From Tarski to Descartes: Formalization of the Arithmetization of Euclidean Geometry

Pierre Boutry, Gabriel Braun, Julien Narboux

From Tarski to Descartes: Formalization of the Arithmetization of Euclidean Geometry

Pierre Boutry, Gabriel Braun, and Julien Narboux

ICube, UMR 7357 CNRS, University of Strasbourg
Pôle API, Bd Sébastien Brant, BP 10413, 67412 Illkirch, France
{boutry, braun, narboux}@unistra.fr

Abstract

This paper describes the formalization of the arithmetization of Euclidean geometry in the Coq proof assistant. As a basis for this work, Tarski's system of geometry was chosen for its well-known metamathematical properties. This work completes our formalization of the two-dimensional results contained in part one of [SST83]. We defined the arithmetic operations geometrically and proved that they verify the properties of an ordered field. Then, we introduced Cartesian coordinates, and provided characterizations of the main geometric predicates. In order to prove the characterization of the segment congruence relation, we provided a synthetic formal proof of two crucial theorems in geometry, namely the intercept and Pythagoras' theorems. To obtain the characterizations of the geometric predicates, we adopted an original approach based on bootstrapping: we used an algebraic prover to obtain new characterizations of the predicates based on already proven ones. The arithmetization of geometry paves the way for the use of algebraic automated deduction methods in synthetic geometry. Indeed, without a "back-translation" from algebra to geometry, algebraic methods only prove theorems about polynomials and not geometric statements. However, thanks to the arithmetization of geometry, the proven statements correspond to theorems of *any* model of Tarski's Euclidean geometry axioms. To illustrate the concrete use of this formalization, we derived from Tarski's system of geometry a formal proof of the nine-point circle theorem using the Gröbner basis method.

Introduction

There are several ways to define the foundations of geometry. In the *synthetic* approach, the axiomatic system is based on some geometric objects and axioms about them. The best-known modern axiomatic systems based on this approach are those of Hilbert [Hil60] and Tarski [SST83]. In the *analytic* approach, a field $\mathbb{F}$ is assumed (usually $\mathbb{R}$) and the space is defined as $\mathbb{F}^n$. In the mixed analytic/synthetic approaches, one assumes both the existence of a field and also some geometric axioms. For example, the axiomatic systems for geometry used for education in north America are based on Birkhoff's axiomatic system [Bir32] in which the field serves to measure distances and angles. This is called the metric approach, a modern development of geometry based on this approach can be found in the books of Millman or Moise [MP91, Moi90]. A similar approach is also used by Chou, Gao and Zhang for the definition of the area method [CGZ94] (a method for automated deduction in geometry). Analogous to

Birkhoff's axiomatic system, here the field serves to measure ratios of signed distances and areas. The axioms and their formalization in Coq can be found in [JNQ12]. Finally, in the modern approach for the foundations of geometry, a geometry is defined as a space of objects and a group of transformations acting on it (Erlangen program [Kle72]).

Although these approaches seem very different, Descartes proved that the analytic approach can be derived from the synthetic approach. This is called arithmetization and coordinatization of geometry. In [Des25] he defined addition, multiplication and square roots geometrically.

Our formalization of geometry consists in a synthetic approach based on Tarski's system of geometry. Readers unfamiliar with this axiomatic system may refer to [TG99] which describes its axioms and their history. Szmielew (a student of Tarski) and Schwabhäuser have produced a systematic development of geometry based on this system. It constitutes the first part of [SST83]. We have formalized the 16 chapters corresponding to this first part in the Coq proof assistant.

In this paper, we report on the formalization of the last three chapters containing the final results: the arithmetization and coordinatization of Euclidean geometry. It represents the culminating result of both [Hil60] and [SST83]. This formalization enables us to put the theory proposed by Beeson in [Bee13] into practice in order to obtain automatic proofs based on geometric axioms using algebraic methods. The arithmetization of geometry is a crucial result because, first, it guarantees that the axiomatic system captures the Euclidean geometry, and then it allows to use algebraic methods.

> As long as algebra and geometry traveled separate paths their advance was slow and their applications limited. But when these two sciences joined company, they drew from each other fresh vitality, and thenceforth marched on at a rapid pace toward perfection.
>
> (Joseph-Louis Lagrange, Leçons élémentaires sur les mathématiques; quoted by Morris Kline, Mathematical Thought from Ancient to modern Times, p. 322)

A formalization of the arithmetization of geometry and characterization of geometric predicates is motivated by the need to exchange geometric knowledge data with a well defined semantics. Algebraic methods for automated deduction in geometry can now be used in dynamic geometry systems such as GeoGebra which is used heavily in classrooms [BHJ+15]. Our formalization paves the way for storing standardized, structured, and rigorous geometric knowledge data [CW13].

Up to our knowledge our library is the first formalization of the arithmetization of Euclidean geometry. However the reverse connection, namely that the Euclidean plane is a model of this axiomatized geometry, has been mechanized by Marić, Petrović, Petrović, and Janičić [MPPJ12]. In [MP15], Marić and Petrović formalized complex plane geometry in the Isabelle/HOL theorem prover. In doing so, they demonstrated the advantage of using an algebraic approach and the need for a connection with a synthetic approach. Some formalization attempts of Hilbert's foundations of geometry have been proposed by Dehlinger, Dufourd and Schreck [DDS00] in the Coq proof assistant, and by Dixon, Meikle and Fleuriot [MF03] using Isabelle/HOL. Likewise, a few developments based on Tarski's system of geometry have been carried out. For example, Richter, Grabowski and Alama have ported some of our Coq proofs to Mizar [RGA14] (46 lemmas). Moreover, Beeson and Wos proved 200 theorems of the first twelve chapters of [SST83] with the Otter theorem prover [BW15]. Finally, Durdevic *et. al.* [SDNJ15] generated automatically some readable proofs in Tarski's system of geometry. None of these attempts went up to Pappus' theorem nor to the arithmetization of geometry.

