A Class of Geometry Statements of Constructive Type and Geometry Theorem Proving*

Shang-Ching Chou and Xiao-Shan Gao†

Department of Computer Science
The Wichita State University, Wichita, KS 67208, USA
e-mail: chou@wiz.wsu.ukans.edu

Abstract. This paper presents a method to generate non-degenerate conditions in geometric form for a class of geometry statements of constructive type, called Class C. We prove a mathematical theorem that in the irreducible case, the non-degenerate conditions generated by our method are *sufficient* for a geometry statement in Class C to be valid in *metric geometry.* About 400 among 600 theorems proved by our computer program are in Class C.

Keywords. Geometry theorem proving, Wu's method, non-degenerate condition, generally true, constructive geometry statement, Euclidean geometry, metric geometry, algebraically closed field.

*The work reported here was supported in part by the NSF Grants CCR–902362 and CCR–917870.

†The current and permanent address: Institute of Systems Science, Academia Sinica, Beijing.

1 Introduction

In the past decade highly successful algebraic methods for mechanical geometry theorem proving have been developed. This began with Wu's pioneering work in 1978 [10]. In the subsequent developments, hundreds of hard theorems in Euclidean and Non–Euclidean geometries have been proved by computer programs based on Wu's method [11], [12], [1], [2]. Inspired by the success of Wu's work, several groups also successfully applied the Gröbner basis (GB) method to prove the same class of geometry statements that Wu's method addresses [7], [8], [9]. In most of the current research, two different but related formulations for geometry statements have been considered.

Formulation (Approach) F1. Introducing parameters and the notion of "generally true" for a geometry statement. The present techniques can prove a statement to be generally true, at the same time giving nondegenerate conditions (usually in algebraic form) automatically.

Formulation (Approach) F2. Giving nondegenerate conditions in geometric form manually (or mechanically) at the beginning as a part of the hypotheses. Then the prover only needs to answer whether the conclusion follows the hypotheses *without adding* any other conditions.

The method originally developed by Wu is for Formulation F1. One of the advantages of Formulation F1 is that non-degenerate conditions can be taken care of without explicitly specifying them. Formulation F1 gives a clear insight into the nature of a geometric statement: if a geometry statement is proved to be generally false, then it cannot be valid no matter how many additional non-degenerate conditions are added so long as the hypotheses keep consistent.

On the other hand, how to translate algebraic nondegenerate conditions into their geometric forms is also an interesting and important topic. This is the aim of this paper. In this paper we presents a method for generating non-degenerate conditions in *geometric form* for a class of geometry statements, called Class C, when using Formulation 1. We prove a mathematical theorem stating that non-degenerate conditions produced by our method are complete in *metric geometry*, i.e., if the statement is not valid under these non-degenerate conditions, then it is generally false.

In Section 2, we discuss difficulties for finding non-degenerate conditions. In Section 3, we present our method for generating *geometric* non-degenerate conditions for Class C. In section 4, we prove that under certain conditions, this method is complete.

2 Difficulties with Non-Degenerate Conditions

Theorems in geometry textbooks often implicitly assume some non-degenerate conditions that are necessary for the theorems to be valid. Finding all non-degenerate conditions sufficient for a geometry statement to be true may be difficult in many cases. The situation becomes more complicated when we use Wu's method (or the GB method) for proving theorems, because these methods are complete only for metric geometry, not for Euclidean geometry. The following are two examples.

Figure 1 Figure 2

Example (2.1). (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 $\triangle ABC$. Let E, F and G be the three feet respectively. Show that E, F and G are collinear (Figure 1).

The obvious non-degenerate condition for this statement seems to be "A, B and C are not collinear". Indeed, in Euclidean geometry, Simson's theorem is valid under this condition. However, if we try to prove Simson's theorem under this condition with Wu's method or the GB method according to Formulation F2 (i.e., without adding any other conditions), then the statement cannot be confirmed. The reason for this phenomenon is that Wu's method (or the GB) method is complete only for metric geometry, not for Euclidean geometry. Here we will not give a detailed account of metric geometry (in Wu's sense), the reader can find the discussion in [3], [12]. We only mention that the theory of metric geometry has many models, among which are Euclidean geometry $\mathbf{R}^2$, unordered metric geometries (e.g., complex geometry $\mathbf{C}^2$), etc.

If we want to decide whether Simson's theorem is a theorem of the theory of metric geometry, then the following additional non-degenerate condition is necessary:

(2.1.1) $$AB, BC, \text{ and } CA \text{ are not isotropic}$$

An isotropic line is a line perpendicular to itself; it does not exist in Euclidean geometry. However, there exist isotropic lines in other models of the theory of metric geometry (e.g., $\mathbf{C}^2$). The result in using Wu's method shows that Simson's theorem can be proved in unordered geometries (i.e., without using the axioms of order) under (2.1.1); however, if we drop (2.1.1) from the hypotheses, then the axioms of order is necessary. Unless we have a clear understanding of the nature of the method and the statements to be proved, it is very hard to come up with non-degenerate conditions such as (2.1.1). This paper gives a satisfactory solution to these problems.

Even in Euclidean geometry, there are some nondegenerate conditions hard to find as illustrated by the following example.

Example (2.2). (the Butterfly Theorem) A, B, C and D are four points on circle (O). E is the intersection of AC and BD. Through E draw a line perpendicular to OE, meeting AD at F and BC at G. Show $FE \equiv GE$ (Figure 2).

Here we need a necessary non-degenerate condition that EO is not perpendicular to AD. The necessity of this condition is hard to perceive.

In the next section, we shall classify a class of geometry statements of constructive type (Class C) and present a mechanical method for producing sufficient number of non-degenerate conditions in geometric form for a statement in this class. In section 4, we shall prove our method is complete.

