

A CLASS OF GEOMETRY STATEMENTS OF CONSTRUCTIVE TYPE AND GEOMETRY THEOREM PROVING*

Shang-Ching Chou and Xiao-Shan Gao[†]

Department of Computer Sciences
University of Texas at Austin
Austin, TX 78712-1188

TR-89-37

November 1989

Abstract

This paper is a technical extension of our previous work not published fully. It describes how to generate non-degenerate conditions in geometric form for a class of geometry statements of constructive type, called Class C, using Wu's method. We reemphasize a mathematical theorem found by us stating 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*. We prove a new theorem: if an irreducible statement in Class C is confirmed to be generally true, then that statement is valid under the geometric non-degenerate conditions generated by our method. As a direct consequence, most of the geometry theorems proved (to be generally true) by our technique based on the Gröbner basis method are valid under those geometric non-degenerate conditions. We also find subclasses of Class C and prove a theorem that the non-degenerate conditions generated by our method are sufficient for a statement in those subclasses to be valid in *Euclidean geometry*.

Keywords: Geometry theorem proving, Wu's method, the Gröbner basis method, non-degenerate condition, generally true, constructive geometry statement, Euclidean geometry, metric geometry Minkowskian geometry, algebraically closed field, Simson's theorem, the Butterfly theorem.

* The work reported here was supported in part by the NSF Grant CCR-8702108.

[†] On leave from Institute of Systems Science, Academia Sinica, Beijing.

1 Introduction

The success in mechanical geometry theorem proving beginning with the pioneering work of Wu Wen-Tsun [18] since 1978 has been widely known. Wu's work and its further developments [19], [20], [1], [2] inspired researchers to apply the Gröbner basis method to the same class of geometry statements that Wu's method addresses. However, Wu's original work, especially Wu's formulation, has been oversimplified by some researchers. In his recent work [13], [14], [15], Kutzler raised two "serious deficiencies or flaws" of Wu's method.

According to Kutzler, the first defect is that Wu's method is incomplete for Euclidean geometry. This is a fact emphasized at the very beginning in Wu's work [18]. Tarski's method was complete but too inefficient. It was Wu who first restricted to a class of geometry statements and found a fast method. In Euclidean geometry that class is a subclass of the class that Tarski's method can solve. However, *there are many unordered geometries, for which Wu's method is complete, but not Tarski's method.* Unordered metric geometry introduced by Wu [19] (see also [2]) is such an example. In [6] we discussed the scopes of the two methods. Even in Euclidean geometry Wu's method is also complete for certain class of geometry statements (e.g., most geometry theorems of equality type encountered in geometry textbooks). This point seems overlooked by Kutzler. Actually, the work done by Chou and Ko [2], [4], [10], was to give a condition to characterize such a class of geometry statements in Euclidean geometry. In Sections 5 and 6 we will present such classes (in Euclidean geometry) *which can be recognized mechanically.*

According to Kutzler, the second defect is Wu's "careless translation technique" and "finding philosophy". In [14], it wrote:

Wu proposed the following formulation of the problem of mechanically proving a geometry theorem ...

Problem (Wu [19]): Given polynomials $h_1, \dots, h_n, c \in \mathbf{Q}[y_1, \dots, y_m]$, find polynomials $d_1, \dots, d_t \in \mathbf{Q}[y_1, \dots, y_m]$ such that

$$\forall a \in \mathbf{R}^m [h_1(a) = 0 \wedge \dots \wedge h_n(a) \wedge d_1(a) \neq 0 \dots d_t(a) \neq 0 \Rightarrow c(a) = 0]$$

and, for all $i = 1, \dots, t$,

$$\neg[\forall a \in \mathbf{R}^m (h_1(a) = 0 \wedge \dots \wedge h_n(a) \Rightarrow d_i(a) = 0)],$$

or report that no such polynomials exist.

This is what Kutzler's called "Wu's finding problem". Actually, what Kutzler's called "Wu's finding problem" [11] is his understanding of Wu's formulation. If studying Wu's work carefully [18], [19], [21], [20], one can come to a completely different formulation. From the very beginning of Chou's work [1], he followed Wu's original formulation which can be found in Wu's first, pioneering paper [18]. This issue was clarified by Chou and Yang in 1986 [8]. Let $V = \text{Zero}(h_1, \dots, h_n)$ be common zeros of h_1, \dots, h_n in an extension field F of \mathbf{Q} ,¹ the field of rational numbers, i.e.,

$$V = \text{Zero}(h_1, \dots, h_n) = \{(a_1, \dots, a_m) \in F^m \mid h_i(a_1, \dots, a_m) = 0, \text{ for } i = 1, \dots, n\}.$$

In [8] it was clarified that the above formulation is equivalent to that the conclusion is valid on one component (case) of the algebraic set V , no matter this case is degenerate or non-degenerate! Note that for most (more than 90%) examples we have encountered in geometry, V has only one non-degenerate component (case), while it can have tens of degenerate components (cases).

¹ Usually, F is \mathbf{R} , the field of real numbers, or \mathbf{C} , the field of complex numbers.

As clarified in [8], there are two formulations dealing with non-degenerate conditions.

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 automatically.

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

If Formulation F1 is what Kutzler claimed a defect because of what he called “finding philosophy”, we will continue to defend this formulation. We don’t see anything wrong with the notion “generally true” and with the methods which can produce proper non-degenerate conditions. Actually these two formulations are closely related, as Theorems (8.1) and (8.2) in this paper show that under certain conditions a geometry statement proved to be generally true is also valid under the geometric non-degenerate conditions generated by our method in Section 3.3. This paper describes a method for generating non-degenerate conditions in geometric form when using Formulation F1. It is a technical extension of the work in [2]. Our result shows that Kutzler’s assertion that almost all results obtained with Wu’s algorithm cannot be regarded as proofs but only as “near proofs” [15], is quite misleading.

In Section 2, we discuss difficulties with non-degenerate conditions. In Section 3, we present our method for generating *geometric* non-degenerate conditions for a class of geometry statements. In section 4, we prove that under certain conditions, this method is complete. In Section 5, we prove this completeness theorem for Euclidean geometry. Section 6 is about experimental results. Our past results show that at least 400 theorems *in geometric form* have been proved fully mechanically, not “nearly”. In Section 7 we prove a theorem stating that, for certain class of geometry statements, even “Kutzler’s finding methods”, KS1 and KS2 in [11] and [13], if succeed, give “full proofs” of those geometry theorems in geometric form.

2. Difficulties with Non-Degenerate Conditions

It is our experience that finding non-degenerate conditions for a geometry statement is not easy. In [14] Kutzler claimed otherwise

“This is certainly true for the careless way of translating a geometry theorem into an algebraic problem, because one would have to explicitly add all nondegeneracy conditions after the translation process. But in our method from Section 3.2, there is no need for this any more, because for each predicate from our geometric specification language we already have determined all necessary conditions.”

In our opinion, this claim is not valid. For example, the last two of the 20 examples in given [14], D–19 and D–20 which were claimed to be theorems by him, *are not theorems* because of missing non-degenerate conditions. Such kinds of mistakes can be made by anyone, no matter how careful that person is in specifying a particular geometry statement. For a given geometric statement, a non-degenerate condition obvious to one person might not be obvious to another, and might even be difficult to accept for a third individual. Since our point of view might not be obvious to others, we use more examples than usual to illustrate our point.

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

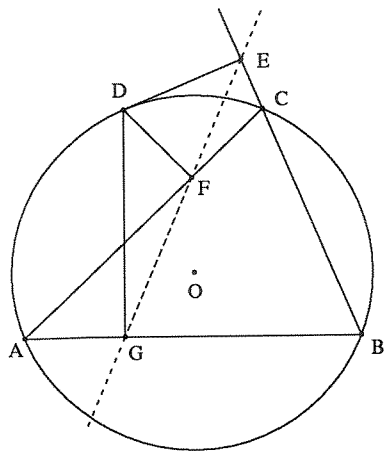


Figure 1

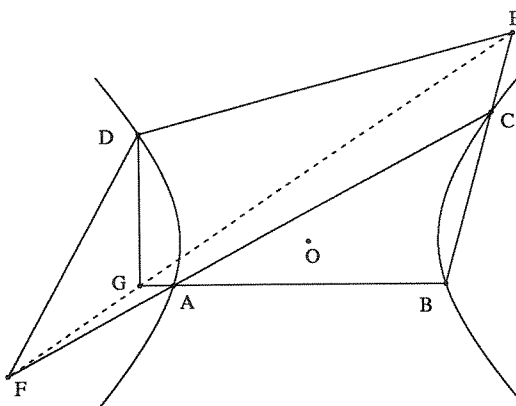


Figure 2

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

In Kutzler's translation, the necessary non-degenerate condition that

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

is missing [14] (an isotropic line is a line perpendicular to itself). Wu's method or the Gröbner basis method cannot confirm this theorem without adding the above condition either as part of the statement or generated during the proof process. But one could argue that Kutzler's translation is correct because in Euclidean geometry isotropic lines do not exist and that problem here is the methods used: Wu's method and the Gröbner basis method are not complete for real closed fields. Do we have to give up proving Simson's theorem when using Kutzler's "careful" translation? A translation technique which takes care of Euclidean geometry and unordered metric geometry is certainly more careful. Since we know that the methods used by us (Wu, Chou, Gao, Kutzler, etc) are complete only for metric geometry, a careful translation technique *should* take this fact into consideration. Such a translation technique will give a much clear insight of the nature of the method used and the theorems proved. For example, Wu's method or the Gröbner basis method can prove Simson's theorem with additional condition (2.1.1) fully, not "nearly". Furthermore, the proofs produced by those methods *do not need the axioms of order*. Even for the field of real numbers \mathbf{R} , there are geometries in which isotropic lines exist. Thus condition (2.1.1) is necessary. Minkowskian geometry is such an example, in which Simson's theorem is valid [3], but only under (2.1.1) (Figure 2).

Whether such conditions are specified as a part of the hypothesis or found by the methods has the same logical result. Logically, we don't see any differences between "proving process" and "finding process" if the "finding process" is properly understood in Wu's or our way, but not in Kutzler's way as stated in Section 1.

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 3).