The formalization presented in this paper is based on the library GeoCoq which contains

a formalization of the first part of [SST83] and some additional results. This includes the proof that Hilbert's axiomatic system (without continuity) can be derived from Tarski's axioms [BN12], some equivalences between different versions of the parallel postulate [BNS15b], some decidability properties of geometric predicates [BNSB14a] and the synthetic proof of some popular high-school geometry theorems. The proofs are mainly manual, but we used the tactics presented in [BNSB14b, BNS15a].

The arithmetization of geometry will allow us to link our formalization to the existing formalization of algebraic methods in geometry. For instance, the Gröbner basis method has already been integrated into Coq by Grégoire, Pottier and Théry [GPT11]. Furthermore, the area method for non-oriented Euclidean geometry has been formalized in Coq as a tactic [CGZ94, Nar04, JNQ12]. Geometric algebras are also available in Coq and can be used to automatically prove theorems in projective geometry [FT11]. Finally, the third author together with Génevaux and Schreck has previously studied the integration of Wu's method [GNS11].

We first describe the formalization of the arithmetization of Euclidean geometry (Sec. 1). Then we provide the characterization of the main geometric predicates (Sec. 2) obtained by an original approach. Finally, we present an example of a proof by computation (Sec. 3).

1 Arithmetization of Geometry

As a basis for this work, Tarski's system of geometry was chosen for its well-known metamathematical properties. We assumed the axioms given in [SST83] excluding the axiom introducing continuity. Tarski's axiom system is based on a single primitive type depicting points and two predicates, namely congruence noted $\equiv$ and betweenness noted $-\!-$. $AB \equiv CD$ states that the line-segments $\overline{AB}$ and $\overline{CD}$ have the same length. $A{-}B{-}C$ means that A, B and C are collinear and B is between A and C (and B may be equal to A or C).

1.1 Definition of Arithmetic Operations

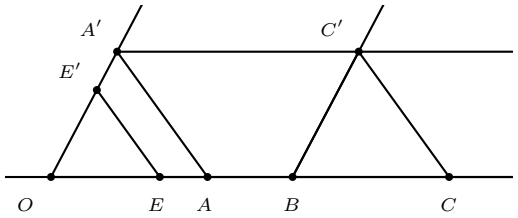
To define the arithmetic operations we first needed to fix the neutral element of the addition O and the neutral element of the multiplication E. The line OE will then contain all the points for which the operations are well-defined as well as their results. Moreover, a third point E' is required for the definitions of these operations. This point, together with points O and E, specifies the geometric constructions corresponding to them. It is to be noticed that these points should not be collinear (collinearity is expressed with the Col predicate defined in Table 1 where all the predicates which are not defined in this paper are listed together with their definition). Indeed, if they were collinear the results of these operations would not be well-defined. These properties are formalized by the definition `Ar2`:

```
Definition Ar2 O E E' A B C := ~ Col O E E' /\ Col O E A /\ Col O E B /\ Col O E C.
```

Definition of Addition

The definition of addition that we adopted is the same as in [SST83] which is expressed in terms of parallel projection. The same definition is given by Hilbert in Chapter V, Section 3 of [Hil60]. `Pj` is a predicate that captures parallel projection. `Pj A B C D` denotes that either lines AB and CD are parallel or $C = D$. The addition is defined as a predicate and not as a function. `Sum O E E' A B C` means that C is the sum of A and B wrt. O, E and E'.

```
Definition Sum O E E' A B C :=
 Ar2 O E E' A B C /\
 exists A', exists C',
 Pj E E' A  A' /\ Col O E' A' /\
 Pj O E  A' C' /\ Pj O E' B  C' /\
 Pj E' E C' C.
```



To prove existence and uniqueness of the last argument of the sum predicate, we introduced an alternative and equivalent definition highlighting the ruler and compass construction presented by Descartes. `Proj P Q A B X Y` states that Q is the image of P by projection on line AB parallel to line XY and `Par A B C D` denotes that lines AB and CD are parallel.

```
Definition Sump O E E' A B C :=
  Col O E A /\ Col O E B /\
  exists A', exists C', exists P', Proj A A' O E' E E' /\ Par O E A' P' /\
                                   Proj B C' A' P' O E' /\ Proj C' C O E E E'.
```

One should note that this definition is in fact independent of the choice of E', and it is actually proved in [SST83]. Furthermore, we could prove it by characterizing the sum predicates in terms of the segment congruence predicate ($\equiv$):

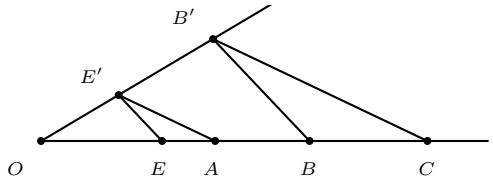
```
Lemma sum_iff_cong : forall A B C,
  Ar2 O E E' A B C -> (O <> C \/ B <> A) ->
 ((Cong O A B C /\ Cong O B A C) <-> Sum O E E' A B C).
```

We used properties of parallelograms to prove this characterization and the properties about `Sum`, contrary to what is done in [SST83] where they are proven using Desargues' theorem[1].

Definition of Multiplication

As for the definition of addition, the definition of multiplication presented in [SST83] uses the parallel projection:

```
Definition Prod O E E' A B C :=
  Ar2 O E E' A B C /\ exists B',
  Pj E E' B B' /\ Col O E' B' /\
  Pj E' A B' C.
```



Similarly to the definition of addition, we introduced an alternative definition which underlines that the definition corresponds to Descartes' ruler and compass construction:

```
Definition Prodp O E E' A B C :=
  Col O E A /\ Col O E B /\ exists B', Proj B B' O E' E E' /\ Proj B' C O E A E'.
```

Using Pappus' theorem, we proved the commutativity of the multiplication and, using Desargues' theorem, its associativity. We omit the details of these well-known facts.

