Automated Production of Readable Proofs for Theorems in Euclidian Geometry

Pedro Quaresma
Mathematics Department
University of Coimbra
Portugal

Predrag Janičić
Faculty of Mathematics
University of Belgrade
Serbia & Montenegro

January 19, 2006

The first approach to the automation of proofs in Geometry focused in synthetic proofs with the attempt to automate the traditional proof method (Gelernter 59, Coelho and Pereira 86, Greeno 79).

— controlling the search space
— guiding the program toward the right deduction.

The first approach to the automation of proofs in Geometry focused in synthetic proofs with the attempt to automate the traditional proof method (Gelernter 59, Coelho and Pereira 86, Greeno 79).

– controlling the search space
– guiding the program toward the right deduction.

Another approach was (is) given by the algebraic approaches (Tarski 30s, Kapur 86, Wang 95, Wu 00).

+ efficient provers
– don't reflect the constructive nature of the problems, are unrelated to any geometric method, and the proofs have only a yes/no conclusion

The first approach to the automation of proofs in Geometry focused in synthetic proofs with the attempt to automate the traditional proof method (Gelernter 59, Coelho and Pereira 86, Greeno 79).

– controlling the search space
– guiding the program toward the right deduction.

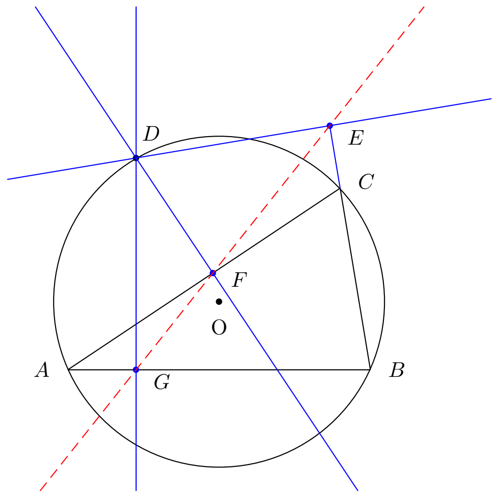
Another approach was (is) given by the algebraic approaches (Tarski 30s, Kapur 86, Wang 95, Wu 00).

+ efficient provers
– don't reflect the constructive nature of the problems, are unrelated to any geometric method, and the proofs have only a yes/no conclusion

In this talk we will speak, briefly, about this two approaches and then we will focus on the automation of the Area Method, a synthetic, efficient method for the automation of proofs in Euclidean Geometry (Zhang 95, Chou 96, Narboux 04).

A construction is expressed in terms of some algebraic quantities and then some property related to the construction is proved by algebraic methods.

Theorem 1 (Simson's theorem) *Let D be a point on the circumscribed circle (O) of triangle ABC. From D three perpendiculars are drawn to the three sides BC, CA, and AB of ΔABC. Let E, F and G be the three feet respectively. Show that E, F, and G are collinear.*



Demonstration

Let $A = (0,0)$, $B = (u_1, 0)$, $C = (u_2, u_3)$, $O = (x_2, x_1)$, $D = (x_3, x_4)$, $E = (x_5, x_4)$, $F = (x_7, x_6)$ and $G = (x_3, 0)$. Then the hypothesis equations are:

h_1	$=$	$2u_2x_2 + 2u_3x_1 - u_3^2 - u_2^2 = 0$	$OA \equiv OC$
h_2	$=$	$2u_1x_2 - u_1^2 = 0$	$OA \equiv OB$
h_3	$=$	$-x_3^2 + 2x_2x_3 + 2u_4x_1 - u_4^2 = 0$	$OA \equiv OD$
h_4	$=$	$u_3x_5 + (-u_2 + u_1)x_4 - u_1u_3 = 0$	Points E, B and C are collinear
h_5	$=$	$(u_2 - u_1)x_5 + u_3x_4 + (-u_2 + u_1)x_3 - u_3u_4 = 0$	$DE \perp BC$
h_6	$=$	$u_3x_7 - u_2x_6 = 0$	Points F, A and C are collinear
h_7	$=$	$u_2x_7 + u_3x_6 - u_2x_3 - u_3u_4 = 0$	$DF \perp AC$

The conclusion is:

$$g = x_4x_7 + (-x_5 + x_3)x_6 - x_3x_4 = 0 \qquad \text{Points } E, F \text{ and } G \text{ are collinear}$$

Now we can triangulate $h_1, h_2, \ldots, h_7$ $(\ldots)$

$$\begin{aligned} &\ldots \\ R_o &= prem(R_1, f_1, x_1) = 0 \end{aligned}$$

Since the final remainder R_0 is 0, by the remainder formula, we have proved Simson's theorem $(\ldots)$.

