系 统 科 学 与 数 学
J. Sys. Sci. & Math. Scis.
4 (3) (1984), 207—235

BASIC PRINCIPLES OF MECHANICAL THEOREM PROVING IN ELEMENTARY GEOMETRIES

WU WENJUN (WU WEN-TSÜN)

(*Institute of Systems Science, Academia Sinica, Beijing*)

Dedicated to Professor Kwan Chao-chih, the late Director of the Institute of Systems Science

§ 1. INTRODUCTION.

By *elementary geometry* we shall mean the one described in Hilbert's *Grundlagen der Geometrie* in which no notion of differentiation is involved, as a contrast to *differential geometry*. It is well known by the theorem of Tarski that the ordinary Euclidean geometry, as one of such elementary geometries, is *decidable*, or in our terminology, *mechanizable* in the following sense: There exists an algorithmic method by which any "theorem" or a geometrical statement meaningful in the geometry in question can be shown, in a *finite* number of steps, to be either true as a real *theorem*, or false so that it is not a *theorem* at all. Any elementary geometry possessing such an algorithmic method will be said to be *mechanizable*, and the theorem in asserting that the geometry in question does possess such an algorithmic method will be called a *Mechanization Theorem*. In the mechanizable case we may program according to the algorithm shown to exist and practise on a computer so that the proof (or disproof) of a theorem in that geometry may be carried out on the computer. This method will be called *mechanical theorem proving* for short. It will lead to what may be called *mechanical theorem discovering* of new theorems. We remark that all these notions can be naturally extended to the case of a given class of theorems or meaningful statements in the geometry in question, not necessarily to the geometry as a whole. In this sense the Theorem of Tarski mentioned may be called the Mechanization Theorem of ordinary Euclidean geometry. However, the algorithmic procedure given by Tarski, even with the great simplifications due to Seidenberg, is too complicated to be feasible. In fact, no theorems of any geometrical interest seem to have been proved in this way up to the present day. On the other hand, the author discovered in 1977 an algorithmic method which leads to Mechanization Theorems of many kinds of elementary geometries including the ordinary Euclidean geometry, as long as we restrict ourselves to the class of theorems involving no order relations. What is important to us is that our method is very efficient. In fact, in the past years we have programed on some small computers and arrived at the proof and discovery of quite nontrivial theorems. Mr. S. C. Chou, now at University of Texas at Austin, USA, has also practised on some computer there, on the basis of our algorithm, and proved some interesting new theorems. The present paper is aimed at explaining the basic principles underlying our method with some illustrative examples about the theorems proved or discovered in this way.

Consider a certain kind of geometry in the sense of Hilbert. As shown in the classical

Received March 12, 1984.

Grundlagen of Hilbert, in starting from the defining axiomatic system of the geometry we may introduce some number system intrinsically associated to that geometry and then to coordinate systems which will turn any geometrical entities and relations into algebraic ones. Let us restrict ourselves to the case that the geometry admits some axiom of infinity as well as some Pascalian axiom so that the number system is a commutative field of characteristic 0. The algebraic relations corresponding to the geometrical relations occuring in a theorem will then be polynomial equations, polynomial inequations, or polynomial inequalities, with coefficients in the associated field, or even with rational or integer coefficients, as is usually the case. Now let us restrict ourselves further to the case that no order relations and axioms are involved in the geometry in question or to a restricted class of theorems in which no order relations are involved. In the algebraic relations above there will appear only polynomial equations and inequations but not any polynomial inequalities. Remark further that all theorems in geometries are actually only *generically* true, or true only under some *non-degeneracy conditions* which are usually not easy to be made explicit and thus only implicitly assumed in the statement of theorems. It turns out that the problem of mechanical theorem proving in the restricted cases mentioned above is algebraically equivalent to the following one:

Problem. Given a system Σ of polynomial equations (or equivalently, system of polynomials) as well as another polynomial g, all in the same finite set of variables $x, y, \cdots,$ decide in a finite number of steps either of the two cases below:

Case 1. A finite set of polynomials D_a is determined such that $g=0$ is a consequence of the system Σ under the non-degeneracy conditions $D_a \neq 0$ such that $D_a=0$ are themselves not consequences of the system Σ.

Case 2. No such set $S=\{D_a\}$ can exist so that Case 1 holds.

In the above formulated problem in the algebraic form the polynomials in Σ correspond to the hypotheses and g the conclusion of the theorem in question whose truth is to be decided. The theorem is seen to be generically true in Case 1 under the non-degeneracy conditions $D_a \neq 0$ found during the procedure but not so in Case 2. The polynomials naturally have their coefficients in the field intrinsically associated to the geometry considered. A solution of the above problem constitutes the Mechanization Theorem of geometries in the algebraic form. The algorithm in furnishing such a solution as well as the proof will be given in Section 4. In Sections 2 and 3 we shall make some preparations. All these depend heavily on the works of J. F. Ritt as exhibited in his two books [2, 3], which seem to be however underservedly little known in the present days.

§ 2. Well-ordering of a Polynomial Set.

In what follows K will be a fixed basic field of characteristic 0. Consider two sets of variables

$$u_1, \cdots, u_e \quad \text{and} \quad x_1, \cdots, x_N,$$

arranged in a fixed order

$$u_1 \prec \cdots \prec u_e \prec x_1 \prec \cdots \prec x_N.$$

We shall consider a linear space K^{e+N} of dimension $e+N$ over the field K, with a basis corresponding to $u_1, \cdots, u_e, x_1, \cdots, x_N$. In what follows by a polynomial we shall always mean one in the variables $u_1, \cdots, u_e, x_1, \cdots, x_N$ with coefficients in K, i.e., an element in the ring $K[u_1, \cdots, u_e, x_1, \cdots, x_N]$.

A monomial

$$\mu = a u_1^{i_1} \cdots u_e^{i_e} x_1^{m_1} \cdots x_N^{m_N} \qquad (a \in K),$$

will sometimes be written in the simple form

$$\mu = aU^I X^M, \ I = (i_1, \cdots, i_e), \quad M = (m_1, \cdots, m_N)$$

or

$$\mu = az^\alpha, \ \alpha = (I, M) = (i_1, \cdots, i_e, m_1, \cdots, m_N).$$

If $a \neq 0$, and the last one $\neq 0$ in the N-tuple $(m_1, \cdots, m_N)$ is m_p, then we say that the monomial is of *class* p; otherwise we say that the class of the monomial $\mu \neq 0$ is 0 .In that case in μ there occurs at most u but not x.

For two sets of non-negative integers

$$\alpha = (a_1, \cdots, a_t), \qquad \beta = (b_1, \cdots, b_t)$$

we say that α *precedes* β, or β *follows* α, which is denoted as

$$\alpha \prec \beta \quad \text{or} \quad \beta \succ \alpha,$$

if there is some k such that

$$a_{k+1} = b_{k+1}, \ \cdots, \ a_t = b_t$$

while $a_k < b_k$. For two non-zero monomials

$$\lambda = a u_1^{i_1} \cdots u_e^{i_e} x_1^{l_1} \cdots x_N^{l_N}, \quad a \neq 0,$$

$$\mu = b u_1^{j_1} \cdots u_e^{j_e} x_1^{m_1} \cdots x_N^{m_N}, \quad b \neq 0,$$

we say that λ *precedes* μ or μ *follows* λ, which is denoted as

$$\lambda \prec \mu \text{ or } \mu \succ \lambda$$

if

$$(i_1, \cdots, i_e, l_1, \cdots, l_N) \prec (j_1, \cdots, j_e, m_1, \cdots, m_N).$$

Any non-zero polynomial F can be written in the form

$$F = a_1 z^{\alpha_1} + a_2 z^{\alpha_2} + \cdots + a_t z^{\alpha_t}$$

in which

$$a_i \in K, \ a_1 \neq 0, \cdots, a_t \neq 0,$$

$$\alpha_1 \succ \alpha_2 \succ \cdots \succ \alpha_t.$$

In that case we say that $a_1 z^{\alpha_1}$ is the *leading term* of F, and the class of z^{α_1} will be called the *class* of F.