[1] We can remark that we proved the parallel case of this theorem without relying on Pappus' theorem but on properties about parallelograms.

1.2 Construction of an Ordered Field

In his thesis [Gup65], Gupta provided an axiom system for the theory of n-dimensional Cartesian spaces over the class of all ordered fields. In [SST83], a n-dimensional Cartesian space over Pythagorean ordered fields is constructed. We restricted ourselves to the planar case.

Field properties

In Tarski's system of geometry, the addition and multiplication are defined as relations capturing their semantics. Afterward these definitions are generalized to obtain total functions. Indeed, the predicates `Sum` and `Prod` only hold if the predicate `Ar2` holds for the same points. All field properties are then proved geometrically. In theory, we could carry out with the relational versions of the arithmetic operators. But in practice, this causes two problems. First, the statements become quickly unreadable. Second, we cannot apply the standard Coq tactics `ring` and `field` because they only operate on rings and fields defined in terms of functions.

Obtaining the function from the functional relation is implicit in [SST83]. In practice, in the Coq proof assistant, we employed the `constructive_definite_description` axiom provided by the standard library:

```
Axiom constructive_definite_description :
  forall (A : Type) (P : A->Prop), (exists! x, P x) -> { x : A | P x }.
```

It allows to turn a relation which has been proved to be functional into a proper Coq function. As the use of the ϵ axiom turn the intuitionist logic of Coq into an almost classical logic [Bel93], we decided to postpone the use of this axiom as much as possible. For example, we defined the sum function relying on the following lemma[2]:

```
Lemma sum_f : forall A B, Col O E A -> Col O E B -> {C | Sum O E E' A B C}.
```

This function is not total, the sum is defined only for points which belong to our ruler (OE). Nothing but total functions are allowed in Coq, hence to define the ring and field structures, we needed a dependent type (a type which depends on a proof), describing the points which belong to the ruler. In Coq's syntax it is expressed as:

```
Definition F : Type := {P: Tpoint | Col O E P}.
```

Here, we chose a different approach than in [SST83], in which, as previously mentioned, the arithmetic operations are generalized to obtain function symbols without having to restrict the domain of the operations. Doing so implies that the field properties only hold under the hypothesis that all considered points belong to the ruler. This has the advantage of enabling the use of function symbols but the same restriction to the points belonging to the ruler is needed.

We defined the equality on `F` with the standard Coq function `proj1_sig` which projects on the first component of our dependent pair, forgetting the proof that the points belong to the ruler:

```
Definition EqF (x y : F) := (proj1_sig x) = (proj1_sig y).
```

[2]We chose to omit the definitions of functions corresponding to the arithmetic operations to avoid technicalities.

This equality is naturally an equivalence relation. One should remark that projecting on the first component is indeed needed. Actually, we showed in [BNSB14a] that the decidability of the equality implies the decidability of every predicate present in [SST83]. The decidability that we assumed was in `Prop` and not in `Set` to avoid assuming a much stronger axiom. By Hedberg's theorem, equality proofs of types which are in `Set` are unique. This allows to get rid of the proof relevance for dependent types. Nevertheless the decidability of the collinearity predicate is in `Prop`, where equality proofs are not unique. Therefore, the proof component is not irrelevant here.

Next, we built the arithmetic functions on the type `F`. In order to employ the standard Coq tactics `ring` and `field` or the implementation of setoids in Coq [Soz10], we proved some lemmas asserting that the operations are morphisms relative to our defined equality. For example, the fact that $A = A'$ and $B = B'$ (= means `eqF`) implies $A + B = A' + B'$ is defined in Coq as:

```
Global Instance addF_morphism : Proper (EqF ==> EqF ==> EqF) AddF.
```

With a view to apply the Gröbner basis method, we also proved that `F` is an integral domain. This would seem trivial, as any field is an integral domain, but we actually proved that the product of any two non-zero elements is non-zero even before we proved the associativity of the multiplication. Indeed, in order to prove this property, one needs to distinguish the cases where some products are null from the general case. Finally, we can prove we have a field:

```
Lemma fieldF : (field_theory OF OneF AddF MulF SubF OppF DivF InvF EqF).
```

Order

We proved that `F` is an ordered field. For convenience we proved it for two equivalent definitions. Namely, that one can define a positive cone on `F` or that `F` is equipped with a total order on `F` which is compatible with the operations. In [SST83], one can only find the proof based on the first definition. The characterization of the betweenness predicate in [SST83] is expressed in terms the order relation and not positivity. The second definition is therefore better suited for this proof than the first one. Nevertheless, for the proof relying on the second definition, we decided to prove the implication between the first and the second definition. Actually an algebraic proof, unlike geometric one, rarely includes tedious case distinctions.

In order to define the positive cone on `F`, we needed to define positivity. A point is said to be positive when it belongs to the half-line OE. `Out O A B` indicates that P belongs to line AB but does not belong to the line-segment $\overline{AB}$.

```
Definition Ps O E A := Out O A E.
```

A point is lower than another one if their difference is positive and the lower or equal relation is trivally defined. `Diff O E E' A B C` denotes that C is the difference of A and B wrt. O, E and E'.

```
Definition LtP O E E' A B := exists D, Diff O E E' B A D /\ Ps O E D.
Definition LeP O E E' A B := LtP O E E' A B \/ A = B.
```

The lower or equal relation is then shown to be a total order compatible with the arithmetic operations.