3 A Class of Geometric Statements of Constructive Type

3.1 Definition of Class C

Most theorems in elementary geometry can be described in a constructive way: given a certain arbitrary points, lines, circles and points on these circles and lines, new points, lines and circles are constructed step by step using geometric constructions such as taking the intersection of two lines, an intersection of a line and a circle, or an intersection of two circles. In this subsection, we use the natural language to give a definition of such a statement. In Section 3.3, we will give the precise formula of such a statement using geometric predicates.

First, let us give "circle" a formal definition. A *circle* h is a pair of a point O and a segment (AB): $h = (O, (AB))$, where A and B are two points. Two circles $(O, (AB))$ and $(P, (CD))$ are equal if and only if $O = P$ and congruent(A, B, C, D) (for the definition of "congruent", see 3.2). O is called the center of the circle and (AB) the radius. A point P is on circle $(O, (AB))$ if congruent(O, P, A, B).

Let Π be a finite set of points. We say line l is constructed *directly* from Π if

(i) The line l joins two points A and B in Π. We denote it by $l = L(AB)$; or

(ii) The line l passes through one point C in Π and parallel to a line joining two points A and B in Π. We denote it by $l = P(C, AB)$; or

(iii) The line l passes through one point C in Π and perpendicular to a line joining two points A and B in Π. We denote it by $l = T(C, AB)$; or

(iv) The line l is the perpendicular–bisector of AB with A and B in Π. We denote it by $l = B(AB)$.

A line l constructed directly from Π is *well defined* if the two points A and B mentioned above are distinct.

Likewise, we say a circle $c = (O, (AB))$ is constructed directly from Π if points O, A and B are in Π. The lines and circles constructed directly from Π are said to be *in* Π, for brevity.

Definition. A geometry statement is of constructive type or in Class C if the points, lines, and circles in the statement can be constructed in a definite prescribed manner using the following ten constructions, assuming Π to be the set of points already constructed so far:

Construction 1. Taking an arbitrary point.

Construction 2. Drawing an arbitrary line. This can be reduced to taking two arbitrary points.

Construction 3. Drawing an arbitrary circle. This can be also reduced to taking two or three arbitrary points.

Construction 4. Drawing an arbitrary line passing through a point in Π. This can be reduced to taking an arbitrary point.

Construction 5. Drawing an arbitrary circle knowing its center in Π. This can be also reduced to taking one or two arbitrary points.

Construction 6. Taking an arbitrary point on a line in Π.

Construction 7. Taking an arbitrary point on a circle in Π.

Construction 8. Taking the intersection of two lines in Π.

Construction 9. Taking an intersection of a line and a circle in Π.

Construction 10. Taking an intersection of two circles in Π.

The conclusion is a certain (equality) relation among the points thus constructed.

In the actual prover [2], [6], we have included more constructions such as taking midpoints and constructions involving angle congruence, the radical axis of two circles, taking a point on a circle knowing three points on the circle, etc.

Example (3.1). Simson's theorem can be specified as a statement in Class C by the following *construction sequence*:

Points A, B, and C are arbitrarily chosen,	construction 1
$O = B(AB) \cap B(AC)$,	construction 8
D is on circle $(O, (OA))$,	construction 7
$E = T(D, BC) \cap L(BC)$,	construction 8
$F = T(D, AC) \cap L(AC)$,	construction 8
$G = T(D, AB) \cap L(AB)$,	construction 8

with conclusion collinear(E, F, G).

The Butterfly theorem can be specified as a statement in Class C by the following construction sequence:

O and A are arbitrarily chosen,	construction 1
B is on $(O, (OA))$,	construction 7
C is on $(O, (OA))$,	construction 7
D is on $(O, (OA))$,	construction 7
$E = L(AC) \cap L(BD)$,	construction 8
$F = L(AD) \cap T(E, OE)$,	construction 8
$G = L(EF) \cap L(BC)$,	construction 8

with conclusion midpoint(F, E, G).

In the above examples, we use a *construction sequence* to express a statement in Class C. We will soon present an algorithm for generating non-degenerate conditions for a statement in Class C, knowing its construction sequence. Before presenting the algorithm, we first specify what exact geometric predicates we use.

3.2 The Basic Predicates

In order to describe the logical formula of a statement in Class C, we need four basic (non-logical) predicates: "collinear(A, B, C)", "parallel(A, B, C, D)", "perpendicular(A, B, C, D)", "congruent(A, B, C, D)".[1] We should emphasize that these predicates do include degenerate cases. To be more precise, let $A = (x_1, y_1)$ $B = (x_2, y_2)$, $C = (x_3, y_3)$ and $D = (x_4, y_4)$.

(1) Predicate "collinear(A, B, C)" means that points A, B and C are on the same line; they are not necessarily distinct. Its corresponding algebraic equation is

$$(x_1 - x_2)(y_2 - y_3) - (x_2 - x_3)(y_1 - y_2) = 0.$$

[1] In our actual prover [2], [3], there are many other predicates. For the complete list of all these predicates and their algebraic equations see pp.97–99 of [3].

(2) Predicate "parallel(A, B, C, D)" means that

$$[(A = B) \vee (C = D) \vee (A, B, C, D \text{ are on the same line}) \vee (AB \parallel CD)].$$

Its algebraic equation is

$$(x_1 - x_2)(y_3 - y_4) - (x_3 - x_4)(y_1 - y_2) = 0.$$

(3) Predicate "perpendicular(A, B, C, D)" means that $[(A = B) \vee (C = D) \vee (AB \perp CD)]$. Its algebraic equation is

$$(x_1 - x_2)(x_3 - x_4) + (y_1 - y_2)(y_3 - y_4) = 0.$$

For convenience, we define a new predicate "isotropic(A, B)" to be perpendicular(A, B, A, B), which means $A = B$ or $L(AB)$ is an isotropic line.