Here we need a necessary non-degenerate condition that $\neg\text{perpendicular}(E, O, A, D)$ (for the precise meaning of the predicate "perpendicular", see Section 3.1). Obviously, Kutzler's trans-

lation technique cannot come up with this condition. In his language there is even no predicate “ \neg perpendicular”. Without this condition, the Butterfly “theorem” is not a theorem. In this example, one might add a Kutzler’s predicate $\text{intersection}(F, F, E, A, D)$ to the hypotheses instead of $\neg\text{perpendicular}(E, O, A, D)$. But this is very artificial because it means to take the intersection of lines EF and AD without actually knowing the line EF .

The above two examples happen to be two of the four elementary examples given in [4]. This is not an accidental phenomenon, but a strong evidence that dealing with non-degenerate conditions is not easy. Furthermore, the borderline between degenerate cases and non-degenerate cases is not so clear as we might think. The following examples illustrate our point.

In many problems, it allows certain “degenerate cases” in the hypotheses. For example, in Simson’s theorem, DE is perpendicular to BC . However, the case $D = E$ is allowed. To overcome this difficulty in Kutzler’s translation, a new predicate “foot” was introduced. Obviously, this predicate is for a special kind of constructions. In his ad hoc language, it is almost impossible to enumerate all such special kinds of constructions. His translation solves this dilemma by introducing the foot construction, but how about “parallel” instead of “perpendicular”?

Example (2.3). Let ABC be a triangle and F be the midpoint of AB . From any point D on AC a parallel line to AB is drawn intersecting BC at E . Show lines CF , AE and BD are concurrent (Figure 4).

Here we have a condition that DE is parallel to AB . According to Kutzler’s translation, D should not be identical to E . But why in Simson’s case he allows $D = E$? There is no reason to exclude the case $D = E$ unless the statement is not valid or becomes meaningless. In this example, when $D = C$, we have $E = D$ and the statement is still meaningful and valid. In our general scheme in Section 3.3, the foot construction and the construction in Example (2.3) are special cases of “taking intersection of a line passing through two points and another line passing through a given point and perpendicular (parallel) to a given line”. Thus $D = E$ is considered non-degenerate in both cases.

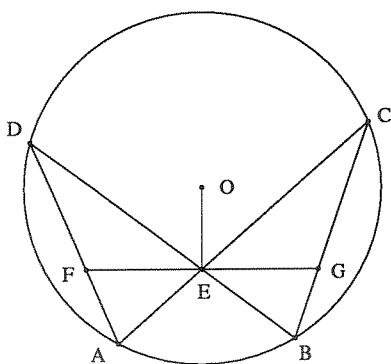


Figure 3

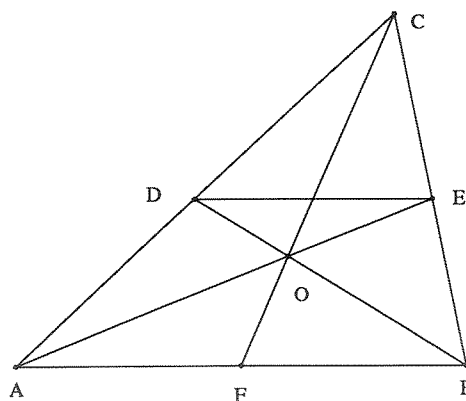


Figure 4

Example (2.4). (D-11 in [14], or Ex8 in [7]). Let ABC be a triangle with $AC \equiv BC$. D is a point on AC ; E is a point on BC such that $AD \equiv BE$. F is the intersection of DE and AB . Show $DF \equiv EF$ (Figure 5).

If we choose AB as the x -axis, then this theorem cannot be confirmed with Kutzler's translation *without adding* any other conditions. However, if we choose AC as the x -axis, then the non-degenerate conditions with Kutzler's translation are enough. The fact that non-degenerate conditions are not coordinate-independent is certainly an unpleasant feature of Kutzler's translation. Something essential must be missing. This can be easily seen in our general scheme in Section 3.3. The following two examples are exact statements from [14], which were claimed to be valid, but actually are not valid in Euclidean geometry with Kutzler's translation.

Example (2.5). (D-19: 5-Star in [14]).

$$\forall R \dots \forall V \forall A \dots \forall E$$

$$[(\text{collinear}(A, R, V) \wedge \text{collinear}(A, S, U) \wedge \text{collinear}(B, S, R) \wedge \text{collinear}(B, T, V) \wedge$$

$$\text{collinear}(C, R, T) \wedge \text{collinear}(C, U, V) \wedge \text{collinear}(D, S, T) \wedge \text{collinear}(D, U, R) \wedge$$

$$\text{collinear}(E, S, V) \wedge \text{collinear}(E, U, T) \wedge \text{collinear}(A, B, C) \wedge \text{collinear}(A, B, D)) \Rightarrow$$

$$\text{collinear}(A, B, E)].$$

This is not a theorem because it is not valid when $R = S = T = V = U$.

Example (2.6). (D-20: the 8_3 configuration problem in [14]).

$$\forall A \dots \forall H$$

$$[(\text{collinear}(A, B, D) \wedge \text{collinear}(A, C, H) \wedge \text{collinear}(A, F, G) \wedge \text{collinear}(B, C, E) \wedge$$

$$\text{collinear}(B, G, H) \wedge \text{collinear}(C, D, F) \wedge \text{collinear}(D, E, G) \wedge \text{collinear}(E, F, H)) \Rightarrow$$

$$\text{collinear}(A, B, C)].$$

This is not a theorem because it is not valid when $A = H$, $B = E$, and $D = F$.

The gap between a geometry problem and its algebraic translation could be bigger than we might think. The following example from [14] (D-13) shows some unexpected gap between an exact geometry problem and its algebraic translation.

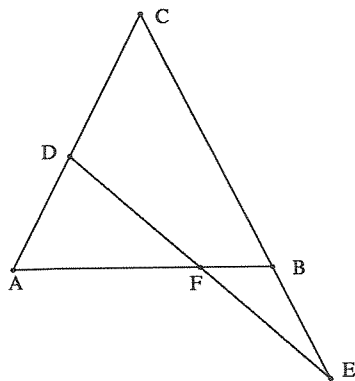


Figure 5

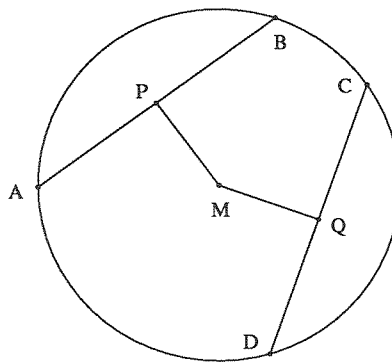


Figure 6

Example 2.7 (D-13 in [14]). In any circle the midpoint of two secants of equal length have equal distances from the circle's center (Figure 5).

In geometry, a circle generally should not be degenerate, i.e., the radius of the circle MA should not be zero. If we look at Kutzler's translation, we have found that his exact (algebraic) statement allows the degenerate circle. In its algebraic form, the conclusion is valid even in this "meaningless" case. Certainly, many people would think such translation is "careless". (For

detailed discussion of this problem see Example (8.4)).

We believe that no careful (mechanical) translation technique which takes care of non-degenerate conditions for all geometry statements can be found. At the best, we can only find a translation technique which is careful enough for a class of geometry statements specified exactly. This is what was exactly done in [2], and in this paper we will present it in detail.

3. A Class of Geometric Statements of Constructive Type

3.1. Our Predicates

In Kutzler's approach [14], a large number of defined predicates are introduced. For example, there are 4 variants (righthangle, normal, onnormal, and foot) for a single predicate "perpendicular". Instead, we use only four basic (non-logical) predicates: "collinear(A, B, C)", "parallel(A, B, C, D)", "perpendicular(A, B, C, D)", "congruent(A, B, C, D)".² The first thing we should emphasize is 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

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

(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

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

(4) Predicate "congruent(A, B, C, D)" includes the cases when $A = B$ or $C = D$. 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 several advantages of using the above predicates.

(1) 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 " $(A \neq B) \wedge (C \neq D) \wedge (A, B, C, D \text{ are not on the same line}) \wedge \neg(AB \parallel CD)$ ". Its corresponding inequation is

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

² In our actual prover [2], [4], there are many other predicates, such as the midpoint, angle congruence, the radical axis of two circles, etc. For the complete list of all those predicates and their algebraic equations see pp.97–99 of [4]. However, for a comparison with Kutzler's predicates and for the class of geometric statements defined in next section, these four predicates are enough.

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 $(A \neq B \wedge C \neq D)$, a redundancy in Kutzler’s translation. In Kutzler’s approach, each predicate usually corresponds to several inequations and one or two equations, thus making its negation hard to use. For example, we want the non-degenerate condition “ \neg perpendicular(E, O, A, D)” in the Butterfly theorem (2.2). The nearest in Kutzler’s predicates is \neg normal(E, O, A, D), which includes the degenerate cases $E = O$ or $A = D$. In our case, we can use the negations of the four predicate in a convenient way. E.g., \neg perpendicular(A, B, A, B) means $A \neq B$ and AB is non-isotropic, i.e., \neg isotropic(A, B), or $(x_1 - x_2)^2 + (y_1 - y_2)^2 \neq 0$. Here we define a new predicate “isotropic(A, B)” to be perpendicular(A, B, A, B).

(2) If these predicates are in the conclusion, e.g., the conclusion is $AB \parallel CD$, do we have to prove that $A \neq B, C \neq D$, and A, B, C, D are not on the same line with Kutzler’s translation?

(3) Our predicates do allow some “degenerate cases” as discussed in Example (2.3) in detail. This provides much more flexibility.

(4) In the case AB is perpendicular or parallel to CD , it is possible that C and D are identical under the previous constructions. If so, then adding the condition that C and D are distinct (as Kutzler’s translation does) will cause inconsistency. Our method in Section 4 can detect such kinds of inconsistency. Here we cite a comment of Wu [21].

Theorems of elementary geometry are usually true only in the generic or non-degenerate case which are implicitly assumed as hypothesis but usually not clearly expressed in the statements of the theorems. In each degenerate case we have to investigate separately whether the theorem is meaningful or not and if it is so whether the theorem remains true or not.

