ON THE MECHANICAL PROOF OF GEOMETRY THEOREMS INVOLVING INEQUALITIES*

Shang-Ching Chou, Xiao-Shan Gao,†
and Dennis S. Arnon††

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

TR-89-31 October 1989

ABSTRACT

In the past decade highly successful algebraic methods for mechanical geometry theorem proving have been developed. The first step in these methods is to assign (variable) coordinates to key points, and then translate the hypotheses and conclusion of a geometric proposition into (multivariate) polynomial equations and inequalities, Next, the algebraic provers apply either the "Wu-Ritt" or "Gröbner Basis" method to analyze zeros of polynomials. To date, the manner in which the Wu-Ritt and Gröbner Basis methods have been employed has limited the algebraic provers to propositions that can be encoded entirely with polynomial equations, i.e. without inequalities. In this paper we explore two techniques for extending Wu-Ritt and Gröbner provers to handle propositions involving inequalities: reduction of polynomials to canonical form modulo a (polynomial) ideal, and the Rabinowitsch/Seidenberg device of converting (polynomial) inequalities to equations by introducing new variables. We illustrate the practical value of these techniques by numerous examples of their use in conjunction with a Wu-Ritt prover.

Keywords: Geometry Theorem Proving, Automated Theorem Proving, Decision Procedures, Wu-Ritt Method, Gröbner Bases, Constructive Polynomial Ideal Theory, Euclidean Geometry, Computational Algebraic Geometry, Computer Algebra, Collins Method, Cylindrical Algebraic Decompositions, Definiteness of Polynomials.

---

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

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

†† Xerox Palo Alto Research Center, Palo Alto, California 94304 USA.

On the Mechanical Proof of Geometry Theorems Involving Inequalities*

SHANG-CHING CHOU and XIAO-SHAN GAO†
Department of Computer Sciences
The University of Texas at Austin, Austin, Texas 78712 USA

DENNIS S. ARNON
Xerox Palo Alto Research Center, Palo Alto, California 94304 USA

Abstract In the past decade highly successful algebraic methods for mechanical geometry theorem proving have been developed. The first step in these methods is to assign (variable) coordinates to key points, and then translate the hypotheses and conclusion of a geometric proposition into (multivariate) polynomial equations and inequalities, Next, the algebraic provers apply either the "Wu–Ritt" or "Gröbner Basis" method to analyze zeros of polynomials. To date, the manner in which the Wu–Ritt and Gröbner Basis methods have been employed has limited the algebraic provers to propositions that can be encoded entirely with polynomial equations, i.e. without inequalities. In this paper we explore two techniques for extending Wu–Ritt and Gröbner provers to handle propositions involving inequalities: reduction of polynomials to canonical form modulo a (polynomial) ideal, and the Rabinowitsch/Seidenberg device of converting (polynomial) inequalities to equations by introducing new variables. We illustrate the practical value of these techniques by numerous examples of their use in conjunction with a Wu–Ritt prover.

Keywords Geometry Theorem Proving, Automated Theorem Proving, Decision Procedures, Wu–Ritt Method, Gröbner Bases, Constructive Polynomial Ideal Theory, Euclidean Geometry, Computational Algebraic Geometry, Computer Algebra, Collins Method, Cylindrical Algebraic Decompositions, Definiteness of Polynomials.

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

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

1. Introduction

In 1977 Wu Wen-tsün introduced an algebraic method that he used to prove quite nontrivial theorems of Euclidean geometry [33]. Wu's own further investigations of the method ([35], [34]), which we refer to as the *Wu–Ritt* method in this paper, have yielded proofs of numerous statements in both unordered and ordered geometries. The method is most clearly applicable to geometric propositions in which all variables are universally quantified, and whose algebraic encodings involve polynomial equations and inequations[1] (as opposed to inequalities). The Wu–Ritt method is complete for such statements in geometries whose associated fields are algebraically closed. In ordered geometries, whose associated fields are not algebraically closed, it can, in general, only confirm such statements. However, it can disprove an ordered geometry assertion provided the hypotheses have certain properties (e.g. they are "irreducible" and have real generic solutions). These properties are explained fully in [13].

All told, hundreds of theorems of Euclidean geometry have been proved to date by the Wu–Ritt method [6], [12], [13], [21], as well as numerous theorems from various non-Euclidean geometries [9], [20], [22]. Yet in spite of its success for "universally quantified, equational" propositions, it has so far seemed unable to handle geometric statements involving existential quantifiers (although [25] offers some preliminary ideas on this issue) or inequalities. A typical example of a statement involving inequalities is: "in a triangle, a greater side is opposite to a greater angle".

Besides the Wu-Ritt method, there are also algebraic methods for geometry theorem proving that rely on Gröbner Bases [3], [5]. The success of these methods for "universally quantified, equational" propositions has been similar to that of the Wu-Ritt method, as has their confinement so far to "universally quantified, equational" propositions. On the other hand, Tarski's decision procedure for real closed fields [31] gives a complete algorithm for proving or disproving (i.e. deciding) any statement (including one involving inequalities and existential quantifiers) in what Tarski called "elementary geometry" [32]. But in spite of the discovery of new and much more efficient decision procedures for real closed fields by G. Collins [18]and others, this approach (and its implementations) needs further improvement to be able to prove a significant number of nontrivial geometry theorems in practice.[2] For the interested reader, we note that [4] provides a "one-stop" survey of the Wu-Ritt method, Gröbner Bases, and the Collins method.

The present paper explores two techniques for proving (universally quantified) geometric statements whose algebraic formulations involve inequalities: reduction of polynomials to canonical form modulo a (polynomial) ideal, and the Rabinowitsch/Seidenberg device of converting (polynomial) inequalities to equations by introducing new variables. Of course, the statements that come within the scope of these techniques are a proper subset of those amenable to a decision procedure for real closed fields. But for those statements that we can treat (and which happen to be true), our techniques often lead to straightforward confirmations, in which the major component (of the actual computation) is an application of the Wu–Ritt method to a subsidiary problem. For example, in an application of the "reduction to canonical form" technique, the work remaining after completion of the Wu–Ritt method may be only to certify the definiteness

---

[1] An *inequation* is an instance of the "$\neq$" relation; an *inequality* is an instance of any of the "$<$", "$>$", "$\leq$", or "$\geq$", relations.

[2] We remark however that solutions of nontrivial quantifier elimination problems of elementary geometry and algebra have recently been obtained using an implementation of Collins' method. See [1], [2], and [19].

of some "critical" polynomial, and this definiteness may turn out either to be evident, or provable by an easy application of a decision procedure (such as Collins').

The present paper seems to be the first published investigation of the reduction to canonical form technique for proving statements involving inequalities. The Rabinowitsch device (for inequations only) originated in the 1930's in the proof of Hilbert's Nullstellensatz. It was first used in geometry theorem proving (for statements involving equations) by Chou and Schelter [16] and Kapur [24]. Most recently it has been used by Kutzler [27]. Seidenberg "extended" Rabinowitsch's device to inequalities, as part of his improved (over Tarski's) decision procedure for real closed fields [30]. Seidenberg's device was first used in geometry theorem proving (for statements with inequalities) by Wu (see Section 5.3 of [34], and also [35]) and Chou (see Chapter 5 of [7] and [11]), and subsequently by Kapur and Mundy [26]. Since for our purposes in this paper we do not need to distinguish the two forms of the device, we call it simply the "Rabinowitsch/Seidenberg" device.

Although the prior work we have cited contains various examples and ruminations relating to proofs of statements involving inequalities, it may fairly be said that the present paper is the first attempt to tackle the subject in a comprehensive fashion. We note that Wu has recently proposed techniques different from those of the present paper, which, though incomplete in many respects, can be used to prove quite a few nontrivial theorems involving inequalities with certain human interactions [37].

The examples and empirical experience we report with our techniques in this paper have all used a Wu–Ritt prover, but there are natural and evident variations of them for use in conjunction with a Gröbner Basis prover. An investigation of the use of (suitable variations of) our techniques in conjunction with a Gröbner Basis prover should be undertaken. In point of fact, since we already use the Gröbner Basis algorithm for reduction of polynomials to canonical form modulo an ideal, the present paper may be said to offer an approach to the mechanical proof of geometry theorems involving inequalities that blends the Wu–Ritt, Gröbner Basis, and Collins methods.

In Section 2 we will specify the geometric statements that fall within our scope and partition them into four categories. These categories are the 2 x 2 possible combinations of (hypotheses are/are not equations) and (conclusion is/is not an equation). We accordingly label them EE, IE, EI, and II. In Section 3 we review the treatment of type EE statements by the original Wu–Ritt method, thereby establishing the framework for the application of our two "inequality" techniques to proofs of EI, IE and IE statements. In Section 4 we illustrate the use of our techniques in proofs of EI and II statements, and in Section 5 we consider IE situations, In Section 6 we offer some concluding observations. The Appendix explains the Wu–Ritt method in some detail. In particular, it presents the modified form of the Wu–Ritt method that is implemented by our computer programs.

2. Categories of Geometric Statements

We wish to address geometric statements of the following algebraic form in ordered geometries:

$$\forall v_i (H \Rightarrow C), \tag{2.1}$$

where H is a set of polynomial equations and inequalities in variables (coordinates) $v_1, ..., v_n$ and

C is a polynomial equation or inequality of rational functions in $v_1, ..., v_n$.[3] Formulation (2.1) is not complete because it does not provide for additional nondegeneracy (or subsidiary) conditions in the hypotheses, which typically are required for the statement to be true. If we distinguish independent variables (parameters) and dependent variables among the v_i, it becomes possible to identify these nondegeneracy conditions. In this paper, as in the Wu–Ritt method, we will so partition variables into these two classes. See Section 3 for details on how this is done, and on certain unavoidable issues that arise in doing so. We will use $u_1, ..., u_d$ to denote the parameters and $x_1, ..., x_r$ to denote the dependent variables.

Definition (2.2). Let S be a geometric statement whose algebraic form is (2.1).

(A) If H contains only equations and C is also an equation, then we call S a geometric statement of type EE.

(B) If H contains only equations and C is an inequality, then we call S a geometric statement of type EI.

(C) If H contains equations as well as inequalities and C is an equation, then we call S a geometric statement of type IE.

(D) If H contains equations as well as inequalities and C is an inequality, then we call S a geometric statement of type II.

In cases (C) and (D), the set of all equations in H is called *the equality (equation) part* of the hypotheses, and the set of all inequalities in H is called *the inequality part* of the hypotheses.

3. Proving Statements of Type EE

In this Section we consider in more detail the algebraic formulation of geometric statements of type EE, and the two essentially different approaches to the mechanical confirmation of such statements that have been utilized to date. The the original method developed by Wu [33] [34] is an instance of the first approach, and so in the course of a discussion of the first approach in Sections 3.1 and 3.2 we will sketch its main features. In Section 3.3 we examine the second approach. The reader who is familiar with these matters can skip this section. The reader who wishes a more detailed presentation can consult [34], [13], [17], [15].

As in Section 2, EE statements are those which can be algebraically expressed as

$$\forall v_i \in F(H \Rightarrow g = 0), \tag{3.1}$$

where $H = \{h_1 = 0, ..., h_q = 0\}$ is a set (conjunction) of polynomial equations in the variables $v_1, ..., v_n$; g is also a polynomial equation in the variables v; F is the field associated with the given geometry. We here assume that F *is an algebraically closed field containing* $\mathbf{Q}$; we will specifically discuss the case when F is the field of real numbers $\mathbf{R}$ in Section 3.2.

Formula (3.1) can be equivalently expressed as:

$$Zero(H) \subset Zero(g), \tag{3.2}$$

[3] The coefficients of all polynomials occurring in H and C are in $\mathbf{Q}$, the field of rational numbers.

Here $Zero(H)$[4] denotes the common zeros of $h_1, ..., h_q$ in the field F, i.e.,

$$Zero(H) = \{(v_1, ..., v_n) \in F^n \mid h_i(v_1, ..., v_n) = 0 \text{ for } i = 1, ..., q\}.$$

As we have already said, (3.1) or (3.2) is generally not an exact algebraic formulation of the original geometric statement, because nondegeneracy conditions usually need still to be added to the hypotheses before the statement can be confirmed. Unfortunately, suitable nondegeneracy conditions are often not obvious. For a given geometric proposition, a nondegeneracy condition that is obvious to one person might not be obvious to a second, and a third person might refuse to accept the condition as relevant or appropriate. A further complication, especially since much of the activity in mechanical geometry theorem proving to date has been directed at propositions that are well known, is the fact that most traditional geometry textbooks ignore the possible need to add nondegeneracy conditions to the hypotheses of the theorems they prove. A final difficulty is the fact that seemingly artificial nondegeneracy conditions must sometimes be added for the Wu–Ritt method, or the Gröbner Bases methods, to confirm a statement.

The two essentially different approaches to confirmation of geometric statements that we mentioned above arise as two different ways of dealing with the matter of nondegeneracy conditions:

Approach (1). Introduce parameters, dependent variables, and the notion of a "generally true" statement, and (attempt to) prove (given) geometric statements to be generally true. Nondegeneracy conditions are automatically constructed during the proof process.

Approach (2). Explicit nondegeneracy conditions must be added to the (hypotheses of the) algebraic formulation of the geometric statement to be proved before the prover is invoked. Thus it is the task of a person, or some program, to "find", i.e. "propose", nondegeneracy conditions. Then the prover need only determine whether the conclusion follows from the hypothesis, and need not produce nondegeneracy conditions. If the prover fails to confirm a statement with the supplied nondegeneracy conditions, one can "go back to the drawing board", propose new nondegeneracy conditions, and run the prover again.

In this paper we use the first approach, which we now examine briefly in the next two Sections. For more detailed discussions and examples, see [17], [13], and [15].

3.1. Approach (1)

With approach (1), the first thing is to determine the parameters. For a given geometric statement, the selection of the parameters is not unique, but is determined by the *meaning* of the statement. The parameters are to be those variables that can take on arbitrary, i.e. general, values. The dependent variables are the remaining variables, whose values for any particular choice of parameter values are determined by the hypothesis set H of polynomial equations.

According to a fundamental theorem in algebraic geometry, the algebraic set $Zero(H)$ can be decomposed into (irredundant) irreducible components:

$$Zero(H) = V_1^* \cup \cdots \cup V_c^* \cup V_1^{\text{dege}} \cup \cdots \cup V_l^{\text{dege}}, \tag{3.3}$$

[4] Here H denotes the polynomial set $\{h_1, ..., h_q\}$. In this paper, the notation for a set of polynomial equations (or inequalities) can denote its corresponding polynomial set (and vice versa), depending on the context.

where $V_1^*, ..., V_c^*$, $c \geq 0$, are all those components of dimension d on which the parameters u are algebraically independent. A necessary algebraic criterion for the correctness of a particular choice of parameters $u_1, ..., u_d$ is:

Criterion (3.4). (1) $c > 0$, i.e., there is at least one component with dimension d on which the parameters $u_1, ..., u_d$ are algebraically independent, and (2) $Zero(H)$ does not contain a component of dimension $> d$ on which the u are algebraically independent.