2 Algebraic Characterization of Geometric Predicates

It is well-known that having algebraic characterizations is very useful. Indeed, if we know a quantifier-free algebraic characterization for every geometric predicate present in the statement of a lemma, the proof can then be seen as verifying that the polynomial(s) corresponding to the conclusion of the lemma belong(s) to the radical of the ideal generated by the polynomials corresponding to the hypotheses of the lemma. Since there are computational ways (for example, the Gröbner basis method) to do this verification, these characterizations allow us to obtain proofs by computations. In this section, we present our formalization of the coordinatization of geometry and the method we employed to automate the proofs of algebraic characterizations.

2.1 Coordinatization of Geometry

To define coordinates, we first defined what is a proper orthonormal coordinate system (`Cs`) as an isosceles right triangle for which the length of the congruent sides equals the unity. `Per A B C` states that A, B and C form a right triangle.

```
Definition Cs O E S U1 U2 := O <> E /\ Cong O E S U1 /\ Cong O E S U2 /\ Per U1 S U2.
```

The predicate `Cd O E S U1 U2 P X Y` denotes that the point P has coordinates X and Y in the coordinate system `Cs O E S U1 U2`. `Cong_3 A B C D E F` designates that the triangles ABC and DEF are congruent and `Projp P Q A B` means that Q is the foot of the perpendicular from P to line AB.

```
Definition Cd O E S U1 U2 P X Y :=
  Cs O E S U1 U2 /\ Coplanar P S U1 U2 /\
  (exists PX, Projp P PX S U1 /\ Cong_3 O E X S U1 PX) /\
  (exists PY, Projp P PY S U2 /\ Cong_3 O E Y S U2 PY).
```

According to Borsuk and Szmielew [BS60], in planar neutral geometry it cannot be proved that the function associating coordinates to a given point is surjective. But assuming the parallel postulate, we can show that there is a one-to-one correspondence between the pairs of points on the ruler representing the coordinates and the points of the plane:

```
Lemma coordinates_of_point : forall O E S U1 U2 P,
  Cs O E S U1 U2 -> exists X, exists Y, Cd O E S U1 U2 P X Y.
Lemma point_of_coordinates : forall O E S U1 U2 X Y,
  Cs O E S U1 U2 ->
  Col O E X -> Col O E Y ->
  exists P, Cd O E S U1 U2 P X Y.
```

2.2 Algebraic Characterization of Congruence

We recall that Tarski's system of geometry has two primitive relations: congruence and betweenness. Following Schwabhäuser, we formalized the characterizations of these two geometric predicates. We have shown that the congruence predicate which is axiomatized is equivalent to the usual algebraic formula stating that the squares of the Euclidean distances are equal[3]:

[3] In the statement of this lemma, `coordinates_of_point_F` asserts the existence for any point of corresponding coordinates in $\mathtt{F}^2$ and the arithmetic symbols denote the operators or relations according to the usual notations.

```
Lemma characterization_of_congruence_F : forall A B C D,
  Cong A B C D <->
  let (Ac, HA) := coordinates_of_point_F A in let (Ax,Ay) := Ac in
  let (Bc, HB) := coordinates_of_point_F B in let (Bx,By) := Bc in
  let (Cc, HC) := coordinates_of_point_F C in let (Cx,Cy) := Cc in
  let (Dc, HD) := coordinates_of_point_F D in let (Dx,Dy) := Dc in
  (Ax - Bx) * (Ax - Bx) + (Ay - By) * (Ay - By) -
  ((Cx - Dx) * (Cx - Dx) + (Cy - Dy) * (Cy - Dy)) =F= 0.
```

The proof relies on Pythagoras' theorem (also known as Kou-Ku theorem). Note that we need a synthetic proof here. It is important to notice that we cannot use an algebraic proof because we are building the arithmetization of geometry. The statements for Pythagoras' theorem that have been proved previously[4] are theorems about vectors: the square of the norm of the sum of two orthogonal vectors is the sum of the squares of their norms. Here we provide the formalization of the proof of Pythagoras' theorem in a geometric context. `Length O E E' A B L` expresses that the length of segment AB can be represented by a point called L in the coordinate system O, E, E'.

```
Lemma pythagoras : forall O E E' A B C AC BC AB AC2 BC2 AB2,
  O<>E -> Per A C B ->
  Length O E E' A B AB -> Length O E E' A C AC -> Length O E E' B C BC ->
  Prod O E E' AC AC AC2 -> Prod O E E' BC BC BC2 -> Prod O E E' AB AB AB2 ->
  Sum  O E E' AC2 BC2 AB2.
```

Our proof of Pythagoras' theorem itself employs the intercept theorem (also known as Thales' theorem). These last two theorems represent crucial theorems in geometry, especially in the education. `Prodg O E E' A B C` designates the generalization of the multiplication which establishes as null the product of points for which `Ar2` does not hold.

```
Lemma thales : forall O E E' P A B C D A1 B1 C1 D1 AD,
  O<>E -> Col P A B -> Col P C D -> ~ Col P A C -> Pj A C B D ->
  Length O E E' P A A1 -> Length O E E' P B B1 ->
  Length O E E' P C C1 -> Length O E E' P D D1 ->
  Prodg O E E' A1 D1 AD ->
  Prodg O E E' C1 B1 AD.
```

2.3 Automated Proofs of the Algebraic Characterizations

In this subsection, we present our formalization of the translation of a geometric statement into algebra adopting the usual formulas. To obtain the algebraic characterizations of the others geometric predicates we adopted an original approach based on bootstrapping: we applied the Gröbner basis method to prove the algebraic characterizations of some geometric predicates which will be used in the proofs of the algebraic characterizations of other geometric predicates. The trick consists in a proper ordering of the proofs of the algebraic characterizations of geometric predicates relying on previously characterized predicates.