The algebraic approach proved already to be very successful — from the quantifier elimination method of Tarski, to the Wu's method and the Gröbner Basis method, many, very efficient, provers have been developed.

Algebraic Proofs

The algebraic approach proved already to be very successful — from the quantifier elimination method of Tarski, to the Wu's method and the Gröbner Basis method, many, very efficient, provers have been developed.

All these methods don't reflect the constructive nature of the problems, are unrelated to any geometric method, and the proofs have only a yes/no conclusion.

Synthetic Proofs

A visual proof

A traditional geometric proof reflects the constructive nature of the problem, uses geometric methods and it is human readable.

We begin by defining a class of geometry statements whose hypotheses can be described constructively and whose conclusions can be represented by polynomials in some geometry quantities, without any relation to a system of coordinates.

The Area Method - Class of Constructive Geometry Statements

We begin by defining a class of geometry statements whose hypotheses can be described constructively and whose conclusions can be represented by polynomials in some geometry quantities, without any relation to a system of coordinates.

Definition 1 (Class of Constructive Geometry Statements) *The class of* Constructive Geometry Statements, $\mathbf{C}$, *is the class of statements defined as follows. A statement in class* $\mathbf{C}$ *is a list* $S = (C_1, C_2, \ldots, C_n, G)$ *where* C_i *for* $1 \leqslant i \leqslant n$ *are constructions such that each* C_i *introduces a new point from the points introduced before; and* $G = (E_1, E_2)$ *where* E_1 *and* E_2 *are polynomials in geometric quantities of the points introduced by the* C_i *and* $E_1 = E_2$ *is the conclusion of the statement.*

The Area Method - Class of Constructive Geometry Statements

We begin by defining a class of geometry statements whose hypotheses can be described constructively and whose conclusions can be represented by polynomials in some geometry quantities, without any relation to a system of coordinates.

Definition 1 (Class of Constructive Geometry Statements) *The class of* Constructive Geometry Statements, $\mathbf{C}$, *is the class of statements defined as follows. A statement in class* $\mathbf{C}$ *is a list* $S = (C_1, C_2, \ldots, C_n, G)$ *where* C_i *for* $1 \leqslant i \leqslant n$ *are constructions such that each* C_i *introduces a new point from the points introduced before; and* $G = (E_1, E_2)$ *where* E_1 *and* E_2 *are polynomials in geometric quantities of the points introduced by the* C_i *and* $E_1 = E_2$ *is the conclusion of the statement.*

Let $S = (C_1, C_2, \ldots, C_n, (E_1, E_2))$ be a statement in $\mathbf{C}$. The ndg condition of S is the set of ndg conditions of the C_is plus the condition that the denominators of the length ratios in E_1 and E_2 are not equal to zero.

The Area Method - Class of Constructive Geometry Statements

We begin by defining a class of geometry statements whose hypotheses can be described constructively and whose conclusions can be represented by polynomials in some geometry quantities, without any relation to a system of coordinates.

Definition 1 (Class of Constructive Geometry Statements) *The class of* Constructive Geometry Statements, $\mathbf{C}$, *is the class of statements defined as follows. A statement in class* $\mathbf{C}$ *is a list* $S = (C_1, C_2, \ldots, C_n, G)$ *where* C_i *for* $1 \leqslant i \leqslant n$ *are constructions such that each* C_i *introduces a new point from the points introduced before; and* $G = (E_1, E_2)$ *where* E_1 *and* E_2 *are polynomials in geometric quantities of the points introduced by the* C_i *and* $E_1 = E_2$ *is the conclusion of the statement.*

Let $S = (C_1, C_2, \ldots, C_n, (E_1, E_2))$ be a statement in $\mathbf{C}$. The ndg condition of S is the set of ndg conditions of the C_is plus the condition that the denominators of the length ratios in E_1 and E_2 are not equal to zero.

This class covers a wide range of geometry theorems.

The Area Method - Basic Geometry Objects & Geometric Quantities

Points are the basic geometry objects, from which we can introduce two other *geometric objects*: **lines** and **circles**.

Points are the basic geometry objects, from which we can introduce two other *geometric objects*: **lines** and **circles**.
The geometric quantities are: *Ratio of Segments*, *Signed Areas*, and *Pythagoras Differences*.

The Area Method - Basic Geometry Objects & Geometric Quantities

Points are the basic geometry objects, from which we can introduce two other *geometric objects*: **lines** and **circles**.
The geometric quantities are: *Ratio of Segments*, *Signed Areas*, and *Pythagoras Differences*.