In the remainder of this paper *we always assume that this algebraic Criterion holds.* Thus, the u are algebraically dependent on all V_i^{dege}. [5]

Now it is clear that all V_i^* correspond to nondegenerate cases of the geometric configuration given by H and all V_i^{dege} correspond to degenerate cases of the given configuration. We call them *nondegenerate components (cases)* and *degenerate components (cases)*, respectively. Usually, we don't care whether the conclusion is valid in the degenerate cases; it may, for example, be the case that the geometric problem is meaningless in the degenerate cases. We can now sharpen (3.1) or (3.2) as follows: what we actually wish to establish, in order to prove a given proposition, is that

$$V_1^* \cup \cdots \cup V_c^* \subset Zero(g),$$

i.e., whether $g = 0$ is valid on all the nondegenerate cases. If it is, then we say that the geometric statement is *generally true.*

As we know, the Ritt–Wu Decomposition Algorithm ([29], [34], [36], [13], [15]; see also Section 4 of the Appendix) realizes Decomposition (3.3). For any ascending chain ASC let us define

$$PD(ASC) = \{g \mid prem(g, ASC) = 0\}$$

(see the Appendix for definitions assumed here). If ASC is irreducible, $PD(ASC)$ is a prime ideal having the ascending chain ASC as an irreducible characteristic set. The Ritt–Wu Decomposition Algorithm realizes Decomposition (3.3) as follows: for $i = 1, .., c, j = 1, ..., l$, there are irreducible ascending chains ASC_i^* and ASC_j, such that $V_i^* = Zero(PD(ASC_i^*))$ and $V_j^{\text{dege}} = Zero(PD(ASC_j^{\text{dege}}))$. Recall that each ascending chain ASC_i^{dege} contains at least a nonzero polynomial involving the u only, and that all ASC_i^* have the following form:

$$\begin{aligned} & f_1(u_1, \ldots, u_d, x_1) \\ & f_2(u_1, \ldots, u_d, x_1, x_2) \\ & \qquad \cdots \\ & f_r(u_1, \ldots, u_d, x_1, \ldots, x_r). \end{aligned} \tag{3.5}$$

Thus we have:

Method (3.6). *for Deciding a Type EE Statement to be Generally True.* To decide whether a type EE geometric statement is generally true is equivalent to deciding whether $Zero(PD(ASC_i^*)) \subset Zero(g)$, i.e., whether $prem(g; ASC_i^*) = 0$ for all $i = 1, ..., c$. .QED.

[5] For completeness, we note here two rather subtle facts: (1) Criterion (3.4) is a necessary but not a sufficient condition for the parameters $u_1, ..., u_d$ to be correctly chosen in geometry. (2) A V_i^{dege} can have dimension $> d$.

Almost all of the 512 examples in [13] are irreducible in the sense that there is only one non-degenerate component (i.e., $c = 1$). We call such a geometric statement *an irreducible statement*, and its equation part H *irreducible*. For a given equation part H, such irreducibility depends on the choice of the parameters. Having chosen parameters, one can quickly obtain their corresponding ascending chains ASC_1^* by the "triangulation" algorithm of the Wu-Ritt method (see Section 3 of the Appendix). Depending on the problems, the computer time for calculation of $prem(g; ASC_1^*)$ usually ranges from a fraction of one second to several minutes.

The general truth of $g = 0$ is independent of the particular algebraically closed field F. In other words, $g = 0$ is generally true in one algebraically closed field F if and only if it is also generally true in any other algebraically closed field. This can be seen from (3.6). Actually we have the following theorem:

Theorem (3.7). $g = 0$ is generally true under H and the parameters u if and only if there is a nonzero polynomial s involving the u only (we will call such a polynomial a u–polynomial), such that $s \cdot g \in Radical(H)$. In that case, we have

$$\forall ux \in E[(H \wedge s \neq 0) \Rightarrow g = 0], \tag{3.8}$$

where E is *any extension* (including $\mathbf{R}$) of $\mathbf{Q}$.

For the proof of the theorem, see p.49 of [13]. Actually, let s_i be a u–polynomial in ASC_i^{dege}, then $s = s_1 \cdots s_l$ is the u–polynomial mentioned in the theorem. Conditions $s_1 \neq 0$,..., $s_l \neq 0$ are usually connected with nondegeneracy. This is an instance of how Approach (1) identifies nondegeneracy conditions automatically.

The following two theorems give methods for deciding the "general validity" of a geometric statements of type EE using Gröbner Bases. Let z be a new variable other than the u and x.

Theorem (3.9). $g = 0$ is generally true under H and the parameters if and only if $\{h_1, ..., h_q, gz - 1\}$ generates the unit ideal in $\mathbf{Q}(u_1, ..., u_d)[x_1, ..., x_r, z]$.

Theorem (3.10). $g = 0$ is generally true under H and the parameters if and only if a Gröbner basis of $\{h_1, ..., h_q, gz - 1\}$ (in $\mathbf{Q}[u_1, ..., u_d, x_1, ..., x_r, z]$) in a compatible ordering $u < x$ contains a u–polynomial.

For the proofs of these two theorems, see Chapter 5 of [13]. The initial idea of the method based on Theorem (3.10) belongs to Kapur [24]. But Kapur's method lacks a proper theoretical basis, and can sometimes "confirm" a geometric statement which is not generally true. Theorem (3.10) is due to Chou. The method based on (3.10) is geometrically sound. In practice, the method based on (3.9) is generally much faster than that based on (3.10). Among 512 theorems in [13] proved by the Wu–Ritt method, 477 have been also proved by the method based on (3.9), with exactly the same polynomial equations and parameters as used with Wu's method.

3.2. The Case When F is $\mathbf{R}$.

Now we generalize Approach (1) to the case when F is the field of real numbers $\mathbf{R}$. Let us begin with two observations. First, if a geometry statement is proved to be generally true when F is an algebraically closed field, then the statement is also generally true when F is $\mathbf{R}$. Second, if (3.8) can be inferred with the methods based on approach (1) when F is an algebraically closed field, then (3.8) can be inferred by the same methods when F is $\mathbf{R}$. For details, see [17]or [13]. Thus the Wu-Ritt method can be used to confirm geometric statements in Euclidean geometry.

We next introduce the following notion.

Definition (3.11). A geometric configuration defined by H and the parameters $u_1, ..., u_d$ is said to be *generic* in $\mathbf{R}$ if each ASC_i^* $(i = 1, ..., c)$ has a zero $\mu = (\tilde{u}_1, \ldots, \tilde{u}_d, \tilde{x}_1, \ldots \tilde{x}_r) \in \mathbf{R}^n$, with $\tilde{u}_1, \ldots, \tilde{u}_d$ algebraically independent. This is equivalent to the requirement that for each ASC_i^* $(i = 1, ..., c)$ there are open intervals O_j $(j = 1, ..., d)$ in $\mathbf{R}$, such that whenever $u_j \in O_j$, $j = 1, ..., d$, ASC_i^* has solutions for $x_1, ..., x_r$ in $\mathbf{R}$.

We have the following theorem.

Theorem (3.12). For a generic statement in $\mathbf{R}$, to decide whether it is generally true is equivalent to deciding whether $prem(g; ASC_i^*) = 0$ for all $i = 1, ..., c$.

Intuitively, a generic configuration corresponds to a diagram which can be drawn on a paper (the real plane). Almost all geometric statements in geometry textbooks are generic. This is the real reason why the Wu–Ritt method, which is complete only for complex geometry, is so successful in confirming theorems in Euclidean geometry. Assuming a geometric configuration to be generic, the Wu–Ritt method is also a method for disproving any assertions on that configuration. However, to decide that a geometric configuration defined by H and parameters $u_1, ..., u_d$ is generic is beyond the Wu–Ritt method.

3.3 Approach (2).

Let $D = \{d_1 \neq 0, ..., d_p \neq 0\}$ be the set of inequations corresponding to nondegeneracy conditions specified manually. Approach (2) is to decide whether

$$\forall v_i \in F[(H \wedge D) \Rightarrow g = 0].$$

Or in a equivalent form:

$$Zero(H/D) \subset Zero(g),$$

where $Zero(H/D)$ denotes $Zero(H) - Zero(D)$. In the light of Approach (1), D should usually be chosen in such way that (1) $V_1^{\text{dege}} \cup \cdots \cup V_l^{\text{dege}} \subset Zero(D)$; and (2) No V_i^* is a subset of $Zero(D)$, unless we want purposely to exclude some general case V_i^*. Indeed, as we will see in Section 5, one way that inequalities arise in geometry statements is precisely for the purpose of excluding some general cases V_i^*. Thus, we can use a mixture of approaches (1) and (2) to exclude such unwanted cases V_i^*. See Example 6 in Section 4.3.

4. Proving Type EI and Type II Statements

We divide this section into three subsections. In Section 4.1 we give a simple working example to illustrate how we employ our techniques. In Section 4.2 we state a general "method" for type EI and II statements that the example motivates. Section 4.3 provides additional examples, presented in detail to enable others to experiment with them.

4.1. An Example of Type EI

Example (4.1). Let $ABCD$ be a parallelogram. Show that points B and D are on either side of diagonal AC (Figure 1).

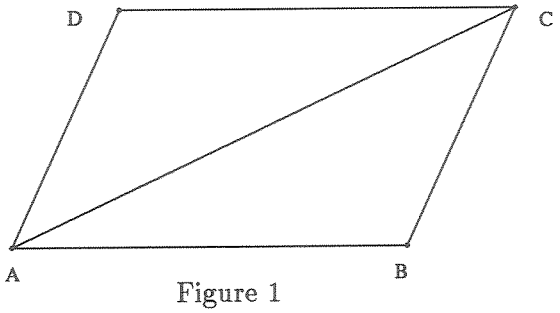

Figure 1

This "trivial" fact is repeatedly used in traditional proofs of the properties of a parallelogram, e.g., in proving the two diagonals bisect each other. However, it seems nontrivial to find a rigorous traditional proof of this fact. (Try it!) This example can be specified as type EI as follows.

Let $A = (0,0)$, $B = (u_1,0)$, $C = (u_2,u_3)$, and $D = (x_2,x_1)$. Then we have two equations for the hypotheses

$$h_1 = u_1x_1 - u_1u_3 = 0 \qquad AB \text{ is parallel to } CD$$
$$h_2 = u_3x_2 - (u_2 - u_1)x_1 = 0 \qquad AD \text{ is parallel to } BC.$$

The conclusion that B and D are on either side of AC is $g < 0$, where $g = (u_3u_1 - u_2 \cdot 0)(u_3x_2 - u_2x_1) = u_1u_3^2x_2 - u_1u_2u_3x_1$. We want to decide whether the following statement is valid under certain nondegeneracy conditions:

$$\forall u_1u_2u_3x_1x_2[(h_1 = 0 \wedge h_2 = 0) \Rightarrow g < 0].$$

Here u_1, u_2, u_3 are selected to be parameters, and x_1 and x_2 are selected to be dependent variables. Reducing g to canonical form modulo the ideal (in $Q(u)[x]$) generated by h_1 and h_2 (this is called the "canonical simplification" of g with respect to this ideal by [5]), we obtain $g = -u_1^2u_3^2$. Alternatively, we could solve $x_1 = u_1$, $x_2 = u_2 - u_1$ and substitute the solution into g to get $g = -u_1^2u_3^2$. This canonical form of g modulo the ideal is only valid under the conditions $u_1 \neq 0$ and $u_3 \neq 0$; in other words, u_1 and u_3 occur in the denominators of the elements c_1 and c_2 of $Q(u)[x]$ such that $g = c_1h_1 + c_2h_2$. Thus we have $g < 0$, under the condition that $u_1u_3 \neq 0$. Note that $u_1u_3 \neq 0$ is indeed connected with nondegeneracy, i.e. to insure that points A, B and C are not collinear.

4.2. Methods for Proving Type EI and II Statements

As for type EE statements, we first give an exact formulation of type EI statements. The algebraic form for EI statements can be

$$\forall v_i \in F(H \Rightarrow g(u,x) \geq 0), \tag{4.2}$$

where $H = \{h_1 = 0, ..., h_q = 0\}$ is a set of polynomial equations, and g is a rational function in the variables u and x.

Definition (4.3). Let the notation be the same as before, let $g(u,x)$ be a rational function with $D_g(u,x)$ as the denominator; let d be a polynomial. g is said to be *generally semi positive definite* under a set of equations H, $d \neq 0$ and the parameters $u_1, ..., u_d$ if

$$(4.4) \qquad \forall ux[((u,x) \in Zero(PD(ASC_i^*)) \wedge D_g(u,x) \neq 0 \wedge d \neq 0 \Rightarrow g(u,x) \geq 0], \text{ for } i = 1,...c.$$

Here zeros of $Zero(PD(ASC_i^*))$ are from $\mathbf{R}$. In that case we say that (4.2) is *generally true* under $d \neq 0$. We can also define "*generally positive definite*", replacing $g \geq 0$ by $g > 0$. Similarly, we can define "*generally semi negative definite*" and "*generally negative definite*".

We now state a general "method" based on the above example. Without loss of generality, we assume that g is asserted to be semi-definite.

Method (4.5). *For Confirming Type EI statements to be generally true.*

First set polynomial d, which eventually expresses the nondegeneracy conditions of the problem, to be 1.

Step 1. Use the Ritt–Wu Decomposition Algorithm (see (4.3) and (4.4) in the Appendix) to decompose H to obtain all nondegenerate components with corresponding ascending chains ASC_1^*,..., ASC_c^*. Since we only need nondegenerate components $Zero(PD(ASC_i^*))$, whenever a u–polynomial U appears in a polynomial set to be further decomposed, we can stop decomposition of that polynomial set further, setting $d = d \cdot U$. This can speedup the original Ritt–Wu algorithm greatly. We call this modified Ritt–Wu Decomposition Algorithm the *General Component Decomposition Algorithm* (see [13]).

Step 2. If the decomposition does not satisfy Criterion (3.4), stop the process. Something might be wrong in specifying the problem or parameters.

Step 3. Reduce $g(u,x)$ to canonical form modulo the ideal generated by each ascending chain ASC_i^* (with respect to the ring $Q(u)[x]$) to obtain a rational function $g_i(u,x)$. Such a reduction of g was called the "evaluation" of $g(u,x)$ on $PD(ASC_i^*)$ in [8]. The g_i are simplifications of g, and hopefully contain the parameters only (this is indeed the case for almost all 35 examples in [10]). We have

$$\forall ux[((u,x) \in Zero(PD(ASC_i^*)) \wedge d_i \neq 0) \Rightarrow g(u,x) = g_i(u,x)], \text{ for all } i = 1,...c,$$

where each d_i is some u–polynomial.

Step 4. To decide whether $g_i(u,x)$ is semi-definite on $Zero(PD(ASC_i^*))$ under $d_i \neq 0$, we can use Collins' method. It is our intuition that the decision of the semi-definiteness of g_i on $Zero(PD(ASC_i^*))$ will be an easier problem. For example, in the linear case of ACS_i^* (i.e., the degree of each leading variable in ASC_i^* is 1), g_i contains the parameters only, so g_i should be globally semi-definite, i.e., $\forall u \in \mathbf{R}(g_i(u) \geq 0)$ or $\forall u \in \mathbf{R}(g_i(u) \leq 0)$. These latter sorts of decisions we can expect to be relatively easy for the Collins method.

