RITT-WU'S DECOMPOSITION ALGORITHM AND GEOMETRY THEOREM PROVING*

Shang-Ching Chou and Xiao-Shan Gao†

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

TR-89-09 March 1989
September 1990 (Revised)

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

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

Ritt-Wu's Decomposition Algorithm and Geometry Theorem Proving*

SHANG-CHING CHOU and XIAO-SHAN GAO†

Department of Computer Sciences
The University of Texas at Austin, Austin, Texas 78712 USA

March, 1989

Abstract An improved Ritt-Wu's decomposition (of an algebraic set into the union of irreducible varieties) algorithm is given. The algorithm has been used to prove geometric theorems that Wu's original method addresses. Unlike Wu's original approach, nondegenerate conditions are given explicitly at the beginning, not generated during the proof process. A program based on this improved version of the algorithm proved more than 500 theorems, including Morley's trisector theorem.

Keywords Wu's method, mechanical theorem proving, prover, elementary geometry, degenerate conditions, Ritt-Wu's principle, algebraic variety, nondegenerate component, ideal, ascending chain, the dimension theorem, Morley's trisector theorem.

---

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

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

1. Introduction

In 1977 Wu Wen-Tsün introduced an algebraic method which could be used to prove quite non-trivial theorems not involving betweenness in Euclidean geometry [11]. Further work [13], [12] showed that the algebraic tools and algorithms of the method were already begun in the work of J. F. Ritt [9], [10]. Wu revised Ritt's work for his own need of mechanically proving geometry theorems. Key to the method is Ritt-Wu's principle [12], and Ritt-Wu's Zero Structure Algorithm [14]. However, there is almost no work touching the improvement of these algorithms. If one implements the algorithms literally according to the description of the work of Ritt and Wu, the sizes of polynomials produced in the process will become larger and larger. People actually use some modifications of Ritt-Wu's original descriptions. Especially, Wu himself uses the notion of ascending chain in weak sense [12] to reduce the sizes of polynomial produced. However, ascending chain in Wu's weak sense still cannot prevent the size growth of polynomials in many cases.

This paper presents another modification of Ritt-Wu's decomposition algorithm, giving its full descriptions and the proof of its correctness (including its termination.) The efficiency of the modification has been demonstrated by the second part of the paper. Besides the improvements of the algorithms, we establish several theorems (especially Theorem (4.4)) of both theoretical and practical interests.

The paper consists of two parts: the improved algorithm and its application to geometry theorem proving.

In the second part we will address the same kinds of geometric statements as Wu's original method addresses. A valid geometric statement is valid only under certain nondegenerate conditions. There are two approaches to dealing nondegenerate conditions:

Approach (1.1). Introducing the notion of "generally (generically) true" and proving geometric statement to be generally true, at the same time giving the nondegenerate conditions automatically during the proof process.

Approach (1.2). Giving nondegenerate conditions manually at the beginning as a part of the hypothesis. Then the prover only needs to answer whether the conclusion follows the hypothesis without adding any additional conditions.

Our prover [2] mainly uses approach (1.1). Approach (1.1) often generates non-degenerate conditions more than actually needed. The second part is to address approach (1.2). Our improved algorithm/program has proved more than 500 theorems according to approach (1.2). Among the work related to approach (1.2), we mention the work of H. P. Ko [8] and D. Kapur [7]. We will discuss their work in Section 8.

Part I. An Improved Ritt-Wu's Decomposition Algorithm

2. Preliminary Definitions and Algorithms

In order to make the paper self contained, we first introduce some definitions briefly, which can be quickly gone through if the reader is already familiar with Ritt-Wu's work [10], [12].

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 non-zero 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 step for most 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 while $m = deg(r, v) \geq k$: $r := b_k r - c_m v^{m-k} f$, 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* $prem(g, f, v) = r = r_0$ 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 $0 < class(f_1) < class(f_2) < \cdots < class(f_r)$.

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 non-negative 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 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 weak sense. The key to our improved version of the algorithm is to use ascending chains in weak sense. As Wu correctly pointed out that 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 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 weak sense simply as 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.[1]

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 $s \le \min(r, m)$ 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 $i \le m$.

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 [12].

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.

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, non-empty polynomial set. The algorithm is to construct a basic set of S.

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, chose 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. ∎

In the original presentation of Ritt-Wu's principle (cf. [10], [12]) 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 to 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

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

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 non-constant 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 non-trivial ascending chain $ASC = f_1, ..., f_r$ and a polynomial g, if W–$prem(g; ASC) = 0$, then $prem(g; ASC) = 0$.

Proof. We use induction on r. Suppose g is not zero and for $r - 1$, the proposition is true. We want to prove it is true for r. According to (2.6) there are 4 cases. In case 1, 0 = W–$prem(g; ASC) = prem(g; ASC)$. Case 2 cannot happen since W–$prem(g; ASC) = g \neq 0$. In case 3, W–$prem(g; ASC)$ = W–$prem(prem(g; f_r); f_1, ..., f_{r-1}) = 0$. By the induction hypothesis, $prem(g; ASC) = prem(prem(g; f_r); f_1, ..., f_{r-1}) = 0$. In case 4, by the induction hypothesis again, $prem(g; ASC) = 0$. ∎

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 Ritt-Wu's principle, when using 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. From Algorithm (2.6) for W–*prem*, it is not hard to see that (i) W–$prem(initial(g); f_1, \ldots, f_r) \neq 0$; (ii) if g and f_k have the same class, say, i, then $deg(g, y_i) < deg(f_k, y_i)$.

If $class(g) \leq class(f_1)$, then g alone forms an ascending chain lower than B. Now suppose $class(g) > class(f_1)$, and let $j = max\{i \mid class(f_i) < class(g)\}$. If $class(f_{j+1}) = class(g)$, then $ld(g) < ld(f_{j+1})$. Thus $f_1, ..., f_j, g$ form an ascending chain lower then B. ∎

3. A Modification of Ritt-Wu's Principle

A complete triangulation algorithm, which was implicitly in Ritt's work ([9], [10]) and was rewritten by Wu in detail ([13], [12]). It was called *Ritt's Principle* and considered the basis of his method by Wu. The following modification is an improvement and used in our prover.

Theorem (3.1). (Ritt-Wu's 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 non-zero 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–$prem(h_j;\ f_1,\ldots,f_r) = 0$ 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 characteristic 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. ∎

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

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

Let G be another polynomial set. Following Wu, we denote $Zero(S/G)$ to be $Zero(S) - \bigcup_{g \in G} Zero(g)$. 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 non-trivial 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, $I = \{I_1, \ldots, I_r\}$ (I is called the *initial set* of ASC) and $J = pfactors(I; ASC)$ (note that J is non-zero).