Definition 2 (Ratio of parallel lines) *Let $ABCD$ be a parallelogram and P, Q be two points on CD. We define the ratio of two parallel line segments as follows:*

$$\frac{\overline{PQ}}{\overline{AB}} = \frac{\overline{PQ}}{\overline{DC}}$$

The Area Method - Basic Geometry Objects & Geometric Quantities

Points are the basic geometry objects, from which we can introduce two other *geometric objects*: **lines** and **circles**.
The geometric quantities are: *Ratio of Segments*, *Signed Areas*, and *Pythagoras Differences*.

Definition 2 (Ratio of parallel lines) *Let $ABCD$ be a parallelogram and P, Q be two points on CD. We define the ratio of two parallel line segments as follows:*

$$\frac{\overline{PQ}}{\overline{AB}} = \frac{\overline{PQ}}{\overline{DC}}$$

Definition 3 (Signed Area) *The* signed area S_{ABC} *of triangle ABC is the usual area with a sign depending on the order of the vertices A, B, and C: if $A - B - C$ rotates counterclockwisely, S_{ABC} is positive, otherwise it is negative.*

The Area Method - Basic Geometry Objects & Geometric Quantities

Points are the basic geometry objects, from which we can introduce two other *geometric objects*: **lines** and **circles**.
The geometric quantities are: *Ratio of Segments*, *Signed Areas*, and *Pythagoras Differences*.

Definition 2 (Ratio of parallel lines) *Let $ABCD$ be a parallelogram and P, Q be two points on CD. We define the ratio of two parallel line segments as follows:*

$$\frac{\overline{PQ}}{\overline{AB}} = \frac{\overline{PQ}}{\overline{DC}}$$

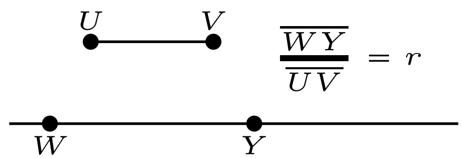
Definition 3 (Signed Area) *The* signed area *S_{ABC} of triangle ABC is the usual area with a sign depending on the order of the vertices A, B, and C: if $A - B - C$ rotates counterclockwisely, S_{ABC} is positive, otherwise it is negative.*

Definition 4 (Pythagoras difference) *For three points A, B, and C, the* Pythagoras difference *P_{ABC} is defined to be:*

$$P_{ABC} = \overline{AB}^2 + \overline{CB}^2 - \overline{AC}^2.$$

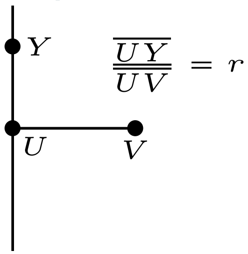
C1 – (POINT[S] Y1, . . . , Yn). Take arbitrary points $Y_1, \ldots, Y_n$ in the plane.

C7 – (PRATIO Y W U V r). Take a point Y on the line (PLINE W U V) such that

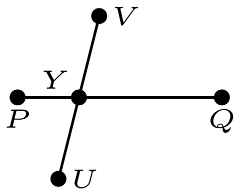


$$\overline{WY} = r\overline{UV}.$$

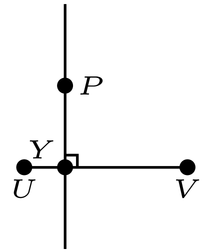
C8 – (TRATIO Y U V r). Take a point Y on line (TLINE U U V) such that



$$r = \frac{4S_{UVY}}{P_{UVU}} \left(= \frac{\overline{UY}}{\overline{UV}}\right).$$



C41 – (INTER Y (LINE U V) (LINE P Q)).

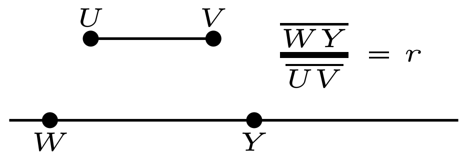


C42 – (FOOT Y P U V). The ndg condition is $U \neq V$.

The Area Method - minimal set of constructions

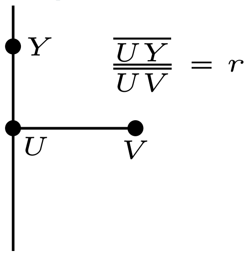
C1 – (POINT[S] Y1, . . . , Yn). Take arbitrary points $Y_1, \ldots, Y_n$ in the plane.