If a non-zero polynomial F has its class $= p \succ 0$, and the leading term $a_1 z^{\alpha_1}$ of F has its degree in $x_p = m$, then F can be written in the form

$$F = C_0 x_p^m + C_1 x_p^{m-1} + \cdots + C_m,$$

in which the C's are all polynomials in u and $x_1, \cdots, x_{p-1}$, containing none of $x_p, x_{p+1}, \cdots, x_N$, with $C_0 \neq 0$ in particular. The polynomial C_0 will then be called the *initial* of F. If the leading term of C_0 is c_0, then the leading term of F is clearly $c_0 x_p^m$.

Consider two non-zero polynomials F and G and any variable x_p. If the highest degree of x_p appearing in F is less than that in G, then we say that F has a *lower rank* than G

$$\mathscr{A}: A_1, A_2, \cdots, A_r.$$

Such a sequence will be called an *ascending set* if either (a) or (b) below holds true:

(a) $r = 1$ and $A_1 \neq 0$;

(b) $r > 1$, $0 < \text{class } (A_1) < \text{class } (A_2) < \cdots < \text{class } (A_r)$, and moreover A_j is reduced with respect to A_i for each pair $j > i$.

It is clear that for an ascending set one always has $r \leqslant N$.

An ascending set will be said to be *contradictory* if $r = 1$, $A_1 \neq 0$ with class $(A_1) = 0$.

Given a second ascending set

$$\mathscr{B}: B_1, B_2, \cdots, B_s,$$

we say that $\mathscr{A}$ has a *higher rank* than $\mathscr{B}$ or $\mathscr{B}$ a *lower rank* than $\mathscr{A}$, which is denoted as

$$\mathscr{A} \succ \mathscr{B} \text{ or } \mathscr{B} \prec \mathscr{A},$$

if either (a)′ or (b)′ below holds true:

(a)′ There is some $j \leqslant \min(r, s)$ such that

$$A_1 \sim B_1, \cdots, A_{j-1} \sim B_{j-1}, \text{ while } A_j \succ B_j;$$

(b)′ $s > r$ and

$$A_1 \sim B_1, \cdots, A_r \sim B_r.$$

If neither of the ascending sets $\mathscr{A}$ and $\mathscr{B}$ is of higher rank than the other, then we say that $\mathscr{A}, \mathscr{B}$ are of the *same rank*, denoted as $\mathscr{A} \sim \mathscr{B}$. In that case we have

$$r = s, \text{ and } A_1 \sim B_1, \cdots, A_r \sim B_r.$$

It is clear that the collection of all ascending sets is partially ordered by the rank. Hence for any set of ascending sets we can speak of the notion of *minimal ascending set*, if it exists. The following lemma, simple as it is, will play an important role in the whole theory.

Lemma 1. *Let*

$$\Phi_1, \Phi_2, \cdots, \Phi_q, \cdots$$

be a sequence of ascending sets Φ_q for which the rank never increases, or for any q we have either $\Phi_{q+1} \prec \Phi_q$ or $\Phi_{q+1} \sim \Phi_q$. Then there is an index q' such that for any $q > q'$ we have

$$\Phi_q \sim \Phi_{q'}.$$

In other words, there is some q' such that any Φ_q for which $q \geqslant q'$ is a minimal ascending set of the above sequence.

Proof. For the ascending set Φ_q let us denote by r_q its number of polynomials and by A_q the first polynomial in the set. Then

$$A_1, A_2, \cdots, A_q, \cdots$$

is a sequence of polynomials for which the rank never increases, or for any q we have either $A_{q+1} \prec A_q$ or $A_{q+1} \sim A_q$. Consequently for any q we have class $(A_{q+1}) \leqslant$ class (A_q) and or G has a *higher rank* than F *with respect to* x_p. We say that F and G have the *same rank with respect to* x_p when neither is of higher rank than the other.

For two non-zero polynomials F and G we say that F has a *lower rank* than G or G a higher rank than F, which is denoted as

$$F \prec G \quad \text{or} \quad G \succ F,$$

if one of the following two cases occurs:

1. class $F <$ class G;

2. class $F =$ class G, say, $= p > 0$, while the degree of x_p in F is less than that of x_p in G, or in other words, F has lower rank than G with respect to x_p.

In the case neither of F and G is of higher or lower rank than the other, F and G will be said to be of the *same rank*, denoted as

$$F \sim G.$$

For example, two non-zero polynomials are of the same rank if both are of class $= 0$. Let F be a polynomial of class $p > 0$. Any polynomial G of rank lower than F with respect to x_p will be said to be *reduced* with respect to F. Clearly the initial of F is of class $<$ p and is already reduced with respect to F.

Let F be of class $p > 0$ written in the form

$$F = f_0 x_p^m + f_1 x_p^{m-1} + \cdots f_m,$$

in which

$$f_i \in K[u_1, \cdots, u_e, x_1, \cdots, x_{p-1}], \quad f_0 \neq 0.$$

Any non-zero polynomial G which has not been reduced with respect to F can then always be written in the form

$$G = g_0 x_p^M + g_1 x_p^{M-1} + \cdots + g_M,$$

in which

$$g_i \in K[u_1, \cdots, u_e, x_1, \cdots, x_{p-1}, x_{p+1}, \cdots, x_N],$$

and

$$g_0 \neq 0, \quad M \geqslant m.$$

By the division algorithm of polynomials, we would get, in dividing G by F, an expression of the form

$$f_0^s G = QF + R,$$

where Q, R are both polynomials with, in the case $R \neq 0$, the degree of x_p in $R < m$ so that R is already reduced with respect to F. The integer s will be determined as the smallest to make possible such an expression that s is unique and is $\leqslant M - m$. If G is already reduced with respect to F, then we can simply take $s = 0$, $Q = 0$, $R = G$ so that the above expression holds true still. In any way, the polynomial R will be called the *remainder* of G with respect to F. The procedure to get the remainder R from G will then be called the *reduction* of G with respect to F.

In what follows we shall consider sequences formed by a finite number of polynomials A_i like the one below.

in the case class (A_{q+1}) = class (A_q), say, $= p > 0$ the degree in x_p of A_{q+1} should be $\leqslant$ the corresponding degree in x_p of A_q. As both class and degree are non-negative integers, there should be some index q_1 such that all A_q are of the same rank for $q \geqslant q_1$.

If there is some $q_1' \geqslant q_1$ such that all $r_q = 1$ for any $q \geqslant q_1'$, then the lemma is clearly true. Suppose the contrary. Then there should be some $q_1' \geqslant q_1$ such that all $r_q \geqslant 2$ for any $q \geqslant q_1'$. Denote the second polynomial in such Φ_q by $A_q^{(1)}$. Then

$$A_{q_1'}^{(1)}, A_{q_1'+1}^{(1)}, \cdots, A_q^{(1)}, \cdots$$

will be a sequence of polynomials with non-increasing ranks. As before there will then be some $q_2 \geqslant q_1'$ such that all $A_q^{(1)}$ are of the same rank for any $q \geqslant q_2 \geqslant q_1' \geqslant q_1$.