(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. For each of I_k, letting $I'_k = prem(I_k; f_1, ..., f_{k-1})$, we have

$$I_1^{s_1} \cdots I_{k-1}^{s_{k-1}} I_k = Q_1 f_1 + \cdots + Q_{k-1} f_{k-1} + I'_k. \tag{3.4.1}$$

For some non-negative integer s_i ($i = 1, ..., k-1$) and polynomials Q_i ($i = 1, ..., k-1$). Therefore, if $a \in Zero(ASC/I)$, then $a \in Zero(ASC/J)$. Conversely, if $a \in Zero(ASC/J)$, then there is k such that $I'_k(a) \neq 0$, where $I'_k = prem(I_k; f_1, ..., f_{k-1})$. By (3.4.1) again, $I_j(a) = 0$ for some $j \leq k$. Thus $a \in Zero(ASC/I)$. Therefore, (i) follows.

Since $ASC \subset Ideal(S)$ and $S \subset PD(ASC)$ (by (2.7) & (3.3)), $Zero(PD(ASC)) \subset Zero(S) \subset Zero(ASC)$. For each $h \in PD(ASC)$ we have $prem(h; ASC) = 0$, thus by the remainder formula, we have:

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

That means $Zero(ASC/I) \subset Zero(PD(ASC))$. Therefore, (ii) follows.

Since $Zero(ASC) = Zero(ASC/I) \cup \bigcup_p \{Zero(ASC \cup \{p\}) \mid p \in I\}$, (iii) follows from (ii) by taking intersection with $Zero(S)$. (iv) is similar. [2] ∎

4. A Modification of Ritt-Wu's Decomposition Algorithm

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

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

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

where each ASC_i is a non-trivial *irreducible* ascending chain,[3] 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 $ASCs$ 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

[2] Since the main purpose of the algorithm in (3.1) is to insure theorem (3.4), we can delete some redundant factors of a polynomial produced by pseudo division. For example, we can replace $y_1 y_2^3 - y_2^2$ by $y_1 y_2^2 - y_2$. Such variable factors y_i is easy to detect and remove. However, it enhances the efficiency greatly in some cases.

[3] For the definition and properties of irreducible ascending chains see the Appendix or [10], [12], [2].

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 (9.4) in the Appendix, 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:

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

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). ∎

Remark. The branches produced in the recursive step 5 can be as many as thousands and most of them are redundant. If $G = 1$, we still don't have a satisfactory strategy to control the growth of the branches and make the termination earlier. In Part II, G is a set of polynomials expressing degenerate cases. We have the following modification to control the growth of branches effectively.

Step 3'. Let I be the set of all initials of ASC,

$$I' = \{p \mid p \in I \text{ and is not a factor of some } g \in G\},$$

$J = pfactors(I'; ASC)$, and $D = \{p \mid p \in I' \text{ and is not a factor of some } g \in G\}$. [4] We have:

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

[4] In our actual implementation, the procedure is more complicated. For example, we can at least use $g' = \text{G--}prem(g; ASC)$ instead of g in G. Here G–*prem* is different from *prem* in that G–*prem* uses only polynomials with constant initials in ASC to take pseudo remainders.

Step $4'$. If $prem(g; ASC) = 0$ for some $g \in G$, then $Zero(ASC/J \cup G)$ is empty. Otherwise, we add this ascending chain to $ASCs$. *More important*, if G–$prem(g; ASC) = 0$ for some $g \in G$, then $Zero(ASC/G)$, hence $Zero(S/G)$ is empty; this branch terminates. Here we use the notation G–*prem* in the previous footnote.

Step $5'$. For each $Zero(S' \cup \{p\}/G)$, do the same recursive process as in step 5. The termination of each branch happens when D in step $3'$ consists of constant polynomials or G–$prem(g; ASC) = 0$ for some $g \in G$ in step $4'$. With a careful arrangement, we can make many branches terminated earlier. This is another key step of our improvements. *End of Remark.*

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

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

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 (see (9.2) in the Appendix). Or alternatively,

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

Step 2 in (4.1) generally needs factorization of polynomials over successive algebraic extensions of the field of rational functions. The actual implementation in our prover can only do (1) factorization of polynomials over the field of rational functions; (2) factorization of polynomials over successive quadratic extensions of the field of rational functions. Even (1) is enough for most problems we found in geometry. The following variant of (4.1) does not need factorization over extension fields.

Algorithm (4.3). Ritt-Wu's 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 needed to change in Algorithm (4.1) is to drop step 2 in the proof of (4.1). However, since multivariate factorization is available in many algebraic systems, we suggest to keep step 2 and check the reducibility of $prem(f_k; f_1, ..., f_{k-1})$. ∎

In the coarse form, $PD(ASC_i)$ even may not 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 $Zero((PD(ASC_i))$ may be contained in others. To remove *all* such redundancy (see Theorem (4.6)) is time-consuming. However, the following theorem, which is important to the second part of the paper, removes some redundancy without any 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 $length(ASC_i) \le n$ for $i \le l$, and $length(ASC_i) > n$ for $i > l$ for some integer $0 \le l \le k$, then $0 < l$ we have the decomposition

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

Proof. First we assume E is algebraically closed and $G = \{1\}$. If $Zero(S)$ is empty (this can be definitely detected using algorithm (4.1)), then nothing is needed to prove. Assume $Zero(S)$ is non-empty. Then we have

$$\begin{aligned} Zero(S) &= \bigcup_{1\leq i\leq l} Zero(ASC_i/I_i) \cup \bigcup_{l<i\leq k} Zero(ASC_i/I_i) \\ &= \bigcup_{1\leq i\leq l} Zero(PD(ASC_i)) \cup \bigcup_{l<i\leq k} Zero(ASC_i/I_i) \end{aligned}$$

By the Affine Dimension Theorem (page 48 in [6]), the dimensions of all *irredundant* (irreducible) components of $Zero(S)$ are greater than or equal to $m-n$, (Remember that m is the number of variables $y_1, ..., y_m$.) By Lemma (4.5) below, $Zero(ASC_i/I_i)$ is contained in the union of irreducible varieties (of $Zero(PD(ASC_i)) \subset Zero(S)$) with dimension $\leq m - length(ASC_i)$. Thus, if $i > l$, $m - length(ASC_i) < m - n$ and each such irreducible variety with dimension $< m-n$ must be in one of the components of $Zero(S)$. Therefore, $l > 0$ and each components of $Zero(S)$ must be contained in some $Zero(PD(ASC_i))$ for $i \leq l$. Hence,

$$Zero(S) = \bigcup_{1\leq i\leq l} Zero(PD(ASC_i)). \tag{4.4.2}$$

Since any extension E of K is contained in an algebraically extension of K, (4.4.2) is valid for any extension E of K. For any polynomial set G, (4.4.1) follows from (4.4.2). (Here we have a little abuse of notations, the l in (4.4.1) and (4.4.2) are different. By the algorithm (4.3) (see Step 3), only those ASC_i in (4.2.2) are kept in (4.4.1) that $prem(g; ASC_i) \neq 0$ for all $g \in G$.) ▮

Remark. 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 $Zero(ASC_i/I_i)$.

Lemma (4.5). Let $ASC = f_1, ..., f_r$ be a non-trivial *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. This is the theorem most difficult to prove in this paper. We write the proof completely in (9.7) in the Appendix. ▮

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) without any cost.