C7 – (PRATIO Y W U V r). Take a point Y on the line (PLINE W U V) such that

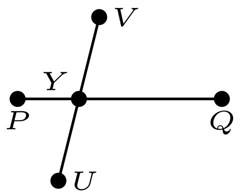


$$\overline{WY} = r\overline{UV}.$$

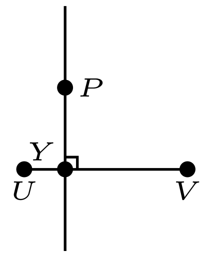
C8 – (TRATIO Y U V r). Take a point Y on line (TLINE U U V) such that



$$r = \frac{4S_{UVY}}{P_{UVU}} \left(= \frac{\overline{UY}}{\overline{UV}}\right).$$



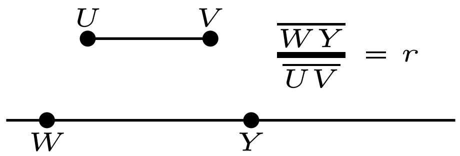
C41 – (INTER Y (LINE U V) (LINE P Q)).



C42 – (FOOT Y P U V). The ndg condition is $U \neq V$.

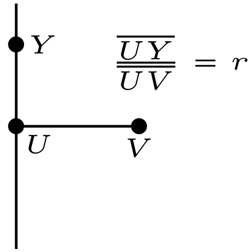
C1 – (POINT[S] Y1, . . . , Yn). Take arbitrary points $Y_1, \ldots, Y_n$ in the plane.

C7 – (PRATIO Y W U V r). Take a point Y on the line (PLINE W U V) such that

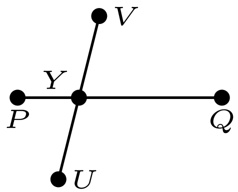


$\overline{WY} = r\overline{UV}$.

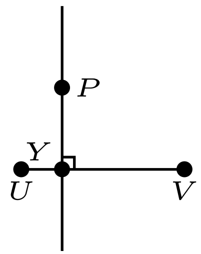
C8 – (TRATIO Y U V r). Take a point Y on line (TLINE U U V) such that



$r = \frac{4S_{UVY}}{P_{UVU}} (= \frac{\overline{UY}}{\overline{UV}})$.



C41 – (INTER Y (LINE U V) (LINE P Q)).

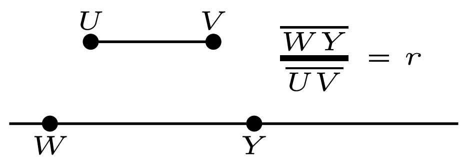


C42 – (FOOT Y P U V). The ndg condition is $U \neq V$.

The Area Method - minimal set of constructions

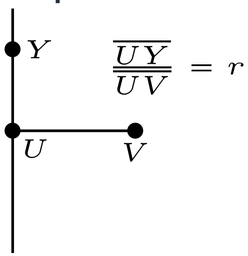
C1 – (POINT[S] Y1, . . . , Yn). Take arbitrary points $Y_1, \ldots, Y_n$ in the plane.

C7 – (PRATIO Y W U V r). Take a point Y on the line (PLINE W U V) such that

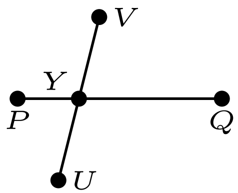


$\overline{WY} = r\overline{UV}$.

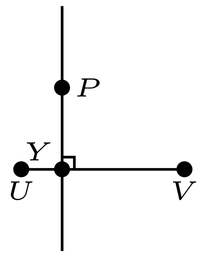
C8 – (TRATIO Y U V r). Take a point Y on line (TLINE U U V) such that



$r = \frac{4S_{UVY}}{P_{UVU}} (= \frac{\overline{UY}}{\overline{UV}})$.



C41 – (INTER Y (LINE U V) (LINE P Q)).

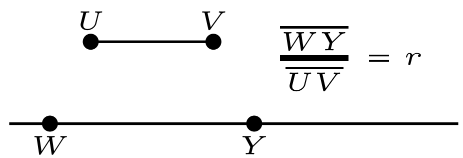


C42 – (FOOT Y P U V). The ndg condition is $U \neq V$.

The Area Method - minimal set of constructions

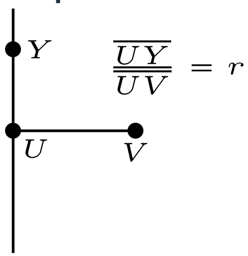
C1 – (POINT[S] Y1, . . . , Yn). Take arbitrary points $Y_1, \ldots, Y_n$ in the plane.

C7 – (PRATIO Y W U V r). Take a point Y on the line (PLINE W U V) such that

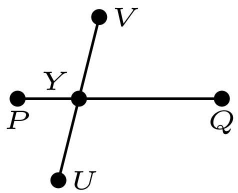


$\overline{WY} = r\overline{UV}$.