If all $r_q \leqslant 2$, then the lemma is proved already. Suppose the contrary. Then there will be some $q_2' \geqslant q_2$ such that all $r_q \geqslant 3$ for any $q \geqslant q_2'$ and we may take the third polynomial $A_q^{(2)}$ in such Φ_q's to form a sequence of polynomials with non-increasing ranks. As for all q we have $r_q \leqslant n$, so proceeding in this way we should stop at some r and some q' such that for all $q \geqslant q'$ we have $r_q = r$ and the r-th polynomials taken from such Φ_q will all have the same rank. It follows that all such Φ_q's will have the same rank and the lemma is proved.

From this lemma we get the following

Lemma 1'. *If in a sequence of ascending sets the ranks are steadily decreasing, then such a sequence can only be composed of a finite number of ascending sets.*

Suppose now we have a non-empty collection $\Sigma = \{F_\alpha\}$ of non-zero polynomials F_α. An ascending set $\mathscr{A}$ of polynomials will be said to *belong* to Σ if each polynomial in $\mathscr{A}$ belongs to Σ. Since each single $F_\alpha \neq 0$ forms by itself an ascending set, such ascending sets belonging to Σ exit naturally. Any minimal ascending set of the collection of all ascending sets belonging to Σ will then be called a *basic set* of Σ.

The following lemma points out not only the existence of such basic sets but also some constructive method of arriving at such basic sets.

Lemma 2. *Let Σ be a finite set of non-zero polynomials. Then Σ has necessarily basic sets and there is a mechanical method in getting such a basic set in a finite number of steps.*

Proof. As Σ is finite, the existence of basic sets is quite evident. So the problem reduces to the mechanical generation of such a basic set.

To show this let us find at the outset a polynomial, say A_1, of lowest rank from $\Sigma = \Sigma_1$. This can clearly be done in a mechanical manner. If class $(A_1) = 0$, then A_1 alone will form already a basic set. Suppose therefore class $(A_1) > 0$. Check whether each polynomial except A_1 in Σ_1 is already reduced with respect to A_1. If no such polynomial exists in Σ_1, then A_1 by itself forms already a basic set of Σ_1. Otherwise let Σ_2 be the subset of Σ_1 formed by all such polynomials except A_1 already reduced with respect to A_1. From the choice of A_1 all polynomials in Σ_2 will have a rank higher than that of A_1. Now let A_2 be a polynomial in Σ_2 of lowest rank. If Σ_2 has not any polynomial which is different from A_2 and is already reduced with respect to A_1, then A_1, A_2 will form a basic set of Σ. Otherwise let Σ_3 be the subset of Σ_2 consisting of all polynomials except A_2 which have already been reduced with respect to A_2. Choose from Σ_3 a polynomial A_3 of lowest rank and proceed as before. As the classes of the polynomials $A_1, A_2, A_3, \cdots$ are steadily increasing and unlikely to become $>N$, we have to stop in a finite number of steps and get finally a basic set in a mechanical manner, Q. E. D.

Lemma 3. *Let Σ be a finite set of non-zero polynomials with a basic set*

$$\mathscr{A}: A_1, A_2, \cdots, A_r$$

of which class $(A_1) > 0$. Let B be a non-zero polynomial reduced with respect to all A's. Then the set Σ' obtained from Σ by adjunction of B will have a basic set of rank lower than that of $\mathscr{A}$.

Proof. If class $(B)=0$, then B alone will form a basic set of Σ' of rank lower than that of $\mathscr{A}$. Suppose therefore class $(B)=p>0$. As B is already reduced with respect to all A's, there should be some $i \geqslant 0$ and $\leqslant r$ such that $p >$ class (A_{i-1}) and $p \leqslant$ class (A_i). Moreover, in the case $p=$ class (A_i), the degree of x_p in B will be less than that of x_p in A_i. Hence

$$A_1, A_2, \cdots, A_{i-1}, B$$

will be an ascending set of Σ' with a rank lower than that of $\mathscr{A}$. The basic set of Σ' will have therefore *à fortiori* a rank lower than that of $\mathscr{A}$, Q. E. D.

Remark. The above lemmas are clearly also true for any infinite set of polynomials and the proofs remain essentially the same as long as the axiom of choice is applied. As the use of axiom of choice will be in opposition to the mechanical thought, the main theme of the whole theory, we have deliberately restrict ourselves to the case of finite sets of polynomials.

Consider now an ascending set

$$\mathscr{A}: A_1, A_2, \cdots, A_r$$

as before with class $(A_1)>0$. Let class $(A_i)=p_i$ and let the initial of A_i be I_i. Then

$$0<p_1<p_2<\cdots<p_r$$

and for each i we have

class $(I_i)<p_i$,

I_i reduced with respect to $A_1, \cdots, A_{i-1}$.

Let B be an arbitrary polynomial. Set $B=R_r$. With respect to the polynomials in $\mathscr{A}$ starting from A_r to A_1 we can form successively the remainders $R_{r-1}, \cdots, R_0$ of R_r so that we get $(s_i \geqslant 0)$:

$$I_r^{s_r} R_r = Q_r A_r + R_{r-1},$$

$$I_{r-1}^{s_{r-1}} R_{r-1} = Q_{r-1} A_{r-1} + R_{r-2},$$

$$\cdots$$

$$I_1^{s_1} R_1 = Q_1 A_1 + R_0.$$

Set $R_0=R$. Then we get an expression of the form

$$I_1^{s_1} \cdots I_r^{s_r} B = Q_1' A_1 + \cdots + Q_r' A_r + R,$$

in which Q' are all polynomials. The polynomial R is determined from B and the ascending set $\mathscr{A}$. We shall call R the *remainder* of B with respect to $\mathscr{A}$. We call also the above formula the *Remainder Formula*.

It is clear that any term of R will have the degree in x_{p_i} less than the corresponding degree in x_{p_i} of A_i. In other words, R is reduced with respect to all polynomials A_i in $\mathscr{A}$. We shall say briefly that R is *reduced* with respect to $\mathscr{A}$ and call the above procedure of getting R from B and $\mathscr{A}$ the *reduction* of B with respect to $\mathscr{A}$. As the determination of one polynomial with respect to the other is done mechanically by the division algorithm, we may state the result in the form of the following

Lemma 4. *Given a non-zero polynomial B and an ascending set $\mathscr{A}$ of which the first polynomial is of class > 0, there is an algorithm which permits to determine the remainder of B with respect to $\mathscr{A}$ in a finite number of steps. Denote the i-th polynomial in $\mathscr{A}$ by A_i and its class by p_i. Then any term in the remainder R will have its degree of x_{p_i} in A_i less than the degree of x_{p_i} in A_i for each i.*

Come now to the *well-ordering* of a polynomial set as follows. For this purpose let us review briefly the notion *zero* of such a set.

Consider any polynomial F. Suppose that there is a set of numbers

$$u_1^0, \cdots, u_e^0, x_1^0, \cdots, x_N^0$$

in K which will turn F into 0 when these numbers are substituted for the variables $u_1, \cdots, u_e, x_1, \cdots, x_N$ in F. Then this set of numbers, which may be considered as the coordinates of a point in the linear space K^{e+N}, is called a *zero* of the polynomial F or alternatively a *solution* of the equation $F = 0$. If the various u^0, x^0 are not numbers of K, but of some extension field $\tilde{K}$ of K, which still turn F into 0 when substituted into it, then, the set of numbers, considered as a point of the linear space $\tilde{K}^{e+N}$ on $\tilde{K}$, will be called an *extended zero* of F or an *extended solution* of $F = 0$. In order to make the involved field $\tilde{K}$ explicit, it will also be called a *$\tilde{K}$-zero* of F or a *$\tilde{K}$-solution* of $F = 0$.

