explicitly.

5   The Proof

In this section, we explain our proof that every field admits an algebraic closure. First, let's recall order theory and Zorn's lemma:

Definition 13 (Partial orders, etc.). *Let $\leq$ be a binary relation on a set S. Then $(S, \leq)$ is a* partial order *if $\leq$ is transitive, reflexive and anti-symmetric. A* chain *of S is a subset of S for which every two of its elements are $\leq$-comparable. A* maximal *element is some $a \in S$ such that for all $x \in S$, if $a \leq x$ then $a = x$, while an* upper bound *of S is some $a \in S$ for which $a \geq x$ for all $x \in S$.*

Lemma 3 (Zorn). *Let $(S, \leq)$ be a partial order where S is nonempty. Suppose every chain $C \subseteq S$ has an upper bound. Then $(S, \leq)$ has a maximal element.*

Let K be a field. We will search for an algebraic closure of K in the set $K[J]$ of multivariate polynomials indexed by a well-chosen set J. Recall that an algebraic closure is a field extension, hence we have to search both for a field L and for a homomorphism $\phi : K \to L$. We make some decisions to reduce the search space. We expect L to be an extension of the induced field $K^{\#}$. Thus, we anticipate ϕ to be the natural homomorphism

$$const^{\#} : K \to K^{\#}$$

from K to the induced field $K^{\#}$. Now, given a field L embedded in $K[J]$ such that $K^{\#}$ is a subfield of L, we can simply put L to be the field extension of K under $const^{\#}$.

Let's simplify further. Let's search for a field $L \subseteq K[J]$ such that $K^{\#}$ is a subfield of L and every polynomial with coefficients in $K^{\#}$ splits in L. Such a field is not necessarily an algebraic closure of $K^{\#}$, with each of its elements

algebraic over $K^{\#}$. Still, exhibiting a field with these simpler characteristics is sufficient to prove the existence of an algebraic closure, as it proves assertion (3) from Proposition 3.

With these considerations in mind, let's define the concrete set of fields.

Definition 14. *For a field K, let $\mathcal{A}_K$ denote the set of fields $L \subseteq K[K[X] \times \mathbb{N}]$ satisfying the following properties:*

1. $K^{\#}$ *is a subfield of L.*
2. *For every $P \in K[X]$ and natural number n, if the pair (P, n) is a nonfree index of L, then the indexed variable $\mathcal{X}_{(P,n)}$ is a root of $P^{const^{\#}}$ in L.*

The intuition for property (2) and for the choice of indexing set, $K[X] \times \mathbb{N}$, follows from the insights developed in the last section. Recall the main idea: to build a chain of fields where at each step we add a new root to the preceding field. What property (2) intuitively does is to ensure that the role of the new root is taken by a formal letter $\mathcal{X}_j$, where j is some previously free index j. Additionally, it also ensures that the choice of index was judicious: if we choose to include the variable $\mathcal{X}_{(P,n)}$ in the next round of the iteration, then it must be in a way such that it is a root of the polynomial $P^{const^{\#}}$. Look both at the statement of Lemma 2 and its proof for an example of similar reasoning.

Next, we equip $\mathcal{A}_K$ with a partial ordering, aiming to use Zorn's lemma.

Definition 15 (Subfield relation). *Let L_1 and L_2 be two fields. We write $L_1 \lesssim L_2$ to denote that L_1 is a subfield of L_2.*

Lemma 4. *If K is a field, then $\mathcal{A}_K$ has a maximal element with respect to the partial order $(\mathcal{A}_K, \lesssim)$.*

Proof. Clearly $(\mathcal{A}_K, \lesssim)$ is a partial order. Consider a chain $C \subseteq \mathcal{A}_K$ and let $E = \bigcup_{F \in C} F$. For any two elements a and b in E, there exists a field $L \in C$ such that $a, b \in L$. We define their composition in E (addition and multiplication) to be the same as in L. Since every two fields in C are comparable, if we obtain another field $L' \in C$ such that the elements a and b belong to L', then the field L' is either a subfield of L or an extension of L. In both cases, the composition of a and b gives the same result either in L or in L'. Hence, their composition is independent on the choice of the field L and therefore well-defined. The set E equipped with these laws satisfies the axioms of a field. It is also straightforward to check that E is an upper bound of C in $\mathcal{A}_K$.

Since every chain has an upper bound, the result holds by Zorn's lemma.   □

Although the usual sequence of fields is hidden by the application of Zorn's lemma, we can think of the maximal element of $\mathcal{A}_K$ as the limit of that sequence. It must be a field that splits every polynomial in $K^{\#}[X]$, since otherwise there would still be space left to add another root, contradicting its maximality:

Theorem 2. *Let K be a field and let M be the maximal element of $\mathcal{A}_K$ for the subfield relation. Every polynomial with coefficients in $K^{\#}$ splits in M.*

Proof. For contradiction, suppose that there exists a polynomial with coefficients in $K^{\#}$ that does not split in M. Since the map $const^{\#}$ is an isomorphism between the fields K and $K^{\#}$ we can suppose that the polynomial which fails to split in M is of the form $P^{const^{\#}}$ where P is some polynomial with coefficients in K.