Now to prove theorems in the usual Euclidean fashion one should incessantly resort to previously proved theorems considered to be already known. As these known theorems are only true under certain non-degeneracy conditions, one should verify whether these non-degeneracy conditions are observed or not each time these theorems are to be applied. One has to consider different cases to exclude each of these degeneracy situations. ...

In the proof process, even we assume the starting figure is in a non-degenerate position, we still cannot insure that the other geometric elements constructed from the starting figure are non-degenerate. ...

In our opinion, the separation of equations from inequations is not a “careless” technique and is a *natural algebraic way* for handling non-degenerate conditions either manually or mechanically.

3.2. Definition of Class C

Now we define a class of statements of constructive type for plane geometry, called Class C. First, let us give “circle” a formal definition. A *circle h* is a pair of a point O and a segment (AB) : $h = (O, (AB))$. Two circles $(O, (AB))$ and $(P, (CD))$ are equal if $O = P$ and congruent(A, B, C, D). 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

- l joins two points A and B in Π or
- l passes through one point in Π ,
and is parallel to a line joining two other points A and B in Π or
- l passes through one point in Π ,
and is perpendicular to a line joining two points A and B in Π or

l is perpendicular bisector of AB with A and B in Π .

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 arbitrary points.

Construction 4. Drawing an arbitrary line 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 an arbitrary point.

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 (not necessarily one of the four predicates). In the actual prover [2], we have included more constructions such as taking midpoints and constructions involving angle congruence, radicals axis of two circles, taking a point on a circle knowing its three points, etc.

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 constructions step by step. Suppose we have already generated a set of non-degenerate conditions DS for 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 3.2. 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 $\text{coll}()$, $\text{perp}()$, $\text{para}()$ and $\text{cong}()$ for predicates $\text{collinear}()$, $\text{perpendicular}()$, $\text{parallel}()$ and $\text{congruent}()$, respectively.

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

(i) A line joining two points A and B in Π . We denote it by $L(AB)$.

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

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

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

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

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

(iv) Perpendicular-bisector of AB with A and B in Π . We denote it by $B(AB)$.

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

By adding the non-degenerate condition $A \neq B$, we cannot insure that the hypothesis is consistent. Our prover checks such kinds of inconsistency, see [2] or Section 4.

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

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

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

This is the most interesting construction, one of our major superiorities over Kutzler's approach. 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 . Kutzler only recognized type LL .

Let the two lines be given by the following equations:

$$\begin{aligned} l_1 : a_1 x + b_1 y + c_1 &= 0, \\ l_2 : a_2 x + b_2 y + c_2 &= 0. \end{aligned}$$

The elegance of our approach is that for all 10 types of intersections, the only non-degenerate condition in algebraic form is $\Delta = a_1 b_2 - a_2 b_1 \neq 0$.

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

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

In the algebraic form, this is equivalent to $\Delta = a_1 b_2 - a_2 b_1 \neq 0$. In Kutzler's translation, a redundant condition

$$D_1 = (a_1 \neq 0 \vee b_1 \neq 0) \wedge (a_2 \neq 0 \vee b_2 \neq 0)$$

was added (in addition to the condition $\Delta \neq 0$). Actually, $\Delta \neq 0 \Rightarrow D_1$.

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

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

Case 8.2.1. If $B = D$, then instead, DS should be $DS := \{\neg\text{coll}(A, B, C)\} \cup DS$.

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

$HS := \{\text{coll}(A, B, I), \text{perp}(C, D, E, I)\} \cup HS$; $DS := \{\neg\text{perp}(A, B, C, D)\} \cup DS$. (See the Butterfly theorem). In the special cases,

Case 8.3.1. If AB is parallel to CD , $\neg\text{perp}(A, B, C, D)$ is reduced to $A \neq B, C \neq D$, line AB is not perpendicular to AB itself. Thus instead, DS should be $DS := \{\neg\text{isotropic}(A, B)\} \cup DS$.

Case 8.3.2. Lines AB and CD are identical. $DS := \{\neg\text{isotropic}(A, B)\} \cup DS$.

Case 8.3.3. $A = C$ and $B = D$. $DS := \{\neg\text{isotropic}(A, B)\} \cup DS$. (See Condition (2.1.1) for Simson's theorem.)

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

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

Case 8.4.1. AB is parallel to CD . $DS := \{\neg\text{isotropic}(A, B)\} \cup DS$.

Case 8.4.2. Lines AB and CD are identical. $DS := \{A \neq B, C \neq D\} \cup DS$.

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

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

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

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

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

Case 8.6.1. lines AB is parallel or identical to CD . $DS := \{\neg\text{isotropic}(A, B)\} \cup DS$.

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

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

Case 8.7.1. lines AB is parallel to CD . $DS := \{\neg\text{isotropic}(A, B)\} \cup DS$.

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

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 step until every construction is processed, finally we have two parts for the hypothesis: one is $HS = \{H_1, \dots, H_r\}$, called the equation part of the hypothesis; the other is $DS = \{\neg D_1, \dots, \neg D_s\}$, called the inequation part of the hypothesis and representing non-degenerate conditions of the statement. Let C be the conclusion of the statement, which is not necessarily one of the four predicates defined in Section 3.1, but whose algebraic form is a polynomial equation in the coordinates of the points in Π . Then the exact statement is³

$$(3.1) \quad \forall P \in \Pi (HS \wedge DS \Rightarrow C).$$

Thus according to our translation, we can denote a statement S in Class C by (HS, DS, C) . In Section 4, we will prove a theorem stating that the mechanically generated non-degenerate conditions by our method are sufficient for an irreducible (to be defined) geometry statement S in Class C , i.e., if (3.1) is not valid in the theory of metric geometry, then it cannot be valid by adding any additional non-degenerate conditions $\neg D_{s+1}, \dots, \neg D_t$ as far as

$$(3.2) \quad \forall P \in \Pi (HS \wedge DS \Rightarrow D_j),$$

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

is not valid for $j = s + 1, \dots, t$.

In Sections 5 and 6, we will prove similar theorems for Euclidean geometry. But first let us look at two examples to see how the method described in this section works.

3.4. A Method for Generation of Constructive Sequences

First we point out that the equation part of the hypothesis of a geometry statement of equality type is always easy to identify and clear cut. If the user misses one and the prover answers “not a theorem”, it is user’s own fault. However, if the user misses one of the necessary non-degenerate conditions and the prover answers “not a theorem”, then the user is probably innocent. Even experts feel hard to deal with non-degenerate conditions, e.g., for D–19 and D–20 in [13]. In this sense, Formulation F1 is better because we don’t have to concern with some very subtle degenerate cases. Besides, if the prover answers “the statement is generally false”, then we know the nature of the statement: it would be useless to search for missing degenerate cases. For Class C, we even have a method for generating the inequation part DS . For a given geometry statement in Class C, the sequence of constructions is not unique. Different construction sequences generally lead to different inequation parts, thus giving slightly different exact versions of the original statement.

The equation parts of Simson’s theorem and the Butterfly theorem are clear. For Simson’s theorem, HS_s is:

$$\begin{aligned}
 &\text{perpendicular}(A, B, D, G), \\
 &\text{perpendicular}(A, C, D, F), \\
 &\text{perpendicular}(B, C, D, E), \\
 &\text{collinear}(A, B, G), \\
 &\text{collinear}(A, C, F), \\
 &\text{collinear}(B, C, E), \\
 &\text{congruent}(O, A, O, B), \\
 &\text{congruent}(O, A, O, C), \\
 &\text{congruent}(O, A, O, D).
 \end{aligned}
 \tag{HS}_s$$

and for the Butterfly Theorem, HS_b is

$$\begin{aligned}
 &\text{congruent}(O, A, O, B), \\
 &\text{congruent}(O, A, O, C), \\
 &\text{congruent}(O, A, O, D), \\
 &\text{collinear}(A, E, C), \\
 &\text{collinear}(B, E, D), \\
 &\text{perpendicular}(O, E, E, F), \\
 &\text{collinear}(E, F, G), \\
 &\text{collinear}(F, A, D), \\
 &\text{collinear}(G, B, C).
 \end{aligned}
 \tag{HS}_b$$

If we know the construction sequence for Simson’s theorem, then generating the inequation part DS is straightforward by the method in the previous subsection. Generally, we cannot generate construction sequence merely by the equation part HS_s . However, to ease the user for specifying the construction sequence, our prover has a method so that the user only needs to specify an order in which the points are constructed. For example for Simson’s theorem, we can arrange the points in the order A, B, C, O, D, E, F, G . Our heuristic to figure out the

construction sequence works as follows:

Check last point (here G) to see which predicates in HS_s involve this point. If there are more than two such predicates, we simply return the answer “the order is not chosen in a proper way or the statement is not of constructive type.” Otherwise, there are only one or two predicates involved. Our prover can figure out whether it is one of constructions 6–10.⁴ In this case, we have two conditions, $\text{perpendicular}(A, B, D, G)$ and $\text{collinear}(A, B, G)$. Thus, the prover figures out that it is construction 8.3: $G = T(D, AB) \cap L(AB)$. Then we delete these two conditions from the set HS_s and go to the next point, i.e., F . Similarly, the prover finds two conditions in the new HS_s involving F and figures out the construction $F = T(D, AC) \cap L(AC)$. Next, $E = T(D, BC) \cap L(BC)$; D is on $(O, (OA))$. Last, for point O , $\text{congruent}(O, A, O, B)$ is recognized by our prover as “ O is on $B(AB)$ ”. Thus the last construction is $O = B(AB) \cap B(AC)$. After that, HS_s is empty, thus the remaining points A , B , and C can be arbitrarily chosen. By the method in previous subsection, our prover then produces the set DS_s of non-degenerate conditions from the above construction sequence:

$$\begin{aligned} &\neg\text{collinear}(A, B, C), \\ &\neg\text{isotropic}(AB), \\ &\neg\text{isotropic}(AC), \\ &\neg\text{isotropic}(BC). \end{aligned}$$

Then the exact statement of Simson’s theorem is:

$$(3.3) \quad \forall A \dots \forall G [HS_s \wedge DS_s \Rightarrow \text{collinear}(E, F, G)].$$

Now let us look at the Butterfly Theorem. If we arrange the points in the order O, A, B, C, D, E, F, G , then we have the 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 8.1
$F = L(AD) \cap T(E, OE)$;	construction 8.3
$G = L(EF) \cap L(BC)$.	construction 8.1

Then the program generates a set DS_b of non-degenerate conditions from the above construction sequence:

$$\begin{aligned} &\neg\text{parallel}(E, F, B, C), \\ &\neg\text{perpendicular}(A, D, O, E), \\ &\neg\text{parallel}(A, C, B, D). \end{aligned}$$

The exact statement of the Butterfly Theorem is:

$$(3.4) \quad \forall A \dots \forall G [HS_b \wedge DS_b \Rightarrow \text{midpoint}(F, E, G)].$$

Note that for the same theorem, the construction sequence is usually not unique. Different construction sequences lead to different non-degenerate conditions and slightly different “the

⁴ As we mentioned before, the prover can figure out more than those constructions. But for the purpose of discussion of basic techniques and principles, constructions 6–10 are enough.

exact statements” of the theorem. For example, we have at least 8 essentially different construction sequences for Simson’s theorem (see Appendix 2). In Section 7, we will give another construction sequence for the Butterfly theorem.