Step 2. Use Theorems (9.5) and (9.6) in the Appendix 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 [3].[5] Having the Gröbner bases, we can decide the inclusionship among these prime ideals, thus removing the remaining redundancy. ∎

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

The method for geometry theorem proving in Part II is based on the following two theorems.

Theorem (4.7). Let the notation be the same as in (4.4) and g be any polynomial. Suppose we have decomposition (4.4.1) (in the coarse or refined form). If $prem(g; ASC_i) = 0$ for all $i = 1, ..., l$, then $Z(S/G) \subset Zero(g)$.

Proof. Since by assumption $prem(g; ASC_i) = 0$, $g \in PD(ASC_i)$. Hence $Zero(PD(ASC_i)) \subset Zero(g)$ for all i. By (4.4.1), $Zero(S/G) \subset Zero(g)$. ∎

Theorem (4.8). Let the notation be the same as in (4.4) and g be any polynomial. Suppose we have decomposition (4.4.1) in the refined form (i.e., all ASC_i are irreducible) and all zeros are taken from an *algebraically closed* extension E of K. Then

(i) Each $Zero(PD(ASC_i)/G)$ is non-empty.

(ii) $Zero(S/G) \subset Zero(g)$ if and *only if* $prem(g; ASC_i) = 0$ for all $i = 1, ..., l$.

Proof. This is an obvious consequence of Theorem (9.3) in the Appendix. ∎

Part II. Applications to Geometry Theorem Proving

5. A Method for Approach (1.2)

Let E be the field associated with a given geometry. Suppose the hypothesis of a geometry statement can be algebraically expressed by a set of polynomial equations $\{h_1(y_1, ..., y_m) = 0, ..., h_n(y_1, ..., y_m) = 0\}$ together with a set of inequations $\{s_1(y_1, ..., y_m) \neq 0, ..., s_q(y_1, ..., y_m) \neq 0\}$ expressing the non-degenerate conditions and the conclusion by a polynomial equation $g(y_1, ..., y_m) = 0$. Then the equivalent algebraic form of the geometry statement is

(5.1) $$\forall y_1 \cdots y_m \in E[(h_1 = 0 \wedge \cdots \wedge h_n = 0 \wedge s_1 \neq 0 \wedge \cdots \wedge s_q \neq 0) \Rightarrow g = 0].$$

Let $S = \{h_1, ..., h_n\}$ and $G = \{s_1 ..., s_q\}$, then the above formula is equivalent to

(5.2) $$Zero(S/G) \subset Zero(g).$$

Method (5.3). Suppose a geometry statement can be given algebraically in the form (5.1), the method is to confirm (5.1), or in the case when the field associated with the geometry is algebraically closed, to decide whether (5.1) is valid.

[5] 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 [3].

Using Algorithms (4.1) or (4.3), and Theorem (4.4) to decompose $Zero(S/G)$ into

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

Each $Zero(PD(ASC_i)/G)$ is called a *component* of $Zero(S/G)$, though it may redundant, reducible or even empty.

Case 1. $prem(g; ASC_i) = 0$ for all $i = 1, ..., l$. Then (5.2), hence formula (5.1) is valid by Theorem (4.7).

Case 2. $prem(g; ASC_i) \neq 0$ for some i. If E is algebraically closed and each ascending chain ASC_i is irreducible, then formula (5.1) is not valid by Theorem (4.8). ∎

In case 2 and when formula (5.1) is disproved, the question that whether the original geometry statement is true remains open up to the interpretation of the person who uses the method. Since nondegenerate conditions are often implicit in a geometry statement and are extremely hard to find in certain cases (see the examples below), some of them may be missing in formula (5.1). In that case formula (5.1) may be false and we don't have any information about the reason why it is false: it is false because the geometry statement is *generally false* or because some nondegenerate conditions are missing. This is why the authors are in favor of approach (1.1) in Section 1 to introduce the notion "*generally (generically) true*", which is inherent to a given geometry statement regardless of how much nondegenerate conditions are added.

6. Examples

Example (6.1) (Pascal's Theorem). Let A, B, C, D, F and E be six points on a circle (O). Let $P = AB \cap DF$, $Q = BC \cap FE$ and $S = CD \cap EA$. Show that P, Q and S are collinear (Figure 1).

The obvious non-degenerate conditions in this problem seem to be "the three pairs of lines, AB and DF, BC and FE, and CD and EA, have normal intersections". Thus we can let $B = (u_1, 0)$, $A = (0, 0)$, $C = (u_2, u_3)$, $O = (x_2, x_1)$, $D = (x_3, u_4)$, $F = (x_4, u_5)$, $E = (x_5, u_6)$, $P = (x_6, 0)$, $Q = (x_8, x_7)$, and $S = (x_{10}, x_9)$. Then the problem can be algebraically specified as follows:

$h_1 = 2u_2x_2 + 2u_3x_1 - u_3^2 - u_2^2 = 0$ — $OA \equiv OC$.

$h_2 = 2u_1x_2 - u_1^2 = 0$ — $OA \equiv OB$.

$h_3 = x_3^2 - 2x_2x_3 - 2u_4x_1 + u_4^2 = 0$ — $OA \equiv OD$.

$h_4 = x_4^2 - 2x_2x_4 - 2u_5x_1 + u_5^2 = 0$ — $OA \equiv OF$.

$h_5 = x_5^2 - 2x_2x_5 - 2u_6x_1 + u_6^2 = 0$ — $OA \equiv OE$.

$h_6 = (u_5 - u_4)x_6 + u_4x_4 - u_5x_3 = 0$ — P, D and F are collinear.

$h_7 = (u_6 - u_5)x_8 - (x_5 - x_4)x_7 + u_5x_5 - u_6x_4 = 0$ — Q, F and E are collinear.

$h_8 = u_3x_8 - (u_2 - u_1)x_7 - u_1u_3 = 0$ — Q, B and C are collinear.

$h_9 = u_6x_{10} - x_5x_9 = 0$ — S, E and A are collinear.

$h_{10} = (u_4 - u_3)x_{10} - (x_3 - u_2)x_9 + u_3x_3 - u_2u_4 = 0$ — S, C and D are collinear.

$s_1 = (u_4 - u_3)x_5 - u_6x_3 + u_2u_6 \neq 0$ — Lines AE and CD have a normal intersection.

$s_2 = u_3x_5 - u_3x_4 - (u_2 - u_1)u_6 + (u_2 - u_1)u_5 \neq 0$ — Lines BC and EF have a normal intersection.

$s_3 = u_1u_5 - u_1u_4 \neq 0$ — Lines AB and DF have a normal intersection.

$g = x_7x_{10} - (x_8 - x_6)x_9 - x_6x_7 = 0$ **Conclusion:** S, Q and P are collinear.

$Zero(S/G)$ has only one component (in 6.9s [6]) whose corresponding ascending chain ASC_1 is just the one obtained from the polynomial set $S = \{h_1, ..., h_{10}\}$ using Ritt-Wu's principle (3.1). Since $prem(g; ASC_1) = 0$ (in 0.4s), theorem has been confirmed. This problem gives an impression that finding non-degenerate conditions is not hard. Let us look at another example.