(4) Predicate "congruent(A, B, C, D)" means [(isotropic(A, B) $\wedge$ isotropic(C, D)) $\vee$ (AB is congruent to CD)]. Its algebraic equation is

$$(x_1 - x_2)^2 + (y_1 - y_2)^2 - (x_3 - x_4)^2 - (y_3 - y_4)^2 = 0.$$

There are many advantages of using the above predicates. Each of the above predicates corresponds to only one equation, thus its negation corresponds to only one inequation. E.g., $\neg$parallel(A, B, C, D) is "$(A \neq B) \wedge (C \neq D) \wedge (A, B, C, D$ are not on the same line$) \wedge \neg(AB \parallel CD)$". Its corresponding inequation is

$$(x_1 - x_2)(y_3 - y_4) - (x_3 - x_4)(y_1 - y_2) \neq 0,$$

which is *the exact non-degenerate condition we want for intersecting two lines* AB and CD: they have only one common point. Note that this condition implies the condition $(A \neq B \wedge C \neq D)$.

3.3 Mechanical Generation of Non-Degenerate Conditions for Class C

For a statement in Class C, we can generate non-degenerate conditions following the construction sequence step by step. Suppose we have already generated a set of non-degenerate conditions DS under the previous constructions. Let HS be the set of the equation hypotheses under the previous constructions, and Π be the set of points constructed so far. The next construction is one of the ten constructions in Section 3.1. First we add the point(s) to be constructed to the set Π. Since the first five constructions are reduced to taking arbitrary points, nothing is added to HS or DS. Thus we assume the next construction is one of constructions 6–10. We use abbreviations coll(), perp(), para() and cong() for predicates collinear(), perpendicular(), parallel() and congruent(), respectively.

Construction 6. Taking an arbitrary point D on a line l in Π. There are four kinds of lines in Π.

Case 6.1. $l = L(AB)$.

$HS := \{\mathrm{coll}(A, B, D)\} \cup HS$; $DS := \{A \neq B\} \cup DS$.

Case 6.2. $l = P(C, AB)$.

$HS := \{\mathrm{para}(A, B, C, D)\} \cup HS$; $DS := \{A \neq B\} \cup DS$.

Case 6.3. $l = T(C, AB)$.

$HS := \{\mathrm{perp}(A, B, C, D)\} \cup HS$; $DS := \{A \neq B\} \cup DS$.

Case 6.4. $l = B(AB)$.

$HS := \{\mathrm{cong}(A, D, B, D)\} \cup HS$; $DS := \{A \neq B\} \cup DS$.

Construction 7. Taking an arbitrary point A on a circle $(B, (CD))$ in Π.

$HS := \{\mathrm{cong}(A, B, C, D)\} \cup HS$.

Construction 8. Taking the intersection I of two lines in Π.

Since there are four types of lines in Π, there are 10 types of intersections: types LL, LP, LT, LB, PP, PT, PB, TT, TB, and BB.

Let the two lines be given by the following equations:

$$\begin{aligned} l_1: &\quad a_1x + b_1y + c_1 = 0, \\ l_2: &\quad a_2x + b_2y + c_2 = 0. \end{aligned}$$

The elegance of our approach is that for all 10 types of intersections, the only non-degenerate condition in algebraic form is $\Delta = a_1b_2 - a_2b_1 \neq 0$.

Case 8.1. Type LL: $I = L(AB) \cap L(CD)$.

$HS := \{\mathrm{coll}(A, B, I), \mathrm{coll}(C, D, I)\} \cup HS$; $DS := \{\neg\mathrm{para}(A, B, C, D)\} \cup DS$.

Note that $\neg\mathrm{para}(A, B, C, D)$ implies $A \neq B$ and $C \neq D$.

Case 8.2. Type LP: $= L(AB) \cap P(E, CD)$.

$HS := \{\mathrm{coll}(A, B, I), \mathrm{para}(C, D, E, I)\} \cup HS$; $DS := \{\neg\mathrm{para}(A, B, C, D)\} \cup DS$. In the special case,

Case 8.2.1. If $B = D$, DS becomes $DS := \{\neg\mathrm{coll}(A, B, C)\} \cup DS$.

Case 8.3. Type LT: $I = L(AB) \cap T(E, CD)$.

$HS := \{\mathrm{coll}(A, B, I), \mathrm{perp}(C, D, E, I)\} \cup HS$; $DS := \{\neg\mathrm{perp}(A, B, C, D)\} \cup DS$. In the special case,

Case 8.3.1. (The foot from E to AB) $A = C$ and $B = D$. $DS := \{\neg\mathrm{isotropic}(A, B)\} \cup DS$.

Case 8.4. Type LB: $I = L(AB) \cap B(CD)$.

$HS := \{\mathrm{coll}(A, B, I), \mathrm{cong}(I, C, I, D)\} \cup HS$; $DS := \{\neg\mathrm{perp}(A, B, C, D)\} \cup DS$. In the special case,

Case 8.4.1. $A = C$ and $B = D$. $DS := \{\neg\mathrm{isotropic}(A, B)\} \cup DS$.[2]

Case 8.5. Type PP: $I = P(E, AB) \cap P(F, CD)$.

$HS := \{\mathrm{para}(A, B, E, I), \mathrm{para}(C, D, F, I)\} \cup HS$; $DS := \{\neg\mathrm{para}(A, B, C, D)\} \cup DS$. In the special case,

Case 8.5.1. $B = D$. $DS := \{\neg\mathrm{coll}(A, B, C)\} \cup DS$.