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 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 [19] (see also [2]). To be more precise, let $S = (HS, DS, C)$ be a statement in Class C, where $HS = \{H_1, \dots, H_r\}$ is the equation part of the hypothesis, $DS = \{D_1, \dots, D_s\}$ is the inequation part of the hypothesis, 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

$$(4.2) \quad \forall P \in \Pi(HS \wedge DS \Rightarrow C),$$

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 $\neg D_{s+1}, \dots, \neg D_t$ as long as it keeps the consistency of the hypothesis, where each D_i is a geometric condition whose algebraic form is an equation. The consistency means that

$$(4.3) \quad \forall P \in \Pi(HS \wedge DS \Rightarrow D_i),$$

is not valid in Ω , for $i = s + 1, \dots, t$. ▮

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, \dots, u_{j-1} , dependent variables x_1, \dots, x_{k-1} , and the equations $h_1 = 0, \dots, h_{k-1} = 0$ corresponding to a part of hypothesis $\{H_1, \dots, H_{k-1}\}$, and an ascending chain of the form:

$$(4.4) \quad \begin{aligned} & f_1(u_1, \dots, u_{j-1}, x_1) \\ & f_2(u_1, \dots, u_{j-1}, x_1, x_2) \\ & \dots \\ & f_{k-1}(u_1, \dots, u_{j-1}, x_1, \dots, x_{k-1}). \end{aligned}$$

Furthermore, we assume the ascending chain (in weak sense) f_1, \dots, f_{k-1} is irreducible.⁵ This is the exact meaning of “irreducible” in the statement of Theorem (4.1). 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 , is

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

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 $R_a = \text{prem}(a; f_1, \dots, f_{k-1})$ and $R_b = \text{prem}(b; f_1, \dots, f_{k-1})$ are zero. (Here *prem* denotes the successive pseudo divisions 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 hypothesis by adding $A \neq B$. Thus our method actually can detect the inconsistency raised in Section 3.1. In that case, we either can say that the hypothesis does not satisfy the dimensionality constraint required by Formulation F1 (see p.47 [4]), or it is a theorem according to Formulation 2 because of the inconsistency of the hypothesis.

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 variable introduced. We have a new irreducible ascending chain f_1, \dots, f_k . Then the condition $a \neq 0$ is equivalent to that the line l is well defined (in our case $A \neq B$).

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 non-zero of leading coefficient ($a \neq 0$) 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 problem later.

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

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

⁵ For the definition of ascending chains, see [19], [4], or [5]. The ascending chain f_1, \dots, f_{k-1} is irreducible if $f(u, x_1, \dots, x_i)$ is irreducible in the field $\mathbf{Q}(u)[x_1, \dots, x_i]/(f_1, \dots, f_{i-1})$, for $i = 1, \dots, k-1$. Here here (f_1, \dots, f_{i-1}) is the ideal generated by f_1, \dots, f_{i-1} .

Our next step is to check whether CD is isotropic, i.e., whether $R = \text{prem}((x_3 - x_4)^2 + (y_3 - y_4)^2; f_1, \dots, f_{k-1})$ is zero. If $R \neq 0$, then let $f_k = h_k$, and f_1, \dots, f_k is irreducible by Lemma (A1.2) in Appendix 1. 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 :

$$\begin{aligned} h_k &= a_1x + b_1y + c_1 = 0, \\ h_{k+1} &= a_2x + b_2y + c_2 = 0. \end{aligned}$$

The first step is to check whether $R = \text{prem}(\Delta; f_1, \dots, f_{k-1})$ is zero, where $\Delta = a_1b_2 - a_2b_1$. 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:

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

where $d = b_2c_1 - b_1c_2$ and $e = a_1c_2 - a_2c_1$. We have an irreducible ascending chain f_1, \dots, f_{k+1} .⁶

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 line l and the circle c :

$$\begin{aligned} h_k &= y^2 + x^2 + ax + by + c = 0, \\ h_{k+1} &= a_1y + b_1x + c_1 = 0. \end{aligned}$$

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

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

Case 9.2. $R \neq 0$. One of $R_a = \text{prem}(a_1; f_1, \dots, f_{k-1})$, $R_b = \text{prem}(b_1; f_1, \dots, f_{k-1})$ 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 , we have

$$\begin{aligned} f_k &= (a_1^2 + b_1^2)x_k^2 + (a_1^2b + 2c_1b_1 - aa_1b_1)x_k + (a_1^2c + c_1^2 - ac_1a_1) = 0, \\ f_{k+1} &= a_1x_{k+1} + b_1x_k + c_1 = 0. \end{aligned}$$

Now we have ascending chain (in weak sense) f_1, \dots, f_{k+1} . We can check whether f_1, \dots, f_{k+1} is irreducible using the algorithm introduced in [1] (also see [2]) 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, \dots, f_{k+1} is irreducible.

⁶ Actually, we can let $f_{k+1} = a_2x_{k+1} + b_2x_k + c_2$ if $\text{prem}(a_2; f_1, \dots, f_{k-1}) \neq 0$. This can save time and space. For details see [2] or Appendix 1.

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 :

$$\begin{aligned} h_k &= y^2 + x^2 + ay + bx + c = 0, \\ h_{k+1} &= y^2 + x^2 + ey + dx + j = 0. \end{aligned}$$

Letting $h_{k+1} := h_k - h_{k+1}$, we have the equation for a l line:

$$h_{k+1} = a_1y + b_1x + c_1 = 0,$$

where $a_1 = a - e$, $b_1 = b - d$, $c_1 = c - j$. This is the line joining the two points of intersection of c_1 and c_2 , if they intersect (this is always 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 called the radical axis of the two circles. Now we check whether $R = \text{prem}(a_1^2 + b_1^2; f_1, \dots, f_{k-1})$ 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 $R_a = \text{prem}(a_1; f_1, \dots, f_{k-1})$ and $R_b = \text{prem}(b_1; f_1, \dots, f_{k-1})$ 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:

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

Now we want to ask whether the formula

$$(4.6) \quad \forall P \in \Pi[(H_1 \wedge \dots \wedge H_r \wedge \neg D_1 \wedge \dots \wedge \neg D_s) \Rightarrow C]$$

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

$$(4.7) \quad \forall \in ux[(h_1 = 0 \wedge \dots \wedge h_r = 0 \wedge d_1 \neq 0 \wedge \dots \wedge d_s \neq 0) \Rightarrow c = 0]$$

is valid in F_Ω , where h_1, \dots, h_r , and c are the polynomials corresponding to H_1, \dots, H_r and C , respectively, d_1, \dots, d_s are polynomials corresponding to D_1, \dots, 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. The predicate (point) “equal” is a logical one as defined in [17]. Let $A = (x_1, y_1)$ and $B = (x_2, y_2)$. The algebraic form of $A \neq B$ is $(a \neq 0 \vee b \neq 0)$, which is the disjunction of two inequations, where $a = x_1 - x_2$ and $b = y_1 - y_2$. However, we can use a new variable z to convert it to an inequation: in any field $(a \neq 0 \vee b \neq 0)$ if and only if $\exists z(az + b \neq 0)$. Then we can move such existentially quantified variables z_k to the outside, and (4.7) becomes

$$(4.7') \quad \forall u_i x_j z_k [(h_1 = 0 \wedge \dots \wedge h_r = 0 \wedge d_1 \neq 0 \wedge \dots \wedge d_s \neq 0) \Rightarrow c = 0].$$

Since new variables z_i are only auxiliary, we will not mention them explicitly.

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

Theorem (4.8). Let the ascending chain $ASC = f_1, \dots, f_r$ in the form of (4.5), which is obtained by the above procedure, be irreducible and the I_i be the initials of the f_i . If F_Ω is algebraically closed, then the following conditions are equivalent:

- (1) $\text{prem}(c; ASC) = 0$;
- (2) The formula

$$(4.9) \quad \forall ux \in F_\Omega [(f_1 = 0 \wedge \dots \wedge f_r = 0 \wedge I_1 \neq 0 \wedge \dots \wedge I_r \neq 0) \Rightarrow c = 0]$$

is valid in F_Ω ;

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

Proof. (1) \Rightarrow (2). Suppose $R = \text{prem}(c; ASC) = 0$. Since we have the remainder formula:

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

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 [4].

(2) \Rightarrow (3). Let J be the set $\{I_i \mid I_i \text{ is not a constant, } i = 1, \dots, r\}$. Let $N = \{d_1, \dots, d_s\}$. We want to show that those I_k in J but not in N can be removed or replaced by d_l 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 $ASC' = f_1, \dots, f_{k-1}, f'_k, f_{k+1}, \dots, f_r$ is also irreducible. By Lemma (A1.1) in Appendix 1, $\text{prem}(c; ASC) = 0$ if and only if $\text{prem}(c; ASC') = 0$. Thus we only need the condition $(a \neq 0 \vee b \neq 0)$. This is the condition for the line to be well defined as given in 3.3.

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.

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

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

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

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

Since ASC is irreducible, (1) and (5) are equivalent. ■

5. The Completeness for Euclidean Geometry: Easy Cases

If Wu's method does not confirm a statement in Euclidean geometry, we generally do not know whether the statement is disproved or not. The work [2] and [10] is to give a condition, i.e., the Condition (GC) below, for geometry statements in Euclidean geometry for which Wu's method is complete, i.e., if Wu's method does not confirm a statement satisfying that condition, the statement is *not* a theorem. However, to decide whether a geometry statement satisfies that condition is beyond Wu's method. In this and next sections, we will present subclasses of Class C that satisfy the condition (GC) and can be recognized by our prover [2].