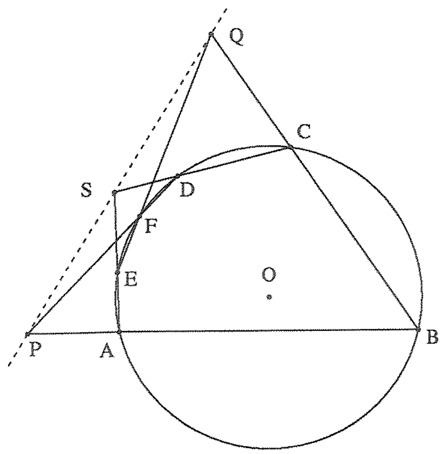

Figure 1

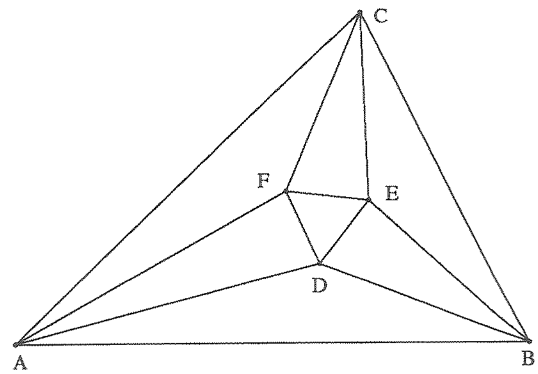

Figure 2

Example (6.2) (Morley's Trisector Theorem.) The points of intersection D, E and F of the adjacent trisectors of the angles of any triangle ABC are the vertices of an equilateral triangle (Figure 2).

We can let $B = (y_1, 0)$, $A = (0, 0)$, $D = (y_2, y_3)$, $C = (y_5, y_4)$, $F = (y_8, y_7)$, and $E = (y_{10}, y_9)$. Then the problem can be algebraically specified as follows:[7]

$h_1 = (y_3^3 + (-3y_2^2 + 6y_1y_2 - 3y_1^2)y_3)y_5 + ((-3y_2 + 3y_1)y_3^2 + y_2^3 - 3y_1y_2^2 + 3y_1^2y_2 - y_1^3)y_4 - y_1y_3^3 + (3y_1y_2^2 - 6y_1^2y_2 + 3y_1^3)y_3 = 0$ $\tan(\angle CBA) - \tan(3\angle DBA) = 0.$

$h_2 = (y_3^3 - 3y_2^2y_3)y_5 + (-3y_2y_3^2 + y_2^3)y_4 = 0$ $\tan(\angle CAB) - \tan(3\angle DAB) = 0.$

$h_3 = y_6^2 - 3 = 0$ $\tan(\pm\pi/3) = \pm\sqrt{3}.$

$h_4 = (((y_3^2 + y_2^2 - y_1y_2)y_5 - y_1y_3y_4)y_6 + y_1y_3y_5 + (y_3^2 + y_2^2 - y_1y_2)y_4)y_8 + ((y_1y_3y_5 + (y_3^2 + y_2^2 - y_1y_2)y_4)y_6 - (y_3^2 + y_2^2 - y_1y_2)y_5 + y_1y_3y_4)y_7 - ((y_3^2 + y_2^2 - y_1y_2)y_5^2 + (y_3^2 + y_2^2 - y_1y_2)y_4^2)y_6 - y_1y_3y_5^2 - y_1y_3y_4^2 = 0$ $\tan(\angle BAD + \angle DBA + \angle ACF) = \pm\sqrt{3}.$

$h_5 = (y_1y_3y_5 - y_1y_2y_4)y_8 + (y_1y_2y_5 + y_1y_3y_4)y_7 = 0$ $\tan(DAB) = \tan(CAF).$

$h_6 = (y_1y_3y_5 + (-y_1y_2 + y_1^2)y_4 - y_1^2y_3)y_{10} + ((y_1y_2 - y_1^2)y_5 + y_1y_3y_4 - y_1^2y_2 + y_1^3)y_9 - y_1^2y_3y_5 + (y_1^2y_2 - y_1^3)y_4 + y_1^3y_3 = 0$ $\tan(ABD) = \tan(EBC).$

[6] Meaning that it took 6.9 seconds to complete the decomposition on a Symbolics 3600, running on Release 7.2.

[7] The specification was due to Wu [12]. There are 18 triangles DEF thus formed. The specification of non-degenerate conditions here was due to us. Our proof applies to all 18 equilateral triangles. The authors first suspected that there might be other non-degenerate conditions. For example, can we insure that all those 18 triangles are well formed under any circumstances (besides those specified)? Is it possible that some trisectors do not intersect under some peculiar conditions?

$h_7 = ((2y_4y_5 - y_1y_4)y_8 + (-y_5^2 + y_1y_5 + y_4^2)y_7 - y_4y_5^2 - y_4^3)y_{10} + ((-y_5^2 + y_1y_5 + y_4^2)y_8 + (-2y_4y_5 + y_1y_4)y_7 + y_5^3 - y_1y_5^2 + y_4^2y_5 - y_1y_4^2)y_9 + (-y_4y_5^2 - y_4^3)y_8 + (y_5^3 - y_1y_5^2 + y_4^2y_5 - y_1y_4^2)y_7 + y_1y_4y_5^2 + y_1y_4^3 = 0$ $\tan(ACF) = \tan(ECB)$.

$s_1 = y_1y_4 \neq 0$ — A, B, C are not collinear.

$s_2 = y_1^2 \neq 0$ — Line AB is non-isotropic.[8]

$s_3 = (y_5 - y_1)^2 + y_4^2 \neq 0$ — Line BC is non-isotropic.

$s_4 = y_5^2 + y_4^2 \neq 0$ — Line AC is non-isotropic.

$g = (y_6y_8 - y_7 - y_2y_6 + y_3)y_{10} + (y_8 + y_6y_7 - y_3y_6 - y_2)y_9 + (-y_2y_6 - y_3)y_8 + (-y_3y_6 + y_2)y_7 + (y_3^2 + y_2^2)y_6 = 0$ **Conclusion:** $\tan(\angle EDF) = \pm\sqrt{3}$.

$Zero(S/G)$ has only one component (in 756.7s) whose corresponding ascending chain ASC_1 is just the one obtained from the polynomial set S using Ritt-Wu's principle. Since $prem(g; ASC_1) = 0$ (in 6.4s), the theorem has been confirmed.

Remark. The authors spent several hours to figure out degenerate cases $s_2 = 0, s_3 = 0$ and $s_4 = 0$. At first, we overlooked these cases and the proof failed;[8] then could we say Morley's trisector theorem is not a "theorem" (in complex geometry)? This is why the authors are in favor of the notion "generally true" introduced by Wu. In Euclidean geometry, $s_i \neq 0$ $(i = 2, 3, 4)$ are consequences of $s_1 \neq 0$; they are redundant. Thus, one could argue that the trouble with $s_2, s_2, and s_3$ is due to the method: if we use the Tarski-Seidenberg-Collins method, then $s_1 \neq 0$ is enough. But Tarski-Seidenberg-Collins method cannot be applied to unordered geometry: in unordered geometry these non-degenerate conditions are all necessary. Let us look at another example.