[2] This is one of the two ways to specify the midpoint.

Case 8.6. Type PT: $I = P(E, AB) \cap T(F, CD)$.

$HS := \{\text{para}(A, B, E, I), \text{perp}(C, D, F, I)\} \cup HS$; $DS := \{\neg\text{perp}(A, B, C, D)\} \cup DS$.

Case 8.7. Type PB: $I = P(E, AB) \cap B(CD)$.

$HS := \{\text{para}(A, B, E, I), \text{cong}(I, C, I, D)\} \cup HS$; $DS := \{\neg\text{perp}(A, B, C, D)\} \cup DS$.

Case 8.8. Type TT: $I = T(E, AB) \cap T(F, CD)$.

$HS := \{\text{perp}(A, B, E, I), \text{perp}(C, D, F, I)\} \cup HS$; $DS := \{\neg\text{para}(A, B, C, D)\} \cup DS$. In the special case,

Case 8.8.1. $B = D$. $DS := \{\neg\text{coll}(A, B, C)\} \cup DS$.

Case 8.9. Type TB: $I = T(E, AB) \cap B(CD)$.

$HS := \{\text{perp}(A, B, E, I), \text{cong}(I, C, I, D)\} \cup HS$; $DS := \{\neg\text{para}(A, B, C, D)\} \cup DS$. In the special case,

Case 8.9.1. $B = D$. $DS := \{\neg\text{coll}(A, B, C)\} \cup DS$.

Case 8.10. Type BB: $I = B(AB) \cap B(CD)$.

$HS := \{\text{perp}(I, A, I, B), \text{cong}(I, C, I, D)\} \cup HS$; $DS := \{\neg\text{para}(A, B, C, D)\} \cup DS$. In the special case,

Case 8.10.1. $B = D$. $DS := \{\neg\text{coll}(A, B, C)\} \cup DS$.

Construction 9. Taking an intersection Q of a line and a circle in Π. Let the line be $L(AB)$, or $P(C, AB)$, or $T(C, AB)$, or $B(AB)$, the circle be $(O, (DE))$. $DS := \{\neg\text{isotropic}(A, B)\} \cup DS$.

If $Q = L(AB) \cap (O, (DE))$, then $HS := \{\text{coll}(A, B, Q), \text{cong}(O, Q, D, E)\} \cup HS$.

If $Q = P(C, AB) \cap (O, (DE))$, then $HS := \{\text{para}(A, B, C, Q), \text{cong}(O, Q, D, E)\} \cup HS$.

If $Q = T(C, AB) \cap (O, (DE))$, then $HS := \{\text{perp}(A, B, C, Q), \text{cong}(O, Q, D, E)\} \cup HS$.

If $Q = B(AB) \cap (O, (DE))$, then $HS := \{\text{cong}(Q, A, Q, B), \text{cong}(O, Q, D, E)\} \cup HS$.

Case 9.1. In the special case when one of the intersections, say S, of the circle and the line is already in Π. $DS := \{\neg\text{isotropic}(A, B), S \neq Q\} \cup DS$.

Construction 10. Taking an intersection Q of two circles in Π. Let the two circles be $(O, (AB))$ and $(P, (CD))$.

$HS := \{\text{cong}(O, Q, A, B), \text{cong}(P, Q, C, D)\} \cup HS$; $DS = \{\neg\text{isotropic}(O, P)\} \cup DS$. In the special case,

Case 10.1. One of the intersections is already in Π, say, S. $DS := \{\neg\text{isotropic}(O, P), S \neq Q\} \cup DS$.

Repeating the above steps until every construction is processed, finally we have two parts for the hypotheses: one is $HS = \{H_1, ..., H_r\}$, called the *equation part* of the hypotheses; the other is $DS = \{\neg D_1, ..., \neg D_s\}$, called the *inequation part* of the hypotheses representing non-degenerate conditions of the statement.

Let C be the conclusion of the statement, whose algebraic form is a polynomial equation in the coordinates of the points in Π. Then the exact statement is[3]

$$\forall P \in \Pi(HS \wedge DS \Rightarrow C). \tag{3.2}$$

Thus according to our translation, we can denote a statement S in Class C by (HS, DS, C).

3.4 Examples

Example (3.4). (Simson's Theorem and the Butterfly theorem). According to the construction sequence of Simson's theorem in Example (3.1), the non-degenerate conditions (the inequation part of the hypotheses) are

$DS_s =$	
$\neg$collinear(A, B, C),	Case 8.10.1
$\neg$isotropic(AB),	Case 8.3.1
$\neg$isotropic(AC),	Case 8.3.1
$\neg$isotropic(BC).	Case 8.3.1

The equation part of the hypotheses is

perpendicular(A, B, D, G),	
perpendicular(A, C, D, F),	
perpendicular(B, C, D, E),	
collinear(A, B, G),	
collinear(A, C, F),	HS_s
collinear(B, C, E),	
congruent(O, A, O, B),	
congruent(O, A, O, C),	
congruent(O, A, O, D).	

Nondegenerate conditions DS_s are exactly what we discussed in Section 2. Then the exact statement of Simson's theorem according to the constructions in (3.1) is:

$$\forall A \cdots \forall G[HS_s \wedge DS_s \Rightarrow \text{collinear}(E, F, G)]. \tag{3.5}$$

Note that for the same theorem, the construction sequence is usually not unique. Different construction sequences lead to different non-degenerate conditions. For example, we have at least 8 essentially different construction sequences for Simson's theorem. However, for all different construction sequences, *the equation part of the hypotheses is always the same*; in this example, it is always HS_s.