From now on, the geometry in discussion is Euclidean geometry, whose associated field is \mathbf{R} .

Definition (5.1). Let f_1, \dots, f_r be an irreducible ascending chain in the form of (4.5), if there are non-empty open intervals U_1, \dots, U_d in \mathbf{R} such that $f_1 = 0, \dots, f_r = 0$ have solutions for x_1, \dots, x_r in \mathbf{R} for all $u_i \in U_i$ ($i = 1, \dots, d$), we say that f_1, \dots, f_r is \mathbf{R} -generic.

Generic Condition (GC). Let a constructive statement $S = (HS, DS, C)$ and h_1, \dots, h_r , and $ASC = f_1, \dots, f_r$ be the same as in Section 4. If f_1, \dots, f_r is \mathbf{R} -generic, then we say S satisfies Generic Condition (GC).

Theorem (5.2). Let the notations be the same as above. If S satisfies Condition (GC), then the following conditions are equivalent

- (1) $\text{prem}(c; f_1, \dots, f_r) = 0$;
- (2) $\forall xu \in \mathbf{R}[(f_1 = 0 \wedge \dots \wedge f_r = 0 \wedge I_1 \neq 0 \wedge \dots \wedge I_r \neq 0) \Rightarrow c = 0]$;
- (3) $\forall xu \in \mathbf{R}[(h_1 = 0 \wedge \dots \wedge h_r = 0 \wedge d_1 \neq 0 \wedge \dots \wedge d_s \neq 0) \Rightarrow c = 0]$;
- (4) $\forall P \in \Pi[HS \wedge DS \Rightarrow C]$ is valid in Euclidean geometry;
- (5) $\text{prem}(d \cdot c; ASC) = 0$ for any polynomial d with $\text{prem}(d; ASC) \neq 0$.

Proof. (2) \Rightarrow (1). This is a theorem proved in [2] (also see [4] or [10]).

The rest of the proof are either obvious or the same as in the proof of Theorem (4.8). ■

Now we enlarge subclasses of class C to satisfy the Generic Condition (GC) in the following two subsequent theorems.

Theorem (5.3). Statements in class C which involve only constructions 1–6, and 8 are either inconsistent or satisfy the Generic Condition (GC). Thus our method is complete for such statements in Euclidean geometry.

Proof. Since for such a statement, either the hypothesis $HS \wedge DS$ is inconsistent, or we have an ascending chain f_1, \dots, f_r with all $\text{deg}(f_i, x_i) = 1$, the condition (GC) is trivially satisfied. ■

Remark. This is what we called linear cases. Wu's method or the Gröbner basis method (used in a proper way) are also decision methods for such linear cases, i.e., they can prove or disprove any assertions on linear configurations in *any geometries*, including Euclidean geometry. About 60% of 512 theorems proved in [4] were specified as linear statements. Nine among the 20 examples (D4–10, 15, 17) in [14] were also specified as linear statements.

Theorem (5.4). For a statement $S = (HS, DS, C)$ in class C which involves only constructions 1-8, either $HS \wedge DS$ is inconsistent with additional conditions that the radius CD is non-zero (i.e., $\neg \text{isotropic}(C, D)$) for construction 7, or it satisfies the Generic Condition (GC).

Proof. Now we are back to construction 7 in the proof of Theorem (4.1). the radius of the circle is given by $r^2 = (x_3 - x_4)^2 + (y_3 - y_4)^2$. $\neg \text{isotropic}(C, D)$ means $r^2 = 0$. We check whether $R = \text{prem}(r^2; f_1, \dots, f_{k-1})$ is 0. If $R = 0$, then the hypothesis $HS \wedge DS$ is inconsistent with $r^2 \neq 0$. In Euclidean geometry $r^2 = 0$ means $C = D$ or the radius is 0. Otherwise, f_1, \dots, f_k is irreducible by Lemma (A1.2), and we have open interval $U_i = (-r, r)$ such that f_1, \dots, f_k have real solutions for the x . Thus, the generic condition (GC) is satisfied. \blacksquare

Remark. Simson's theorem and the Butterfly theorem in Section 3.4 belong to this subclass. Thus we can prove and *disprove* any assertions about these two configurations. E.g., if we want to decide whether $AF \equiv BG$ for the Butterfly configuration, *this is not a theorem in Euclidean geometry* as shown by our prover. Our result is much stronger than any results based on Formulation F2: it cannot be a theorem by adding any consistent non-degenerate conditions.

6. The Completeness for Euclidean Geometry: More General Cases

To enlarge our subclass further we first need some new tools. From now on we are free to use any tools and notions developed by Ritt, Wu and us [16], [19], [22], [4], [5]. The coefficients of all polynomials mentioned are in \mathbf{Q} . Let F be a field containing \mathbf{Q} . Let PS be a polynomial set in variables y_1, \dots, y_m . We denote $F\text{-Zero}(PS)$ or $\text{Zero}(PS)$ (if F is clear in the context) the common zeros of polynomials of PS in the field F , i.e.,

$$\text{Zero}(PS) = \{(a_1, \dots, a_m) \in F^m \mid f(a_1, \dots, a_m) = 0 \text{ for all } f \in PS\}.$$

Let GS be another polynomial set. We denote $\text{Zero}(PS/GS)$ the set difference $\text{Zero}(PS) - \text{Zero}(GS)$.

Definition (6.1). Let $HS, DS, h_1, \dots, h_r, d_1, \dots, d_s$ be the same as before. A statement $S = (HS, DS, C)$ in Class C is said to be irreducible in weak sense if $Z = \text{Zero}(h_1, \dots, h_r / d_1, \dots, d_s)$ has only one non-degenerate component, i.e., $Z = \text{Zero}(PD(f_1, \dots, f_r) / d_1, \dots, d_r)$, where f_1, \dots, f_r is an irreducible ascending chain in the form of (4.5). Here $PD(f_1, \dots, f_r)$ denotes the prime ideal with f_1, \dots, f_r as a characteristic set, i.e., the set $\{g \mid \text{prem}(g; f_1, \dots, f_r) = 0\}$.

Obviously, an irreducible statement defined in Section 4 is an irreducible statement in weak sense. Example (2.3) is not irreducible, but irreducible in weak sense.

Theorem (6.2). There is an algorithm to decide whether $S = (HS, DS, C)$ is irreducible in weak form.

Proof. Using the method in Chapter 3 of [2] and Ritt–Wu's decomposition algorithm in [5]. \blacksquare

Theorem (6.3). Let the notations be the same as the above. If $S = (HS, DS, C)$ is an

irreducible statement in weak sense, then the formula (4.6) is valid iff $prem(c; f_1, \dots, f_r) = 0$.

Proof. Straightforward. ▮

Generic Condition (GCW). Let $S = (HS, DS, C)$ an irreducible statement in weak form. If f_1, \dots, f_r is \mathbf{R} -generic, then S is said to satisfy the Condition (GCW).

Theorem (6.4). If $S = (HS, DS, C)$ satisfies (GCW), then the following conditions are equivalent:

- (1) $prem(c; f_1, \dots, f_r) = 0$;
- (2) $\forall xu \in \mathbf{R}[(f_1 = 0 \wedge \dots \wedge f_r = 0 \wedge I_1 \neq 0 \wedge \dots \wedge I_r \neq 0) \Rightarrow c = 0]$;
- (3) $\forall xu \in \mathbf{R}[(h_1 = 0 \wedge \dots \wedge h_r = 0 \wedge d_1 \neq 0 \wedge \dots \wedge d_s \neq 0) \Rightarrow c = 0]$;
- (4) $\forall P \in \Pi[HS \wedge DS \Rightarrow C]$ is valid in Euclidean geometry;
- (5) $prem(d \cdot c; ASC) = 0$ for any polynomial d with $prem(d; ASC) \neq 0$.

Now we define the class of geometry statements involving constructions 1–8, 9.1 and 10.1 to be Class C_E . Our main result in this Section is that Wu's method is complete for Class C_E in Euclidean geometry.

Theorem (6.5). A geometry statement S in Class C_E is either inconsistent or irreducible in weak sense. In the latter case, S satisfies the Condition (GCW).

Proof. The proof also provides a method to decide whether S is inconsistent, and in the consistent case to compute the irreducible ascending chain f_1, \dots, f_r and decide whether S is valid or not in Euclidean geometry. The elegance of the method is that we don't need any factorization. For a complete proof see Appendix 1. ▮

Theorem (6.6). Let $S = (HS, DS, C)$ be a statement in Class C_E . If (4.6) is not valid in Euclidean geometry, then it cannot be a theorem in Euclidean geometry by adding any set of additional conditions $\{\neg D_{s+1}, \dots, \neg D_t\}$ as long as it keeps the consistency of the hypothesis, i.e., $\forall P \in \Pi(HS \wedge DS \Rightarrow D_i)$ is not valid in Euclidean geometry, for $i = s + 1, \dots, t$.

Proof. This is a direct consequence of Theorems (6.4) and (6.5). ▮

7. Examples and Experiment Results

Example (7.1). (Simson's Theorem). By the calculation of Example (2.1) in Chapter 1 of [4], we actually proved the exact geometric statement of Simson's theorem (3.3).

Example (7.2). (the Butterfly Theorem). If we introduce points in the order E, O, A, B, C, D, F, G , then our prover produces the following construction sequence for HS_b (Section 3.4).

E, O , and A are arbitrarily chosen	construction 1
B is on circle $(O, (O, A))$	construction 7
$C = L(AE) \cap (O, (OA))$	construction 9.1
$D = L(BE) \cap (O, (OA))$	construction 9.1
$F = T(E, EO) \cap L(AD)$	construction 8.3
$G = L(EF) \cap L(BC)$	construction 8.1

Then DS_{b1} is

\neg isotropic(AE), $A \neq C$,
 \neg isotropic(BE), $B \neq D$,
 \neg perpendicular(A, D, O, E),
 \neg parallel(A, C, B, D).