Given a set of polynomials Σ, if a set of numbers as given above is a zero (or extended zero, or $\tilde{K}$-zero) of every polynomial in Σ, then it will be called simply a *zero* (resp. an *extended zero* or a *$\tilde{K}$-zero*) of Σ or a *solution* (resp. an *extended solution* or a *$\tilde{K}$-solution*) of $\Sigma = 0$.

Consider now a set $\Sigma = \Sigma_1$ of non-zero polynomials, supposed to be finite in number. By Lemma 2, Σ_1 will have some basic set, say Φ_1. If Φ_1 is a contradictory set, then Φ_1 consists of a single polynomial A_1 belonging to Σ_1 for which class $(A_1) = 0$. Suppose on the contrary that Φ_1 is not contradictory so that the first polynomial in Φ_1 has its class > 0. For polynomials B, which belong to Σ_1 but not to Φ_1, let us form the remainders R_B of B with respect to Φ_1 supposed not all 0. Adjoin all such remainders R_B, whenever non-zero, to the set Σ_1 to get an enlarged set of non-zero polynomials Σ_2. From the formula about remainders each R_B, when non-zero, will be a linear sum of polynomials in Φ_1 as well as the polynomial B, with polynomials as coefficients. It follows that the set Σ_2 will have the same set of zeros (or extended zeros, or $\tilde{K}$-zeros for any extended field $\tilde{K}$) as the original set Σ_1. Form now the basic set Φ_2 of Σ_2. By Lemma 3 Φ_2 will have a rank lower than that of Φ_1. If Φ_2 is not a contradictory ascending set then we can proceed as before. In this way we shall get either a contradictory ascending set after a finite number of steps or a sequence of finite sets of polynomials

$$\Sigma_1 \subset \Sigma_2 \subset \cdots \subset \Sigma_q \subset \cdots,$$

where all Σ_i have same set of zeros (or extended zeros or $\tilde{K}$-zeros for any extended field $\tilde{K}$) with the corresponding non-contradictory basic sets Φ_i having steadily decreasing ranks:

$$\Phi_1, \Phi_2, \cdots, \Phi_q, \cdots.$$

Now by Lemma 1, such a sequence can have only a finite number of terms. In other words, if the last one of such a sequence of finite sets of polynomials is Σ_q, with Φ_q as the corresponding basic set, then the remainder of any polynomial of Σ_q not in Φ_q with respect to Φ_q will be equal to 0.

Let Φ_q be

$$\Phi_q: F_1, F_2, \cdots, F_r,$$

in which each F_i is either belonging originally to Φ_{q-1}, or is the non-zero remainder of some polynomial in Σ_{q-1} with respect to Φ_{q-1}. By the remainder formula each F_i is thus a linear sum of polynomials in Φ_{q-1} with polynomials as coefficients. It follows that any zero of Σ_{q-1} and thus any zero of Σ is also a zero of Φ_q.

On the other hand let the initials of polynomials in Φ_q be $I_1, I_2, \cdots, I_r$. From the construction we know that for any polynomial G in Σ_q, there should be non-negative integers $s_i \geqslant 0$ such that

$$I_1^{s_1} \cdots I_r^{s_r} G = Q_1 F_1 + \cdots + Q_r F_r.$$

It follows that any zero of Φ_q, if not a zero of any one of the initials $I_1, \cdots, I_r$, is necessarily also a zero of Σ_q and thus a zero of $\Sigma = \Sigma_1$. The same is clearly true for extended zeros or $\tilde{K}$-zeros for any extended field $\tilde{K}$.

Denote Φ_q by Φ. Then what we have proved may be reformulated as the theorem below:

Theorem (Ritt). *There is an algorithm which permits to get, after mechanically a finite number of steps, either a polynomial A of class 0, i. e. one in variables $u_1, \cdots, u_e$ so that any zero of Σ is also a zero of A, or a non-contradictory ascending set*

$$\Phi: F_1, \cdots, F_r,$$

with initials $I_1, \cdots, I_r$ such that any zero of Σ is also a zero of Φ, and any zero of Φ which is not zero of any of the initials I_i, will also be a zero of Σ. The same is ture for extended zeros and $\tilde{K}$-zeros.

We shall call the mechanical procedure which permits to determine Φ from Σ a *well-ordering* of Σ and the above theorem will be called the *Well-Ordering Theorem*. The theorem is due to Ritt and forms the basis of our method. We shall call the theorem *Ritt Principle* accordingly. The polynomial set Φ in the theorem is called a *characteristic set* of Σ.

§ 3. A Constructive Theory of Algebraic Varieties

As before, let K be the basic field of characteristic 0 and

$$x_1 \prec x_2 \prec \cdots \prec x_N$$

be a set of variables arranged in a definite order with $u_1, \cdots, u_e$ neglected. A polynomial will always be understood as one in $K[x_1, \cdots, x_N]$.

A finite set of non-zero polynomials will simply be called a *polynomial set*. The polynomial set Σ obtained from putting together the polynomials in two polynomial sets Σ_1 and Σ_2 will be denoted as $\Sigma_1 + \Sigma_2$. For polynomials F, G, etc., $\Sigma + \{F\}$ will also be denoted as $\Sigma + F$, and $\Sigma + \{F, G\}$ as $\Sigma + F + G$, etc.

We say that a polynomial set Σ defines an *algebraic variety* or simply *a variety*, to be denoted as $|\Sigma|$, with Σ as its *defining set*. For two polynomial sets Σ_1 and Σ_2, if any extended zero of Σ_1 is also an extended zero of Σ_2, then we say that the algebraic variety defined by Σ_1 is a *subvariety* of that defined by Σ_2, to be denoted as

$$\Sigma_2 = 0 | \Sigma_1, \quad \text{or} \quad |\Sigma_1| \subset |\Sigma_2|.$$

If, further, we have $|\Sigma_2| \subset |\Sigma_1|$ so that Σ_1, Σ_2 have the same set of extended zeros, then we say that Σ_1, Σ_2 are *equivalent*, denoted as

$$\Sigma_1 \approx \Sigma_2, \quad \text{or} \quad |\Sigma_1| = |\Sigma_2|.$$

If $|\Sigma_1| \subset |\Sigma_2|$ but $|\Sigma_1| \neq |\Sigma_2|$, or $|\Sigma_1| \subsetneqq |\Sigma_2|$, then we say that the variety defined by Σ_1 is a *true* subvariety of that defined by Σ_2.

Given a polynomial F, if any extended zero of Σ is also one of F, i.e.

$$\{F\} = 0 | \Sigma \text{ or } |\Sigma| \subset |\{F\}|,$$

then we say that $F = 0$ on Σ, denoted as $F = 0 | \Sigma$. Otherwise we denote this as

$$F \neq 0 | \Sigma.$$

Given $k + 1$ polynomial sets Σ, $\Sigma_1, \cdots, \Sigma_k (k > 1)$ having the following property: Any extended zero of Σ is also an extended zero of at least one of the sets $\Sigma_1, \cdots, \Sigma_k$, and conversely, any extended zero of any Σ_i is also one of Σ, then we say that $\Sigma_1, \cdots, \Sigma_k$ are a *decomposition* of Σ, or the corresponding algebraic varieties $|\Sigma_1|, \cdots, |\Sigma_k|$ are a *decomposition* of $|\Sigma|$, denoted as

$$|\Sigma| = |\Sigma_1| \cup \cdots \cup |\Sigma_k| \qquad (k > 1).$$

If for any i, $|\Sigma_i|$ cannot be omitted in the above decomposition, then the decomposition is said to be *uncontractible*. In this case the variety defined by each Σ_i is a true subvariety of the variety defined by Σ, but not a subvariety defined by the union of other Σ_j's.