Step 5. If the problem is to decide whether g is generally definite under some inequation $d \neq 0$, we can first confirm whether g is generally semi-definite. If so, we try to further confirm that $d \neq 0 \Rightarrow g_i \neq 0$ under $PD(ASC_i^*)$ and $d_i \neq 0$.

Method (4.6). *For Confirming Type II Statements to be Generally True.*

Step 1. Use Seidenberg's device for converting inequalities to equations by introducing new auxiliary variables y_k. For example, in a real closed field, $h > 0$ if and only if $\exists y(y^2h - 1 = 0)$. After this transformation, the hypothesis becomes a set of equations H', and the statement (2.1)

becomes

$$\forall u_i x_j [(\exists y_k H') \Rightarrow g > 0].$$

This is equivalent to

$$\forall y_k u_i x_j [H' \Rightarrow g > 0],$$

because g is free of the variables y.

Step 2. We arrange the variables in the order $y < u < x$. In this order, some independent variables will become dependent variables, renamed as x. The last formula is of type EI, and we can use the method (4.5) for type EI to tackle it.

Methods (4.5) and (4.6) turn out to be practical as illustrated by many nontrivial examples. In the next subsection, we will give 6 examples to show how the methods can be used to prove nontrivial theorems.

In Method (4.6), i.e. the II case, there is a potential concern that the new auxiliary variables y_k, which obviously have no geometric significance, might appear in some nontrivial fashion in nondegeneracy conditions. In our experience so far, e.g. as illustrated by the examples in Section 4.3, this has not occurred.

4.3. Examples of Type EI and II

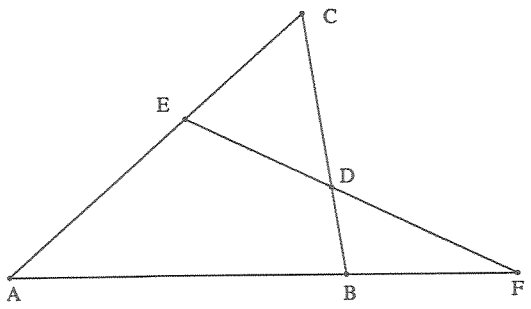

Figure 2

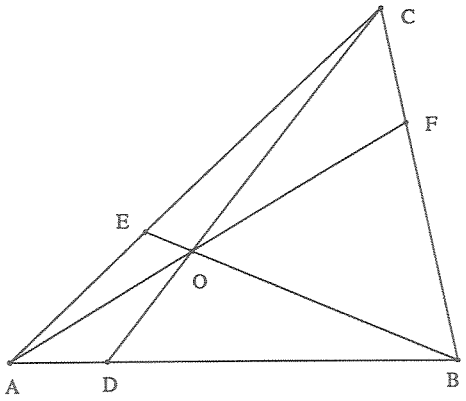

Figure 3

Example 1 (Pasch). A line intersects the three sides BC, CA and AB of triangle ABC at D, E and F, respectively. If D is between B and C, and E is between C and A, then F is outside segment AB (Figure 2).

We can let $A = (0,0)$, $B = (u_1,0)$, $C = (u_2,u_3)$, $D = (x_2,x_1)$, $E = (x_4,x_3)$, and $F = (x_5,0)$. To express "D is between B and C" and "E is between C and A", we can use (besides the equation part) the inequalities $(x_2 - u_1)(u_2 - x_2) > 0$ and $(x_4 - u_2)(0 - x_4) > 0$, or alternatively, the inequalities $(x_2 - u_1)/(u_2 - x_2) > 0$ and $(x_4 - 0)/(u_2 - x_4) > 0$, respectively. [6] Introducing two new variables y_1 and y_2, we can convert these two inequalities into equations $(x_2 - u_1)y_1^2/(u_2 - x_2) - 1 =$

[6] We prefer the latter because it can reduce the degrees of leading variables. Besides, cancellation of some common factors of denominators and numerators can reduce the sizes of polynomials.

0 and $(x_4 - 0)y_2^2/(u_2 - x_4) - 1 = 0$, or in polynomial forms, $(y_1^2 + 1)x_2 - u_2 - y_1^2 u_1 = 0$ and $(y_2^2 + 1)x_4 - u_2 = 0$. [7] Thus the hypotheses can be expressed by the following set of equations H:

$$
\begin{aligned}
h_1 &= u_3x_2 + (-u_2 + u_1)x_1 - u_1u_3 = 0 && B, D \text{ and } C \text{ are collinear}\\
h_2 &= (y_1^2 + 1)x_2 - u_2 - y_1^2u_1 = 0 && D \text{ is between } B \text{ and } C\\
h_3 &= u_3x_4 - u_2x_3 = 0 && A, E \text{ and } C \text{ is collinear}\\
h_4 &= (y_2^2 + 1)x_4 - u_2 = 0 && E \text{ is between } C \text{ and } A\\
h_5 &= (-x_3 + x_1)x_5 - x_1x_4 + x_2x_3 = 0 && D, E \text{ and } F \text{ are collinear.}
\end{aligned}
$$

The conclusion that point F is outside segment AB can be expressed as $x_5(x_5 - u_1) > 0$, or $g = x_5/(x_5 - u_1) > 0$.

Using the General Component Decomposition Algorithm as stated in (4.5), [8] we find that $Zero(H)$ has only one nondegenerate component with the corresponding ascending chain $ASC_1^* =$

$$
\begin{aligned}
f_1 &= ((y_1^2 + 1)u_2 - (y_1^2 + 1)u_1)x_1 - (u_2 - u_1)u_3\\
f_2 &= (y_1^2 + 1)x_2 - u_2 - y_1^2u_1\\
f_3 &= (y_2^2 + 1)u_2x_3 - u_2u_3\\
f_4 &= (y_2^2 + 1)x_4 - u_2\\
f_5 &= (x_3 - x_1)x_5 + x_1x_4 - x_2x_3.
\end{aligned}
$$

Evaluating the expression g on $PD(ASC_1^*)$, we have $g = y_1^2/y_2^2$. The problem has been reduced to deciding the positive-definiteness of expression y_1^2/y_2^2, which obviously can be detected manually since $y_1 \neq 0$ and $y_2 \neq 0$.

Remark. In the above proof, we have used some nondegeneracy conditions, i.e., the leading coefficients of the f_k are nonzero. In this particular case, they are $(y_1^2 + 1)u_2 - (y_1^2 + 1)u_1 \neq 0$, $y_1^2 + 1 \neq 0$, $y_2^2 + 1 \neq 0$, and $x_3 - x_1 \neq 0$. We will not list nondegeneracy conditions in the subsequent examples, since we are using Approach (1) to confirm a geometric statement to be generally true.

Example 2. Let O be a point, $D = AB \cap OC$, $E = AC \cap OB$, and $F = BC \cap OA$. If D is between B and A, and E is between C and A, then F is between B and C (Figure 3).

Letting $A = (0,0)$, $B = (u_1,0)$, $C = (u_2,u_3)$, $D = (x_1,0)$, $E = (x_2,x_3)$, $O = (x_4,x_5)$, and $F = (x_6,x_7)$, we have the following set of 7 equations: $H = \{h_1 = 0, ..., h_7 = 0\}$ for the hypotheses and one inequality $g < 0$ for the conclusion.

$$
\begin{aligned}
h_1 &= (y_1^2 + 1)x_1 - u_1 = 0 && D \text{ is between } A \text{ and } B\\
h_2 &= (y_2^2 + 1)x_2 - u_2 = 0 && E \text{ is between } A \text{ and } C\\
h_3 &= u_2x_3 - u_3x_2 = 0 && E, A \text{ and } C \text{ are collinear}\\
h_4 &= (x_2 - u_1)x_5 - x_3x_4 + u_1x_3 = 0 && O \text{ is on } BE\\
h_5 &= (x_1 - u_2)x_5 + u_3x_4 - u_3x_1 = 0 && O \text{ is on } CD\\
h_6 &= x_4x_7 - x_5x_6 = 0 && F \text{ is on } AO\\
h_7 &= (-u_2 + u_1)x_7 + u_3x_6 - u_1u_3 = 0 && F \text{ is on } BC\\
g &= x_6^2 + (-u_2 - u_1)x_6 + u_1u_2 < 0 && \text{Conclusion: } F \text{ is between } B \text{ and } C.
\end{aligned}
$$

---

[7] Under $u_2 - x_2 \neq 0$ and $u_2 - x_4 \neq 0$, respectively.

[8] Selecting y_1, y_2, u_1, u_2 and u_3 to be parameters. In all subsequent examples, we have already renamed the variables as stated in step 2 of Method (4.6). The y and u always form the parameter set.

Using the General Component Decomposition Algorithm as stated in (4.5), we find $Zero(H)$ has only one nondegenerate component with the corresponding ascending chain $ASC_1^* =$

$$
\begin{aligned}
f_1 &= (y_1^2+1)x_1 - u_1 \\
f_2 &= (y_2^2+1)x_2 - u_2 \\
f_3 &= u_2x_3 - u_3x_2 \\
f_4 &= ((-x_1+u_2)x_3 - u_3x_2 + u_1u_3)x_4 + (u_1x_1 - u_1u_2)x_3 + u_3x_1x_2 - u_1u_3x_1 \\
f_5 &= (x_1-u_2)x_5 + u_3x_4 - u_3x_1 \\
f_6 &= ((u_2-u_1)x_5 - u_3x_4)x_6 + u_1u_3x_4 \\
f_7 &= (-u_2+u_1)x_7 + u_3x_6 - u_1u_3.
\end{aligned}
$$

On $PD(ASC_1^*)$, g is evaluated to be $g_1 = -y_2^2y_1^2(u_2-u_1)^2/(y_1^2+y_2^2)^2$, which is negative definite if $u_2 - u_1 \neq 0$.

Example 3. Let CD and CE be the internal bisector and the median on the side AB of triangle ABC, respectively. Then $CD \leq CE$ (Figure 4).

Let $C = (0,0)$, $A = (u_1,u_2)$, $B = (x_1,x_2)$, $D = (x_3,0)$, and $E = (x_4,x_5)$. We can specify the problem as follows

$$
\begin{aligned}
h_1 &= u_1x_2 + u_2x_1 = 0 && CA \text{ is the reflection of } CB \text{ wrpt } x\text{-axis} \\
h_2 &= -y_1^2x_2 - u_2 = 0 && B \text{ and } A \text{ are on either side of } x\text{-axis} \\
h_3 &= (-x_2+u_2)x_3 + u_1x_2 - u_2x_1 = 0 && D \text{ is on } AB \\
h_4 &= -2x_4 + x_1 + u_1 = 0 && \\
h_5 &= -2x_5 + x_2 + u_2 = 0 && E \text{ is the midpoint of } A \text{ and } B \\
g &= -x_5^2 - x_4^2 + x_3^2 \leq 0 && \text{Conclusion: } CD \leq CE.
\end{aligned}
$$

Using the General Component Decomposition Algorithm as stated in (4.5), we find $Zero(H)$ has only one nondegenerate component with the corresponding ascending chain $ASC_1^* =$

$$
\begin{aligned}
f_1 &= -y_1^2u_2x_1 + u_1u_2 \\
f_2 &= -y_1^2x_2 - u_2 \\
f_3 &= (-x_2+u_2)x_3 + u_1x_2 - u_2x_1 \\
f_4 &= -2x_4 + x_1 + u_1 \\
f_5 &= -2x_5 + x_2 + u_2.
\end{aligned}
$$

Evaluating g on $PD(ASC_1^*)$, we have

$$
g_1 = -\frac{((y_1^4+2y_1^2+1)u_2^2 + (y_1^4+6y_1^2+1)u_1^2)(y_1-1)^2(y_1+1)^2}{4y_1^4(y_1^2+1)^2},
$$

which is negative definite if $y_1^2 \neq 1$, meaning $CA \neq CB$.

To decide whether $g_1 \leq 0$, we used factorization. Since factorization is generally faster than decision of definiteness. The following theorem is useful for deciding the definiteness of polynomials produced by our method.

Theorem (4.7) Suppose polynomial f in variables $u_1, ..., u_d$ can be expressed as the product of two polynomials g and h, where g and h have no common factors.

(1) $\forall u \in \mathbf{R}^d(f > 0)$ if and only if
$[\forall u \in \mathbf{R}^d(h > 0)$ and $\forall u \in \mathbf{R}^d(g > 0)]$ or
$[\forall u \in \mathbf{R}^d(h < 0)$ and $\forall u \in \mathbf{R}^d(g < 0)]$.

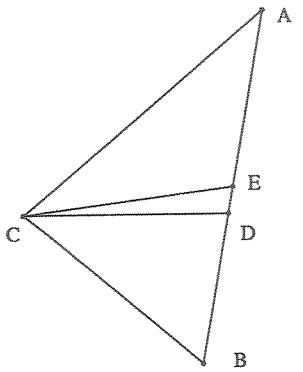

Figure 4

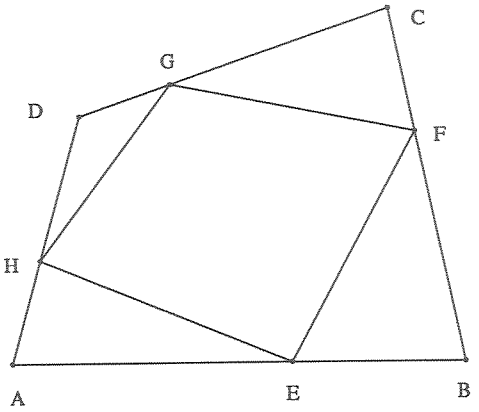

Figure 5

(2) $\forall u \in \mathbf{R}^d(f \geq 0)$ if and only if
$[\forall u \in \mathbf{R}^d(h \geq 0)$ and $\forall u \in \mathbf{R}^d(g \geq 0)]$ or
$[\forall u \in \mathbf{R}^d(h \leq 0)$ and $\forall u \in \mathbf{R}^d(g \leq 0)]$.

Proof. The proof of (1) is easy. Suppose $\forall u \in \mathbf{R}^d(h > 0)$ or $\forall u \in \mathbf{R}^d(h < 0)$ is not true. Then there are some $u'$and $u'' \in \mathbf{R}^d$ such that $h(u') < 0$ and $h(u'') > 0$. Thus we can also conclude that $h(u) = 0$ for some $u \in \mathbf{R}^d$ by continuity of h. Thus $f(u) = 0$ for some $u \in \mathbf{R}^d$, contradicting the hypothesis.

To prove (2), we require that h and g be polynomials. First, note that the real algebraic variety defined by $z = h(u_1, ..., u_d)$ has no singular points, and hence is a smooth manifold of dimension d in $\mathbf{R}^{d+1}$. (Actually the variety defined by $z = h(u_1, ..., u_d)$ is homeomorphic to R^d.)