Then the exact geometric statement of the Butterfly theorem, $HS_b \wedge DS_{b1} \Rightarrow C$, was proved by the calculation in Example (2.4) of Chapter 1 of [4] because of Theorems (6.4) and (6.5) (see also Example 36 of [4]). Furthermore, we can prove and disprove any assertions about the Butterfly configuration in Euclidean geometry. For example, if we want to ask whether $AF \cdot CG = BG \cdot DF$, then it was easily disproved in Euclidean geometry in about 36 seconds. Our disproof is much stronger than any results based on Formulation F2: it cannot be a theorem in Euclidean geometry by adding any number of additional conditions as far as $HS \wedge DS$ is consistent with those additional conditions.

For 413 among 512 theorems proved in [4], non-degenerate conditions are generated all in geometric form using the method described in this paper or in [2] with some extensions to include other constructions. In [4], we gave full geometric specifications of these 413 theorems and proved them under such specifications without adding any other conditions.

8. On Proving Methods Based on the Gröbner Basis Method

The Gröbner basis (GB) method has been also successfully used for the same class of geometry theorems which Wu's method originally addresses. Chou-Schelter [7] and Kapur [9] were the first to use the GB method to prove theorems according to Formulation F2. Chou-Schelter [7] and Kutzler-Stifter [12] were the first to use the GB method according to Formulation F1.

However, Kutzler is obviously unaware of Formulation F1 as he claims that his methods KS1 and KS2 are for "Wu's finding problem" and only give "near proofs" [14], [13]. In the light of our analysis of statements of Class C, we can see that method KS2 actually did more than Kutzler thought in [13], [14].⁷ At least for a subclass of Class C, method KS2 actually gives "full proofs" of those statements, instead of "near proofs" if we add *geometric non-degenerate conditions generated by our method*.

First we give a brief review of Formulation F1. (The reader can find detailed discussions in [4] or [8]).

Formulation F1. Let the equation part of the hypothesis of a geometry statement be given algebraically by a set of equations $h_1(u_1, \dots, u_d, x_1, \dots, x_r) = 0, \dots, h_n(u_1, \dots, u_d, x_1, \dots, x_r) = 0$, where the u are parameters and the x are dependent variables (such as given by the method in Section 4). Let the conclusion is also an equation $c(u, x) = 0$. The statement is *generally true* if there is a non-zero polynomial s containing the u only, called a u -polynomial, such that $sc \in \text{Radical}(h_1, \dots, h_n)$.

Theorem (8.1). For an irreducible statement S in weak sense, S is generally true if and only if the geometric form of S, (4.6), is valid.

⁷ Method KS2 is very similar to the method used by Chou-Schelter. However, KS2 is incomplete in the sense it cannot derive $x^2 = 0 \Rightarrow x = 0$ in contrast to Chou-Schelter's method. For historical reason it should be more appropriate to denote it by CS-KS2.

Proof. By the definition of “irreducible in weak form”,

$$\text{Zero}(h_1, \dots, h_r / d_1, \dots, d_s) = \text{Zero}(PD(f_1, \dots, f_r) / d_1, \dots, d_s).$$

Since $\text{Zero}(PD(f_1, \dots, f_r) / d_1, \dots, d_s) \subset \text{Zero}(c)$ iff $\text{prem}(c; f_1, \dots, f_r) = 0$. This is in turn equivalent to there is a u -polynomial s such that $sc \in \text{Ideal}(f_1, \dots, f_r) \subset \text{Radical}(h_1, \dots, h_r)$.

Now suppose there is a u -polynomial s such that $sc \in \text{Radical}(h_1, \dots, h_r)$. Thus sc vanishes on $\text{Zero}(PD(f_1, \dots, f_r) / d_1, \dots, d_r)$; hence $\text{prem}(sc; f_1, \dots, f_r) = 0$. Since s is a u -polynomial, $\text{prem}(c; f_1, \dots, f_r) = 0$. By Theorem (6.3), (4.6) is valid. \blacksquare

Theorem (8.2). For a statement $S = (HS, DS, C)$ in Class C which involves construction 1–8, 9.1 and 10.1, $(HS \wedge DS \Rightarrow C)$ is valid if and only if Method CS–KS2 proves S to be generally true, or in Kutzler’s term, KS2 confirms the “finding problem”. In the case that $HS \wedge DS \Rightarrow C$ is not valid, it cannot be valid by adding any number of non-degenerate conditions $\neg D_{s+1}, \dots, \neg D_t$ as long as each D_i is consistent with $HS \wedge DS$, i.e., $(HS \wedge DS \Rightarrow D_i)$ is not valid in Euclidean geometry for $i = s + 1, \dots, t$.

Proof. This is a direct consequence of Theorems (6.4), (6.5) and (8.1). \blacksquare

Example (8.3). Theorem (8.2) can apply to all 12 examples given in Chou–Schelter’s work on the Gröbner basis method [7] but one, Ex8, which is D–11 in [14]. Thus our Gröbner basis method in [7] actually gave full proofs of these 11 examples instead of “near proofs”.

Example (8.4). By a close look at the selection of parameters and equations of 20 examples given in [14], we have found that 12 involve only constructions 1–8, (D–2, D4–10, D14–17), thus Kutzler’s results of method KS2 actually gave “full proofs” of these 12 statements instead of “near proofs” if he uses our translation technique for generating non-degenerate conditions instead of his. Especially we would like to mention examples D–14, D–16 and D–17 because Kutzler’s results show that no “full proofs” have ever been given by any methods (provers) he discussed. By Theorem (8.2), Kutzler’s method KS2 itself actually gave full proofs of these theorems in 21.4, 3151.6, and 73.9 seconds, respectively. Here we have made this assertion even *without knowing the explicit forms of the u -polynomials found by his prover*. However, we do require the non-degenerate conditions be those generated by our method. Note that the equation parts are the same for both our translation and Kutzler’s translation.

Example (8.5). Examples D–11 and D–13 in [14], according to Kutzler’s specifications, belong to Class C, but involve constructions 9 or 10. Thus we cannot arrive at the same conclusion as in (8.4) without further decomposition. Using our prover, we have confirmed that D–11 is irreducible in weak sense and D–13 is irreducible. Thus together with our results, the results by method KS2 in [14] also gave full proofs of these two theorems under geometric non-degenerate conditions produced by our method, not by Kutzler’s translation. According to Kutzler’s results, no “full proof” of D–11_a has never been found by all methods he discussed. This was due to the omission of the non-degenerate condition “ \neg isotropic(B, C)” required by our construction 9.

D–13 was proved by some methods in [14] even *without adding any non-degenerate conditions*. According to Kutzler’s specification, point D is obtained by our construction 10: $D = (C, (AB)) \cap (M, (MA))$. Thus we do require that CM be non-isotropic. There are differences between Kutzler’s “full proof” by method KS1 and the “full proof” by method KS2. The former does not require any non-degenerate conditions, but the latter needs \neg isotropic(C, M) generated by our method. It is worthwhile to explain this very subtle situation in more detail.

The equation part of the theorem HS is

$$\{\text{cong}(A, M, M, B), \text{cong}(A, M, M, C), \text{cong}(A, M, M, D), \\ \text{cong}(A, B, C, D), \text{midpoint}(A, P, B), \text{midpoint}(C, Q, D)\}.$$

The conclusion C is congruent(P, M, M, Q). The result by KS1 shows

$$(8.6) \quad \forall A \dots \forall Q (HS \Rightarrow DS)$$

is valid. Our result (without knowing the concrete form of u -polynomial found by method KS2) shows that method KS2 actually proved

$$(8.7) \quad \forall A \dots \forall Q (HS \wedge \neg \text{isotropic}(C, M) \Rightarrow C).$$

To see this delicate difference, following Kutzler's selection of parameters and dependent variables, we let $A = (u_1, 0)$, $M = (0, 0)$, $B = (u_2, x_1)$, $C = (u_3, x_2)$, $D = (x_3, x_4)$, $P = (x_5, x_6)$, and $Q = (x_7, x_8)$. Then we have

$$\begin{aligned} h_1 &= x_1^2 + u_2^2 - u_1^2 = 0 & MA \equiv MB. \\ h_2 &= x_2^2 + u_3^2 - u_1^2 = 0 & MA \equiv MC. \\ h_3 &= x_4^2 - 2x_2x_4 + x_3^2 - 2u_3x_3 + x_2^2 - x_1^2 + u_3^2 - u_2^2 + 2u_1u_2 - u_1^2 = 0 & AB \equiv CD. \\ h_4 &= x_4^2 + x_3^2 - u_1^2 = 0 & MA \equiv MD. \\ h_5 &= 2x_5 - u_2 - u_1 = 0 & \\ h_6 &= 2x_6 - x_1 = 0 & P \text{ is the midpoint of } B \text{ and } A. \\ h_7 &= 2x_7 - x_3 - u_3 = 0 & \\ h_8 &= 2x_8 - x_4 - x_2 = 0 & Q \text{ is the midpoint of } D \text{ and } C. \\ c &= x_8^2 + x_7^2 - x_6^2 - x_5^2 = 0 & \text{Conclusion: } MP \equiv MQ. \end{aligned}$$

By the Ritt-Wu's algorithm described in [5], we have

$$\text{Zero}(h_1, \dots, h_8) = \text{Zero}(PD(ASC_1)) \cup \text{Zero}(PD(ASC_2)),$$

where the ascending chain $ASC_1 =$

$$\begin{aligned} &2x_8 - x_4 - x_2 \\ &2x_7 - x_3 - u_3 \\ &2x_6 - x_1 \\ &2x_5 - u_2 - u_1 \\ &x_2x_4 + u_3x_3 - u_1u_2 \\ &u_1x_3^2 - 2u_2u_3x_3 + u_1u_3^2 + u_1u_2^2 - u_1^3 \\ &x_2^2 + u_3^2 - u_1^2 \\ &x_1^2 + u_2^2 - u_1^2, \end{aligned}$$

and $ASC_2 =$

$$\begin{aligned} &2x_8 - x_4 - x_2 \\ &2x_7 - x_3 - u_3 \\ &2x_6 - x_1 \\ &2x_5 - u_2 \\ &x_2x_4 + u_3x_3 \\ &x_2^2 + u_3^2 \\ &x_1^2 + u_2^2 \\ &u_1. \end{aligned}$$

ASC_1 corresponds to the non-degenerate case, while ASC_2 corresponds to the degenerate case. Thus, $prem(c; ASC_1)$ must be zero if D-13 is a theorem. However, it happens that $prem(c; ASC_2) = 0$. This explains why (8.6) is also valid. But generally, we need $\neg isotropic(C, M)$ to exclude ASC_2 . This also explains why D-11_a is not a theorem without adding condition $\neg isotropic(B, C)$.