[3]Depending on the context, HS can also denote the conjunction of its elements, i.e., $HS = H_1 \wedge \cdots \wedge H_r$. The same convention is for DS and other sets of geometric conditions.

The non-degenerate conditions for the Butterfly theorem according to the construction sequence in (3.1) are

$$
\begin{array}{ll}
DS_b = & \\
\neg\text{parallel}(A, C, B, D), & \text{Case 8.1} \\
\neg\text{perpendicular}(A, D, O, E), & \text{Case 8.3} \\
\neg\text{parallel}(E, F, B, C). & \text{Case 8.1}
\end{array}
$$

The equation part of the hypotheses is

$$
\left.
\begin{array}{l}
\text{congruent}(O, A, O, B), \\
\text{congruent}(O, A, O, C), \\
\text{congruent}(O, A, O, D), \\
\text{collinear}(A, E, C), \\
\text{collinear}(B, E, D), \\
\text{perpendicular}(O, E, E, F), \\
\text{collinear}(E, F, G), \\
\text{collinear}(F, A, D), \\
\text{collinear}(G, B, C).
\end{array}
\right. \qquad HS_b
$$

The exact statement of the Butterfly theorem according to the constructions in (3.1) is:

$$\forall A \cdots \forall G[HS_b \wedge DS_b \Rightarrow \text{midpoint}(F, E, G)]. \tag{3.6}$$

The results in Section 4 and 5 show that either (3.5) (or (3.6)) is a theorem in the theory of metric geometry, or it cannot be a theorem no matter how many additional non-degenerate conditions are added as long as the hypotheses keep consistent.

4 The Completeness of Non-Degenerate Conditions for Metric Geometry

The completeness of our method for generating non-degenerate conditions DS can be stated as the following theorem.

Theorem (4.1). For an irreducible (to be defined later) statement in Class C, our mechanically generated non-degenerate conditions are sufficient for the statement to be valid in the theory of metric geometry. To be more precise, let S = (HS, DS, C) be a statement in Class C, where $HS = \{H_1, ..., H_r\}$ is the equation part of the hypotheses, $DS = \{\neg D_1, ..., \neg D_s\}$ is the inequation part of the hypotheses, and C is the conclusion. Let Π be the set of all points involved in S. If S is irreducible (to be defined later) and the formula

$$\forall P \in \Pi(HS \wedge DS \Rightarrow C), \tag{4.2}$$

is not valid in a model Ω of the theory of metric geometry whose associated field F_Ω is *algebraically closed*, then (4.2) cannot be a theorem in Ω by adding any set of additional conditions $\neg D_{s+1}, ..., \neg D_t$ as long as they keep the consistency with the hypotheses, where each D_i is a geometric condition whose algebraic form is an equation. The consistency means that

$$\forall P \in \Pi(HS \wedge DS \Rightarrow D_i), \tag{4.3}$$

are not valid in Ω, for $i = s+1, ..., t$. .QED.

Now we are proving Theorem (4.1). Our final goal is to prove Theorem (4.8) which is the algebraic form of Theorem (4.1). The proof here was originally in [2].

We use the algebraic approach. Following Hilbert and Wu, we use two kinds of variables: the parameters u_j and the dependent variables x_k. *Our proof also provides a method to choose the parameters u, the dependent variables x, the variable order in x, and a method to decide whether (4.2) is valid in Ω.*

After adopting an appropriate coordinate system, each point P in the statement S corresponds to a pair of coordinates: $P = (x_p, y_p)$. We introduce new parameters u, dependent variables x, and equations according to the steps of constructions. Under the previous constructions, suppose we have already introduced parameters $u_1, ..., u_{j-1}$, dependent variables $x_1, ..., x_{k-1}$, and the equations $h_1 = 0, ..., h_{k-1} = 0$ corresponding to a part of hypotheses $\{H_1, ..., H_{k-1}\}$, and an ascending chain of the form:

$$\begin{array}{l} f_1(u_1, \ldots, u_{j-1}, x_1) \\ f_2(u_1, \ldots, u_{j-1}, x_1, x_2) \\ \ldots \\ f_{k-1}(u_1, \ldots, u_{j-1}, x_1, \ldots, x_{k-1}). \end{array} \tag{4.4}$$

Furthermore, we assume the ascending chain (in week sense) $f_1, ..., f_{k-1}$ is irreducible.[4] Let Π be the set of points constructed so far. First we add the next point(s) to be constructed to Π. Since Constructions 1–5 introduce only arbitrarily chosen points, we only need to assign new parameters to the coordinates of the points. E.g., for construction 1 (taking any point A), we can let $A = (u_j, u_{j+1})$. Thus we assume that the next construction is one of Constructions 6–10.

Construction 6. Taking an arbitrary point D on a line l in Π. Let the corresponding condition in HS be H_k. Let the line equation $h_k = 0$ for l, which is the algebraic form of H_k, be

$$ax + by + c = 0.$$

[4]For the definition of ascending chains, see [11], [3], or [4]. The ascending chain $f_1, ..., f_{k-1}$ is irreducible if $f(u, x_1, ..., x_i)$ is irreducible in the field $\mathbf{Q}(u)[x_1, ..., x_i]/(f_1, ..., f_{i-1})$, for $i = 1, ..., k-1$. Here here $(f_1, ..., f_{i-1})$ is the ideal generated by $f_1, ..., f_{i-1}$.

Here a, b, and c are polynomials in coordinates of the previously constructed points. E.g., if $l = T(C, AB)$ and $C = (x_2, y_2)$, $A = (x_3, y_3)$, $B = (x_4, y_4)$, then the equation is:

$$(x - x_2)(x_3 - x_4) + (y - y_2)(y_3 - y_4) = 0,$$

i.e., $a = x_3 - x_4$, $b = y_3 - y_4$ and $c = -x_2(x_3 - x_4) - y_2(y_3 - y_4)$.