We say that the polynomial set Σ is *reducible* if it has some uncontractible decomposition and the variety defined by it is also said to be *reducible*. In the contrary case we say that Σ as well as the variety defind by it is *irreducible*. If in a certain decomposition of Σ each Σ_i is irreducible, then we say that this decomposition is an *irreducible decomposition* of Σ; the same for the variety defined by Σ. In this case each Σ_i or the variety defined by it is called an *irreducible component* of Σ or the variety defined by it.

We consider now the problem of reducibility of a polynomial set or its defining algebraic variety. The following two lemmas give some well-known criteria for their *irreducibility*.

Lemma 1. *A necessary and sufficient condition for a polynomial set Σ to be irreducible is that there cannot exist two non-zero polynomials G and H such that*

$$GH = 0|\Sigma,$$

while

$$G \neq 0|\Sigma,\ H \neq 0|\Sigma.$$

For the second criterion let us first introduce the important notion of the so-called *generic point* of a variety. Consider two extension fields $\tilde{K}$ and K' of K and two points $\xi = (\tilde{x}_1, \cdots, \tilde{x}_N)$, $\tilde{x}_i \in \tilde{K}$, and $\xi' = (x'_1, \cdots, x'_N)$, $x'_i \in K'$, in the N-dimensional linear spaces $\tilde{K}^N$ and K'^N on $\tilde{K}$ and K' respectively. Suppose that these two points possess the following property:

For any polynomial $F(x_1, \cdots, x_N)$ in $K[x_1, \cdots, x_N]$, that ξ is an extended zero of F would imply that ξ' is also an extended zero of F; in other words, $F(x'_1, \cdots, x'_N) = 0$ as long as $F(\tilde{x}_1, \cdots, \tilde{x}_N) = 0$.

In this case ξ' will be called a *specialization* of ξ with respect to K, or simply a specialization of ξ if no misunderstanding can occur.

Suppose the polynomial set Σ has a certain extended zero ξ such that any extended zero of Σ is a specialization of ξ with respect to K, then we say that ξ is a *generic point* of the polynomial set Σ or one of the algebraic variety $|\Sigma|$ defined by it. The following lemma gives the second irreducibility criterion of polynomial sets or algebraic varieties:

Lemma 2. *A necessary and sufficient condition for a polynomial set Σ or its variety to be irreducible is that Σ has generic points.*

The two lemmas above give some necessary and sufficient conditions which are however merely existential in character and not constructive at all. Given a polynomial set Σ, there is no means to ascertain in a finite number of steps whether the conditions in the lemmas can be satisfied or not. For the purpose of mechanical theorem proving, we have to devise some mechanical procedure which permits to decide in a finite number of steps whether a given polynomial set is irreducible or not, and in the case it is reducible, to give in a finite number of steps the various irreducible components of the decomposition. Such a mechanization may be considered as constituting a *constructive theory* of algebraic geometry. It was given in details in the two books of J. F. Ritt[2,3] and we shall give some outlines in somewhat revised form of this theory below.

Consider an ascending set

$$\Phi: A_1, A_2, \cdots, A_n$$

in which the class of A_i is p_i with

$$0 < p_1 < p_2 < \cdots < p_n.$$

We shall change the notations in setting

$$x_{p_1} = y_1, \cdots, x_{p_n} = y_n$$

and denote the other x's in the original order as $u_1, \cdots, u_d$. We call

$$d = N - n$$

the *dimension* of the ascending set Φ, denoted as

$$d = \dim \Phi.$$

Write now the polynomials A_i in Φ in the following form:

$$\Phi:\begin{cases} A_1 = C_{10}y_1^{m_1} + C_{11}y_1^{m_1-1} + \cdots + C_{1m_1}, \\ A_2 = C_{20}y_2^{m_2} + C_{21}y_2^{m_2-1} + \cdots + C_{2m_2}, \\ \cdots \\ A_n = C_{n0}y_n^{m_n} + C_{n1}y_n^{m_n-1} + \cdots + C_{nm_n}. \end{cases}$$

In the expressions $C_{i0} \neq 0$ are initials of A_i, and each C_{ij} is a polynomial in $u_1, \cdots, u_d, y_1, \cdots, y_{i-1}$ with coefficients in K. Furthermore each A_i has already been reduced with respect to $A_1, \cdots, A_{i-1}$ so that the degrees of $y_1, \cdots, y_{i-1}$ in C_{ij} are less than $m_1, \cdots, m_{i-1}$ respectively. The first problem to be considered is to give conditions for Φ to be the basic set of a certain irreducible polynomial set.

For this problem let us suppose that the ascending set Φ possesses the following property:

Let the transcendental extension field $K(u_1, \cdots, u_d)$ of K got by adjoining $u_1, \cdots, u_d$ be denoted by K_0; then A_1, as a polynomial in $K_0[y_1]$ with coefficients in K_0, is irreducible in $K_1[y_1]$.

Let the algebraic extension field of K_0 got by adjoining an extended zero η_1 of $\tilde{A}_1 = 0$ be denoted by $K_0(\eta_1) = K_1$; then the polynomial $\tilde{A}_2$ in $K_1[y_2]$ obtained by substituting η_1 for y_1 in A_2 is irreducible in $K_1[y_2]$.

Let the algebraic extension field of K_1 got by adjoining an extended zero η_2 of $\tilde{A}_2 = 0$ be denoted by $K_1(\eta_2) = K_2$; then the polynomial $\tilde{A}_3$ in $K_2[y_3]$ obtained by substituting η_1 for y_1 and η_2 for y_2 in A_3 is irreducible in $K_2[y_3]$.

Suppose that proceeding in the same manner we get successively algebraic extensions $K_i = K_{i-1}(\eta_i)$, polynomials $\tilde{A}_i$ obtained by substituting $\eta_1, \cdots, \eta_{i-1}$ for $y_1, \cdots, y_{i-1}$ in A_i, and some extended zeros η_i of $\tilde{A}_i = 0$, where each $\tilde{A}_i$ is irreducible in $K_{i-1}[y_i]$ for $i = 1, 2, \cdots, n$. Under these conditions we say that the ascending set Φ is *irreducible*. By known methods there exist some mechanical procedures which permit to decide in a finite number of steps whether Φ is irreducible or not.

Let Φ be irreducible and satisfy the conditions above. Then u_i, η_j are all elements in $\tilde{K} = K_n$ and $\tilde{\eta} = (u_1, \cdots, u_d, \eta_1, \cdots, \eta_n)$ can be considered as a point of the linear space $\tilde{K}^{d+n} = \tilde{K}^N$. We shall call $\tilde{\eta}$ a *generic point* of Φ and $\tilde{K}$ a *generating field* of Φ.

The following lemma is quite important for the theory.

Lemma 3. *If the ascending set Φ is irreducible with*

$$\tilde{\eta} = (u_1, \cdots, u_d, \eta_1, \cdots, \eta_n)$$

a generic point as above, then for a polynomial $F \in K[u_1, \cdots, u_d, y_1, \cdots, y_n]$ to have the remainder $R = 0$ with respect to Φ, it is necessary and sufficient that $\tilde{\eta}$ is an extended zero of F.