[4] A list of statements of previous formalizations of Pythagoras' theorem can be found on Freek Wiedijk webpage: http://www.cs.ru.nl/~freek/100/.

For example, we characterized parallelism in terms of midpoints and collinearity using the famous midpoint theorem that we previously proved synthetically[5]. `Midpoint M A B` means that M is the midpoint of A and B.

```
Lemma characterization_of_parallelism_F_aux : forall A B C D,
  Par A B C D <->
  A <> B /\ C <> D /\
  exists P, Midpoint C A P /\ exists Q, Midpoint Q B P /\ Col C D Q.
```

In the end, we only proved the characterizations of congruence, betweenness, equality and collinearity manually. The first three were already present in [SST83] and the last one was fairly straightfoward to formalize from the characterization of betweenness. However, it is impossible to obtain the characterizations of collinearity from the characterizations of betweenness by bootstrapping. Indeed, only a characterization of a geometric predicate G with subject $\overline{x}$ of the form $G(\overline{x}) \Leftrightarrow \bigwedge_{k=1}^{n} P_k(x) = 0 \wedge \bigwedge_{k=1}^{m} Q_k(x) \neq 0$ for some m and n and for some polynomials $\left(P_k\right)_{1\leq k\leq n}$ and $\left(Q_k\right)_{1\leq k\leq m}$ in the coordinates x of the points can be handled by either Wu's method or Gröbner basis method. Nevertheless, in theory, the characterization of the betweenness predicate could be employed by methods such as the quantifier elimination algorithm for real closed fields formalized by Cohen and Mahboubi in [CM12]. Then we obtained automatically the characterizations of midpoint, right triangles, parallelism and perpendicularity (in this order). The characterization of midpoint can be easily proven from the fact that if a point is equidistant from two points and collinear with them, either this point is their midpoint or these two points are equal. To obtain the characterization of right triangles, we exploited its definition which only involves midpoint and segment congruence. One should notice that the existential quantifier can be handled using a lemma asserting the existence of the symmetric of a point with respect to another one. To obtain the characterization of perpendicularity, we employed the characterizations of parallelism, equality and right triangle. The characterization of parallelism is used to produce the intersection point of the perpendicular lines which is needed as the definition of perpendicular presents an existential representing this point. Proving that the lines are not parallel allowed us to avoid producing the point of intersection by computing its coordinates. This was more convenient as these coordinates cannot be expressed as a polynomial but only as a rational function. Having a proof in Coq highlights the fact that the usual characterizations include degenerated cases. For example, the characterization of perpendicularity in Table 2 entails that lines AB and CD are non-degenerated.

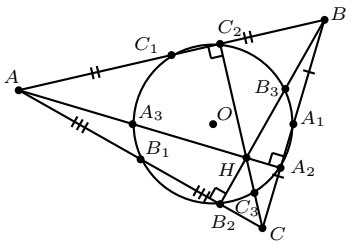
3 An Example of a Proof by Computation

To show that the arithmetization of geometry is useful in practice, we proved automatically one example applying the `nsatz` tactic developed by Grégoire, Pottier and Théry [GPT11]. This tactic corresponds to an implementation of the Gröbner basis method. Our example is the nine-point circle theorem which states that the following nine points are concyclic: the midpoints of each side of the triangle, the feet of each altitude and the midpoints of the line-segments from each vertex of the triangle to the orthocenter[6]:

[5] Note that it is important that we have a synthetic proof, because we cannot use an algebraic proof to obtain the characterization of parallelism since an algebraic proof would depend on the characterization of parallelism.

[6] In fact, many well-known points belong to this circle and this kind of properties can easily be proved formally using barycentric coordinates [NB16].

```
Lemma nine_point_circle:
forall A B C A1 B1 C1 A2 B2 C2 A3 B3 C3 H O,
  ~ Col A B C ->
  Col A B C2 -> Col B C A2 -> Col A C B2 ->
  Perp A B C C2 -> Perp B C A A2 -> Perp A C B B2 ->
  Perp A B C2 H -> Perp B C A2 H -> Perp A C B2 H ->
  Midpoint A3 A H -> Midpoint B3 B H -> Midpoint C3 C H ->
  Midpoint C1 A B -> Midpoint A1 B C -> Midpoint B1 C A ->
  Cong O A1 O B1 -> Cong O A1 O C1 ->
  Cong O A2 O A1 /\ Cong O B2 O A1 /\ Cong O C2 O A1 /\
  Cong O A3 O A1 /\ Cong O B3 O A1 /\ Cong O C3 O A1.
```



Compared to other automatic proofs using purely algebraic methods (either Wu's method or Gröbner basis method), the statement that we proved is syntactically the same, but the definitions and axioms are completely different. We did not prove a theorem about polynomials but a geometric statement. This proves that the nine-point circle theorem is true in *any* model of Tarski's Euclidean geometry axioms (without continuity) and not only in a specific one. We should remark that to obtain the proof automatically with the `nsatz` tactic, we had to clear the hypotheses that the lines appearing as arguments of the `Perp` predicate are well-defined. In theory, this should only render the problem more difficult to handle, but in practice the `nsatz` tactic can only solve the problem without these additional (not needed) assumptions. Non-degeneracy conditions represent an issue as often in geometry. Here we have an example of a theorem where they are superfluous but, while proving the characterizations of the geometric predicates, they were essential.

Moreover we should notice that Wu's or the Gröbner basis methods rely on the *Nullstellensatz* and are therefore only complete in an algebraically closed field. Hence, we had to pay attention to the characterization of equality. Indeed, as the field `F` is not algebraically closed one can prove that $x_A = x_B$ and $y_A = y_B$ is equivalent to $(x_A - x_B)^2 + (y_A - y_B)^2 = 0$ but this is not true in an algebraically closed one. Therefore, the tactic `nsatz` is unable to prove this equivalence.