Let $Q \in M[X]$ be an irreducible polynomial with degree greater than 1 such that Q divides P and let n be a natural number such that the pair (P, n) is a free index in M. They both exist, otherwise the polynomial $P^{const^{\#}}$ would split in M. We are able to apply Lemma 1 instantiated with the polynomial Q and the index (P, n) to obtain a field L such that M is a subfield of L and the indexed variable $\mathcal{X}_{(P,n)}$ is a root of Q in L.

Clearly L belongs to $\mathcal{A}_K$, and the indexed variable $\mathcal{X}_{(P,n)}$ is an element of L that does not belong to M. So L is an element of $\mathcal{A}_k$ that is strictly greater than M for the subfield relation, contradicting the maximality of M.   □

Corollary 2. *Every field K admits an algebraic closure.*

Proof. The maximal element of $\mathcal{A}_K$ is a field extension of K under the homomorphism $const^{\#}$ satisfying statement 3 from Proposition 3. By the same theorem, this is sufficient to prove the existence of an algebraic closure of K.   □

6   Related Work

We believe that our formalisation of the existence of an algebraic closure of any field is novel. There is, however, a closely related theorem formalised by Gonthier as part of the Mathematical Components library: every *countable* field admits an algebraic closure [4]. Since the proof is carried out in Coq, it is especially interesting for its computational content. The restriction to countable fields, on the other hand, precludes the application of the theorem to uncountable fields such as the reals or the p-adic numbers. This is a problem for the p-adic numbers, which (unlike the reals) has no well-known algebraic closure construction.

Also in Coq, Mathematical Components supports multivariate polynomials [7] over finite indexing sets of the form $\{1, \ldots, n\}$.

Schwarzweller proved in Mizar that the real numbers and finite fields are not algebraically closed [11]. The Lean community maintains an online "Algebraic closure roadmap" [1].

Work has also been done in Isabelle. An AFP entry [12] uses lists to represent monomials and polynomials. This choice has a drawback—permutations of a list might denote the same monomial—but it allows the operations to be executable. Haftmann et al. [5] discuss different options for multivariate polynomials in Isabelle. They propose two representations, an abstract and a concrete one, and establish a formal correspondence between them. The abstract representation resembles ours; it is based on finite maps. However, the authors fix the indexing set to the integers. We cannot use either of these Isabelle libraries because both rely on type classes to model the algebraic properties satisfied by polynomial coefficients. While type classes are fine for sharing algebraic theorems between types [10], they are too restrictive for abstract algebra, forcing algebraic objects to be types.

7   Conclusion

We formalised a proof that every field has an algebraic closure, a fundamental theorem. We have given a precise description of the known proofs and demonstrated how Zorn's lemma led to a straightforward formalisation.

The obstacles that we encountered during the realisation of this work led to the investigation of some problems related to Isabelle's type-discipline: that all the objects in a collection must have the same type. As a solution, we introduced the notion of an *induced structure*, which yields an isomorphic algebraic structure having a specified type.

To complete this project, we had to extend the HOL-Algebra library with general-purpose topics such as multivariate polynomials, arithmetic on arbitrary rings, arithmetic on the ring of polynomials, subfields, finite extensions and much of the content in Section 2. The fundamental nature of this material strengthens the library as the basis for further developments of algebra in Isabelle/HOL. It comprises nearly 15,000 nonempty lines of code in 21 new theories. The work was completed within 5 months.

Availability Our development is part of HOL-Algebra. It is included in the distribution of Isabelle, directory `src/HOL/Algebra` and can also be browsed online.[3]

8   Acknowledgments

References

1. Algebraic closure roadmap. Online at `https://github.com/leanprover-community/mathlib/wiki/Algebraic-closure-roadmap`.
2. Kevin Buzzard. Existence of algebraic closure of a field. Online at `https://wwwf.imperial.ac.uk/~buzzard/maths/teaching/15Aut/M3P11/algclosure.pdf`, 2016.
3. Kevin Buzzard. Motivation for algebraically closed fields?, October 2019. Email dated 2019-10-05.
4. Georges Gonthier. Closed fields. Online at `https://math-comp.github.io/htmldoc/mathcomp.field.closed_field.html`.
5. Florian Haftmann, Andreas Lochbihler, and Wolfgang Schreiner. Towards abstract and executable multivariate polynomials in isabelle. In *Isabelle Workshop*, 2014. Online at `https://www3.risc.jku.at/publications/download/risc_5012/IWS14.pdf`.

---

[3] `https://isabelle.in.tum.de/dist/library/HOL/HOL-Algebra/`

6. David Hernandez and Yves Laszlo. *Introduction à la Théorie de Galois*. Polytechnique, 2014.
7. Florent Hivert and Laurent Thery. Multinomials. Online at `https://github.com/math-comp/multinomials`.
8. Zbigniew Jelonek. A simple proof of the existence of the algebraic closure of a field. *Universitatis Iagellonicae Acta Mathematica*, 30:131–132, 1993. Online at `http://www2.im.uj.edu.pl/actamath/PDF/30-131-132.pdf`.
9. Serge Lang. *Algebra*. Springer, 2002.
10. Lawrence C. Paulson. Organizing numerical theories using axiomatic type classes. *J. Automated Reasoning*, 33(1):29–49, 2004.
11. Christoph Schwarzweller. On roots of polynomials and algebraically closed fields. *Formalized Mathematics*, 25, 10 2017.
12. Christian Sternagel and René Thiemann. Executable multivariate polynomials. *Archive of Formal Proofs*, 08 2010.