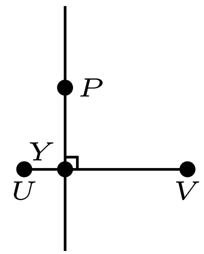
C8 – (TRATIO Y U V r). Take a point Y on line (TLINE U U V) such that



$r = \frac{4S_{UVY}}{P_{UVU}} (= \frac{\overline{UY}}{\overline{UV}})$.



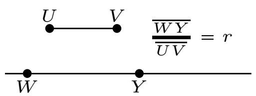
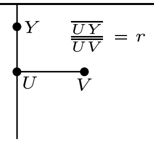
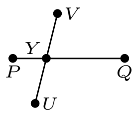
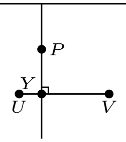
C41 – (INTER Y (LINE U V) (LINE P Q)).



C42 – (FOOT Y P U V). The ndg condition is $U \neq V$.

The Area Method - Elimination Procedures

The key step of the method is to *eliminate points from geometry quantities*. The point are introduced naturally and are eliminated from the conclusion in the reverse order. Considering only the minimal set of constructions: C1, C7, C8, C41, and C42 we need only to eliminate points introduced by four constructions from three kinds of geometry quantities.

		Construction	ndc	Elimination formulas				
				P_{AYB}	P_{ABY} P_{ABCY}	S_{ABY} S_{ABCY}	$\frac{\overline{AY}}{\overline{CD}}$	$\frac{\overline{AY}}{\overline{BY}}$
1	C7		$U \neq V$, if $r = \frac{r1}{r2}$ then $r2 \neq 0$	EL6	EL2		EL12	
2	C8		$U \neq V$, if $r = \frac{r1}{r2}$ then $r2 \neq 0$	EL9	EL8	EL7	EL13	
3	C41			EL5	EL3		EL10	EL1
4	C42		$U \neq V$	EL5	EL4		EL11	
				A	B	C	D	E

The Area Method — The Algorithm

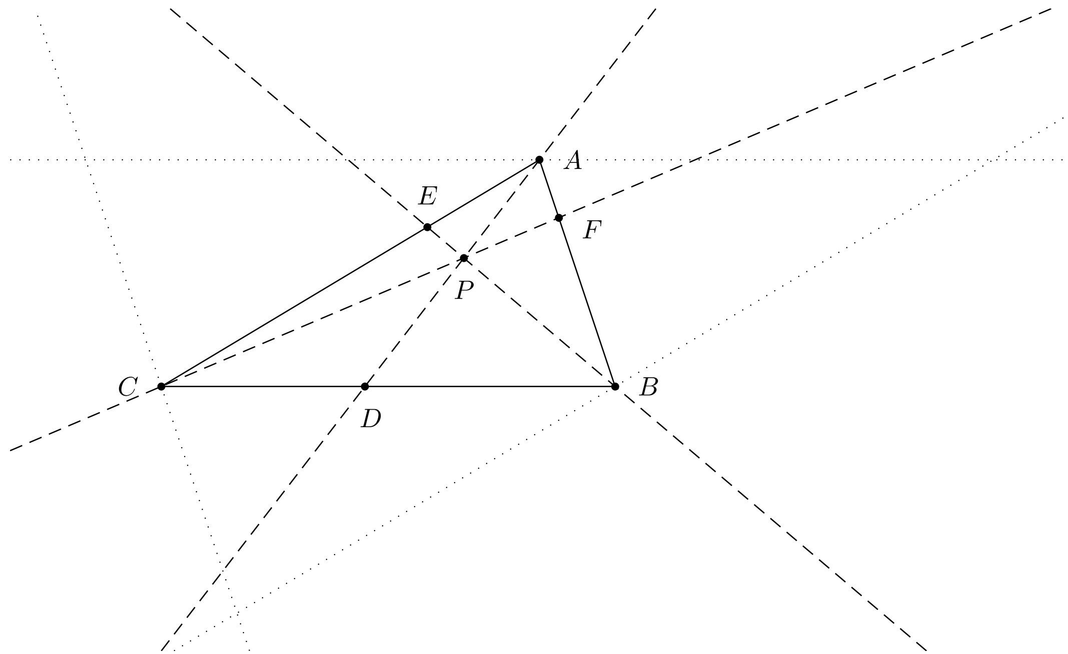
The automation of the area method by Chou et. al. gives us the possibility of developing efficient provers capable of producing short and readable proofs for many geometric theorems.