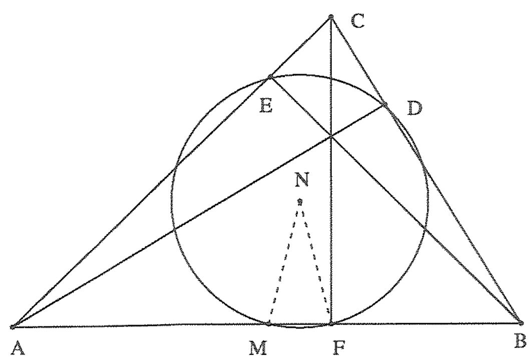

Figure 3

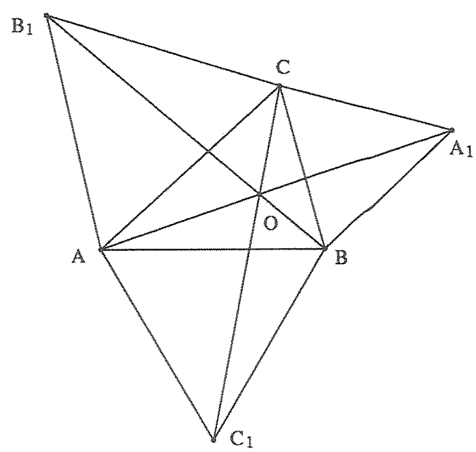

Figure 4

Example (6.3). (One Form of the Nine Point Circle Theorem). The circle (N) passing though the feet D, E and F of the three altitudes of a triangle ABC also passes through the midpoints of its sides (Figure 3).

We can let $B = (y_1, 0)$, $A = (0, 0)$, $C = (y_2, y_3)$, $D = (y_5, y_4)$, $E = (y_7, y_6)$, $F = (y_2, 0)$, $M = (y_8, 0)$, and $N = (y_{10}, y_9)$. Then the problem can be algebraically specified as follows:

[8] An isotropic line is a line perpendicular to itself. It does not exist in Euclidean geometry, but exists in general metric geometries.

[8] $Zero(S/\{s_1\})$ has 7 components (in 846.2s) and g vanishes *only* on the first component (in 9.2s).

$h_1 = y_3y_5 + (-y_2 + y_1)y_4 - y_1y_3 = 0$ — D, B and C are collinear.
$h_2 = (y_2 - y_1)y_5 + y_3y_4 = 0$ — $AD \perp CB$.
$h_3 = y_3y_7 - y_2y_6 = 0$ — E, A and C are collinear.
$h_4 = y_2y_7 + y_3y_6 - y_1y_2 = 0$ — $EB \perp CA$.
$h_5 = 2y_8 - y_1 = 0$ — M is the midpoint of A and B.
$h_6 = (2y_7 - 2y_2)y_{10} + 2y_6y_9 - y_7^2 - y_6^2 + y_2^2 = 0$ — $NF \equiv NE$.
$h_7 = (2y_5 - 2y_2)y_{10} + 2y_4y_9 - y_5^2 - y_4^2 + y_2^2 = 0$ — $NF \equiv ND$.
$s_1 = y_1y_3 \neq 0$ — A, B, C are not collinear.
$s_2 = y_3^2 + y_2^2 - y_1y_2 \neq 0$ — Line AC is not perpendicular to line CB.
$s_3 = y_1y_2 \neq 0$ — Line BA is not perpendicular to line AC.
$s_4 = y_1y_2 - y_1^2 \neq 0$ — Line AB is not perpendicular to line BC.
$g = (2y_8 - 2y_2)y_{10} - y_8^2 + y_2^2 = 0$ — **Conclusion**: $NF \equiv NM$.

Again, $Zero(S/G)$ has only one component (in 11.9s) whose corresponding ascending chain ASC_1 is just the one obtained from the polynomial set S using Ritt-Wu's principle. Since $prem(g; ASC_1) = 0$ (in 0.3s), the theorem has been confirmed.

Here $s_1 \neq 0 \wedge s_2 \neq 0 \wedge s_3 \neq 0$ means that triangle ABC is not a right triangle. If we drop them, then $Zero(G/\{s_4\})$ has 4 components (in 37.3s) and $g = 0$ is valid *only* on the first component (in 0.4s). Thus they are all necessary even in Euclidean geometry. For this theorem, one can still argue that only $s_1 \neq 0$ is necessary if we don't introduce point N and change the conclusion to be "D, E, F and M are on the same circle." Now we give another "trouble" example.

Example (6.4). Let ABC be a triangle. Three equilateral triangles A_1BC, AB_1C and ABC_1 are erected (either all outside or all "inside" the triangle) on the three respective sides BC, CA and AB. Show that (i) lines AA_1, BB_1 and CC_1 are concurrent at a point, say, O; (ii) $\angle B_1OC_1 \equiv \angle C_1OA_1$ (Figure 4).

We can let $B = (y_1, 0)$, $A = (0, 0)$, $C = (y_2, y_3)$, $C_1 = (y_5, y_4)$, $B_1 = (y_7, y_6)$, $A_1 = (y_9, y_8)$, and $O = (y_{11}, y_{10})$. Then the problem can be algebraically specified as follows:

$h_1 = 2y_1y_5 - y_1^2 = 0$ — $C_1A \equiv C_1B$.
$h_2 = y_5^2 + y_4^2 - y_1^2 = 0$ — $AC_1 \equiv AB$.
$h_3 = (y_1y_3y_5 - y_1y_2y_4)y_7 - (y_1y_2y_5 + y_1y_3y_4)y_6 = 0$ — $\tan(C_1AB) = \tan(CAB_1)$.
$h_4 = 2y_2y_7 + 2y_3y_6 - y_3^2 - y_2^2 = 0$ — $B_1C \equiv B_1A$.
$h_5 = (y_1y_3y_5 + (y_1y_2 - y_1^2)y_4)y_9 + ((-y_1y_2 + y_1^2)y_5 + y_1y_3y_4)y_8 - y_1^2y_3y_5 + (-y_1^2y_2 + y_1^3)y_4 = 0$
$\tan(C_1AB) = \tan(A_1BC)$.
$h_6 = (2y_2 - 2y_1)y_9 + 2y_3y_8 - y_3^2 - y_2^2 + y_1^2 = 0$ — $A_1B \equiv A_1C$.
$h_7 = y_8y_{11} - y_9y_{10} = 0$ — O, A_1 and A are collinear.
$h_8 = y_6y_{11} + (-y_7 + y_1)y_{10} - y_1y_6 = 0$ — O, B_1 and B are collinear.
$s_1 = y_1y_3 \neq 0$ — A, B, C are not collinear.
$s_2 = y_1^2 \neq 0$ — Line AB is non-isotropic.
$s_3 = y_3^2 + y_2^2 - 2y_1y_2 + y_1^2 \neq 0$ — Line BC is non-isotropic.
$s_4 = y_3^2 + y_2^2 \neq 0$ — Line AC is non-isotropic.
$g_1 = (y_4 - y_3)y_{11} - (y_5 - y_2)y_{10} + y_3y_5 - y_2y_4 = 0$
Conclusion 1: C_1, C and O are collinear.
$g_2 = a_3y_{11}^3 + a_2y_{11}^2 + a_1y_{11} + a_0 = 0$ [9]

[9] We omit the explicit forms of the huge polynomials a_3, a_2, a_1, and a_0.