We prove the assertion by contradiction. Suppose there exist $u' \in \mathbf{R}^d$ and $u'' \in \mathbf{R}^d$ such that $h(u') > 0$ and $h(u'') < 0$. Then there must exist neighborhoods, say, U' and U'' for u' and u'', respectively, such that $h(u) > 0$ for all $u \in U'$ and $h(u) < 0$ for all $u \in U''$. Thus the intersection of $z = h$ and $z = 0$ must be a $d-1$-dimensional variety, since $z = h$ is a smooth manifold, and contains d-dimensional neighborhoods on both sides of $z = 0$.

We shall prove that $h = 0$ and $g = 0$ intersect in a $d-1$-dimensional variety. Since $z = h$ passes through $z = 0$ from $z < 0$ to $z > 0$ in a $d-1$-dimensional variety, then there exists a point Q in $\mathbf{R}^d$ such that $h(Q) = 0$ and for each neighborhood W of Q there exist two points $Q' \in W$ and $Q'' \in W$ satisfying $h(Q') > 0$ $h(Q'') < 0$. (I.e. h changes sign at Q.) Then there is a suitably small neighborhood W' of Q such that $h = 0$ and W' intersect in a $d-1$-dimensional set U and each point in U has the same property as the above Q (i.e., h changes sign at it). We have: g must be zero on U. If this is not true then there is a point P in U such that $g(P) \neq 0$. Without loss of generallity, assume $g(P) > 0$. Then there is a neighborhood V for P such that $g > 0$ on V. By the assumption of U, we have points $P' \in V$ and $P'' \in V$ such that $h(P') > 0$ and $h(P'') < 0$. Thus we get a contradiction: $h(P'')g(p'') < 0$. Thus the intersection of $h = 0$ and $g = 0$ contains a $d-1$-dimensional set U. Then the $d-1$ variety containing U must be conatained in the intersection of $h = 0$ and $g = 0$, i.e., $h = 0$ and $g = 0$ have a $d-1$-dimensional variety in common.

Since h and g have no common divisors, then their resultant $r(u_1, ..., u_{d-1})$ with respect to u_d is not zero. This implies there is an algebraic relation among $u_1, ..., u_{d-1}$. We can prove similarly

that there is a nontrivial algebraic relation among every $d-1$ distinct variables of $u_1, ..., u_d$. Hence the dimension of the intersection of $h=0$ and $g=0$ is less than $d-1$. This contradicts the fact that $h=0$ and $g=0$ intersect in a $d-1$-dimensional variety. .QED.

Example 4. An inscribed quadrilateral of a convex quadrilateral is convex (Figure 5).

Let $A=(0,0)$, $B=(u_1,0)$, $C=(u_2,u_3)$, $D=(x_1,x_2)$, $E=(x_3,0)$, $F=(x_4,x_5)$, $G=(x_6,x_7)$, and $H=(x_8,x_9)$. Then the hypotheses and conclusion can be specified as follows

$$h_1 = -u_2x_2 + u_3x_1 + y_1^2u_1u_3 = 0 \qquad B \text{ and } D \text{ are on either side of } AC$$
$$h_2 = (-u_2 + (y_2^2+1)u_1)x_2 + u_3x_1 - u_1u_3 = 0 \qquad A \text{ and } C \text{ are on either side of } BD$$
$$h_3 = (y_3^2+1)x_3 - u_1 = 0 \qquad E \text{ is between } A \text{ and } B$$
$$h_4 = (y_4^2+1)x_4 - u_2 - y_4^2u_1 = 0 \qquad F \text{ is between } B \text{ and } C$$
$$h_5 = (u_2-u_1)x_5 - u_3x_4 + u_1u_3 = 0 \qquad F \text{ is on } BC$$
$$h_6 = (y_5^2+1)x_6 - x_1 - y_5^2u_2 = 0 \qquad G \text{ is between } C \text{ and } D$$
$$h_7 = (x_1-u_2)x_7 + (-x_2+u_3)x_6 + u_2x_2 - u_3x_1 = 0 \qquad G \text{ is on } CD$$
$$h_8 = (y_6^2+1)x_8 - y_6^2x_1 = 0 \qquad H \text{ is between } D \text{ and } A$$
$$h_9 = -x_1x_9 + x_2x_8 = 0 \qquad H \text{ is on } DA$$
$$g = ((-x_6+x_3)x_9 + x_7x_8 - x_3x_7)/((-x_6+x_4)x_9 + (x_7-x_5)x_8 - x_4x_7 + x_5x_6) > 0$$
Conclusion: E and F are on the same side of HG.

Using the General Component Decomposition Algorithm as stated in (4.5), we find $Zero(H)$ has only one nondegenerate component with the corresponding ascending chain $ASC_1^* =$

$$f_1 = (((y_2^2+1)u_1)u_3)x_1 + (((-y_1^2-1)u_1)u_2 + ((y_2^2+1)y_1^2)u_1^2)u_3$$
$$f_2 = (-u_2 + (y_2^2+1)u_1)x_2 + u_3x_1 - u_1u_3$$
$$f_3 = (y_3^2+1)x_3 - u_1$$
$$f_4 = (y_4^2+1)x_4 - u_2 - y_4^2u_1$$
$$f_5 = (u_2-u_1)x_5 - u_3x_4 + u_1u_3$$
$$f_6 = (y_5^2+1)x_6 - x_1 - y_5^2u_2$$
$$f_7 = (x_1-u_2)x_7 + (-x_2+u_3)x_6 + u_2x_2 - u_3x_1$$
$$f_8 = (y_6^2+1)x_8 - y_6^2x_1$$
$$f_9 = -x_1x_9 + x_2x_8.$$

Evaluating g on $PD(ASC_1^*)$, we have

$$g_1 = \frac{(((y_6^2y_5^2y_3^2 + y_6^2y_5^2)y_2^2 + y_6^2y_5^2y_3^2 + 1)y_1^2 + (y_6^2+1)y_5^2y_2^2 + y_5^2 + 1)(y_4^2+1)}{(((y_6^2y_5^2y_4^2+1)y_2^2 + y_4^2 + 1)y_1^2 + (y_6^2+1)y_5^2y_4^2y_2^2 + (y_5^2+1)y_4^2)(y_3^2+1)},$$

which is positive definite.

We have also used Methods (4.5) and (4.6) to prove many other theorems such as the altitude (median) on a greater side of a triangle is smaller than the altitude (median) on a smaller side. However, all these examples are linear in the sense that for the characteristic set $f_1, ..., f_r$ obtained from $h_1, ..., h_r$, $deg(f_i, x_i) = 1$. Thus we can reduce the problems to deciding definiteness or semi-definiteness of polynomials of the form $g(y_1, \ldots, y_s, u_1, \ldots, u_d)$, such that there are no relations among th y's and u's (neither equations nor inequalities). This is closely related to Hilbert's 17th problem. If, on the other hand, for some f_i, $deg(f_i, x_i) \geq 2$ and f_i is irreducible in x_i, then the situation may be complicated. However, in many such cases, the final results from our program are still polynomials of the form $g(y_1, ..., y_s, u_1, ..., u_d)$ which are definite. Now let us look at some such examples.

Example 5. Let AB and CD be two chords of circle (O), M and N be their midpoints, respectively. If $AB > CD$, then $OM < ON$ (Figure 6).

Let $M = (0,0)$, $A = (u_1,0)$, $B = (x_1,0)$, $O = (0,u_2)$, $C = (u_3,x_2)$, $D = (x_3,x_4)$, and $N = (x_5,x_6)$. Then the problem can be specified as follows

$$
\begin{aligned}
h_1 &= x_1 + u_1 = 0 && M \text{ is the midpoint of } A \text{ and } B\\
h_2 &= x_2^2 - 2u_2x_2 + u_3^2 - u_1^2 = 0 && OC = OA\\
h_3 &= y_1^2x_4^2 - 2y_1^2x_2x_4 + y_1^2x_3^2 - 2y_1^2u_3x_3 + y_1^2x_2^2 - y_1^2x_1^2 + 2y_1^2u_1x_1\\
&\quad + y_1^2u_3^2 - y_1^2u_1^2 + 1 = 0 && AB > CD\\
h_4 &= x_4^2 - 2u_2x_4 + x_3^2 - u_1^2 = 0 && OD = OA\\
h_5 &= 2x_5 - x_3 - u_3 = 0\\
h_6 &= 2x_6 - x_4 - x_2 = 0 && N \text{ is the midpoint of } C \text{ and } D\\
g &= -x_6^2 + 2u_2x_6 - x_5^2 < 0 && \text{Conclusion: } OM < ON.
\end{aligned}
$$

Using the General Component Decomposition Algorithm as stated in (4.5), we find $Zero(H)$ has only one nondegenerate component with the corresponding ascending chain $ASC_1^* =$

$$
\begin{aligned}
f_1 &= x_1 + u_1\\
f_2 &= x_2^2 - 2u_2x_2 + u_3^2 - u_1^2\\
f_3 &= (-4y_1^6x_2^2 + 8y_1^6u_2x_2 - 4y_1^6u_3^2 - 4y_1^6u_2^2)x_3^2 + (4y_1^6u_3x_2^2 - 8y_1^6u_2u_3x_2 - 4y_1^6u_3x_1^2 + 8y_1^6u_1u_3x_1 +\\
&\quad 4y_1^6u_3^3 + (8y_1^6u_2^2 + 4y_1^4)u_3)x_3 - y_1^6x_2^4 + 4y_1^6u_2x_2^3 + (2y_1^6x_1^2 - 4y_1^6u_1x_1 - 2y_1^6u_3^2 - 4y_1^6u_2^2 + 4y_1^6u_1^2 - 2y_1^4)x_2^2 +\\
&\quad (-4y_1^6u_2x_1^2 + 8y_1^6u_1u_2x_1 + 4y_1^6u_2u_3^2 + (-8y_1^6u_1^2 + 4y_1^4)u_2)x_2 - y_1^6x_1^4 + 4y_1^6u_1x_1^3 + (2y_1^6u_3^2 + 4y_1^6u_2^2 -\\
&\quad 4y_1^6u_1^2 + 2y_1^4)x_1^2 + (-4y_1^6u_1u_3^2 - 8y_1^6u_1u_2^2 - 4y_1^4u_1)x_1 - y_1^6u_3^4 + (-4y_1^6u_2^2 - 2y_1^4)u_3^2 + (4y_1^6u_1^2 - 4y_1^4)u_2^2 - y_1^2\\
f_4 &= (2y_1^2x_2 - 2y_1^2u_2)x_4 + 2y_1^2u_3x_3 - y_1^2x_2^2 + y_1^2x_1^2 - 2y_1^2u_1x_1 - y_1^2u_3^2 - 1\\
f_5 &= 2x_5 - x_3 - u_3\\
f_6 &= 2x_6 - x_4 - x_2.
\end{aligned}
$$

Evaluating g on $PD(ACS_1^*)$, we have $g_1 = -1/4y_1^2$, which is negative definite.

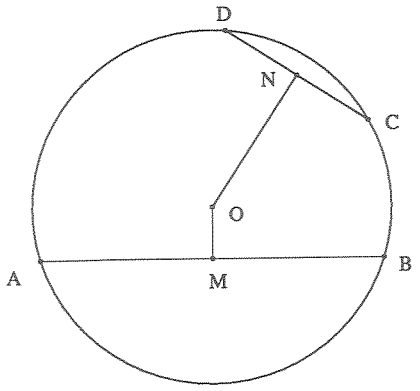

Figure 6

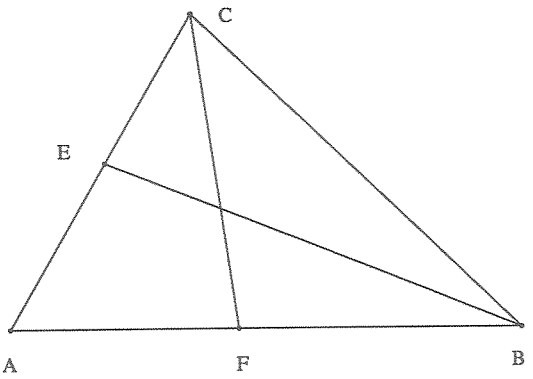

Figure 7

Example 6. Let CF and BE be two *internal* bisectors of a triangle ABC. If $AB > AC$, then $BE > CF$ (Figure 7).

This problem is known to be difficult. A famous problem in elementary geometry, "A triangle with two equal internal bisectors is an isosceles triangle", is a direct consequence of this theorem.

Let $A = (0,0)$, $B = (x_1, 0)$, $C = (x_2, x_3)$, $F = (x_4, 0)$, and $E = (x_5, x_6)$. The problem can be specified as follows:

$h_1 = -y_1^2x_3^2 - y_1^2x_2^2 + y_1^2x_1^2 - 1 = 0$ $\qquad AB > AC$

$h_2 = (2x_2 - x_1)x_3x_4^2 + (-2x_3^3 - 2x_2^2x_3)x_4 + x_1x_3^3 + x_1x_2^2x_3 = 0$ $\qquad tan(ACF) = tan(FCB)$

$h_3 = (y_2^2 + 1)x_4 - x_1 = 0$ $\qquad F$ is between A and B

$h_4 = (y_3^2 + 1)x_5 - x_2 = 0$ $\qquad E$ is between A and C

$h_5 = x_1x_3x_6^2 + ((2x_1x_2 - 2x_1^2)x_5 - 2x_1^2x_2 + 2x_1^3)x_6 - x_1x_3x_5^2 + 2x_1^2x_3x_5 - x_1^3x_3 = 0$ $\qquad tan(ABE) = tan(EBC)$

$h_6 = x_2x_6 - x_3x_5 = 0$ $\qquad E$ is on AC

$g = x_6^2 + x_5^2 - 2x_1x_5 - x_4^2 + 2x_2x_4 - x_3^2 - x_2^2 + x_1^2 > 0$ $\qquad$ Conclusion: $BE > CF$.

Using the General Component Decomposition Algorithm as stated in (4.5), we find $Zero(H/x_5)$ has only one nondegenerate component with the corresponding ascending chain $ASC_1^* =$

$f_1 = (y_2^4 - y_3^4)y_1^2x_1^2 - y_2^4$
$f_2 = 2y_1^2x_1x_2 + (y_2^4 - 2)y_1^2x_1^2 - y_2^4 + 1$
$f_3 = x_2^2x_3^2 + x_2^4 - 2x_1x_2^3 + (-y_3^4 + 1)x_1^2x_2^2$
$f_4 = (y_2^2 + 1)x_4 - x_1$
$f_5 = (y_3^2 + 1)x_5 - x_2$
$f_6 = x_2x_6 - x_3x_5$.

Evaluating g on $PD(ASC_1^*)$, we have

$$g_1 = \frac{((y_3^2 + 1)y_2^2 + y_3^2)((y_3^4 + y_3^2)y_2^4 + (y_3^4 + 3y_3^2 + 1)y_2^2 + y_3^2)}{y_1^2(y_2^2 + y_3^2)(y_2^2 + 1)^2(y_3^2 + 1)^2},$$

which is positive definite.