Our first step is to check whether $R_a = prem(a; f_1, ..., f_{k-1})$ and $R_b = prem(b; f_1, ..., f_{k-1})$ are zero. (Here *prem* denotes the successive pseudo remainder of a polynomial by an ascending chain, see [4] for details.)

Case 6.1. $R_a = 0$ and $R_b = 0$. Then the line l is not well defined. We detect the inconsistency of the hypotheses with adding $A \neq B$. In that case, we either can say that the hypotheses do not satisfy the dimensionality constraint required by Formulation F1 (see p.47 [3]), or it is a theorem according to Formulation 2 because of the inconsistency of the hypotheses.

Case 6.2. One of R_a and R_b, say R_b, is zero. We can let $D = (x_k, u_j)$ and have a new equation:

$$f_k = ax_k + bu_j + c = 0,$$

where u_j and x_k are the new parameter and dependent variables introduced. We have a new irreducible ascending chain $f_1, ..., f_k$. Then the condition $a \neq 0$ is equivalent to that the line l is well defined ($A \neq B$).

Case 6.3. Both R_a and R_b are not zero. We can do the same as in case 6.2. The only difference is that the condition $a \neq 0$ is no longer equivalent to $A \neq B$. But $(a \neq 0 \vee b \neq 0)$ is equivalent to $A \neq B$. We will come back to this in the proof of (4.8).

Construction 7. Taking an arbitrary point A on a circle $(B, (CD))$ in Π. Let $A = (x_k, u_j)$, $B = (x_2, y_2)$, $C = (x_3, y_3)$ and $D = (x_4, y_4)$. Then the algebraic form of the corresponding hypothesis H_k in HS is the equation

$$h_k = (x_k - x_2)^2 + (u_j - y_2)^2 - (x_3 - x_4)^2 - (y_3 - y_4)^2 = 0.$$

Our next step is to check whether CD is isotropic, i.e., whether $R = prem((x_3 - x_4)^2 + (y_3 - y_4)^2; f_1, ..., f_{k-1})$ is zero. If $R \neq 0$, then let $f_k = h_k$, and $f_1, ..., f_k$ is irreducible (see [2]). We always assume this is the case.

Construction 8. Taking the intersection I of two lines l_1 and l_2 in Π. We have two corresponding hypotheses H_k and H_{k+1} in HS, whose algebraic forms are two equations for lines l_1 and l_2:

$$h_k = a_1 x + b_1 y + c_1 = 0,$$
$$h_{k+1} = a_2 x + b_2 y + c_2 = 0.$$

First, we check whether $R = prem(\Delta; f_1, ..., f_{k-1})$ is zero, where $\Delta = a_1 b_2 - a_2 b_1$. Note that $\Delta \neq 0$ is the algebraic form of the non-degenerate condition generated in cases 8.1–8.10 of Section 3.3.

Case 8.1. $R = 0$. Then adding condition $\Delta \neq 0$ causes inconsistency with the previous constructions.

Case 8.2. $R \neq 0$. Letting $I = (x_{k+1}, x_k)$, then we have two new equations:

$$f_k = \Delta x_k + d = 0,$$
$$f_{k+1} = \Delta x_{k+1} + e = 0.$$

where $d = a_1c_2 - a_2c_1$ and $e = b_2c_1 - b_1c_2$. We have an irreducible ascending chain $f_1, ..., f_{k+1}$.

Construction 9. Taking an intersection Q of a line l and a circle c in Π. We have two corresponding hypotheses H_k and H_{k+1} in HS whose corresponding algebraic forms are the equations for the circle c and the line l:

$$h_k = y^2 + x^2 + by + ax + c = 0$$
$$h_{k+1} = b_1y + a_1x + c_1 = 0.$$

First, we check whether $R = prem(a_1^2 + b_1^2; f_1, ..., f_{k-1})$ is zero.

Case 9.1. $R = 0$. Then the hypothesis $HS \wedge DS$ is inconsistent.

Case 9.2. $R \neq 0$. One of $R_a = prem(a_1; f_1, ..., f_{k-1})$, $R_b = prem(b_1; f_1, ..., f_{k-1})$ is zero, say, R_b. (They cannot be both zero, otherwise R would be zero). Then $a_1 \neq 0$ means that the line l is well defined. We introduce two dependent variables x_k, x_{k+1} and let $Q = (x_{k+1}, x_k)$. Eliminating y from equation h_k, we have

$$f_k = (b_1^2 + a_1^2)x_k^2 + (2a_1c_1 + ab_1^2 - a_1bb_1)x_k + (c_1^2 - bb_1c_1 + b_1^2c) = 0,$$
$$f_{k+1} = a_1x_{k+1} + b_1x_k + c_1 = 0.$$

Now we have an ascending chain (in weak sense) $f_1, ..., f_{k+1}$. We can check whether f_1, ..., f_{k+1} is irreducible using the algorithm introduced in [1] and implemented in our prover. *If it is reducible, generally it is still open whether non-degenerate conditions DS are sufficient.* In the statement of Theorem (4.1), we assume $f_1, ..., f_{k+1}$ is irreducible.

Case 9.3. $R \neq 0$, and both R_a and R_b are non-zero. We can do the same as in Case 9.2. The only difference is that $(a_1 \neq 0 \vee b_1 \neq 0)$ is the condition that the line l is well defined. We will come back to this condition in the proof of Theorem (4.8).