Conclusion 2: $\tan(B_1OC_1) = \tan(C_1OA_1)$.

$Zero(S/G)$ has only two components (in 101.9s) whose corresponding ascending chains are:

$$\begin{array}{l} 4y_4^2 - 3y_1^2 \\ 2y_5 - y_1 \\ 2y_1y_6 + 2y_2y_4 - y_1y_3 \\ 2y_2y_7 + 2y_3y_6 - y_3^2 - y_2^2 \\ 2y_1y_8 + (-2y_2 + 2y_1)y_4 - y_1y_3 \\ (2y_2 - 2y_1)y_9 + 2y_3y_8 - y_3^2 - y_2^2 + y_1^2 \\ (y_6y_9 + (-y_7 + y_1)y_8)y_{10} - y_1y_6y_8 \\ y_6y_{11} + (-y_7 + y_1)y_{10} - y_1y_6, \end{array} \qquad ASC_1$$

and

$$\begin{array}{l} 2y_2 - y_1 \\ 4y_3^2 - 3y_1^2 \\ y_4 - y_3 \\ 2y_5 - y_1 \\ y_6 \\ y_7 - y_1 \\ y_8 \\ y_9. \end{array} \qquad ASC_2$$

Since $prem(g_1; ASC_1) = 0$ and $prem(g_1; ASC_2) = 0$ (in 0.2s), conclusion 1 has been confirmed. Thus we would think that the non-degenerate conditions $s_1 \neq 0, \ldots, s_4 \neq 0$ are enough. However, $prem(g_2; ASC_1) = 0$ (in 0.1s), but $prem(g_2; ASC_2) \neq 0$ (in 21.2s). Is conclusion 2 not valid? Obviously, ASC_2 corresponds to the case when triangle ABC is equilateral, $A = A_1$, $B = B_1$, and $C = C_1$. This is certainly a degenerate case. In that case, the conclusions "AA_1, BB_1, and CC_1 are concurrent" and "$\angle B_1OC_1 \equiv \angle C_1OA_1$" become meaningless. (Question: why $g_1 = 0$ is still true in this meaningless case !?)

We can give more examples with hidden degenerate cases hard to find even for geometry experts. How can one exclude such kinds of hidden degenerate cases without enormous human efforts? The answer was already in Wu's work: to introduce the notion of "generally true". For details, see [5]. For example, y_1, y_2 and y_3 can obviously be chosen as parameters. Thus, $Zero(PD(ASC_2))$ corresponds to a degenerate case because the parameters are algebraically dependent on it. We only need to check whether g_1 and g_2 vanish on non-degenerate case $Zero(PD(ASC_1))$, but *not* on $Zero(PD(ASC_2))$.

7. Experimental Results

We have implemented Method (5.3) in our prover [1]. More than 500 theorems have been proved in this way. In particular, we have experimented with the same set of 512 theorems in [2] (using the same coordinates and equations). The prover described in [1] is based on approach (1.1) and can generate non-degenerate conditions in geometric forms for a large class of geometry statements. For 413 of the 512 theorems, the prover can generate non-degenerate conditions *all* in geometric forms. For most of those 413 theorems, we use such machine generated non-degenerate conditions in geometric forms as the inputs to our new method. We have paid particular attentions to a few problems among these 413 theorems, specifying non-degenerate conditions manually. For example, we proved Feuerbach's theorem under the only non-degenerate that "the vertices of the triangle are not collinear."

For the rest 91 theorems, some non-degenerate conditions in polynomial inequations were generated by our previous method. First we simply deleted these algebraic inequations, using the rest machine generated non-degenerate conditions in geometric forms as inputs. About half of these 91 theorems were confirmed this way. We have to pay more attentions to the rest half, adding more non-degenerate conditions in geometric forms manually.

In this way, we have proved 493 of the 512 theorems.[9] Among the 493 theorems proved, 471 were proved within 300 seconds; 12 within one hours. We list the following typical timing samples (besides the four examples in Section 6.)

Theorem	Sources	Decomp	Redu	Total Time
Parallelogram	Section 2 [4]	0.23	0.02	0.25
Theorem of Centroid	Ex1 in [4]	0.33	0.02	0.35
Simson's Theorem	Ex2 in [4]	0.6	0.1	0.7
Brahmagupta's Theorem	Section 4 [4]	3.8	0.1	3.9
Butterfly Theorem	Ex5 in [4]	56.5	0.3	56.8
Pappus' Theorem	Ex6 in [4]	2.5	0.2	2.7
Pappus Point Theorem	Ex7 in [4]	8.0	1.9	9.9
Isosceles Midpoint	Ex8 in [4]	3.7	0.1	3.8
Gauss' Theorem	Ex9 in [4]	0.15	0.05	0.2
Gauss Point Theorem	Ex10 in [4]	5.7	0.6	6.3
Gauss Conic Theorem	Ex17 in [2]	101.3	1502.1	1603.4
Feuerbach's Theorem	Ex204 in [2]	27.3	1.0	28.3

The timing is specified in seconds. Here "Decomp" means the time spent on decomposition; "Redu" means the time spent on checking whether $prem(g; ASC_i) = 0$. The Pappus point theorem was not proved by the program in [7]. It was easily proved by our program, under nondegenerate conditions that each of the following pairs of lines have a normal intersection: A_1B and CC_1, AC_1 and BB_1, A_1C and BB_1, AA_1 and BC_1, AC_1 and A_1C, AB_1 and A_1B, and EF and GH.

Here the theorem of parallelogram, the theorem of centroid, and Feuerbach's theorem were proved under the only non-degenerate condition that "points A, B and C are not collinear."

[9] We had trouble with the rest 19 theorems within the time or space limit of the computer.

The authors still have difficulty understanding why the non-degenerate condition that "the three sides of the triangle ABC are non-isotropic" is necessary for Morley's theorem, Simson's theorem, but *not* for Feuerbach's theorem.

8. Related Work

Now let us first sum up Wu's method as we understand. Let S, G and g be the same as in Section 5. If $Zero(S) \subset Zero(g)$, then $g = 0$ follows from $h_1 = 0 \wedge \cdots \wedge h_n = 0$. However, this is usually not the case because some nondegenerate conditions are missing. We can decompose $Zero(S)$ according to theorem (4.2):

$$Zero(S) = \bigcup_{1 \le i \le k} Zero(PD(ASC_i)).$$

Some components $Zero(PD(ASC_i))$ correspond to degenerate cases, and others to nondegenerate cases. Of course, $g = 0$ does not have to be valid on degenerate components. How to identify these non-degenerate components?