5. Proving Type IE Statements

In Section 4 we used the Rabinowitsch/Seidenberg device to convert hypothesis inequalities to equations, i.e. to "get rid" of the hypothesis inequalities. What is different about the examples in this section is not so much that they are of type IE rather than EI or II, but that the equality part of their hypotheses consists of several cases, and the conclusion holds on some, but not all, of those cases. In these situations, far from wanting to "get rid" of the inequality part of the hypothesis, it has a crucial "job" to do: to "select" for us exactly those cases in the equality part on which the conclusion holds.

First, however, we note that there are some type IE statements that are actually of type EE, in the sense that we can simply delete the inequalities from the hypotheses and still have a true statement. In such cases we obtain (the proof of) a more general statement than we started with. Let us consider the following example.

Example. (Feuerbach's theorem). The nine-point circle (N) of a triangle is tangent to the incircle of the triangle (Figure 8).

Let us recall that the incircles and excircles of a triangle are the four circles which are each tangent to the three lines that are coincident with the sides of the triangle. The incircle is the unique one of these circles whose center is inside the triangle; the other three are the excircles.

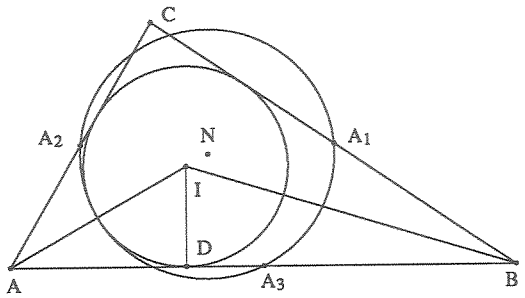

Figure 8

To select the incircle and not an excircle in Feuerbach's theorem, i.e., to specify that the incenter I is inside the triangle, we need inequalities. Hence this statement seems to be of type IE. However, let us study it more closely.

We can let $A = (0,0)$, $B = (u_1,0)$, $C = (u_2,u_3)$, $I = (x_1,x_2)$, $D = (x_1,0)$, $A_1 = (x_3,x_4)$, $A_2 = (x_5,x_6)$, $A_3 = (x_7,0)$, and $N = (x_8,x_9)$. Then the equality part of the statement can be expressed by the following set of equations H:

$$h_1 = u_1u_3x_2^2 + ((2u_1u_2 - 2u_1^2)x_1 - 2u_1^2u_2 + 2u_1^3)x_2 - u_1u_3x_1^2 + 2u_1^2u_3x_1 - u_1^3u_3 = 0$$

$\tan(CBI) = \tan(IBA)$

$$h_2 = u_1u_3x_2^2 + 2u_1u_2x_1x_2 - u_1u_3x_1^2 = 0$$

$\tan(CAI) = \tan(IAB)$

$$h_3 = 2x_3 - u_2 - u_1 = 0$$

A_1 is the midpoint of B and C

$$h_4 = 2x_4 - u_3 = 0$$

$$h_5 = 2x_5 - u_2 = 0$$

A_2 is the midpoint of A and C

$$h_6 = 2x_6 - u_3 = 0$$

A_3 is the midpoint of A and B

$$h_7 = 2x_7 - u_1 = 0$$

$NA_3 \equiv NA_2$

$$h_8 = 2x_6x_9 + (-2x_7 + 2x_5)x_8 + x_7^2 - x_6^2 - x_5^2 = 0$$

$NA_3 \equiv NA_1$.

$$h_9 = 2x_4x_9 + (-2x_7 + 2x_3)x_8 + x_7^2 - x_4^2 - x_3^2 = 0$$

The conclusion that circle (N) is tangent to circle (I) is equivalent to $g = ((8x_2x_7 - 8x_1x_2)x_8 - 4x_2x_7^2 + 4x_1^2x_2)x_9 + (-4x_7^2 + 8x_1x_7 + 4x_2^2 - 4x_1^2)x_8^2 + (4x_7^3 - 4x_1x_7^2 + (-8x_2^2 - 4x_1^2)x_7 + 4x_1^3)x_8 - x_7^4 + (4x_2^2 + 2x_1^2)x_7^2 - x_1^4 = 0$. Using the Ritt–Wu Decomposition Algorithm, there is only one nondegenerate component with the following corresponding ascending chain $ASC_1^* =$

$$f_1 = 4x_1^4 - 8u_1x_1^3 + (-4u_3^2 - 4u_2^2 + 4u_1u_2 + 4u_1^2)x_1^2 + (4u_1u_3^2 + 4u_1u_2^2 - 4u_1^2u_2)x_1 - u_1^2u_3^2$$
$$f_2 = (2x_1 + 2u_2 - 2u_1)x_2 - 2u_3x_1 + u_1u_3$$
$$f_3 = 2x_3 - u_2 - u_1$$
$$f_4 = 2x_4 - u_3$$
$$f_5 = 2x_5 - u_2$$
$$f_6 = 2x_6 - u_3$$
$$f_7 = 2x_7 - u_1$$
$$f_8 = 4x_8 - 2u_2 - u_1$$
$$f_9 = 4u_3x_9 - u_3^2 + u_2^2 - u_1u_2.$$

The fact that $deg(f_1,x_1) = 4$ means that there are four solutions for I: one is the incenter and the other three are the excenters. Now f_1 is irreducible in x_1. Since x_1 has real solutions for some real generic values of u_1, u_2 and u_3 (i.e., there are some open intervals for u_1, u_2 and u_3

in $\mathbf{R}$, in which x_1 has solutions in $\mathbf{R}$),[9] $prem(g; f_1, ..., f_9)$ must be zero by Theorem (3.12) (also see Theorem (3.6) of Chapter 4 in [13]) if the conclusion that the incircle touches the nine-point circle is true. $Prem(g; f_1, ..., f_9) = 0$ can be easily checked by our prover in a few seconds. Thus, the prover actually proves that the nine point circle (N) is tangent to the incircle as well as to the three excircles.

If the circle (N) were not tangent to one of the three excircles, f_1 would be reducible and the inequality part that I is inside the triangle ABC would be essential. Thus we come to the following important principle:

Principle (5.1). *If the equality part of the hypothesis in a geometric statement of type IE is irreducible (see Section 3.1). and has real general solutions, then the inequality part is redundant and the statement can be confirmed by the original Wu–Ritt method, treated as a statement of type EE. Conversely, if the inequality part is not redundant, then the problem is reducible.*

The reader can find many such examples in [13]. Now let us look at an example of type IE in which the inequality part is essential. In the remainder of Section 5 we will repeat the three-part format of Section 4: first a motivating example in Section 5.1, then the statement of a general "method" in Section 5.2, followed by more examples in Section 5.2.

5.1. A Working Example

Example (5.2). On the two sides AC and BC of triangle ABC, two squares $ACDE$ and $BCFG$ are drawn. M is the midpoint of AB. Line CM intersects FD at H. Show that $DF \perp CH$ (Figure 9a).

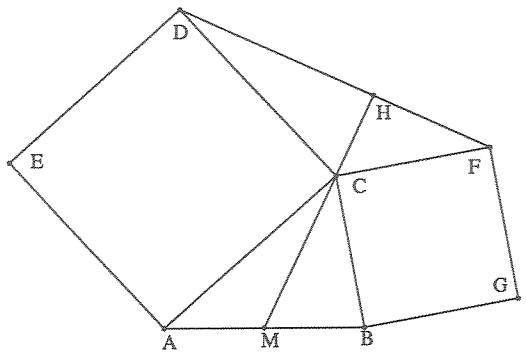

Figure 9a

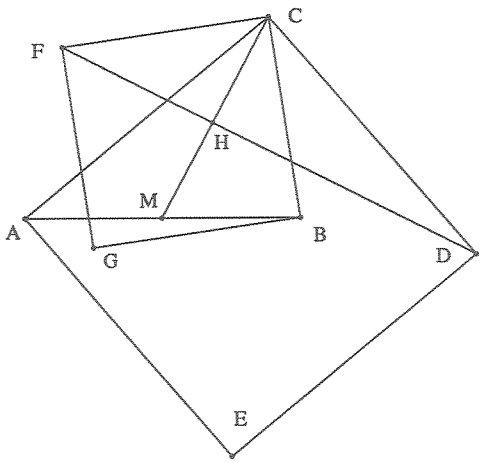

Figure 9b

Figure 9a implicitly suggests that the statement is true in the case when both squares are "outside" the triangle. Actually, in the other cases (Figures 9b and 10ab) the statement may

[9] Proving this fact, which is implicitly assumed in textbooks of geometry, is beyond the Wu–Ritt method. Intuitively, this corresponds to the fact that given any three points A, B and C in certain general position, one can always draw the incenter.

or may not be true. This is a statement of type IE where the inequality part is essential if we specify the equality part as follows.

Let $A = (u_1, 0)$, $B = (u_2, u_3)$, $C = (0,0)$, $D = (0, u_1)$, $F = (x_1, x_2)$, $M = (x_3, x_4)$, and $H = (x_5, x_6)$, then the equality part of the hypotheses and the conclusion can be expressed by the following equations:

$$
\begin{aligned}
h_1 &= x_2^2 + x_1^2 - u_3^2 - u_2^2 = 0 && CF = BC \\
h_2 &= u_3 x_2 + u_2 x_1 = 0 && CF \perp BC \\
h_3 &= 2x_3 - u_2 - u_1 = 0 && \\
h_4 &= 2x_4 - u_3 = 0 && M \text{ is the midpoint of } A \text{ and } B \\
h_5 &= x_3 x_6 - x_4 x_5 = 0 && H \text{ is on } CM \\
h_6 &= x_1 x_6 - (x_2 - u_1)x_5 - u_1 x_1 = 0 && H \text{ is on } DF \\
g &= (x_2 - u_1)x_6 + x_1 x_5 && \text{Conclusion: } DF \perp CH.
\end{aligned}
$$

The ascending chain (characteristic set) obtained from $h_1, ..., h_6$ is

$$
\begin{aligned}
f_1 &= (u_3^2 + u_2^2)x_1^2 - u_3^4 - u_2^2 u_3^2 \\
f_2 &= u_3 x_2 + u_2 x_1 \\
f_3 &= 2x_3 - u_2 - u_1 \\
f_4 &= 2x_4 - u_3 \\
f_5 &= (x_1 x_4 - (x_2 - u_1)x_3)x_5 - u_1 x_1 x_3 \\
f_6 &= x_1 x_6 - (x_2 - u_1)x_5 - u_1 x_1.
\end{aligned}
$$

To confirm whether $g = 0$ follows from $h_1, = 0, ..., h_6 = 0$ under some nondegeneracy conditions, we can compute $prem(g; f_1, ..., f_6)$ to see whether it is zero. The result on computers shows that $prem(g; f_1, ..., f_6) \neq 0$. If the conclusion is true, this means that $h_1, ..., h_6$ is reducible and the inequality part is essential. Actually, $f_1 = (u_3^2 + u_2^2)(x_1 - u_3)(x_1 + u_3)$, and we have two nondegenerate components for $h_1, ..., h_6$ with corresponding ascending chains $ASC_1^* = f_1', f_2, f_3, f_4, f_5, f_6$ and $ASC_2^* = f_1'', f_2, f_3, f_4, f_5, f_6$, where $f_1' = x_1 - u_3$, $f'' = x_1 + u_3$. The results on computers show that $prem(g; ASC_1^*) = 0$, but $prem(g; ASC_2^*) \neq 0$. To decide which component corresponds to the case in the original statement, we need the inequality part:

$[(D \text{ and } B \text{ on either side of } AC) \wedge (A \text{ and } F \text{ on either side of } BC)] \vee$
$[(D \text{ and } B \text{ on the same side of } AC) \wedge (A \text{ and } F \text{ on the same side of } BC)]$.

Let $\delta = -u_1^2 u_3^2 / (u_1^2 u_2 x_2 - u_1^2 u_3 x_1)$, then $\delta > 0$ is its algebraic form.

Much as we did in Section 4, we can reduce δ modulo the ideals generated by ASC_1^* and ASC_2^*, respectively. $\delta = u_3^2/(u_3^2 + u_2^2)$ for component ASC_1^*, and $\delta = -u_3^2/(u_3^2 + u_2^2)$ for component ASC_2^*. We can "see" $\delta > 0$ for component ASC_1^* if we assume A, B and C are not collinear (i.e., $u_1 u_3 \neq 0$). Thus component ASC_1^* corresponds to the cases in Figures 9a and 9b, in which the conclusion is generally true. The proof is completed. However, to decide whether $u_3^2/(u_3^2 + u_2^2) > 0$ (under $u_1 u_3 \neq 0$) is beyond the scope of our method.

Remark. As we have seen, component ASC_2^* corresponds to the cases in Figures 10a and 10b, in which the conclusion is generally false. Decomposition of the equality part of H (when it is reducible) into several nondegenerate components is actually a case study approach. It gives more insight into the given geometric configuration defined by the equality part. For example, if we want to decide whether, say, $DF \perp AB$, we will find that it is generally true for component ASC_1^* (Figures 10ab) but not true for component $PD(ASC_2^*)$.

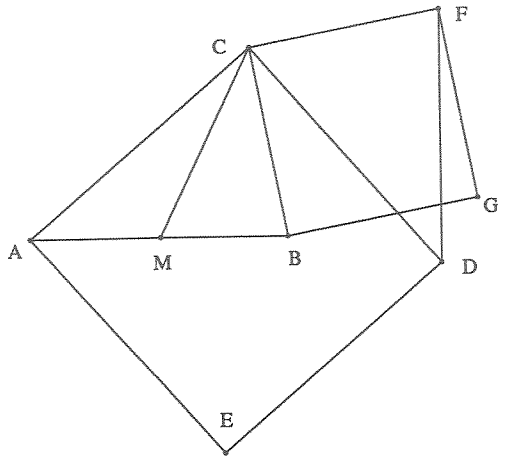

Figure 10a

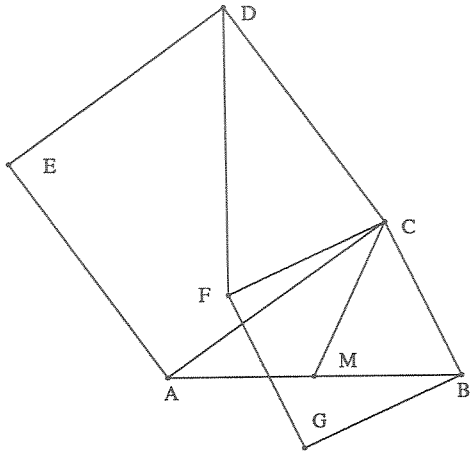

Figure 10b

5.2. The Method for Type IE Statements

Let the conclusion C be $g = 0$ where g is a polynomial in the parameters u_i and dependent variables x_j.

Step 1. Triangulize the equation part of the hypothesis H to obtain an ascending chain ASC.

Step 2. Calculate $prem(g; ASC)$, the successive pseudo remainder of g with respect to ASC. If $prem(g; ASC)$ is zero, then the statement is confirmed to be generally true and the inequality part of H is redundant (as in the example of Feuerbach's theorem). The statement is actually a statement of type EE.

Now assume $prem(g; ASC) \neq 0$. Then we check whether triangular form ASC is irreducible.

Step 3. Suppose ASC is irreducible. If we assume that the real generic zeros exist for ASC (as most problems in textbooks of geometry implicitly assume), then the statement is generally false.