Proof. Denote the ascending set formed by the first k terms in Φ by

$$\Phi_k: A_1, A_2, \cdots, A_k \ (1 \leqslant k \leqslant n).$$

Denote by K_k the $(d+k)$-dimensional linear space over K with basis $u_1, \cdots, u_d, y_1, \cdots,$

y_k. Similarly for the others. Then Φ_k is clearly irreducible, and

$$\bar{\eta}_k = (u_1, \cdots, u_d, \eta_1, \cdots, \eta_k),$$

when considered as a point in K_k^{d+k}, is a generic point of Φ_k while K_k is the generating field of Φ_k.

We shall prove by induction on k the following two assertions:

1_k. $\bar{\eta}_{k-1}$ is not an extended zero of C_{k0}.

2_k. If $R_k \in K[u_1, \cdots, u_d, y_1, \cdots, y_k]$ is already reduced with respect to Φ_k and $\bar{\eta}_k$ is an extended zero of R_k, then R_k is identically 0.

As $C_{k+1,0} \in K[u_1, \cdots, u_d, y_1, \cdots, y_k]$ is known to be reduced with respect to Φ_k and is $\neq 0$, so 1_{k+1} is a consequence of 2_k.

Suppose 2_{k-1} has already been proved. Consider any R_k satisfying the conditions in 2_k. Write R_k as a polynomial in y_k,

$$R_k = S_0 y_k^r + S_1 y_k^{r-1} + \cdots + S_r,$$

in which $S_i \in K[u_1, \cdots, u_d, y_1, \cdots, y_{k-1}]$ with $r < m_k$. Substitute $y_1, \cdots, y_{k-1}$ in S_i by $\eta_1, \cdots, \eta_{k-1}$ with the resulting S_i as $\tilde{S}_i \in K_{k-1}$. Set

$$\tilde{R}_k = \tilde{S}_0 y_k^r + \tilde{S}_1 y_k^{r-1} + \cdots + \tilde{S}_r \in K_{k-1}[y_k].$$

By hypothesis η_k is an extended zero of $\tilde{R}_k = 0$. As $r < m_k$ and η_k is an extended zero of the irreducible polynomial $\tilde{A}_k$ in K_{k-1}, $\tilde{R}_k$ should be identically 0 and so $\tilde{S}_0 = 0, \cdots, \tilde{S}_r = 0$. As R_k is reduced with respect to Φ_k so that each S_i is reduced with respect to Φ_{k-1}, by induction hypothesis 2_{k-1} we have necessarily $S_i = 0$ so that $R_k = 0$, i. e., 2_k holds true. It follows that 1_{k+1} is also true. The above proof is clearly valid for 2_1 while 1_1 is quite evident. Consequently 1_k and 2_k are true for $k = 1, 2, \cdots, n$.

It is now easy to complete the proof of Lemma 3 as follows.

Let the remainder of F with respect to $\Phi_n = \Phi$ be R; then we have the following remainder formula

$$C_{10}^{s_1} \cdots C_{n0}^{s_n} F = Q_1 A_1 + \cdots + Q_n A_n + R.$$

Suppose $R = 0$. Since $\bar{\eta}$ is an extended zero of all A_i's while by 1_k it is not an extended zero of any C_{k0}, so by the formula above it should be an extended zero of F. Conversely, if $\bar{\eta}$ is an extended zero of F, then by the same formula $\bar{\eta}$ should also be an extended zero of R. By 2_n we have necessarily $R = 0$. This completes the proof.

Lemma 4. *Let the ascending set*

$$\Phi: A_1, A_2, \cdots, A_n$$

be irreducible with a generic point

$$\bar{\eta} = (u_1, \cdots, u_d, \eta_1, \cdots, \eta_n)$$

as before. If the polynomial $F \in K[u_1, \cdots, u_d, y_1, \cdots, y_n]$ has its remainder $\neq 0$ with respect to Φ, then in $K[u_1, \cdots, u_d, y_1, \cdots, y_n]$ there are polynomials G and Q_i, $i = 1, \cdots, n$ such that

$$GF-(Q_1A_1+\cdots+Q_nA_n)\in K[u_1,\cdots,u_d]$$

and that

$$G(\bar{\eta})\neq 0.$$

Proof. Omitted.

Given an irreducible set Φ as above, let Ω be the set of all polynomials in $K[u_1,\cdots,u_d,y_1,\cdots,y_n]$ for which the remainder with respect to Φ is 0. By Lemma 3, this set will form clearly a module. By the Hilbert basis theorem, there will be a finite number of polynomials in Ω, such that any polynomial of Ω is a linear combination of these polynomials with polynomial coefficients. We may add the A_i's of Φ into this finite set and denote the enlarged finite set by Ω_Φ. By Lemma 3 this polynomial set will have clearly Φ as its basic set and $\bar{\eta}$ as an extended zero.

Let G be any polynomial with $\bar{\eta}$ as an extended zero; then by Lemma 3 G has its remainder $=0$ with respect to Φ. By the construction of Ω_Φ, G is a linear sum of polynomials in Ω_Φ so that $G=0/\Omega_\Phi$. It follows that any extended zero of Ω_Φ is a specialization of $\bar{\eta}$ or that Ω_Φ is an irreducible polynomial set with $\bar{\eta}$ as a generic point. We thus get the following

Theorem 1. *Any irreducible ascending set Φ is the basic set of some irreducible polynomial set Ω_Φ,*

The above proof showing how to get an irreducible polynomial set Ω_Φ from a given irreducible ascending set Φ is based on the use of the finite basis theorem of Hilbert. As Ω is transfinite, and the existence of a finite basis depends on the axiom of choice, only the existence of such an irreducible polynomial set Ω_Φ has been actually proved. However, there does exist some mechanical procedure to produce in a finite number of steps such an irreducible polynomial set Ω_Φ consisting of a finite number of polynomials. In other words, we may strengthen the above theorem to the following form:

Theorem 1'. *There exists some mechanical procedure for any irreducible ascending set Φ which will permit to determine in a finite number of steps a finite number of polynomials including those of Φ that form an irreducible polynomial set Ω_Φ with any generic point of Φ as its generic point.*

The proof of the constructive Theorem 1' is not a simple one. As in applications the mere existence of such an irreducible polynomial set Ω_Φ will already be sufficient, as guaranteed by the Hilbert basis theorem, we shall satisfy ourselves in merely stating the theorem while putting aside the proof.

The next problem to be studied is the decomposition of a polynomial set or the corresponding algebraic variety into irreducible components. For this purpose let Φ, $\bar{\eta}$ and Ω_Φ be as before. we have shown that the irreduciblity of Φ is a sufficient condition for Φ to be the basic set of some irreducible polynomial set Ω_Φ with the same generic point $\bar{\eta}$ as Φ which can even be determined in a mechanical manner in a finite number of steps. To this we now give the following supplement:

Lemma 5. *Let the basic set Φ of a polynomial set Λ be irreducible with the class of each polynomial A_i in Φ being >0. Denote the initial of A_i by I_i, $i=1,\cdots,n$. If any polynomial in Λ has its remainder 0 with respect to Φ, then Λ has a decomposition*

$$|\Lambda| = |\Omega_\Phi| \cup |\Lambda + I_1| \cup \cdots \cup |\Lambda + I_n|,$$

in which Ω_Φ or the corresponding algebraic variety $|\Omega_\Phi|$ is irreducible.