$\rightarrow$ $S = (C_1, C_2, \ldots, C_m, (E, F))$ is a statement in $\mathbf{C}$.
$\leftarrow$ The algorithm tells whether S is true, or not, and if it is true, produces a proof for S.

```
for (i=m;i==1;i--) {
  if (the ndg conditions of Ci is satisfied) exit;
  // Let G1,...,Gn be the geometric quantities in E and F
  for (j=1;j<=n,j++) {
    Hj<-eliminating the point introduced by construction Ci from Gj
    E <- E[Gj:=Hj]
    F <- F[Gj:=Hj]
  }
}
if (E==F) S <- true
else S<-false
```

Theorem 2 (Ceva's Theorem) *Let ΔABC be a triangle and P be any point in the plane. Let $D = AP \cap CB$, $E = BP \cap AC$, and $F = CP \cap AB$. Show that:*

$$\frac{\overline{AF}}{\overline{FB}} \cdot \frac{\overline{BD}}{\overline{DC}} \cdot \frac{\overline{CE}}{\overline{EA}} = 1$$



Demonstration

Prover actual output:

$$\left(\left(\frac{\overrightarrow{AF}}{\overrightarrow{FB}} \cdot \frac{\overrightarrow{BD}}{\overrightarrow{DC}}\right) \cdot \frac{\overrightarrow{CE}}{\overrightarrow{EA}}\right) = 1 \quad \text{by the statement} \quad (1)$$

$$\left(-1 \cdot \left(\frac{S_{APC}}{S_{BPC}} \cdot \left(\frac{\overrightarrow{BD}}{\overrightarrow{DC}} \cdot \frac{\overrightarrow{CE}}{\overrightarrow{EA}}\right)\right)\right) = 1 \quad \text{by Lemma 8 (point } F \text{ eliminated)} \quad (2)$$

$$\frac{\left(S_{APC} \cdot \left(\frac{\overrightarrow{BD}}{\overrightarrow{DC}} \cdot \frac{S_{CPB}}{S_{APB}}\right)\right)}{S_{BPC}} = 1 \quad \text{by Lemma 8 (point } E \text{ eliminated)} \quad (3)$$

$$\frac{\left(S_{APC} \cdot \frac{S_{BPA}}{S_{CPA}}\right)}{S_{APB}} = 1 \quad \text{by Lemma 8 (point } D \text{ eliminated)} \quad (4)$$

$$1 = 1 \quad \text{by algebraic simplifications} \quad (5)$$

Q.E.D.

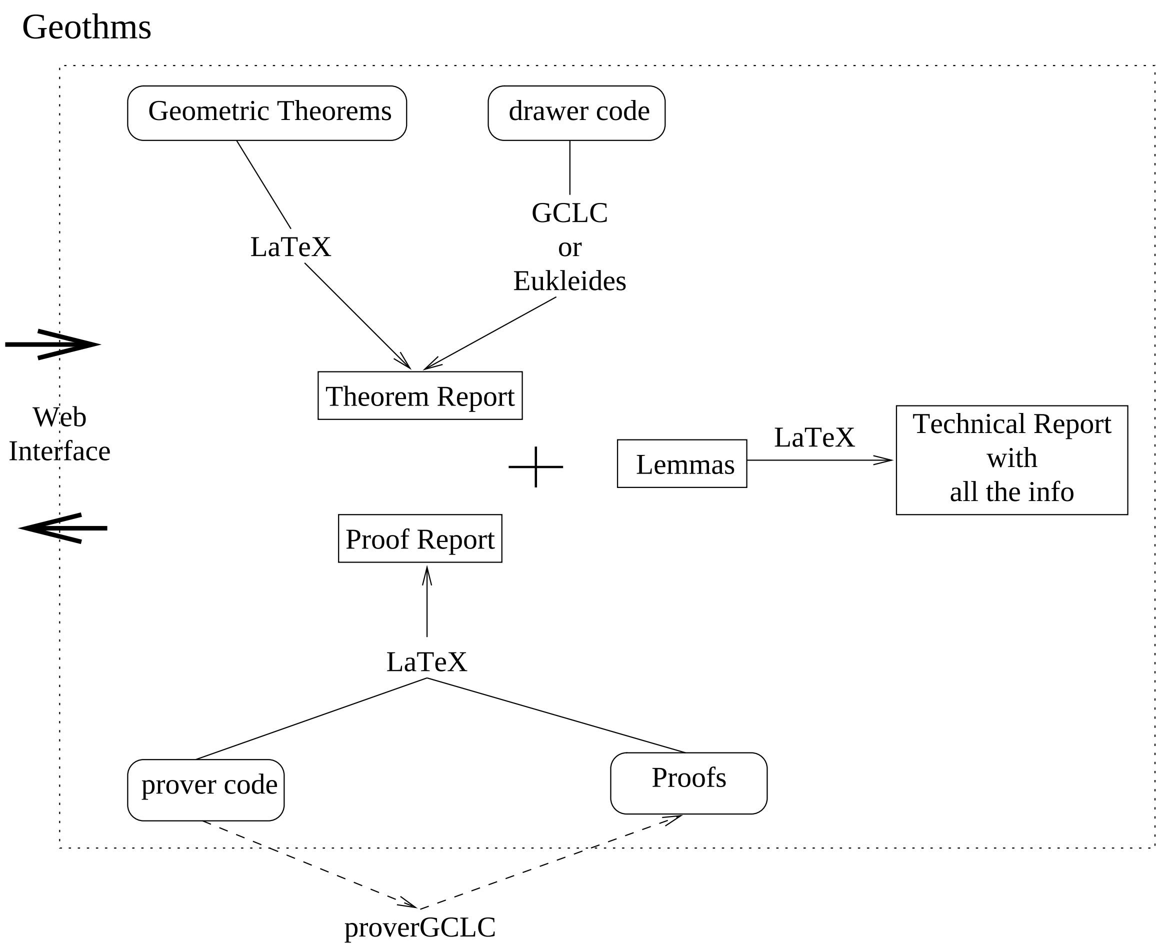
The theorem prover is based on Chou's area method, it produces traditional geometric, readable proofs. The proofs are expressed in terms of higher-level geometry lemmas and expression simplifications. Apart from required geometric elimination lemmas, the prover use a range of rewrite rules for simplification of expression involved.