Step 4. Otherwise, decompose the equation part of H into several nondegenerate (irreducible) components with corresponding ascending chains ASC_1^*,...,ASC_c^* $(c \geq 2)$. Assume the real generic zeros exist for the equation part of H.

(1) If $prem(g; ASC_i^*) \neq 0$ for all i, then the statement is generally false.

(2) If $prem(g; ASC_i^*) = 0$ for all i, then the statement is generally true.

(3) Suppose $prem(g; ASC_i^*) = 0$ for some i and $prem(g; ASC_{i'}^*) \neq 0$ for the other i'. Let $\delta > 0$ be a member of the inequality part of the hypothesis H. Then evaluate δ for each $ASC_{i'}$ to see whether $\delta \leq 0$ is valid for each δ, using the method for type EI. If it is, then the statement is generally true.

5.3. More Examples for Type IE Statements

Example 7. Let A and B be two points on a side of an angle $\angle O$. Let A_1 and B_1 be two points on the two sides OA and OB of the angle such that $OA_1 \equiv OA$ and $OB_1 \equiv OB$. I is the intersection of AB_1 and A_1B. Show that OI is the bisector of $\angle O$ (Figure 11a).

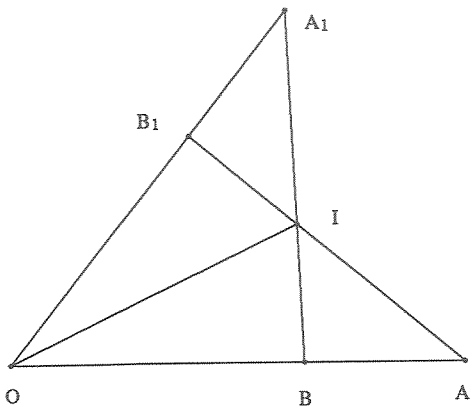

Figure 11a

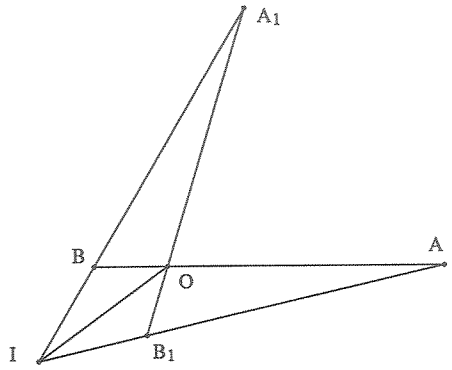

Figure 11b

Let $O = (0,0)$, $A = (u_1, 0)$, $B = (u_2, 0)$, $A_1 = (u_3, x_1)$, $B_1 = (x_2, x_3)$, and $I = (x_4, x_5)$. Then the equality part of the problem can be specified as follows:

$$\begin{aligned} h_1 &= x_1^2 + u_3^2 - u_1^2 = 0 && OA_1 = OA \\ h_2 &= x_3^2 + x_2^2 - u_2^2 = 0 && OB_1 = OB \\ h_3 &= u_3 x_3 - x_1 x_2 = 0 && B_1 \text{ is on } OA_1 \\ h_4 &= (x_2 - u_1)x_5 - x_3 x_4 + u_1 x_3 = 0 && I \text{ is on } AB_1 \\ h_5 &= (u_3 - u_2)x_5 - x_1 x_4 + u_2 x_1 = 0 && I \text{ is on } BA_1 \\ g &= u_1 x_1 x_5^2 + 2u_1 u_3 x_4 x_5 - u_1 x_1 x_4^2 = 0 \end{aligned}$$

Conclusion: $\tan(AOI) = \tan(IOA_1)$.

Triangulizing $h_1, ..., h_5$, we have the following ascending chain

$$\begin{aligned} f_1 &= x_1^2 + u_3^2 - u_1^2 \\ f_2 &= (x_1^2 + u_3^2)x_2^2 - u_2^2 u_3^2 \\ f_3 &= u_3 x_3 - x_1 x_2 \\ f_4 &= ((u_3 - u_2)x_3 - x_1 x_2 + u_1 x_1)x_4 + (-u_1 u_3 + u_1 u_2)x_3 + u_2 x_1 x_2 - u_1 u_2 x_1 \\ f_5 &= (u_3 - u_2)x_5 - x_1 x_4 + u_2 x_1. \end{aligned}$$

However, $prem(g; f_1, ..., f_5) \neq 0$. Actually, the polynomial set $h_1, ..., h_5$ can be decomposed into two nondegenerate components: $ASC_1^* = f_1, f_2', f_3, f_4, f_5$ and $ASC_2^* = f_1, f_2'', f_3, f_4, f_5$, where $f_2' = u_1 x_2 + u_2 u_3$ and $f_2'' = u_1 x_2 - u_2 u_3$.

Computer results have shown that $prem(g; ASC_1^*) \neq 0$ and $prem(g; ASC_2^*) = 0$. The inequality part of the statement is

[O is between A and B, and between A_1 and B_1] $\vee$
[O is not between A and B, and not between A_1 and B_1],

or in its algebraic form, $\delta = u_1 u_3/(u_2 x_2) > 0$ (see Figures 11ab). Evaluating δ, we have $\delta = -u_1^2/u_2^2$ for component ASC_1^* and $\delta = u_1^2/u_2^2$ for component ASC_2^*. Thus the statement has been confirmed.

Example 8. Three equilateral triangles A_1BC, AB_1C and ABC_1 are erected on the three respective sides, BC, CA and AB, of a triangle ABC, then lines AA_1, BB_1 and CC_1 are concurrent (Figure 12a).

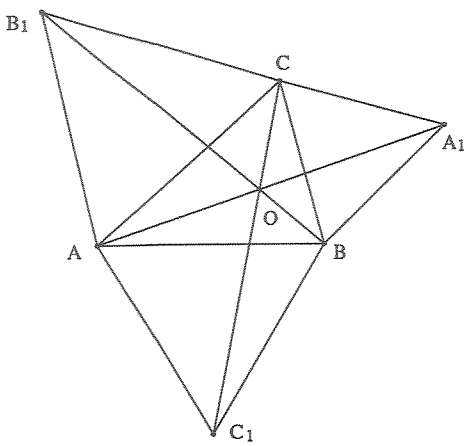

Figure 12a

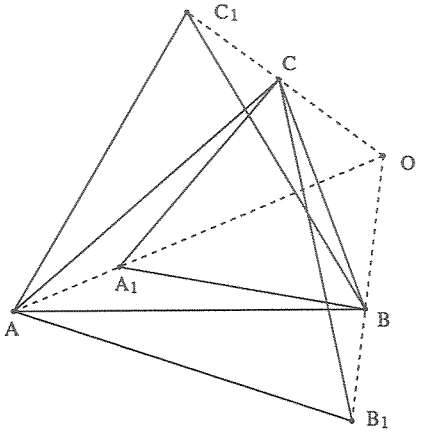

Figure 12b

Let $A = (0,0)$, $B = (u_1, 0)$, $C = (u_2, u_3)$, $B_1 = (x_1, x_2)$, $C_1 = (x_3, x_4)$, $A_1 = (x_5, x_6)$, and $O = (x_7, x_8)$. If we specify the equality part of the problem as follows, then it is a statement of type IE. [10]

$$h_1 = -x_2^2 - x_1^2 + u_3^2 + u_2^2 = 0$$
$$h_2 = 2u_3x_2 + 2u_2x_1 - u_3^2 - u_2^2 = 0$$
$$h_3 = 2u_1x_3 - u_1^2 = 0$$
$$h_4 = -x_4^2 - x_3^2 + u_1^2 = 0$$
$$h_5 = -x_6^2 - x_5^2 + 2u_1x_5 + u_3^2 + u_2^2 - 2u_1u_2 = 0$$
$$h_6 = 2u_3x_6 + (2u_2 - 2u_1)x_5 - u_3^2 - u_2^2 + u_1^2 = 0$$
$$h_7 = (-x_3 + u_2)x_8 + (x_4 - u_3)x_7 - u_2x_4 + u_3x_3 = 0$$
$$h_8 = (-x_1 + u_1)x_8 + x_2x_7 - u_1x_2 = 0$$
$$g = x_5x_8 - x_6x_7 = 0$$

$AC = AB_1$
$B_1A = B_1C$
$C_1A = C_1B$
$AB = AC_1$
$BC = A_1B$
$A_1B = A_1C$

C, C_1 and O are collinear
B, B_1 and O are collinear
Conclusion: A, O and A_1 are collinear.

The polynomial set $h_1, ..., h_8$ can be decomposed into four nondegenerate components: $ASC_1^* = f_1, f_2, f_3, f_4', f_5', f_6, f_7, f_8$, $ASC_2^* = f_1, f_2, f_3, f_4'', f_5', f_6, f_7, f_8$, $ASC_3^* = f_1, f_2, f_3, f_4', f_5'', f_6, f_7, f_8$, and $ASC_4^* = f_1, f_2, f_3, f_4'', f_5'', f_6, f_7, f_8$, where

$$f_1 = (-4u_3^2 - 4u_2^2)x_1^2 + (4u_2u_3^2 + 4u_2^3)x_1 + 3u_3^4 + 2u_2^2u_3^2 - u_2^4$$
$$f_2 = 2u_3x_2 + 2u_2x_1 - u_3^2 - u_2^2$$
$$f_3 = 2u_1x_3 - u_1^2$$
$$f_4' = (2u_3^3 + 2u_2^2u_3)x_4 + (2u_1u_3^2 + 2u_1u_2^2)x_1 - u_1u_2u_3^2 - u_1u_2^3$$
$$f_4'' = (2u_3^3 + 2u_2^2u_3)x_4 + (-2u_1u_3^2 - 2u_1u_2^2)x_1 + u_1u_2u_3^2 + u_1u_2^3$$
$$f_5' = (2u_3^2 + 2u_2^2)x_5 + (2u_3^2 + 2u_2^2)x_1 + (-2u_2 - u_1)u_3^2 - 2u_2^3 - u_1u_2^2$$
$$f_5'' = (2u_3^2 + 2u_2^2)x_5 + (-2u_3^2 - 2u_2^2)x_1 - u_1u_3^2 - u_1u_2^2$$
$$f_6 = 2u_3x_6 + (2u_2 - 2u_1)x_5 - u_3^2 - u_2^2 + u_1^2$$
$$f_7 = ((-x_1 + u_1)x_4 + x_2x_3 - u_2x_2 + u_3x_1 - u_1u_3)x_7 + (u_2x_1 - u_1u_2)x_4 + (-u_1x_2 - u_3x_1 + u_1u_3)x_3 + u_1u_2x_2$$
$$f_8 = (-x_1 + u_1)x_8 + x_2x_7 - u_1x_2.$$

we found that $prem(g; ASC_2^*) = 0$, but $prem(g; ASC_1^*) \neq 0$, $prem(g; ASC_3^*) \neq 0$, and $prem(g; ASC_4^*) \neq 0$. In order to clarify the situation, we introduce two inequality conditions.

[10] This problem can also specified as an type EE statement. See example (5.7) of Chapter 4 in [13].

Let Δ_1 be

[(C and C_1 are on the same side of AB) $\wedge$ (B and B_1 are on the same side of AC)] $\vee$ [(C and C_1 are on either side of AB) $\wedge$ (B and B_1 are on either side of AC)].

and let Δ_2 be

[(C and C_1 are on the same side of AB) $\wedge$ (A and A_1 are on the same side of BC)] $\vee$ [(C and C_1 are on either side of AB) $\wedge$ (A and A_1 are on either side of BC)].

Let $\delta_1 = -u_1^2u_3^2/(u_1u_2x_2 - u_1u_3x_1)x_4$, and let $\delta_2 = u_1^2u_3^2/((u_1u_2 - u_1^2)x_4x_6 - u_1u_3x_4x_5 + u_1^2u_3x_4)$.

Δ_1 holds iff $\delta_1 > 0$; Δ_2 holds iff $\delta_2 > 0$. Thus, the inequality part of the hypotheses is $\delta_1 > 0 \wedge \delta_2 > 0$.

Evaluating δ_1 and δ_2 for various components, we have

For component ASC_1^*, $\delta_1 = -4u_3^2/(3u_3^2 + 3u_2^2)$ and $\delta_2 = -4u_3^2/3(u_3^2 + u_2^2 - 2u_1u_2 + u_1^2)$.

For component ASC_2^*, $\delta_1 = 4u_3^2/(3u_3^2 + 3u_2^2)$ and $\delta_2 = 4u_3^2/3(u_3^2 + u_2^2 - 2u_1u_2 + u_1^2)$.

For component ASC_3^*, $\delta_1 = -4u_3^2/(3u_3^2 + 3u_2^2)$ and $\delta_2 = 4u_3^2/3(u_3^2 + u_2^2 - 2u_1u_2 + u_1^2)$.

For component ASC_4^*, $\delta_1 = 4u_3^2/(3u_3^2 + 3u_2^2)$ and $\delta_2 = -4u_3^2/3(u_3^2 + u_2^2 - 2u_1u_2 + u_1^2)$.

Thus, the conclusion is true if and only if the three triangles are all outside $\triangle ABC$ (Figure 12a), or all "inside" $\triangle ABC$ (Figure 12b). The proof is completed.

A Challenge Problem for IE Type Statements. *A triangle with two equal internal bisectors is an isosceles triangle.*

This problem was solved in [34] with more human-machine interaction than was required for the examples in this paper. We would like to see a mechanical solution with less human involvement.

6. Conclusions

It is a curious fact that, to date, all examples in which we have successfully applied the Rabinowitsch/Seidenberg device have been of type II, and not of type IE[11] It is equally curious that, to date, all examples with an essential inequality part in their hypotheses have been of type IE, and not II. Recall from Section 5 that these are statements whose equality part has multiple irreducible nondegenerate components, the conclusion holds on some but not on others, and the inequality part of the hypothesis serves to identify those components of the equality part on which the conclusion holds. So far we know of no reason why we have had this experience.

Proofs of all examples appearing in this paper have been produced mechanically, except possibly for steps in which it is necessary to decide the definiteness or semi-definiteness of some polynomial (possibly in the presence of additional polynomial constraints). The problem whose proof (exclusive of such definiteness decisions) required the most time was Example 6, which

[11] We note that there seems to be an intrinsic obstacle to applying the Rabinowitsch/Seidenberg device to the conclusion of a type EI statement, for there seems no way to remove the new existential quantifier we would then have there, nor any way to convert it to a universal quantifier. However see [26] for some thoughts on this subject.

took about 5 minutes on a Symbolics 3600. The reader can find a collection of 35 examples in [10], including nearly all those in the present paper. Deciding the definiteness of polynomials is, strictly speaking, outside the scope of the techniques presented in this paper. As we have seen in all examples in this paper, however, we have found it often possible to accomplish definiteness decisions by inspection. For certain examples in [10]this is not the case. However, the third author was able to use an implementation of the Collins method to carry out the definiteness decisions for all those examples except one (which does not appear in the present paper) in five minutes or less on a Vax 785.