Proof. For such a polynomial G in Λ or not with its remainder 0 with respect to Φ we would have, for some $s_i \geqslant 0$ and $Q_i \in K[u_1, \cdots, u_d, y_1, \cdots, y_n]$,

$$I_1^{s_1} \cdots I_n^{s_n} G = Q_1 A_1 + \cdots + Q_n A_n.$$

By the construction of Ω_Φ, G should be a linear sum of polynomials in Ω_Φ so that any extended zero of Ω_Φ should be an extended zero of G and hence an extended zero of Λ. Conversely, any extended zero of Λ may be considered an extended zero of A_i's. Hence by the above formula it should be an extended zero of either any such G or some I_i. In other words, it should be an extended zero of Ω_Φ or some $\Lambda + I_i$. Thus we have the decomposition as shown in the lemma.

Lemma 6. *Let Λ, Φ be as in Lemma 5 with Λ being irreducible. Then*

$$\Lambda \approx \Omega_\Phi \text{ or } |\Lambda| = |\Omega_\Phi|.$$

Proof. Let the initials of the polynomials in Φ be I_i, $i = 1, \cdots, n$. Then it is clear by definition that

$$|\Lambda + I_1| \cup \cdots \cup |\Lambda + I_n| = |\Lambda + I_1 \cdots I_n|.$$

The decomposition given in Lemma 5 can therefore be written in the form

$$|\Lambda| = |\Omega_\Phi| \cup |\Lambda + I_1 \cdots I_n|.$$

As the generic point of Φ is also a generic point of Ω_Φ but cannot be any extended zero of $I_1 \cdots I_n$, so $|\Omega_\Phi| \not\subset |\Lambda + I_1 \cdots I_n|$. If Λ has some extended zero which is not an extended zero of Ω_Φ, it should be an extended zero of $\Lambda + I_1 \cdots I_n$ so that we shall have $|\Lambda + I_1 \cdots I_n| \not\subset |\Omega_\Phi|$. In this way $|\Lambda|$ would have an uncontractible decomposition contrary to the irreducibility hypothesis of Λ. Hence we should have $|\Lambda| \subset |\Omega_\Phi|$. As conversely we should have $|\Omega_\Phi| \subset |\Lambda|$, so $|\Lambda| = |\Omega_\Phi|$, Q. E. D.

Consider now an ascending set Φ as before but with Φ not necessarily irreducible. Then there will be some k such that

$$\Phi_{k-1}: A_1, A_2, \cdots, A_{k-1}$$

is irreducible, with

$$\bar\eta_{k-1} = (u_1, \cdots, u_d, \eta_1, \cdots \eta_{k-1})$$

as a generic point, and that the polynomial $\tilde A_k$ got from A_k by substituting $\eta_1, \cdots, \eta_{k-1}$ for $y_1, \cdots, y_{k-1}$ is reducible in $K_{k-1}[y_k]$, where $K_{k-1} = K_0(\eta_1, \cdots, \eta_{k-1})$. Let the irreducible factorization of $\tilde A_k$ in $K_{k-1}[y_k]$ be given by

$$\tilde A_k = g_1 \cdots g_h,$$

in which each $g_i \in K_{k-1}[y_k]$ is irreducible, and $h \geqslant 2$. As in g_i the coefficients of powers of y_k are all elements of K_{k-1} and can thus be expressed as the quotients of two polynomials in $u_1, \cdots, u_d, \eta_1, \cdots, \eta_{k-1}$, multiplying by a common multiple of the denominators we would get an expression of the form

$$\tilde{D}\tilde{A}_k = \tilde{G}_1\cdots\tilde{G}_h,$$

in which $D \in K[u_1, \cdots, u_d, y_1, \cdots, y_{k-1}]$, $G_i \in K[u_1, \cdots, u_d, y_1, \cdots, y_k]$, while $\tilde{D}$, $\tilde{G}_i$ are got from D, G_i by substituting $\eta_1, \cdots, \eta_{k-1}$ for $y_1, \cdots, y_{k-1}$ and are polynomials in $K_{k-1}[y_k]$. We may also consider D as already reduced with respect to Φ_{k-1}. Similarly we may consider G_i as already reduced with respect to Φ_k.

Write the polynomial $G_1\cdots G_h - DA_k$ in a form according to powers of y_k, say,

$$G_1\cdots G_h - DA_k = \sum_j B_j y_k^j,$$

in which $B_j \in K[u_1, \cdots, u_d, y_1, \cdots, y_{k-1}]$. Denote by b_j the element in $K_{k-1} = K_0(\eta_1, \cdots, \eta_{k-1})$ got from B_j by substituting $\eta_1, \cdots, \eta_{k-1}$ for $y_1, \cdots, y_{k-1}$. Then we have $b_j = 0$ since $\tilde{D}\tilde{A}_k = \tilde{G}_1\cdots\tilde{G}_k$. In other words, each B_j will have $\bar{\eta}_{k-1}$ as an extended zero. It follows from the proof of Lemma 5 that each B_j will have its remainder 0 with respect to the *irreducible* ascending set Φ_{k-1}, so that there are non-negative integers $s_{j1}, \cdots, s_{j,k-1}$ and polynomials $Q_{ji} \in K[u_1, \cdots, u_d, y_1, \cdots, y_{k-1}]$ verifying the relation $(C_{i_0} = I_i)$

$$I_1^{s_{j1}}\cdots I_{k-1}^{s_{j,k-1}} B_j = \sum_{i=1}^{k-1} Q_{ji} A_i.$$

Set $s_i = \max_j (s_{ji})$; we then get

$$I_1^{s_1}\cdots I_{k-1}^{s_{k-1}}(G_1\cdots G_h - DA_k) = \sum_{i=1}^{k-1} Q_i A_i$$

or

$$I_1^{s_1}\cdots I_{k-1}^{s_{k-1}} G_1\cdots G_h = \sum_{i=1}^{k} Q_i A_i,$$

in which Q_i are polynomials in $u_1, \cdots, u_d, y_1, \cdots, y_k$.

From the above it is easy to get the following

Lemma 7. *Let the polynomial set Λ have Φ as basic set, and let the class of term A_i be > 0 and the initial of A_i be I_i, $i = 1, \cdots, n$. Suppose that Φ is reducible, so that there is some k for which the ascending set Φ_{k-1} formed by the first $k-1$ terms of Φ is irreducible with $\bar{\eta}_{k-1} \in K_{k-1}$ as a generic point, while the polynomial got from A_k by substituting $\bar{\eta}_{k-1}$ for the corresponding variables is reducible with an irreducible factorization into polynomials $G_1, \cdots, G_h$. Then there is a decomposition of the form*

$$|\Lambda| = |\Lambda + I_1| \cup \cdots \cup |\Lambda + I_{k-1}| \cup |\Lambda + G_1| \\ \cup \cdots \cup |\Lambda + G_h|.$$

Proof. Any extended zero of either a $\Lambda + I_i$ or a $\Lambda + G_j$ on the right-hand side of the above expression is clearly also an extended zero of Λ. Conversely, any extended zero of Λ is also an extended zero of all A_i's. From the expression just before the lemma it is also an extended zero of some I_i or some G_j, i. e. one of some $\Lambda + I_i$ or $\Lambda + G_j$. This proves the decomposition formula.

Lemma 8. *Let Λ be a polynomial set with Φ as basic set as in Lemma 5 or Lemma*

7. Then the basic set of any polynomial set $\Lambda + I_i$ or $\Lambda + G_j$ appearing in the right-hand side of the decompositions of these lemmas will have its rank lower than that of Φ.