Construction 10. Taking an intersection Q of two circles c_1 and c_2 in Π. We have two corresponding hypotheses H_k and H_{k+1} in HS whose corresponding algebraic forms are the equations for the circles c_1 and c_2:

$$h_k = y^2 + x^2 + by + ax + c = 0,$$
$$h_{k+1} = y^2 + x^2 + ey + dx + j = 0.$$

We can replace h_{k+1} by $h_k - h_{k+1}$. Then $h_{k+1} = 0$ is a line equation:

$$h_{k+1} = b_1 y + a_1 x + c_1 = 0,$$

where $a_1 = a - d,\ b_1 = b - e,\ c_1 = c - j$. This is the line l joining the two intersection points of c_1 and c_2, if they intersect (this is generally the case if the field associated with the geometry Ω is algebraically closed.) But line l exists in Ω even if c_1 and c_2 do not have common points in Ω. This is the radical axis of the two circles. Note that l is non-isotropic iff the line joining the two centers is non-isotropic. Now we check whether $R = prem(a_1^2 + b_1^2; f_1, ..., f_{k-1})$ is zero.

Case 10.1. $R = 0$. This means the radical is isotropic, hence the line OP joining the two centers is isotropic and the hypothesis $HS \wedge DS$ is inconsistent.

Case 10.2. $R \neq 0$, and one of $R_a = prem(a_1; f_1, ..., f_{k-1})$ and $R_b = prem(b_1; f_1, ..., f_{k-1})$ is zero. Then we have exactly the same situation as in case 9.2.

Case 10.3. $R \neq 0$, and both R_a and R_b are not zero. Then we have exactly the same situation as in case 9.3.

Repeating this process until we complete all constructions. Finally, we have an irreducible ascending chain:

$$\begin{aligned} & f_1(u_1, \ldots, u_d, x_1) \\ & f_2(u_1, \ldots, u_d, x_1, x_2) \\ & \qquad \ldots \\ & f_r(u_1, \ldots, u_d, x_1, \ldots, x_r). \end{aligned} \tag{4.5}$$

This is the *definition* of a statement in Class C to be *irreducible* in the statement of Theorem (4.1). Now we want to ask whether the formula, i.e., the exact statement of S

$$\forall P \in \Pi[(H_1 \wedge \cdots \wedge H_r \wedge \neg D_1 \wedge \cdots \wedge \neg D_s) \Rightarrow C] \tag{4.6}$$

is valid in Ω, or in its equivalent algebraic form, whether the formula

$$\forall \in ux[(h_1 = 0 \wedge \cdots \wedge h_r = 0 \wedge d_1 \neq 0 \wedge \cdots \wedge d_s \neq 0) \Rightarrow c = 0] \tag{4.7}$$

is valid in F_Ω, where $h_1, ..., h_r$, and c are the polynomials corresponding to $H_1, ..., H_r$ and C, respectively, $d_1, ...d_s$ are polynomials corresponding to $D_1, ..., D_s$.

Remark. Each $\neg D_i$ is one of the negations of the four predicates: collinear, parallel, perpendicular, and point equal. The algebraic form for each of the first three is a polynomial equation. For the last one, it is also an equation if we introduce a new variable z (see [5]).

We have the following theorem which is the algebraic form of Theorem (4.1).

Theorem (4.8). For a statement S $= (HS, DS, C)$ in Class C, let the ascending chain $ASC = f_1, ..., f_r$ of the form of (4.5), obtained by the above procedure, be irreducible and $I_i = lc(f_i)$ be the initial of f_i. If F_Ω is algebraically closed, then the following conditions are equivalent:

(1) $prem(c; ASC) = 0$;
(2) The formula

$$\forall ux \in F_\Omega[(f_1 = 0 \wedge \cdots \wedge f_r = 0 \wedge I_1 \neq 0 \wedge \cdots \wedge I_r \neq 0) \Rightarrow c = 0] \tag{4.9}$$

is valid in F_Ω;
(3) Formula (4.7) is valid in F_Ω;
(4) Formula (4.6) is valid in Ω;
(5) $prem(d \cdot c; ASC) = 0$ for any polynomial d with $prem(d; ASC) \neq 0$.

Proof. (1) $\Rightarrow$ (2). Suppose $R = prem(c; ASC) = 0$. Since we have the remainder formula (p.13 of [3]):

$$I_1^{s_1} \cdots I_r^{s_r} c = Q_1 f_1 + \cdots + Q_r f_r + R,$$

and $R = 0$, formula (4.9) is valid.

(2) $\Rightarrow$ (1). Since F_Ω is algebraically closed and the ascending chain ASC is irreducible, (1) follows from (2) by Theorem (3.7) on p.30 in [3].

(2) $\Rightarrow$ (3). Let J be the set $\{I_i \mid I_i \text{ is not a constant}, i = 1, ..., r\}$. Let $N = \{d_1, ..., d_s\}$. Note that $N \subset J$. We want to show that those I_k in J but not in N can be removed in (4.9). Such an I_k can be only the following three cases:

Case 1: Case 6.3. In this case $I_k = a$. We can let $A = (u_j, x_k)$ and $f'_k = bx_k + au_j + c = 0$. The ascending chain $ASC' = f_1, ..., f_{k-1}, f'_k, f_{k+1}, ..., f_r$ is also irreducible. By Lemma (A1.1) [5], $prem(c; ASC) = 0$ if and only if $prem(c; ASC') = 0$. Thus we only need the condition $(a \neq 0 \vee b \neq 0)$. This is equivalent to one of the conditions $d_i \neq 0$.

Case 2: Case 9.3. In this case $I_k = a_1$. Using the same technique as in Case 1, we can come to the conclusion that $I_k \neq 0$ can be replaced by a weaker condition $(a_1 \neq 0 \vee b_1 \neq 0)$, which is implied by $a_1^2 + b_1^2 \neq 0$, i.e., by the condition that line l is non-isotropic.