Conclusion

In this paper, we produced the first *synthetic and formal* proofs of the intercept and Pythagoras' theorems. Furthermore, we obtained the arithmetization of geometry in the Coq proof assistant. This completes the formalization of the two-dimensional results contained in part one of [SST83]. To obtain the algebraic characterizations of some geometric predicates, we adopted an original approach based on bootstrapping. Our formalization of the arithmetization of geometry paves the way for the use of algebraic automated deduction methods in synthetic geometry within the Coq proof assistant.

Statistics

In a formalization effort, it is always interesting to know the value of the so-called De Bruijn factor. This factor is defined as the ratio between the size of the formalization and of the informal proof. This number is difficult to define precisely. Actually the length of the formalization depends heavily on the style of the author, and can vary in a single formalization. Likewise the length of the textbook proof fluctuates a lot depending on the author. For example, one can see below that Hilbert's description of the arithmetization of geometry is much shorter than

the one produced by Schwabhäuser. Moreover, during our formalization effort, we noticed that the De Bruijn factor is not constant in a single book. Indeed, the proofs in the first chapters of [SST83] are more precise than the last chapters which leave more room for implicit arguments and cases.

The GeoCoq library currently consists (as of Nov 2015) of about 2800 lemmas and more than 100klines of code. In the next table, we provide statistics about the part of the development described in this paper. In this table, we compare the size of the formalization to the number of pages in the two books [SST83] and [Hil60]. Our formalization follows mainly [SST83], however, we proved some additional results about the characterization of geometric predicates. The proofs of these characterizations can be found in [Wu94] but the lengths of the formal and informal proofs cannot really be compared because we proved these characterizations mainly automatically (see Sec. 2.3). Actually, we first proved the characterization of the midpoint predicate manually and then automatically and the script of the proof by computation was eight times shorter than our original one.

	Coq formalization		[SST83]		[Hil60]	
	#lemmas	#loc	Chapter	#pages	Chapter	#pages
Construction of an ordered field:			Ch. 14	17	Ch. V.3	8
Sum	100	4646				
Product	50	3310				
Order	38	1944				
Length of segments	39	3212	Ch. 15	3	-	-
Coordinates and some characterizations	33	2494	Ch. 16	9	-	-
Instantiation of `fieldF` and other characterizations	46	1206	-	-	-	-
Total		*16812*		*29*		*8*

Perspectives

A simple extension of this work would be to define square root geometrically following Descartes. This definition will require an axiom of continuity, such as line-circle intersection.

A more involved extension of this work consists in verifying the constructive version of the arithmetization of geometry introduced by Beeson in [Bee15]. This would necessitate to remove our axiom of decidability of equality and to replace it with Markov's principle for congruence and betweenness. We will have to reproduce Beeson's importation of the negative theorems present in [SST83] by implementing the Gödel double-negation interpretation and to formalize Beeson's proofs of existential theorems. Finally, to recover all ordered field properties, we will have to prove that Beeson's definitions of addition and multiplication are equivalent to the definitions presented in this paper, in the sense that they produce the same points (but without performing case distinctions).

Finally, another possible extension of this work is the formalization of the arithmetization of hyperbolic geometry. For this goal we could reuse the large portion of our formalization which is valid in neutral geometry.

Availability The Coq development is available here: `http://geocoq.github.io/GeoCoq/`.

References

[Bee13] Michael Beeson. Proof and computation in geometry. In Tetsuo Ida and Jacques Fleuriot, editors, *Automated Deduction in Geometry (ADG 2012)*, volume 7993 of *Springer Lecture Notes in Artificial Intelligence*, pages 1–30, Heidelberg, 2013. Springer.

[Bee15] Michael Beeson. A constructive version of Tarski's geometry. *Annals of Pure and Applied Logic*, 166(11):1199–1273, 2015.

[Bel93] John L Bell. Hilbert's ϵ-operator in intuitionistic type theories. *Mathematical Logic Quarterly*, 39(1):323–337, 1993.

[BHJ+15] Francisco Botana, Markus Hohenwarter, Predrag Janičić, Zoltán Kovács, Ivan Petrović, Tomás Recio, and Simon Weitzhofer. Automated Theorem Proving in GeoGebra: Current Achievements. *Journal of Automated Reasoning*, 55(1):39–59, 2015.

[Bir32] George D Birkhoff. A set of postulates for plane geometry, based on scale and protractor. *Annals of Mathematics*, pages 329–345, 1932.

[BN12] Gabriel Braun and Julien Narboux. From Tarski to Hilbert. In Tetsuo Ida and Jacques Fleuriot, editors, *Post-proceedings of Automated Deduction in Geometry 2012*, volume 7993 of *LNCS*, pages 89–109, Edinburgh, United Kingdom, September 2012. Jacques Fleuriot, Springer.

[BNS15a] Pierre Boutry, Julien Narboux, and Pascal Schreck. A reflexive tactic for automated generation of proofs of incidence to an affine variety. October 2015.

[BNS15b] Pierre Boutry, Julien Narboux, and Pascal Schreck. Parallel postulates and decidability of intersection of lines: a mechanized study within Tarski's system of geometry. submitted, July 2015.

[BNSB14a] Pierre Boutry, Julien Narboux, Pascal Schreck, and Gabriel Braun. A short note about case distinctions in Tarski's geometry. In Francisco Botana and Pedro Quaresma, editors, *Automated Deduction in Geometry 2014*, Proceedings of ADG 2014, pages 1–15, Coimbra, Portugal, July 2014.