The program can prove a range of non-trivial theorems, including theorems due to Ceva, Menelaus, Gauss, Pappus, Tales, etc.

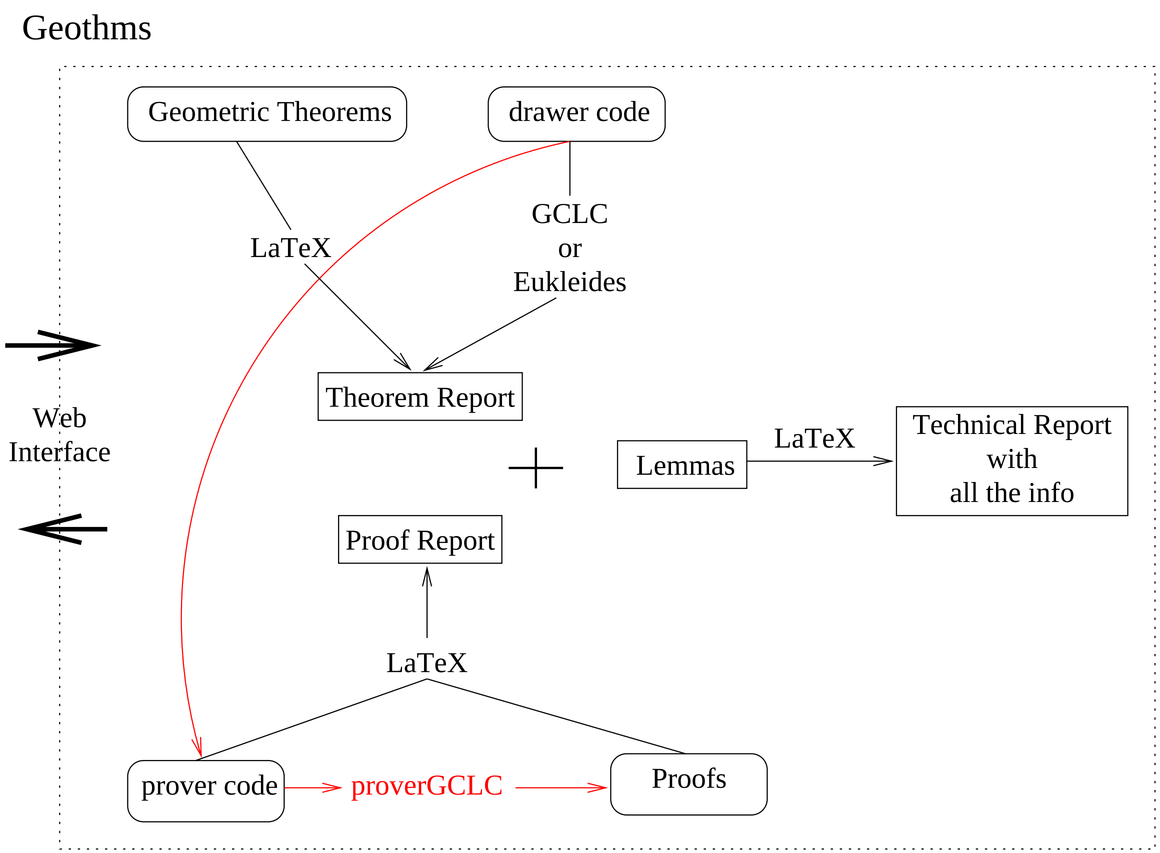
It is tightly integrated into GCLC. This means that one can use the prover to reason about a construction (i.e., about objects introduced in it), without any required adaptations required for the deduction process. Of course, only the conjecture itself has to be added.