Approach (1.1) is to use parameters and can identify non-degenerate components easily. Then one only needs to check whether $g = 0$ is valid on those non-degenerate components. If it is, then one says the geometry statement is generally true.

Approach (1.2) is to specify a set of degenerate conditions (its algebraic form is the polynomial set G) manually. In the context of Wu's method, $Zero(S/G)$ is expected to contain only non-degenerate components, hopefully, no more no less. As we have seen in Section 6, this task is sometimes very difficult.

Though people knew at the very beginning that Wu's method could prove theorems with approach (1.2), no attempt was ever made at that time, because people (including the first author) thought it tended to be much slower than the method based on approach (1.1). Wu, Chou, Gao and others proved hundreds of theorems based on approach (1.1).

The first author experimented with approach (1.2) during 1984–1985 using the Gröbner basic method. He was able to prove about 10 theorems and found it very slow [4]. The hardest one proved by him was perhaps Simson's theorem.

D. Kapur, on the other hand, was successful in proving more and harder theorems using the same approach. By careful arrangements of the order of production and reduction of S-polynomials, he proved about 25 theorems [7]. The hardest perhaps were Pascal's theorem and the Butterfly theorem. However, he was unable to prove harder theorems such as Pappus point thoerem in [4], which was easily proved by our Method (5.3) in about 10 seconds. With our method/program, we can prove much harder theorems such as Morley's trisector theorem, etc. It does not need factorization with the Gröbner basis method. According to Kapur, this is an advantage of using that method. Our program cannot do factorization over algebraic extensions other than quadratic successive extensions. Thus it is incomplete. Incompleteness does not bother us very much so long as the program can prove much more and much harder theorems than other programs. We could implement a factorization (over extension fields) algorithm (but not necessarily efficient) in our program and still use current program mainly, thus making the program complete.

H. P. Ko [8] was the first to use Ritt-Wu's method to prove geometry theorems according to approach (1.2). She was able to prove at least 25 theorems. The hardest were perhaps also Pascal's theorem and the Butterfly theorem. Our work is in the same direction as Ko's work and has similarities and differences with her work [8]. Our method/algorithms/program are faster than hers. Especially, we establish a deep theorem (Theorem (4.4)) which makes the proof procedure faster and *much clearer*. For example, in Example (6.1) (Pascal's theorem), Ko produced four components with ascending chains T_1, T_2, T_3, and T_4.[10] According to Theorem (4.4), $Zero(T_i/I_i \cup G) \subset Zero(PD(T_1)/G)$ (for $i = 2,3,4$; here the I_i are the initial sets of ascending chains T_i) and $Zero(S/G) = Zero(PD(T_1)/G)$.[11] So we only need to check whether $prem(g;T_1) = 0$, or in the terminology of approach (1.1), to check whether g vanishes on non-degenerate components. Another example: in Appendix C of [8], Ko listed 6 ascending chains $T_1, ..., T_6$ for $Zero(S/G)$ of Simson's theorem. According to our Theorem (4.4), only T_1, T_2, and T_3 are necessary simply because $length(T_i) > 7$ for $i = 4, 5, 6$.

Appendix

9. Properties of Irreducible Ascending Chains

Let $ASC = f_1, \ldots, f_r$ be an ascending chain, not consisting of a constant. After a *suitable renaming* [12] 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:

$$f_1(u_1, \ldots, u_d, x_1)$$

$$f_2(u_1, \ldots, u_d, x_1, x_2)$$

$$\ldots \qquad (9.0)$$

$$f_r(u_1, \ldots, u_d, x_1, \ldots, x_r).$$

(9.1)–(9.4) can be found in Ritt-Wu's original work; (9.5)–(9.7) are new.

Definition (9.1). An ascending chain $f_1, ..., f_r$ of the form (9.0) 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.[13]

[10] See pp110–112 of [8]. We use exactly the same coordinates and equations and inequations as Ko's. Her T_1 is equivalent to our ASC_1 in Example (6.1). If we didn't use Theorem (4.4), our program produced two ascending chains.

[11] $Zero(T_1/I_1 \cup G)$ is a proper subset of $Zero(S/G)$. This is the advantage to introduce the notation $PD(T_1)$ and the zero set $Zero(PD(T_1)/G)$.

[12] 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.

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

Theorem (9.2). Let ascending chain ASC of the form (9.0) 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) $Zero(PD(ASC)) \subset Zero(g)$, here the zeros are taken from an algebraically closed extension of the field K.

Proof. See lemma 3, page 234 in [12] and Theorem (3.7) on page 31 of [2]. ∎

Theorem (9.3). 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 non-zero polynomial $h \in K[u]$ such that $pg - h \in Ideal(ASC)$.

Proof. By (9.2), polynomial g, considered as an element in the filed F_r, is non-zero. Thus there exists a polynomial p' in F_r such that $p'g = 1$. Considered as polynomials in $K[u, x]$ and clearing the denominators we have $pg - h \in Ideal(ASC)$ for some polynomial p and some non-zero polynomial polynomial $h \in K[u]$. ∎

Theorem (9.4). 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 [2]. Furthermore, by (9.3), if we wish, we can chose g and h in such a way that the initials of g and h contain parameters only. This fact will be needed in (9.7). ∎

The following theorems can help us remove certain redundancy in decomposition (4.4.2).

Theorem (9.5). Suppose irreducible ascending chains ASC_1 and ASC_2 have the same parameter set. If $prem(g; ASC_1) = 0$ for all g in ASC_2, then $PD(ASC_1) = PD(ASC_2)$.

Proof. Since ASC_1 and ASC_2 have the same parameter set, they have the same length. Let it be r; let $ASC_1 = f_1, ..., f_r$ and $ASC_2 = g_1, ..., g_r$. Then $lv(f_i) = lv(g_i)$ $(i = 1, ..., r)$. We use the abbreviation $PD_i = PD(ASC_i)$, $ID_i = Ideal(ASC_i)$ for $i = 1, 2$ and use J_1 and J_2 to denote any power products (changed in different contexts) of the initials ASC_1 and ASC_2, respectively.

First, suppose $g \in PD_2$. We have $J_2 g \in ID_2$ for some J_2. Thus by the hypothesis of the theorem, there is a J_1 such that $J_1 J_2 g \in ID_1 \subset PD_1$. Since $J_1 \notin PD_1$ and PD_1 is a prime ideal, we have

(9.5.1) $$J_2 g \in PD_1.$$

Since $J_2 \notin PD_2$, by (9.3) there are a polynomial p and a non-zero polynomial $h \in K[u]$ such that $pJ_2 - h \in ID_2$. Thus, there is another J_1' such that $J_1'(pJ_2 - h) \in ID_1$. Thus $pJ_2 - h \in PD_1$. Since $h \notin PD_1$, pJ_1 is also not in PD_1, thus $J_2 \notin PD_1$. Since PD_1 is a prime ideal, by (9.5.1) we have $g \in PD_1$. Thus $PD_2 \subset PD_1$.