[BNSB14b] Pierre Boutry, Julien Narboux, Pascal Schreck, and Gabriel Braun. Using small scale automation to improve both accessibility and readability of formal proofs in geometry. In Francisco Botana and Pedro Quaresma, editors, *Automated Deduction in Geometry 2014*, Proceedings of ADG 2014, pages 1–19, Coimbra, Portugal, July 2014.

[BS60] Karol Borsuk and Wanda Szmielew. *Foundations of geometry*. North-Holland, 1960.

[BW15] Michael Beeson and Larry Wos. Finding proofs in Tarskian geometry. *Journal of Automated Reasoning*, submitted, 2015.

[CGZ94] Shang-Ching Chou, Xiao-Shan Gao, and Jing-Zhong Zhang. *Machine Proofs in Geometry*. World Scientific, Singapore, 1994.

[CM12] Cyril Cohen and Assia Mahboubi. Formal proofs in real algebraic geometry: from ordered fields to quantifier elimination. *Logical Methods in Computer Science*, 8(1:02):1–40, February 2012.

[CW13] Xiaoyu Chen and Dongming Wang. Formalization and Specification of Geometric Knowledge Objects. *Mathematics in Computer Science*, 7(4):439–454, 2013.

[DDS00] Christophe Dehlinger, Jean-François Dufourd, and Pascal Schreck. Higher-Order Intuitionistic Formalization and Proofs in Hilbert's Elementary Geometry. In *Automated Deduction in Geometry*, pages 306–324, 2000.

[Des25] René Descartes. *La géométrie*. Open Court, Chicago, 1925. first published as an appendix to the *Discours de la Méthode* (1637).

[FT11] Laurent Fuchs and Laurent Théry. A Formalisation of Grassmann-Cayley Algebra in Coq. In *Post-proceedings of Automated Deduction in Geometry (ADG 2010)*, 2011.

[GNS11] Jean-David Genevaux, Julien Narboux, and Pascal Schreck. Formalization of Wu's simple method in Coq. In Jean-Pierre Jouannaud and Zhong Shao, editors, *CPP 2011 First*

International Conference on Certified Programs and Proofs, volume 7086 of *Lecture Notes in Computer Science*, pages 71–86, Kenting, Taiwan, December 2011. Springer-Verlag.

[GPT11] Benjamin Grégoire, Loïc Pottier, and Laurent Théry. Proof Certificates for Algebra and their Application to Automatic Geometry Theorem Proving. In *Post-proceedings of Automated Deduction in Geometry (ADG 2008)*, number 6701 in Lecture Notes in Artificial Intelligence, 2011.

[Gup65] Haragauri Narayan Gupta. *Contributions to the axiomatic foundations of geometry*. PhD thesis, University of California, Berkley, 1965.

[Hil60] David Hilbert. *Foundations of Geometry (Grundlagen der Geometrie)*. Open Court, La Salle, Illinois, 1960. Second English edition, translated from the tenth German edition by Leo Unger. Original publication date, 1899.

[JNQ12] Predrag Janicic, Julien Narboux, and Pedro Quaresma. The Area Method : a Recapitulation. *Journal of Automated Reasoning*, 48(4):489–532, 2012.

[Kle72] Felix C. Klein. *A comparative review of recent researches in geometry*. PhD thesis, 1872.

[MF03] Laura Meikle and Jacques Fleuriot. Formalizing Hilbert's Grundlagen in Isabelle/Isar. In *Theorem Proving in Higher Order Logics*, pages 319–334, 2003.

[Moi90] E.E. Moise. *Elementary Geometry from an Advanced Standpoint*. Addison-Wesley, 1990.

[MP91] Richard S Millman and George D Parker. *Geometry: a metric approach with models*. Springer Science & Business Media, 1991.

[MP15] Filip Marić and Danijela Petrović. Formalizing complex plane geometry. *Annals of Mathematics and Artificial Intelligence*, 74(3-4):271–308, 2015.

[MPPJ12] Filip Marić, Ivan Petrović, Danijela Petrović, and Predrag Janičić. Formalization and implementation of algebraic methods in geometry. In Pedro Quaresma and Ralph-Johan Back, editors, Proceedings First Workshop on *CTP Components for Educational Software,* Wrocław, Poland, 31th July 2011, volume 79 of *Electronic Proceedings in Theoretical Computer Science*, pages 63–81. Open Publishing Association, 2012.

[Nar04] Julien Narboux. A Decision Procedure for Geometry in Coq. In Slind Konrad, Bunker Annett, and Gopalakrishnan Ganesh, editors, *Proceedings of TPHOLs'2004*, volume 3223 of *Lecture Notes in Computer Science*. Springer-Verlag, 2004.

[NB16] Julien Narboux and David Braun. Towards A Certified Version of the Encyclopedia of Triangle Centers. In J. Rafael Sandra, Dongming Wang, and Jing Yang, editors, *Special Issue on Geometric Reasoning*, pages 1–17. Springer, 2016. to appear.

[RGA14] William Richter, Adam Grabowski, and Jesse Alama. Tarski geometry axioms. *Formalized Mathematics*, 22(**2**):167–176, 2014.

[SDNJ15] Sana Stojanović Durdević, Julien Narboux, and Predrag Janičić. Automated Generation of Machine Verifiable and Readable Proofs: A Case Study of Tarski's Geometry. *Annals of Mathematics and Artificial Intelligence*, page 25, 2015.

[Soz10] Matthieu Sozeau. A new look at generalized rewriting in type theory. *Journal of Formalized Reasoning*, 2(1):41–62, 2010.

[SST83] Wolfram Schwabhäuser, Wanda Szmielew, and Alfred Tarski. *Metamathematische Methoden in der Geometrie*. Springer-Verlag, Berlin, 1983.