The details work as follows. Assuming the order $x_{k+1} < x_k$, we have an ascending chain: $ASC' = f_1, ...f_{k-1}, f'_k, f'_{k+1}, f_{k+2}, ..., f_r$, where

$$f'_k = (b_1^2+a_1^2)x_{k+1}^2+(2b_1c_1-aa_1b_1+a_1^2b)x_{k+1}+(c_1^2-aa_1c_1+a_1^2c) = 0,$$
$$f'_{k+1} = b_1x_k + a_1x_{k+1} + c_1 = 0.$$

The discriminate of f'_k is $-a^2\delta$ and that of f_k is $-b^2\delta$, where $\delta = 4c_1^2-4bb_1c_1-4aa_1c_1 + 4b_1^2c + 4a_1^2c - a^2b_1^2 + 2aa_1bb_1 - a_1^2b^2$. Thus f'_k is irreducible over $F = \mathbf{Q}(u)[x]/(f_1, ..., f_{k-1})$ iff f_k is irreducible over F. Hence f'_k is irreducible. Since $prem(f_i; ASC') = 0$ and $prem(f'_i; ASC) = 0$ (for $i = k, k+1$), $prem(c; ASC) =$

0 iff $prem(c; ASC') = 0$ by Lemma (A1.1) [5]. Thus we only need the condition $(a_1 \neq 0 \vee b_1 \neq 0)$ which is implied by $d_i = a_1^2 + b_1^2 \neq 0$.

Case 3: Case 10.3. The same as in Case 2.

Thus the formula $\forall ux[(f_1 = 0 \wedge \cdots \wedge f_r \wedge d_1 \neq 0 \wedge \cdots \wedge d_s \neq 0) \Rightarrow c = 0]$ is valid. Since $(h_1 = 0 \wedge \cdots \wedge h_r = 0) \Rightarrow (f_1 = 0 \wedge \cdots \wedge f_r = 0)$, (3) follows from (2).

(3) $\Rightarrow$ (2). By the remainder formula, $(f_1 = 0 \wedge \cdots \wedge f_r = 0 \wedge I_1 \neq 0 \wedge \cdots \wedge I_r \neq 0) \Rightarrow h_i = 0$. Also $I_1 \neq 0 \wedge \cdots \wedge I_r \neq 0 \Rightarrow d_i \neq 0$. Thus (4.9) follows from (4.7).

(3) $\Leftrightarrow$ (4). Since (4.7) is the algebraic form of (4.6), (3) and (4) are equivalent.

(1) $\Leftrightarrow$ (5). Since ASC is irreducible, (1) and (5) are equivalent. .QED.

Remark. Interpreted in other way, (5) is the completeness of the non-degenerate conditions $\neg D_i$. Suppose $\neg D$ is another non-degenerate condition whose algebraic form is $d \neq 0$. If $(HS \wedge DS) \Rightarrow C$ is not valid in Ω, then $prem(c; ASC) \neq 0$. By (5), $prem(d \cdot c; ASC) \neq 0$, i.e., $(HS \wedge DS \wedge \neg D) \Rightarrow C$ is still not valid.

References

[1] S.C. Chou, "Proving Elementary Geometry Theorems Using Wu's Algorithm",in *Automated Theorem Proving: After 25 years*, Ed. By W.W. Bledsoe and D. Loveland, AMS Contemporary Mathematics Series **29** (1984), 243-286.

[2] S.C. Chou, "Proving and Discovering Theorems in Elementary Geometries Using Wu's Method", PhD Thesis, Department of Mathematics, University of Texas, Austin (1985).

[3] S.C. Chou, *Mechanical Geometry Theorem Proving,* D. Reidel Publishing Company, Dordrecht, Netherlands, 1988.

[4] S.C. Chou and X. S. Gao, "Ritt-Wu's Decomposition Algorithm and Geometry Theorem Proving", in *Proceedings of CADE–10, Lecture Notes in AI*, Vol.449, 207–220, 1990.

[5] S.C. Chou and X.S. Gao, "A Class of Geometry Statements of Constructive Type and Geometry Theorem Proving", TR-89-37, Computer Sciences Department, The University of Texas at Austin, November, 1989.

[6] S.C. Chou, "A Geometry Theorem Prover for Macintoshes", TR-91-8, Computer Sciences Department, The University of Texas at Austin, March, 1991.

[7] S.C. Chou and W.F. Schelter, "Proving Geometry Theorems with Rewrite Rules", *Journal of Automated Reasoning,* **2(4)** (1986), 253–273.

[8] D. Kapur, "Geometry Theorem Proving Using Hilbert's Nullstellensatz", in Proceedings of the 1986 Symposium on Symbolic and Algebraic Computation, 202-208.

[9] B. Kutzler and S. Stifter, "Automated Geometry Theorem Proving Using Buchberger's Algorithm", in Proceedings of the 1986 Symposium on Symbolic and Algebraic Computation, 209-214.

[10] Wu Wen-tsün, "On the Decision Problem and the Mechanization of Theorem Proving in Elementary Geometry", *Scientia Sinica* **21** (1978), 157-179.

[11] Wu Wen-tsün, "Basic Principles of Mechanical Theorem Proving in Geometries", *J. of Sys. Sci. and Math. Sci.* **4(3)**, 1984, 207-235, republished in *Journal of Automated Reasoning* **2(4)** (1986), 221-252.

[12] Wu Wen-tsün,*Basic Principles of Mechanical Theorem Proving in Geometries,* (in Chinese) Peking 1984.