Remark (8.8). The first 18 of 20 examples in [14] can be easily specified as statements in Class C and proved with the method (prover) presented in [2] or in this paper entirely automatically, all in geometric form (all in less than 5 seconds on a Symbolics 3600). The last two cannot be specified in a constructive way.

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 and H.P. Ko, "On Mechanical Theorem Proving in Minkowskian Plane Geometry", *Proc. of Symp. of Logic in Computer Science*, pp187-192, 1986.
- [4] S.C. Chou, *Mechanical Geometry Theorem Proving*, D. Reidel Publishing Company, Dordrecht, Netherlands, 1988.
- [5] S. C. Chou and X. S. Gao, "Ritt-Wu's Decomposition Algorithm and Geometry Theorem Proving", Technical Report 89-09, Department of Computer Sciences, University of Texas at Austin, 1989.
- [6] S.C. Chou, X.S. Gao and Nicholas McPhee, "A Combination of Ritt-Wu's Method and Collins' Method", TR-89-28, Computer Sciences Department, The University of Texas at Austin, October, 1989.
- [7] S.C. Chou and W.F. Schelter, "Proving Geometry Theorems with Rewrite Rules", *Journal of Automated Reasoning*, **2(4)** (1986), 253-273.
- [8] S.C. Chou and G.J. Yang, "On the Algebraic Formulation of Certain Geometry Statements and Mechanical Geometry Theorem Proving", *Algorithmica*, Vol. 4, 1989, 237-262.
- [9] D. Kapur, "Geometry Theorem Proving Using Hilbert's Nullstellensatz", in Proceedings of the 1986 Symposium on Symbolic and Algebraic Computation, 202-208.
- [10] H.P. Ko and S.C. Chou, "A Decision Method for Certain Algebraic Geometry Problems", Preprint, to appear in *Rocky Mountain Journal of Mathematics*.
- [11] B. Kutzler and S. Stifter, "On The Application of Buchberger's Algorithm to Automated Geometry Theorem Proving", *Journal of Symbolic Computation*, (1986) **2**, 389-397.
- [12] 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.
- [13] B. Kutzler, "Careful Algebraic Translations of Geometry Theorems", in *Proceedings of the 1989 Symposium on Symbolic and Algebraic Computation*, 254-263.
- [14] B. Kutzler, "Algebraic Approaches to Automated Geometry Theorem Proving", PhD Thesis, Johannes Kepler University, Linz, Austria, 1988.
- [15] B. Kutzler, "Deciding a Class of Euclidean Geometry Theorems with Buchberger's Al-

- gorithm”, in *Proceedings of Conference on Computer Aided Proofs in Analysis*, March 22–25, 1989, Cincinnati, USU, IMA series, to appear.
- [16] J. F. Ritt, *Differential Algebra*, AMS Colloquium Publications, New York, 1950.
 - [17] J. R. Shoenfield, *Mathematical Logic*, Addison-Wesley Publishing Company, (1967).
 - [18] Wu Wen-tsün, “On the Decision Problem and the Mechanization of Theorem Proving in Elementary Geometry”, *Scientia Sinica* **21** (1978), 157-179.
 - [19] 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.
 - [20] Wu Wen-tsün, *Basic Principles of Mechanical Theorem Proving in Geometries*, (in Chinese) Peking 1984.
 - [21] Wu Wen-tsün, “Toward Mechanization of Geometry – Some Comments on Hilbert’s ‘Grundlagen der Geometrie’”, *Acta Mathematica Scientia* **2** (1982), 125-138.
 - [22] Wu Wen-tsün, “On Zeros of Algebraic Equations –An Application of Ritt’s Principle”, *Kexue Tongbao* **31(1)** (1986), 1-5.

Appendix 1. Proof of Theorem (6.5)

Lemma (A1.1). Let $ASC_1 = f_1, \dots, f_k$ and $ASC_2 = g_1, \dots, g_j$ be two irreducible ascending chains in variables y_1, \dots, y_m . If $\text{prem}(f_i; g_1, \dots, g_j) = 0$ and $\text{prem}(lc(f_i); g_1, \dots, g_j) \neq 0$ for $i = 1, \dots, k$, then $PD(ASC_1) \subset PD(ASC_2)$.

Proof. See Appendix of [5]. ▮

Lemma (A1.2). Let $ASC = f_1(u, x_1), \dots, f_k(u, x_1, \dots, x_k)$ be an irreducible ascending chain, and $f_{k+1}(u, x_1, \dots, x_{k+1}) = x_{k+1}^2 + y^2 - r$, where r is a polynomial in u and x_1, \dots, x_k and y is a new variable other than the u and x . If $\text{prem}(r; f_1, \dots, f_k) \neq 0$, then the ascending chain f_1, \dots, f_{k+1} is irreducible.

Proof. See Chapter 3 of [2]. ▮

Lemma (A1.3). Let $ASC = f_1(u, x_1), \dots, f_k(u, x_1, \dots, x_k)$ be an irreducible ascending chain, I_k be the initial of f_k , and DS be a polynomial set. Then

$$\text{Zero}(PD(ASC)/DS \cup \{I_k\}) = \text{Zero}(PD(f_1, \dots, f_{k-1}) \cup \{f_k\}/DS \cup \{I_k\}).$$

Proof. Since $PD(f_1, \dots, f_{k-1}) \subset PD(ASC)$ and $f_k \in PD(ASC)$, one direction is obvious. Let h be any polynomial in $PD(ASC)$, then $I_k^s h = Qf_k + R$ for some integer $s \geq 0$, and polynomials Q and $R \in PD(f_1, \dots, f_{k-1})$. Thus a zero of $PD(f_1, \dots, f_{k-1})$ and f_k , which is not a zero of I_k , is a zero of $PD(ASC)$. ▮

Lemma (A1.4). For any polynomial sets PS and DS and polynomials a and b , if DS contains a polynomial of form $za + b$ or $a^2 + b^2$ then

$$\text{Zero}(PS/DS) = \text{Zero}(PS/DS) \cup \text{Zero}(PS/DS \cup \{a\}) \cup \text{Zero}(PS/DS \cup \{b\}).$$

Proof. Since $\text{Zero}(PS/DS) = \text{Zero}(HS \cup \{ab\}/DS) \cup \text{Zero}(HS/DS \cup \{a\}) \cup \text{Zero}(HS/DS \cup \{b\})$, and the first is empty, the equation follows. ▮

(A1.5) Proof of Theorem (6.5). Let $S = (HS, DS, C)$ be a statement in Class C involving constructions 1–8, 9.1 and 10.1 only. We want to show either $HS \wedge DS$ is inconsistent by adding the condition that the radius of the circle is non-isotropic in construction 7, or S is irreducible in weak sense. Furthermore, in the case of consistency, S satisfies (GCW).

The proof is essentially the same as the proof of Theorem (4.1). But with advanced tools, the proof is much clearer for those who are familiar with Ritt–Wu’s decomposition.

As in proof of (4.1), we use induction. Suppose under the previous construction, we have an irreducible ascending chain:

$$\begin{aligned} & f_1(u_1, \dots, u_{j-1}, x_1) \\ & f_2(u_1, \dots, u_{j-1}, x_1, x_2) \\ (A1.6) \quad & \dots \\ & f_{k-1}(u_1, \dots, u_{j-1}, x_1, \dots, x_{k-1}), \end{aligned}$$

and $\text{Zero}(h_1, \dots, h_{k-1}/d_1, \dots, d_{l-1}) = \text{Zero}(PD(f_1, \dots, f_{k-1})/d_1, \dots, d_{l-1})$. Since constructions 1–5 only introduces new parameters, we can assume that the next step is one of the following constructions.

Construction 6. We have a new equation $h_k = f_k = ax + by + c = 0$. $d_l = az_l + b$. We consider $R_a = \text{prem}(a; f_1, \dots, f_{k-1})$ and $R_b = \text{prem}(b; f_1, \dots, f_{k-1})$.

Case 6.1. If both R_a and R_b are zero, then $\text{prem}(az_l + b; f_1, \dots, f_{k-1})$ is zero. Hence

$$\text{Zero}(h_1, \dots, h_{k-1}/d_1, \dots, d_l) = \text{Zero}(PD(f_1, \dots, f_{k-1})/d_1, \dots, d_l)$$

is empty. The hypothesis is inconsistent.

Case 6.2. One of R_a and R_b , say R_b , is zero. If we let $D = (x_k, u_j)$, then $f_k = h_k = ax_k + bu_j + c$. Then

$$\text{Zero}(PD(f_1, \dots, f_{k-1})/d_1, \dots, d_l) = \text{Zero}(PD(f_1, \dots, f_{k-1})/d_1, \dots, d_{l-1}, a).$$

By Lemma (A1.3),

$$\text{Zero}(PD(f_1, \dots, f_{k-1}), h_k/d_1, \dots, d_{l-1}, a) = \text{Zero}(PD(f_1, \dots, f_k)/d_1, \dots, d_{l-1}, a).$$

Hence $\text{Zero}(h_1, \dots, h_k/d_1, \dots, d_l) = \text{Zero}(PD(f_1, \dots, f_k)/d_1, \dots, d_l)$.

Case 6.3. For f_1, \dots, f_k is an irreducible ascending chain, considering the order $y < x$. If can consider the order $x < y$, then it is another ascending chain $f_1, \dots, f_{k-1}, f'_k$. Here f'_k is identical to f_k . We use different notations to emphasize the different leading variables. By Lemma (A1.4),

$$\begin{aligned} \text{Zero}(PD(f_1, \dots, f_{k-1}), f_k/d_1, \dots, d_l) = \\ \text{Zero}(PD(f_1, \dots, f_{k-1}), f_k/d_1, \dots, d_l, a) \cup \text{Zero}(PD(f_1, \dots, f_{k-1}), f'_k/d_1, \dots, d_l, b). \end{aligned}$$

Hence

$$\text{Zero}(h_1, \dots, h_k/d_1, \dots, d_l) = \text{Zero}((PD(f_1, \dots, f_k)/d_1, \dots, d_l, a) \cup \text{Zero}((PD(f_1, \dots, f'_k)/d_1, \dots, d_l, b)).$$

By Lemma (A1.1), $PD(f_1, \dots, f_k) = PD(f_1, \dots, f'_k)$. Hence $\text{Zero}(h_1, \dots, h_k/d_1, \dots, d_l) = \text{Zero}(PD(f_1, \dots, f_k)/d_1, \dots, d_l)$ by Lemma (A1.4).