It is implemented in C++ programming language.

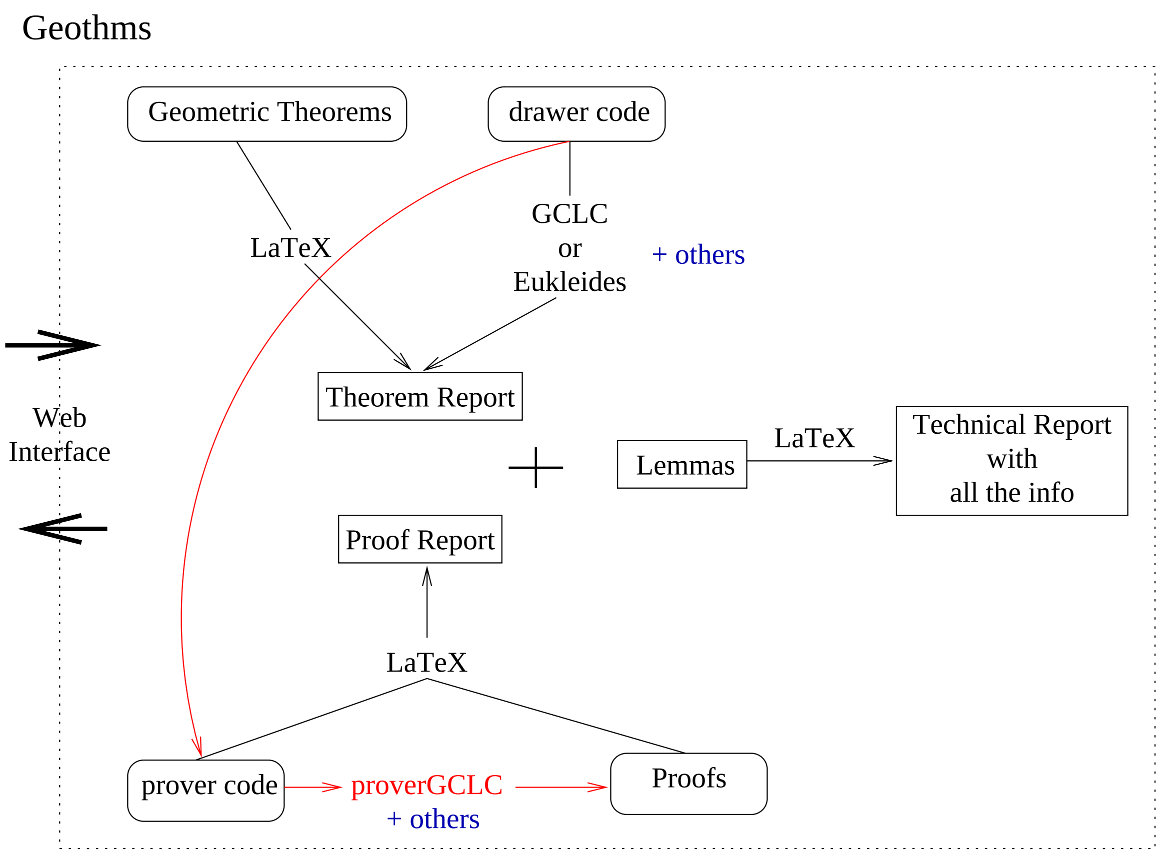
GeoThms — a data base of geometric theorems aiming to provide a common repository of theorems and proofs in the area of constructive problems in Euclidean Geometry (http://hilbert.mat.uc.pt/~geothms).



GeoThms — a data base of geometric theorems aiming to provide a common repository of theorems and proofs in the area of constructive problems in Euclidean Geometry (http://hilbert.mat.uc.pt/~geothms).



GeoThms — a data base of geometric theorems aiming to provide a common repository of theorems and proofs in the area of constructive problems in Euclidean Geometry (http://hilbert.mat.uc.pt/~geothms).



This theorem prover and this database, together with dynamic geometry programs, such as GCLC or Eukleides, will constitute a framework for describing geometry constructions, visualizing them, storing and searching them, proving properties about constructions, teaching and studying geometry, etc.