In the future we will seek new blends of the Collins decision procedure, the Wu–Ritt method, the Gröbner basis method, and the techniques presented in this paper, that can mechanically prove geometry theorems involving inequalities, and that fruitfully balance speed with generality.

Acknowledgment The authors wish to thank Dianne King for careful editing of the manuscript and Professor Christoph Hoffmann and the referees for helpful advice.

References

[1] D. Arnon and M. Mignotte, "On Mechanical Quantifier Elimination for Elementary Algebra and Geometry", *J. Symbolic Computation*, **5** (1988), 237-259.

[2] D. Arnon, "Geometric Reasoning with Logic and Algebra", *Artificial Intelligence Journal*, V. **37** (1988), pp. 37-60.

[3] B. Buchberger, "Gröbner Bases: An Algorithmic Method in Polynomial Ideal Theory", Chapter 6 in *Recent Trends in Multidimensional Systems Theory*, N.K. Bose (ed.) D. Reidel Publ. Comp. 1985.

[4] B. Buchberger, G. E. Collins, and B. Kutzler, "Algebraic Methods for Geometric Reasoning", *Ann. Rev. Comput. Sci.* **3** (1988), 85-119.

[5] B. Buchberger, "Applications of Göbner Bases in Non-linear Computational Geometry", in *Geometric Reasoning*, Ed. by D. Kapur and J. Mundy, MIT Press, 413–446, 1989.

[6] 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.

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

[8] S.C. Chou, "A Method for Mechanical Derivation of Formulas in Elementary Geometry", *Journal of Automated Reasoning*, **3**(1987), 291–299.

[9] 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.

[10] S.C. Chou, "A Step toward Mechanically Proving Geometry Theorems Involving Inequality – Experimental Results", Preprint, March 1986, Institute for Computing Science, University of Texas at Austin.

[11] S.C. Chou, "A Step toward Mechanical Proving Geometry Theorems Involving Inequality", Preprint, May 1986, Institute for Computing Science, University of Texas at Austin, circulated at Oxford Workshop on Geometric Reasoning (June 30–July 3, 1986).

[12] S.C. Chou, " Proving Geometry Theorems Using Wu's Method: A Collection of Geometry Theorems Proved Mechanically", Technical Report 50, Institute for Computing Science, University of Texas at Austin, July 1986.

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

[14] S.C. Chou, W. F. Schelter, and J. G. Yang, "An Algorithm for Constructing Gröbner Bases from Characteristic Sets and its Application to Geometry", 1987, to appear in *Algorithmica.*

[15] 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.

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

[17] 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.

[18] G.E. Collins, "Quantifier Elimination for Real Closed Fields by Cylindrical Algebraic Decomposition", *Lecture Notes In Computer Science*, **33** (1975), Springer-Verlag, Berlin, 134-183.

[19] G.E. Collins and J. Johnson, "Quantifier Elimination and the Sign Variation Method for Real Root Isolation", in *Proceedings of the 1989 Symposium on Symbolic and Algebraic Computation*, ACM, New York, 264-271.

[20] Xiaoshan Gao, "Transcendental functions and Mechanical Theorem Proving in Elementary Geometries", 1988, to appear in *Journal of Automated Reasoning.*

[21] Xiaoshan Gao and Dongming Wang, "Geometry Theorems Proved Mechanically Using Wu's Method", Part on Elementary Geometries, MM preprint No2, 1987.

[22] Xiaoshan Gao, "Constructive Methods for Polynomial Set and Their Application", PhD Thesis, Institute of Systems Science, Academia Sinica, 1988 (in Chinese).

[23] R. Hartshorne, *Algebraic Geometry,* Springer-Verlag, 1978.

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

[25] D. Kapur, "A Refutational Approach to Geometry Theorem Proving", *Artificial Intelligence*, Vol. 37, (1988) pp61–93.

[26] D. Kapur and J. Mundy, "Wu's Method and its Application to Perspective Viewing", *Artificial Intelligence Journal*, V. **37** (1988), pp.15-36.

[27] B. Kutzler, "Algebraic Approaches to Automated Geometry Theorem Proving", PhD Thesis, Johannes Kepler University, Linz, Austria, 1988.

[28] J. F. Ritt, *Differential Equation from Algebraic Standpoint,* AMS Colloquium Publications Volume 14, New York, 1938.

[29] J. F. Ritt, *Differential Algebra,* AMS Colloquium Publications, New York, 1950.

[30] A. Seidenberg, "A New Decision Method for Elementary Algebra", *Annals of Math.*, **60** (1954), 365–371.

[31] A. Tarski, *A Decision Method for Elementary Algebra and Geometry,* Report R-109, second revised ed. Santa Monica, CA: The Rand Corporation, 1951.

[32] A. Tarski, "What is Elementary Geometry", Proc. Intl. Symp. on the Axiomatic Method, with special reference to geometry and physics, University of California, Berkeley, Dec. 1957 - Jan. 1958, Studies in Logic and the Foundations of Mathematics, (L. Henkin, P. Suppes, A. Tarski eds.), North Holland, Amsterdam, 1959, pp. 16-29.

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

[34] 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.

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

[36] Wu Wen-tsün, "On Zeros of Algebraic Equations –An Application of Ritt's Principle", *Kexue Tongbao* **31(1)** (1986), 1-5.

[37] Wu Wen-tsün, "A Mechanization Method of Geometry and its Applications – III: Mechanical Proving of Polynomial Inequality and Equations-Solving", *Research Preprints, Mathematics–Mechanization* No.2, pp1–17, 1987.

Appendix: The Ritt–Wu Decomposition Algorithm

1. Introduction

As already mentioned, Wu introduced his algebraic method for geometry theorem proving in 1977 [33]. In the course of Wu's further investigation of the method [35], [34] he discovered that its algebraic "machinery" was already known in the work of J. F. Ritt [28], [29]. Wu revised Ritt's work to better suit his objective of mechanically proving geometry theorems. The outcome of this developmental process is that we may describe the Wu–Ritt geometry theorem proving method as consisting of two key components: the Ritt–Wu Principle [34], and the Ritt–Wu Decomposition Algorithm [34], [36]. If one implements these algorithms directly from the descriptions of Ritt and Wu, one encounters drastic growth in the sizes of polynomials generated. Thus researchers have developed modified versions of Wu's and Ritt's original algorithms. Wu himself uses the notion of an "ascending chain in the weak sense" [34]However, ascending chains in Wu's weak sense still do not definitively deal with the problem of size growth.

This appendix is an abbreviation of the first part of the paper [15]. It presents yet another modified version of Ritt–Wu decomposition algorithm, essentially the version implemented in our computer programs. We omit many proofs, which can be found in [15].

2. Preliminary Definitions and Algorithms

Let K be a computable field such as $\mathbf{Q}$, the field of rational numbers, and $y = y_1, y_2, \ldots, y_m$ be indeterminates. Unless stated otherwise, all polynomials mentioned in this section are in $A = K[y_1, \ldots, y_m] = K[y]$. We fix the order of the indeterminates as $y_1 < y_2 < \cdots < y_m$, which is essential for the subsequent discussion. Unless stated otherwise, we assume this order among the variables $y_1, ..., y_m$.

Let f be a polynomial. Denote the degree of f in the variable y_i, i.e., the highest degree of y_i occurring in f, by $deg(f, y_i)$. The *class* of f is the smallest integer c such that f is in $K[y_1, \ldots, y_c]$. We denote it by $class(f)$. If f is in K we define $class(f) = 0$. Let $c = class(f)$ be nonzero and $lv(f)$ denote the *leading variable* y_c of f. Considering f as a polynomial in y_c, we can write f as

$$a_n y_c^n + a_{n-1} y_c^{n-1} + \cdots + a_0$$

where $a_n, ..., a_0$ are in $K[y_1, \ldots, y_{c-1}]$, $n > 0$, and $a_n \neq 0$. We call a_n the *initial* or leading coefficient of f and n the leading degree of f, denoting them as $lc(f)$ and $ld(f)$, respectively.

Now we present the pseudo division algorithm, a basic building block for subsequent algorithms. Let f and g be in $K[y]$ and v be one of the $y_1, ..., y_m$. Suppose that $deg(f, v) > 0$. Considering f and g as polynomials in v, we can write g and f as $g = a_n v^n + \cdots + a_0$, $f = b_k v^k + \cdots + b_0$. First set $r = g$. Then repeat the following process until $m = deg(r, v) < k$: $r := b_k r - c_m v^{m-k} f$, where c_m is the leading coefficient of r in the variable v. It is easy to see that m strictly decreases after each iteration. Thus the process terminates. At the end, we have the *pseudo remainder* $r = r_0$, which we write as $prem(g, f, v)$, and the following formula:

$$b_k^s g = qf + r_0, \quad \text{where } s \leq n - k + 1 \text{ and } deg(r_0, v) < deg(f, v).$$

Let f and g be two polynomials. A polynomial g is *reduced with respect to* f if $deg(g, y_c) <$

$deg(f, y_c)$, where $c = class(f) > 0$. Let $c = class(f) > 0$, then $prem(g, f, y_c)$ is reduced with respect to f; we denote $prem(g, f, y_c)$ simply by $prem(g; f)$.

Definition (2.1). Let $C = f_1, f_2, \ldots, f_r$ be a sequence of polynomials in $K[y]$. We call it a *quasi–ascending chain* or a triangular form if either $r = 1$ and $f_1 \neq 0$, or $r > 1$ and $0 < class(f_1) < class(f_2) < \cdots < class(f_r)$ for $i < j$.

Let $f_1, \ldots, f_r$ be a quasi–ascending chain with $class(f_1) > 0$. We define $prem(g; f_1, \ldots, f_r)$ inductively to be $prem((prem(g; f_2, \ldots, f_r); f_1)$. Let it be R. Then we have the following important *Remainder Formula*:

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

where the I_i are the initials of the f_i, $s_1, \ldots, s_r$ are some nonnegative integers, $Q_1, \ldots, Q_r$ are polynomials. Furthermore, $deg(R, x_i) < deg(f_i, x_i)$, for $i = 1, \ldots, r$, where $x_i = lv(f_i)$.

(i) A quasi–ascending chain is called an *ascending chain in Ritt's sense* if f_j are reduced with respect to f_i for $i < j$.

(ii) A quasi–ascending chain is called an *ascending chain in Wu's sense* if the initials I_j of the f_j are reduced with respect to f_i for $i < j$.

(iii) A quasi–ascending chain is called an *ascending chain in the weak sense* if $prem(I_i; f_1, ..., f_r) \neq 0$, for $i = 1, ..., r$.

Obviously, an ascending chain in Ritt's sense is an ascending chain in Wu's sense; an ascending chain in Wu's sense is an ascending chain in the weak sense. The key to our improved version of the algorithm is to use ascending chains in the weak sense. As Wu correctly pointed out, using quasi–ascending chains without any restrictions, one cannot insure the termination of algorithms (3.1), and (4.1) or (4.3). One of the main tasks of our improved version is to use ascending chains in the weak sense in a proper way, insuring both the termination of the algorithm and the reduction of the size growth of polynomials. From now on, *we will call an ascending chain in the weak sense simply an 'ascending chain'.*

We define a partial order $<$ in $K[y]$: $f < g$ (g is of *higher rank* or *higher* than f) if $class(f) < class(g)$ or $class(f) = class(g) > 0$ and $ld(f) < ld(g)$. If neither $f < g$ nor $g < f$, then we say f and g are of the same rank. Obviously, this partial order is well founded, i.e., every nonempty polynomial set S has a minimal element, i.e., the one which is not higher than any other element in S.[12]

Definition (2.2). Let $C = f_1, ..., f_r$ and $C_1 = g_1, ..., g_m$ be two ascending chains. We define $C < C_1$ if there is an s such that $s \leq \min(r, m)$ and f_i and g_i are of the same rank for $i < s$ and that $f_s < g_s$, or $m < r$ and f_i and g_i are of the same rank for $i \leq m$.

Proposition (2.3). The partial order $<$ among the set of all ascending chains is well-founded, i.e, there are no infinite, strictly decreasing sequences of ascending chains $C_1 > C_2 > \cdots > C_k > \cdots$.

Proof. See Lemma 1 of [34].

Definition (2.4). Let S be a nonempty polynomial set. A minimal ascending chain in the set of all chains formed from polynomials in S is called a *basic set* of S.

[12] In practice, one can further order polynomials with the same rank to enhance the efficiency while preserving the well foundedness.

Unless stated otherwise, whenever we talk about a finite polynomial set S, we assume S does not contain zero. By (2.3), every nonempty polynomial set S has a basic set.

Algorithm (2.5). Let S be a finite, nonempty polynomial set. The algorithm is to construct a basic set of S.

Algorithm and Proof. Let f_1 be a polynomial with minimal rank in S. If f_1 is of class zero, then it is a basic set of S. Now let f_1 be of positive class. Let S_1 be the set of all polynomials in S, whose classes are higher than $class(f_1)$ and whose initials I are such that $prem(I; f_1) \neq 0$. If S_1 is empty, then f_1 forms a basic set of S. Now suppose S_1 is nonempty. Continuing this way, at step k, we have an ascending chain $C = f_1,...,f_k$ in S. Let S_k be the set of all polynomials in S, whose classes are higher than $class(f_k)$ and whose initials I are such that $prem(I; f_1, ..., f_k) \neq 0$. If S_k is empty, then $f_1, ..., f_k$ is a basic set of S. Otherwise, choose an element f_{k+1} with minimal rank in S_k. $f_1, ..., f_k, f_{k+1}$ form an ascending chain again. Eventually, we arrive at a basic set of S in no more than m steps. .QED.

In the original presentation of the Ritt–Wu Principle (cf. [29], [34]) the key operation $prem(f; ASC)$ is repeatedly used. Since the main purpose of triangulation is to reduce the class or the leading degree of f, we need only take fewer pseudo remainders than $prem(f; ASC)$ takes. This can reduce the size growth of polynomials produced. The following W–*prem* is one of our key steps to control the size growth of polynomials.

Algorithm W–*prem* (2.6). Given a polynomial g and an ascending chain $ASC = f_1, ..., f_r$ with nonconstant f_1. We define W–$prem(g; ASC)$ to be:

Case 1. $prem(g; f_1, ..., f_r)$ if $prem(initial(g); f_1, ..., f_r) = 0$.

Case 2. g if $class(f_r) < class(g)$.

Case 3. W–$prem(prem(g; f_r); f_1, ..., f_{r-1})$ if $class(f_r) = class(g)$.

Case 4. W–$prem(g; f_1, ..., f_{r-1})$ if $class(f_r) > class(g)$.

The remainder formula is still valid for W–*prem*, except $deg(R, x_i) < deg(f_i, x_i)$ (where $x_i = lv(f_i)$) is not necessarily true.

Proposition (2.7). For a nontrivial ascending chain $ASC = f_1, ..., f_r$ and a polynomial g, if W–$prem(g; ASC) = 0$, then $prem(g; ASC) = 0$.

Proof. See [15]. .QED.

We introduce a new notation extremely important for the rest of the paper:

$$PD(ASC) = \{g \mid prem(g; ASC) = 0\}.$$

Thus, (2.7) says that if W–$prem(g; ASC) = 0$, then $g \in PD(ASC)$. The following proposition insures the termination of the triangulation procedure of the Ritt–Wu Principle, when it uses W–*prem*.