[TG99] Alfred Tarski and Steven Givant. Tarski's system of geometry. *The bulletin of Symbolic Logic*, 5(2), June 1999.

[Wu94] Wen-Tsün Wu. *Mechanical Theorem Proving in Geometries*. Springer-Verlag, 1994.

A Definitions of the Geometric Predicates

Coq	Notation	Definition
`Bet A B C`	$A{-}B{-}C$	
`Cong A B C D`	$AB \equiv CD$	
`Cong_3 A B C A' B' C'`		$AB \equiv A'B' \wedge AC \equiv A'C' \wedge BC \equiv B'C'$
`Col A B C`	$\mathrm{Col}\ A\,B\,C$	$A{-}B{-}C \vee B{-}A{-}C \vee A{-}C{-}B$
`Out O A B`		$O \neq A \wedge O \neq B \wedge (O{-}A{-}B \vee O{-}B{-}A)$
`Midpoint M A B`	$A{-}M{-}B$	$A{-}M{-}B \wedge AM \equiv BM$
`Per A B C`	$\llcorner A\,B\,C$	$\exists C', C{-}B{-}C' \wedge AC \equiv AC'$
`Perp_in P A B C D`	$AB \underset{P}{\perp} CD$	$A \neq B \wedge C \neq D \wedge \mathrm{Col}\ P\,A\,B \wedge \mathrm{Col}\ P\,C\,D \wedge (\forall U\,V, \mathrm{Col}\ U\,A\,B \Rightarrow \mathrm{Col}\ V\,C\,D \Rightarrow \llcorner U\,P\,V)$
`Perp A B C D`	$AB \perp CD$	$\exists P, AB \underset{P}{\perp} CD$
`Coplanar A B C D`	$\mathrm{Cp}\ A\,B\,C\,D$	$\exists X, (\mathrm{Col}\ A\,B\,X \wedge \mathrm{Col}\ C\,D\,X) \vee (\mathrm{Col}\ A\,C\,X \wedge \mathrm{Col}\ B\,D\,X) \vee (\mathrm{Col}\ A\,D\,X \wedge \mathrm{Col}\ B\,C\,X))$
`Par_strict A B C D`	$AB \parallel_s CD$	$A \neq B \wedge C \neq D \wedge \mathrm{Cp}\ A\,B\,C\,D \wedge \neg\exists X, \mathrm{Col}\ X\,A\,B \wedge \mathrm{Col}\ X\,C\,D$
`Par A B C D`	$AB \parallel XY$	$AB \parallel_s CD \vee (A \neq B \wedge C \neq D \wedge \mathrm{Col}\ A\,C\,D \wedge \mathrm{Col}\ B\,C\,D)$
`Proj P Q A B X Y`		$A \neq B \wedge X \neq Y \wedge \neg AB \parallel XY \wedge \mathrm{Col}\ A\,B\,Q \wedge (PQ \parallel XY \vee P = Q)$
`Pj A B C D`		$AB \parallel CD \vee C = D$
`Opp O E E' A B`		$\mathtt{Sum}\ O\ E\ E'\ B\ A\ O$
`Diff O E E' A B C`		$\exists B', \mathtt{Opp}\ O\ E\ E'\ B\ B' \wedge \mathtt{Sum}\ O\ E\ E'\ A\ B'\ C$
`Projp P Q A B`		$A \neq B \wedge ((\mathrm{Col}\ A\,B\,Q \wedge AB \perp PQ) \vee (\mathrm{Col}\ A\,B\,P \wedge P = Q))$
`Length O E E' A B L`		$O \neq E \wedge \mathrm{Col}\ O\,E\,L \wedge \mathtt{LeP}\ O\ E\ E'\ O\ L \wedge OL \equiv AB$
`Prodg O E E' A B C`		$\mathtt{Prod}\ O\ E\ E'\ A\ B\ C \vee (\neg \mathtt{Ar2}\ O\ E\ E'\ A\ B\ B \wedge C = O)$

Table 1: Definitions of the geometric predicates.

B Algebraic Characterizations of Geometric Predicates

Geometric predicate	Algebraic Characterization
$A{-}B{-}C$	$\exists t, 0 \leq t \leq 1 \wedge \begin{array}{r} t(x_C - x_A) = x_B - x_A \\ t(y_C - y_A) = y_B - y_A \end{array} \wedge$
$A{-\!\!\!\cdot}I{-\!\!\!\cdot}B$	$\begin{array}{rcl} 2x_I - (x_A + x_B) & = & 0 \\ 2y_I - (y_A + y_B) & = & 0 \end{array} \wedge$
$\triangle A\,B\,C$	$(x_A - x_B)(x_B - x_C) + (y_A - y_B)(y_B - y_C) = 0$
$AB \perp CD$	$\begin{array}{rcl} (x_A - x_B)(y_C - y_D) - (y_A - y_B)(x_C - x_D) & = & 0 \wedge \\ (x_A - x_B)(x_A - x_B) + (y_A - y_B)(y_A - y_B) & \neq & 0 \wedge \\ (x_C - x_D)(x_C - x_D) + (y_C - y_D)(y_C - y_D) & \neq & 0 \end{array}$
$AB \parallel CD$	$\begin{array}{rcl} (x_A - x_B)(x_C - x_D) + (y_A - y_B)(y_C - y_C) & = & 0 \wedge \\ (x_A - x_B)(x_A - x_B) + (y_A - y_B)(y_A - y_B) & \neq & 0 \wedge \\ (x_C - x_D)(x_C - x_D) + (y_C - y_D)(y_C - y_D) & \neq & 0 \end{array}$

Table 2: Algebraic Characterizations of geometric predicates.

In this table we denoted by x_P the abscissa of a point P and by y_P its ordinate.