Construction 7. We check whether CD is isotropic. If it is then $(HS \wedge DS)$ is inconsistent with $\neg\text{isotropic}(C, D)$. Otherwise we have an irreducible ascending chain f_1, \dots, f_k by Lemma (A1.2) and

$$\text{Zero}(h_1, \dots, h_k/d_1, \dots, d_{l-1}) = \text{Zero}(PD(f_1, \dots, f_k)/d_1, \dots, d_{l-1}).$$

Construction 8. Now $d_l = \Delta$. Let $R = \text{prem}(\Delta; f_1, \dots, f_{k-1})$.

Case 8.1. $R = 0$, then $\text{Zero}(h_1, \dots, h_{k-1}/d_1, \dots, d_l) = \text{Zero}(PD(f_1, \dots, f_{k-1})/d_1, \dots, d_l)$ is empty. $(HS \wedge DS)$ is inconsistent.

Case 8.2. $R \neq 0$. Let $R_a = \text{prem}(a_2; f_1, \dots, f_{k-1})$ and $R_b = \text{prem}(b_2; f_1, \dots, f_{k-1})$. Then one of R_a and R_b , say R_a , is non-zero. Let $I = (x_{k+1}, x_k)$ and

$$f_k = \Delta x_k + d$$

$$f_{k+1} = f_{k+1} = a_2 x_{k+1} + b_2 x_k + c_2.$$

where $d = b_2 c_1 - b_1 c_2$. Then we can use the same technique as in Cases 6.2–3 to prove

$$\text{Zero}(h_1, \dots, h_k, h_{k+1}/d_1, \dots, d_l) = \text{Zero}(PD(f_1, \dots, f_k, f_{k+1})/d_1, \dots, d_l).$$

Construction 9.1. Then $d_l = a_1^2 + b_1^2$ and $d_{l+1} = a z_l + b$. Let $Q = (x, y)$ and $S = (x', y')$. Then they both satisfy the following two equations.

$$\begin{aligned} h_k(x, y) &= y^2 + x^2 + ax + by + c = 0 \\ h_{k+1}(x, y) &= a_1 y + b_1 x + c_1 = 0. \end{aligned}$$

$h_k(x, y) = 0$ and $h_{k+1}(x, y) = 0$ are new hypothesis from construction 9.1. But $h_k(x', y') = 0$ and $h_{k+1}(x', y') = 0$ are the assumption, i.e.,

$$\begin{aligned} \text{prem}(h_k(x', y'); f_1, \dots, f_{k-1}) &= 0 \\ \text{prem}(h_{k+1}(x', y'); f_1, \dots, f_{k-1}) &= 0. \end{aligned}$$

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

Case 9.1. $R = 0$. Then $\text{Zero}(PD(f_1, \dots, f_{k-1})/d_1, \dots, d_l)$ is empty. $HS \wedge DS$ is inconsistent.

Case 9.2. $R \neq 0$. One of $R_a = \text{prem}(a_1; f_1, \dots, f_{k-1})$, $R_b = \text{prem}(b_1; f_1, \dots, f_{k-1})$ must be non-zero, say, R_b . (They cannot be both zero, otherwise R would be zero). We introduce two dependent variables x_k, x_{k+1} and let $Q = (x_{k+1}, x_k)$. Eliminating y in equations h_k and h_{k+1} :

$$\begin{aligned} f_k(x_k) &= d_l x_k^2 + (a_1^2 b + 2c_1 b_1 - a a_1 b_1) x_k + (a_1^2 c + c_1^2 - a c_1 a_1) \\ f_{k+1} &= a_1 x_{k+1} + b_1 x_k + c_1 = 0. \end{aligned}$$

Thus $\text{prem}(f_k(x'); f_1, \dots, f_{k-1}) = 0$. Dividing $f_k(x_k)$ by $x_k - x'$ we have

$$f_k = (x_k - x')((a_1^2 + b_1^2)x_k + c') + r(u, x_1, \dots, x_{k-1}).$$

Thus $\text{prem}(r; f_1, \dots, f_{k-1}) = 0$ and $f_k = (x_k - x')((a_1^2 + b_1^2)x_k + c')$ on $\text{Zero}(h_1, \dots, h_{k-1}/d_1, \dots, d_{l-1})$; the ascending chain f_1, \dots, f_k is reducible. Now

$$\begin{aligned} \text{(A1.7)} \quad \text{Zero}(h_1, \dots, h_{k+1}/d_1, \dots, d_{l+1}) &= \\ \text{Zero}(PD(f_1, \dots, f_{k-1}), d_l x_k + c', f_{k+1}/d_1, \dots, d_{l+1}) &\cup \\ \text{Zero}(PD(f_1, \dots, f_{k-1}), x_k - x', f_{k+1}/d_1, \dots, d_{l+1}). & \end{aligned}$$

By the same technique as used in cases 6.2–6.3

$$\text{Zero}(PD(f_1, \dots, f_{k-1}), d_l x_k + c', f_{k+1}/d_1, \dots, d_l, d_{l+1}) = \text{Zero}(PD(ASC_1)/d_1, \dots, d_l, d_{l+1})$$

$$\text{Zero}(PD(f_1, \dots, f_{k-1}), x_k - x', f_{k+1}/d_1, \dots, d_l, d_{l+1}) = \text{Zero}(PD(ASC_2)/d_1, \dots, d_l, d_{l+1}).$$

where $ASC_1 = f_1, \dots, f_{k-1}, d_l x_k + c', f_{k+1}$ and $ASC_2 = f_1, \dots, f_{k-1}, x_k - x', f_{k+1}$. By calculation, $\text{prem}(d_{l+1}; f_{k+1}, x_k - x') = a_1 x' + b_1 y' + c_1$, Thus $\text{prem}(d_{l+1}; ASC_2) = 0$ by the assumption. Thus the second set on the right hand side of (A1.7) is empty and we have:

$$\text{Zero}(h_1, \dots, h_{k+1}/d_1, \dots, d_{l+1}) = \text{Zero}(PD(f_1, \dots, f_{k+1})/d_1, \dots, d_{l+1}).$$

where $f_k = d_1 x_k + c'$. Since $\deg(f_k, x_k) = \deg(f_{k+1}, x_{k+1}) = 1$, f_1, \dots, f_{k+1} is irreducible.

Construction 10.1. This is similar to Construction 10 in the proof of (4.1) and Construction 9.1 above.

Repeating the above process, we finally have an irreducible ascending chain (or the hypothesis is inconsistent) of form (4.5) and prove that S is irreducible in weak sense.

To prove S satisfies (GCW), we note that all $\deg(f_k, x_k) = 1$ but for Construction 7. By the same argument as in (5.4), f_1, \dots, f_r is \mathbf{R} -generic. ■

Appendix 2. More Construction Sequences for Simson's Theorem

As we pointed out in (7.2) of [4], "here are at least eight essentially different construction sequences for the configuration of Simson's theorem that preserve the original meaning of the theorem." Here we list them in detail.

(A.1) The construction sequence is the one listed in Section 3.4.

(A.2)
 Points O and A are arbitrarily chosen;
 B is on circle $(O, (OA))$;
 C is on circle $(O, (OA))$;
 D is on circle $(O, (OA))$;
 $E = T(D, BC) \cap L(BC)$;
 $F = T(D, AC) \cap L(AC)$;
 $G = T(D, AB) \cap L(AB)$.

(A2.3)
 Points A and B are arbitrarily chosen;
 O is on $B(AB)$;
 C is on circle $(O, (OA))$;
 D is on circle $(O, (OA))$;
 $E = T(D, BC) \cap L(BC)$;
 $F = T(D, AC) \cap L(AC)$;
 $G = T(D, AB) \cap L(AB)$.

(A2.4)
 Points A, B , and E are arbitrarily chosen;
 O is on $B(AB)$;
 $C = L(EB) \cap (O, (OA))$;
 $D = T(E, BC) \cap (O, (OA))$;
 $F = T(D, AC) \cap L(AC)$;
 $G = T(D, AB) \cap L(AB)$.

The construction of point C is reducible, though the prover confirmed the theorem. The reader is strongly recommended to study this case.

(A2.5)
 Points A and B are arbitrarily chosen;

O is on $B(AB)$;
 C is on circle $(O, (OA))$;
 E is on $L(BC)$;
 $D = T(E, BC) \cap (O, (OA))$;
 $F = T(D, AC) \cap L(AC)$;
 $G = T(D, AB) \cap L(AB)$.

(A2.6)

Points A, B , and E are arbitrarily chosen;
 C is on $L(EB)$;
 $O = B(AB) \cap B(AC)$;
 $D = T(E, BC) \cap (O, (OA))$;
 $F = T(D, AC) \cap L(AC)$;
 $G = T(D, AB) \cap L(AB)$.

(A2.7)

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

(A2.8)

Points A and G are arbitrarily chosen;
 B is on $L(GA)$;
 D is on $T(G, AB)$;
 O is on $B(AD) \cap B(AB)$;
 C is on circle $(O, (OA))$;
 $E = T(D, BC) \cap L(BC)$;
 $F = T(D, AC) \cap L(AC)$.

Note that the equation part, i.e., HS_s in Section 4.3, is the same for all these 8 constructive statements of Simson's theorem. But the 8 different constructive sequences give slightly different the inequation parts (non-degenerate conditions). The simplest is perhaps the non-degenerate condition that $DS = \neg\text{isotropic}(AB) \wedge \neg\text{isotropic}(AC) \wedge \neg\text{isotropic}(BC)$ ". Here we see that the conclusion " A, B and C are collinear" follows from HS_s and DS even without assuming " $\text{points } A, B, \text{ and } C \text{ are not collinear}$ ".