Proposition (2.8). Let $B = f_1, \ldots, f_r$ be a basic set of polynomial set S with $0 < class(f_1)$, and h be a polynomial. Suppose $g =$ W–$prem(h; f_1, \ldots, f_r)$ is not zero. Then the set $S_1 = S \cup \{g\}$ has a basic set lower than B.

Proof. See [15]. .QED.

Let $ASC = f_1, \ldots, f_r$ be an ascending chain, not consisting of a constant. After a *suitable renaming* [13] of the y_j, we may assume that $class(f_1) = d + 1$ and $m = d + r = class(f_r)$, where $d \geq 0$. We distinguish the y_i for $i \leq d$ by calling them u_i and use x_i to denote $lv(f_i)$. We call $\{u_1, ..., u_d\}$ the parameter set of the ascending chain ASC.

Thus ASC has the following "triangular" form:

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

Definition (2.9). An ascending chain $f_1, ..., f_r$ of the form (T) is called *irreducible* if each f_i is irreducible in the polynomial ring $K(u)[x_1, \ldots, x_i]/(f_1, ..., f_{i-1})$. Thus the sequence $F_0 = K(u)$, $F_1 = F_0[x_1]/(f_1)$, ..., $F_r = F_{r-1}[x_r]/(f_r) = F_0[x]/(f_1, ..., f_r)$ is a tower of field extensions.[14]

Theorem (2.10). Let ascending chain ASC of the form (T) be irreducible, g be a polynomial. Then $PD(ASC)$ is a prime ideal and the following are equivalent:

(i) $g \in PD(ASC)$, i.e., $prem(g; ASC) = 0$.

(ii) $Zero(PD(ASC)) \subset Zero(g)$, where the zeros are taken to lie in an algebraically closed extension of the field K.

Proof. See lemma 3, page 234 in [34] and Theorem (3.7) on page 31 of [13]. .QED.

Theorem (2.11). Let ascending chain ASC be irreducible and g be a polynomial. If $prem(g; ASC) \neq 0$ then there exist a polynomial p and a nonzero polynomial $h \in K[u]$ such that $pg - h \in Ideal(ASC)$.

Proof. See [15]. .QED.

Theorem (2.12). Let $f_1, \ldots, f_r$ be an ascending chain. Suppose that $f_1, \ldots, f_{k-1}$ $(0 < k \leq r)$ is irreducible, but $f_1, \ldots, f_k$ is ıreducible. Then there are polynomials g and h in $K[u, x]$ reduced with respect to $f_1, \ldots, f_r$ such that $class(g) = class(h) = class(f_k)$ and $gh \in$ the ideal generated by $f_1, ..., f_k$.

Proof. See Theorem (3.6) on page 30 of [13]. .QED.

3. The Ritt–Wu Principle (a Modified Version)

A complete triangulation algorithm was implicit in Ritt's work ([28], [29]) and was rewritten by Wu in detail ([35], [34]). It was called *Ritt's Principle*, and considered the basis of his own

---

[13] This renaming changes the ordering of the y in a way that the variables y_i not occurring in ASC are less than the variables occurring in ASC. The ordering among the variables occurring in ASC are the same as before; The variables not occurring in ASC can be in any order.

[14] Here $(f_1, ..., f_r)$ etc. denotes the polynomial ideal of $K(u)[x]$ (not of $K[u, x]$), generated by $f_1, ..., f_r$.

method, by Wu. The following modification is an improvement and used in our prover.

Theorem (3.1). (Ritt–Wu Principle). Let $S = \{h_1, ..., h_n\}$ be a finite nonempty polynomial set in $A = K[y_1, \ldots, y_m]$, and I be the ideal $(h_1, ..., h_n)$ of A. The algorithm is to construct an ascending chain ASC such that either

(3.2). ASC consists of a nonzero constant in $K \cap I$.

(3.3). $ASC = f_1, \ldots, f_r$ with $class(f_1) > 0$ and such that $f_i \in I$ and W–$prem(h_j; f_1, ..., f_r) = 0$ for all $i = 1, \ldots, r$ and $j = 1, \ldots, n$.

Proof. By (2.5), we can construct a basic set B_1 of $S_1 = S$. If B_1 consists of only one nonzero constant, then we have (3.2). Otherwise, we can expand S_1 to S_2 by adding nonzero W–$prem(g; B_1)$ of all g elements of S_1. If $S_2 = S_1$, then we have (3.3). Otherwise, we can construct a basic set B_2 of S_2. By (2.8), $B_1 > B_2$. If B_2 does not consist of one nonzero constant, then we can expand S_2 to S_3 using the same procedure. Thus we have a strictly increasing sequence of polynomial sets:

$$S_1 \subset S_2 \subset \cdots,$$

with the corresponding strictly decreasing sequence of characteristicnon sets

$$B_1 > B_2 > \cdots .$$

By (2.3), this decreasing sequence can be only finite. Thus, there is an integer $k \geq 1$ such that either B_k consists of a nonzero constant or $S_k = S_{k+1}$; then we have either (3.2) or (3.3), respectively. .QED.

Now let us fix an extension E of the base field K. We denote $Zero(S)$ the common zeros of polynomials in S, i.e., the set

$$\{(a_1, ..., a_m) \in E^m \mid h(a_1, ..., a_m) = 0, \text{ for all } h \in S\}.$$

Let G be another polynomial set. Following Wu, we denote $Zero(S/G)$ to be $Zero(S) - Zero(G)$. Note that *all zeros are taken from the (fixed) extension E*. Unless essential, we will not mention this field explicitly. We have $Zero(S/\{1\}) = Zero(S)$.

Let ASC be a nontrivial ascending chain and G be a polynomial set. We introduce a new notation $pfactors(G; ASC) =$

Case 1. 0 if $prem(g; ASC) = 0$ for some $g \in G$.

Case 2. $\bigcup\{$all prime factors of $prem(g; ASC)|$ for all $g \in G\}$.

In the case of (3.2), the polynomial set S is said to be contradictory and does not have a common zeros. Otherwise we have the following:

Theorem (3.4). Suppose S in (3.1) is not contradictory. Let $ASC = f_1, \ldots, f_r$ be the ascending chain obtained in (3.3), I_k be the initials of the f_k, $I = \{I_1, ..., I_r\}$ (I is called the *initial set* of ASC) and $J = pfactors(I; ASC)$ (note that J is nonzero).

(i) $Zero(ASC/I) = Zero(ASC/J)$.

(ii) $Zero(ASC/I) \subset Zero(PD(ASC)) \subset Zero(S) \subset Zero(ASC)$.

(iii) $Zero(S) = Zero(ASC/I) \cup \bigcup_p \{Zero(S \cup \{p\}) \mid p \in I\}$.

(iv) $Zero(S) = Zero(ASC/J) \cup \bigcup_p \{Zero(S \cup \{p\}) \mid p \in J\}$.

Proof. See [15]. .QED.

4. The Ritt–Wu Decomposition Algorithm (a Modified Version)

Algorithm (4.1). Ritt–Wu Zero Decomposition Algorithm (Refined Form). Let S and G be two nonempty polynomial sets. The algorithm is either to detect the emptiness of $Zero(S/G)$ or to decompose $Zero(S/G)$ in the following form:

$$Zero(S/G) = \bigcup_{1 \le i \le k} Zero(ASC_i/I_i \cup G) \tag{4.1.1}$$

$$Zero(S/G) = \bigcup_{1 \le i \le k} Zero(PD(ASC_i)/G) \tag{4.1.2}$$

where each ASC_i is a nontrivial *irreducible* ascending chain, the I_i are the initial sets of the ascending chains ASC_i, and $prem(g; ASC_i) \neq 0$ for all $g \in G$ and $i = 1, ..., k$.

Proof. Let $ASCs$ be a set of ascending chains, initialized to be empty at the beginning.

Step 1. According to (3.1) we can construct an ascending chain having the property of either (3.2) or (3.3). In the case of (3.2), $Zero(S/G)$ is empty. In the case of (3.3), we have an ascending chain ASC and a polynomial set S' (i.e., S_k in the proof of (3.1)) having ASC as one of its basic sets. $Zero(S) = Zero(S')$.

Step 2. Check whether the ascending chain $ASC = f_1, ..., f_r$ is reducible. If it is, then there is an integer $k > 0$ such that $f_1, ..., f_{k-1}$ is irreducible, but $f_1, ..., f_k$ is reducible. By (2.12), we can find two polynomials g and h with $class(f_k) = class(g) = class(h)$ and $gh \in Ideal(f_1, ..., f_k)$. We have decomposition: $Zero(S') = Zero(S' \cup \{g\}) \cup Zero(S' \cup \{h\})$. Obviously, $S' \cup \{g\}$ and $S' \cup \{h\}$ have basic sets strictly lower than that of S'. We can take each of $S' \cup \{g\}$ and $S' \cup \{h\}$ as a new S, and go to step 1.

Step 3. Let I be the initial set of of ASC. By (3.4) we have:

$$Zero(S/G) = Zero(S'/G) = Zero(ASC/I \cup G) \cup \bigcup_p \{Zero(S' \cup \{p\}/G) : p \in I\}. \tag{4.1.3}$$

Step 4. If $prem(g; ASC) = 0$ for some $g \in G$, then $Zero(ASC/I \cup G)$ is empty. Otherwise, we add this ascending chain to $ASCs$.

Step 5. For each p in I, let $p' = prem(p; ASC)$. Note that $p' \neq 0$. For each $Zero(S' \cup \{p\}/G) = Zero(S' \cup \{p, p'\}/G)$ in (4.1.3), take $S' \cup \{p, p'\}$ as a new S, then go to step 1. Repeat this process recursively. Since $S' \cup \{p, p'\}$ has a basic set *strictly* lower than that of S' by (2.8), this recursive process will finally terminate. For otherwise, we would have a strictly decreasing

sequence of ascending chains, contradicting to (2.3). The termination of each branch happens when I consists of constant polynomials.

Upon termination, we have two cases:

(i) $ASCs$ is empty. This means that S does not have common zeros.

(ii) $ACSs = \{ASC_1, ..., ASC_k\}$ $(1 \leq k)$, then we have the decomposition (4.1.1). Since $Zero(ASC_i/I_i) \subset Zero(PD(ASC_i)) \subset Zero(S)$, (4.1.2) follows from (4.1.1). .QED.

Remark. The branches produced in the recursive step 5 can number in the thousands and *most* of them are redundant. For $G = 1$, we still lack a satisfactory strategy to control the growth of the branches and make termination occur earlier. In Part II of [15], G is a set of polynomials expressing degenerate cases. In this case we do have a modification to control the growth of branches effectively (see [15]).

Theorem (4.2). Let E be an algebraically closed extension of the base field K and $G = \{1\}$. Then (4.1.2) becomes

$$Zero(S) = \bigcup_{1 \leq i \leq k} Zero(PD(ASC_i)) \tag{4.2.1}$$

which is a decomposition of algebraic set $Zero(S)$ into the union of the irreducible varieties $Zero(PD(ASC_i))$. Here each $PD(ASC_i)$ is a prime ideal by (2.10). Or alternatively,

$$Radical(S) = \bigcap_{1 \leq i \leq k} PD(ASC_i). \tag{4.2.2}$$

Step 2 in (4.1) generally requires factorization of polynomials over successive algebraic extensions of the field of rational functions. The following variant of (4.1) does not require factorization over extension fields.

Algorithm (4.3). Ritt–Wu Zero Decomposition Algorithm (Coarse Form). The same statement as in (4.1), except we do not require that each ascending ASC_i be irreducible.

Proof. The only thing needing change in Algorithm (4.1) is to drop step 2 in the proof of (4.1). However, since multivariate factorization is available in many computer algebra systems, we suggest keeping step 2 and checking the reducibility of $prem(f_k; f_1, ..., f_{k-1})$. .QED.

In the coarse form, $PD(ASC_i)$ may not even be an ideal. Thus, decomposition (4.2.2) is generally not valid.

The decompositions in (4.1)–(4.3) are generally redundant, i.e., some components may be contained in others. To remove *all* such redundancy is time-consuming. However, the following theorem removes some redundancy at no cost.

Theorem (4.4). Let $n = length(S)$ be the number of polynomials in S. Suppose that the emptiness of $Zero(S)$ is not detected in algorithm (4.1) or (4.3) and the set unions in (4.1.1) and (4.1.2) (either in the refined form or in the coarse form) are arranged in such a way that $length(ASC_i) \leq n$ for $i \leq l$, and $length(ASC_i) > n$ for $i > l$ for some integer $0 \leq l \leq k$, then $0 < l$ we have the decomposition

$$Zero(S/G) = \bigcup_{1 \leq i \leq l} Zero(PD(ASC_i)/G). \tag{4.4.1}$$

Proof. The theorem is based on the Affine Dimension Theorem (page 48 in [23]) and the following Lemma (4.5). For details see the proof of Theorem (4.4) in [15]. .QED.

Lemma (4.5). Let $ASC = f_1, ..., f_r$ be a nontrivial *quasi* ascending chain, I_i be the initials of f_i, and $J = \{I_1, ..., I_r\}$. Then $Zero(ASC/J)$ is contained in the union of varieties $\subset Zero(PD(ASC))$ with dimensions $\leq m - r$.

Proof. See (9.7) of [15]. .QED.

Remark. This lemma is difficult to prove. Notice that if ASC is irreducible, then Lemma (4.5) is obviously true by the Affine Dimension Theorem. Thus Theorem (4.4) under the refined form (4.1) is true, independently of Lemma (4.5). The practical importance of Lemma (4.5) is that we can use Theorem (4.4) without factorization. Notice also that the formula:

$$Zero(S/G) = \bigcup_{1 \leq i \leq l} Zero(ASC_i/I_i \cup G)$$

is generally not true even for the refined form. This is the key advantage to use $Zero(PD(ASC_i))$ instead of $Zero(ASC_i/I_i)$.

Theorem (4.6). There is an algorithm to remove the redundancy in the decomposition (4.2.1) *completely.*

Proof.

Step 1. First we can use Theorem (4.4) to remove some redundancy in (4.2.1) at no cost.

Step 2. Use Theorems (9.5) and (9.6) in [15] to remove further redundancy.

Step 3. For each remaining prime ideal $PD(ASC_i)$, we can obtain its Gröbner basis from the ascending chain ASC_i, using the algorithm in [14].[15] Having the Gröbner bases, we can decide the inclusionship among these prime ideals, thus removing the remaining redundancy. .QED.

Remark. Steps 1 and 2 are not necessary, but they are much cheaper than step 3. Thus the algorithm is more efficient based on Theorem (4.4), and Theorems (9.5) an (9.6) in [15].

[15] Let $ASC = f_1, ..., f_r$ be an irreducible ascending chain. Then $GB(PD(ASC)) = K[y] \cap GB(f_1, ..., f_r, I \cdot z - 1)$, where I is the product of all initials of ASC and z is a new variable. Here the compatible ordering among monomials can be any ordering satisfying $u^i x^j < z$. For details, see [14].