Suppose $g \notin PD_2$. We can use the same argument after (9.5.1) to infer that $g \notin PD_1$. Thus $PD_1 \subset PD_2$. ∎

$f_1, ..., f_r$.

Remark. Actually, it is not hard to prove that $deg(f_i, x_i) = deg(g_i, x_i)$. Furthermore, the converse of (9.5) is also true.

Theorem (9.6). Let $ASC_1 = f_1, ..., f_r$ and $ASC_2 = g_1, ..., g_s$ be two ascending chains and ASC_1 be irreducible. Suppose that $prem(g_i; ASC_1) = 0$ for all g_i $(i = 1, ..., s)$ and $prem(I_i; ASC_1) \neq 0$ for all initials I_i of g_i $(i = 1, ..., s)$. Then $PD(ASC_2) \subset PD(ASC_1)$.

Proof. We use the same notation as in the previous proof. Suppose $g \in PD_2$. We can use the same argument in the second paragraph of the previous proof to get (9.5.1): i.e., $J_2 g \in PD_1$ for some J_2. From the hypothesis, $J_2 \notin PD_1$, thus $g \in PD_1$ and $PD_2 \subset PD_1$. ∎

(9.7) *Proof of Lemma* (4.5).

Proof. We use induction on $m - r$.

(1) Base case: $m - r = 0$. In that case, the parameter set of ASC is empty.

Case (1.1) ASC is not in weak sense, i.e., $prem(I_j; f_1, ..., f_{j-1}) = 0$ for some $j > 1$, then $Zero(ASC/J)$ is empty.

Case (1.2) ASC is irreducible. Then $Zero(ASC/J)$ is contained in the (irreducible) variety $Zero(PD(ASC))$, the dimension of which is $m - r = 0$. The theorem is true.

Case (1.3) ASC is reducible. Suppose $f_1, .., f_{k-1}$ is irreducible, and $f_1, ..., f_k$ is reducible $(1 \leq k)$. For simplicity and without loss of generality, we can assume f_k has only two irreducible factors, i.e., there are two polynomials f_k' and f_k'' with the same class as $class(f_k)$ such that $f_1, ..., f_k'$ and $f_1, ..., f_k''$ are irreducible, $f_k' f_k'' \in Ideal(f_1, ..f_k)$, $prem(f_k; f_1, ..., f_k') = 0$ and $prem(f_k; f_1, ..., f_k'') = 0$. Furthermore, we can chose f_k' and f_k'' in such a way that the initials $I_k' = lc(f_k')$ and $I_k'' = lc(f_k'')$ *contain parameters only.* Thus

$$(4.5.1) \qquad \begin{aligned} Zero(ASC/J) = {} & Zero(ASC'/J \cup \{I_k'\}) \bigcup Zero(ASC''/J \cup \{I_k''\}) \\ & \bigcup Zero(ASC \cup \{I_k'\}/J) \bigcup Zero(ASC \cup \{I_k''\}/J), \end{aligned}$$

where

$$ASC' = f_1, ..., f_{k-1}, f_k', f_{k+1}, ..., f_r,$$
$$ASC'' = f_1, ..., f_{k-1}, f_k'', f_{k+1}, ..., f_r.$$

In this base case, since parameter set is empty, I_k' and I_k'' are constants. Thus (4.5.1) actually is

$$(4.5.2) \qquad Zero(ASC/J) = Zero(ASC'/J \cup \{I_k'\}) \bigcup Zero(ASC''/J \cup \{I_k''\}).$$

For quasi ascending ASC' (or ASC'') we have three cases:

Case (1.3.1) ASC' is not in the weak sense, i.e., $prem(I_j; ASC') = 0$ for some $j > k$, then $Zero(ASC'/J \cup \{I_k'\})$ is empty. We can delete it from the union (4.5.2).

Case (1.3.2) ASC' is irreducible. Then

$$(4.5.3) \qquad prem(f_j; ASC') = 0 \text{ for all } i = 1, ..., r.$$

Thus $PD(ASC) \subset PD(ASC')$ by Lemma (2.3) below. Hence

$$Zero(ASC'/J \cup \{I'_k\}) \subset Zero(PD(ASC')) \subset Zero(PD(ASC)).$$

$Zero(PD(ASC'))$ is a variety of dimension $m - r$.

Case (1.3.3) ASC' is reducible. We recursively repeat the same procedure of $Zero(ASC/J)$ as for $Zero(ASC'/J')$, until either case (1.3.1) or case (1.3.2) happen, here $J' = \{I_1, ..., I'_k, ..., I_r\}$. When case (1.3.2) happens, (4.5.3) is still valid.

Thus we conclude that $Zero(ASC/J)$ is contained in the union of those components of the algebraic set $Zero(PD(ASC))$ whose dimension is $m - r = 0$.

(2) Induction case: suppose the theorem is true for quasi ascending chains $g_1, ..., g_d$ with $m - d < m - r$. We want to show it is also true for $f_1, ..., f_r$. We can use the same argument as in the base case.

Case (2.1) ASC is not in weak sense, then $Zero(ASC/J)$ is empty.

Case (2.2) ASC is irreducible. Then as before, the theorem is true.

Case (2.3) ASC is reducible. We can repeat the same argument as in case (1.3) and also have 3 cases for each of ascending chains ASC' and ASC''. Here we emphasize that I'_k and I''_k contain only the parameters of ASC. Decomposition (4.5.1) is valid, but (4.5.2) is no longer valid. Instead, we can decompose (using Algorithm (4.1) or (4.3)), say, $Zero(\{I'_k\})$, into

$$Zero(\{I'_k\}) = \bigcup_i Zero(ASC'_i/I'_{k,i}).$$

Here for each i, $I'_{k,i}$ is the initial set of the ascending chain ASC'_i. Then

$$Zero(ASC \cup \{I'_k\}/J) = \bigcup_i Zero(ASC'_i \cup ASC/I'_{k,i} \cup J).$$

Note that $ASC'_i \cup ASC$ forms another quasi ascending chain since ASC'_i involves only the parameters of ASC. For each $Zero(ASC'_i \cup ASC/I'_{k,i} \cup J)$, we now can use the induction hypothesis to conclude that it is contained in the union of varieties (with dimension $\leq m-r+1$) $\subset Zero(PD(ASC'_i \cup ASC)) \subset Zero(PD(ASC))$. Thus the proof is completed. ∎

References

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

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

[3] 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.*

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

[5] S.C. Chou and G.J. Yang, "On the Algebraic Formulation of Certain Geometry Statements and Mechanical Geometry Theorem Proving", Preprint, May 1986, revised in July 1987, to appear in *Algorithmica.*

[6] R. Hartshorne, *Algebraic Goemetry,* Springer-Verlag, 1978.

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

[8] H.P. Ko, "Geometry Theorem Proving by Decomposition of Quasi-Algebraic Sets: An Application of the Ritt-Wu Principle", *Artificial Intelligence,* Vol. 37, pp95-122 (1988).

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

[10] R. F. Ritt, *Differential Algebra,* AMS Colloquium Publications, New York, 1950.

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

[12] 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.

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

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