Proof. As each I_i is already reduced with respect to Φ and each G_j is assumed to be reduced with respect to Φ_k and hence also reduced with respect to Φ, the present lemma is an immediate consequence of Lemma 3 of Section 2.

Lemma 9. *Let the polynomial set Λ be irreducible with an irreducible ascending set Φ as its basic set. Suppose also that any polynomial in a polynomial set Λ' or Λ has its remainder 0 with respect to Φ. Then*

$$|\Lambda| \cup |\Lambda'| = |\Lambda'|,$$

or the decomposition $|\Lambda| \cup |\Lambda|'$ is contractible.

Proof. By Lemma 6 we have $|\Omega_\Phi| = |\Lambda|$. By hypothesis any polynomial G' in Λ' has its remainder 0 with respect to Φ. It follows therefore that the generic point of Φ, or the generic point of Ω_Φ, is an extended zero of G', whence $G' = 0/\Omega_\Phi$. Consequently $\Lambda' = 0/\Omega_\Phi$, or $|\Omega_\Phi| \subset |\Lambda'|$, or $|\Lambda| \subset |\Lambda'|$. This proves the lemma.

From the above lemmas and also the preceding section we get the following mechanical procedure for getting the uncontractible irreducible decomposition of a polynomial set.

Let the given polynomial set be Σ. By the well-ordering theorem given in the preceding section, we can, in following some mechanical procedure, successively enlarge the given set Σ to get a sequence of polynomial sets steadily increasing as shown below:

$$\Sigma = \Sigma_1 \subset \Sigma_2 \subset \cdots \subset \Sigma_q = \Lambda.$$

These polynomial sets are actually mutually equivalent, viz.

$$\Sigma = \Sigma_1 \approx \Sigma_2 \approx \cdots \approx \Sigma_q = \Lambda.$$

Two cases may appear. In the first case Λ turns out, in a certain step, to be a contradictory set consisting of a single term which is a non-zero element in K. In this case Σ itself is a contradictory set with no extended zeros. Hence it is only necessary to consider the second case. In that case Λ has a basic set

$$\Phi: A_1, A_2, \cdots, A_n,$$

with $I_1, \cdots, I_n$ as initials and class of $A_1 > 0$. Moreover, Λ will possess the following properties: Any polynomial in Λ will have its remainder 0 with respect to Φ, any extended zero of Σ is also one of Φ, and any extended zero of Φ, if not one of any initial I_i, is also an extended zero of Σ.

Now according to the beginning part of this section, there is some mechanical procedure to verify whether Φ is reducible, or whether A_i's are reducible in the successively extended fields K_{i-1}. We have two subcases again.

In the first subcase Φ is irreducible. By Lemma 5 there is a decomposition

$$|\Lambda| = |\Omega_\Phi| \cup |\Lambda + I_1| \cup \cdots \cup |\Lambda + I_n|,$$

in which Ω_Φ is irreducible while all $\Lambda + I_i$ have some basic sets of ranks lower than that of Λ. We may then consider each $\Lambda + I_i$ as a new polynomial set Σ and proceed again as in the beginning.

In the second subcase Φ is reducible. Then we have by Lemma 7 some decomposition

$$|A| = |A + I_1| \cup \cdots \cup |A + I_{k-1}| \cup |A + G_1| \cup \cdots \cup |A + G_h|,$$

in which each $A + I_i$ or $A + G_j$ has some basic set of a rank lower than that of A. We may then consider each $A + I_i$ or $A + G_j$ as a new polynomial set and proceed again as before.

Whatever the subcase may be, we may take each $A + I_i$ or $A + G_j$ as a new polynomial set Σ' in succession and proceed as before to get a sequence

$$\Sigma' = \Sigma'_1 \approx \Sigma'_2 \approx \cdots \Sigma'_{q'} = A'.$$

In the case that A' has a basic set consisting of a single term which is a non-zero element of the field K, we may remove $|A'|$ or the original $|A + G_j|$ or $|A + I_i|$ from the decomposition. In the contrary case $|A'|$ will be decomposed further into several algebraic varieties with basic sets of rank lower than the preceding ones for the corresponding polynomial set, plus possibly one with corresponding irreducible polynomial set $\Omega_{\Phi'}$ having an irreducible ascending set Φ' as a basic set. In this way we will get a further decomposition of $|A|$ or $|\Sigma|$ itself. In the decomposition there will appear irreducible polynomial sets of the form Ω_Φ, $\Omega_{\Phi'}$ as well as those of the form $A' + I'$ or $A' + G'$. For the latter ones we may decompose them further as before.

As in each step for further decomposition the polynomial sets $A' + I'$ or $A' + G'$ involved have their basic sets of ranks lower than the preceding ones, the decomposition should stop in a finite number of steps owing to the well-ordering theorem of Section 2. Consequently, in a finite number of steps we shall arrive at a decomposition of the following form:

$$|\Sigma| = |\Omega_{\Phi_1}| \cup |\Omega_{\Phi_2}| \cup \cdots \cup |\Omega_{\Phi_t}|,$$

in which each Φ_i is an irreducible ascending set, and Ω_{Φ_i} is the irreducible polynomial set got from Φ_i as described in Theorem 1.

According to the above construction, each $|\Omega_{\Phi_i}|$ cannot be a subvariety of any $|\Omega_{\Phi_j}|$, $j > i$, but we cannot say that some $|\Omega_{\Phi_i}|$ cannot be a subvariety of any $|\Omega_{\Phi_j}|$, $j < i$. This is because we apply only Theorem 1 which asserts the mere existence of Ω_{Φ_i} from Φ_i. If we take into account Theorem 1' which asserts a mechanical procedure for the concrete determination of Ω_{Φ_i} from Φ_i, then we may use Lemma 9 to prove if any $|\Omega_{\Phi_i}|$ is a subvariety of a preceding $|\Omega_{\Phi_j}|$, $j < i$, or not. It follows that, on the basis of Theorem 1', we can get a noncontractible irreducible decomposition of $|\Sigma|$ in a mechanical manner.

In a word, we get finally the following

Theorem 2. *There is a mechanical procedure which permits to determine for a polynomial set Σ, in a finite number of steps, a noncontractible irreducible decomposition of the form*

$$|\Sigma| = |\Omega_{\Psi_1}| \cup \cdots \cup |\Omega_{\Psi_s}|,$$

in which each Ψ_i is an irreducible ascending set of Ω_{Ψ_i}.

For the application to mechanical theorem proving, it is however actually not necessary to carry out the decomposition up to the end to arrive at a noncontractible one. In fact, it is usually sufficient to have an irreducible decomposition which may be a contractible one.

Hence for the applications the existential Theorem 1, but not necessarily the constructive Theorem 1′, will be quite sufficient to meet the purpose.

§4. Proof of the Algebraic Mechanization Theorem

We give below the proof of the Mechanization Theorem in the algebraic form as described in Section 1. For this we first make some preparations.

Given a set of variables $x_1, \cdots, x_N$ arranged in a definite order:

$$x_1 \prec x_2 \prec \cdots \prec x_N,$$

and given a basic field K of characteristic 0 and an ascending set of polynomials in $K[x_1, \cdots, x_N]$,

$$\Phi: A_1, A_2, \cdots, A_n,$$

for which the classes satisfy the relations

$$0 < p_1 < p_2 < \cdots < p_n,$$

we rewrite each x_{p_i} as y_i and the other x's as $u_1, \cdots, u_d$ with $d = N - n$. Then A_i's cen be put in the form

$$A_i = C_{i0} y_i^{m_i} + C_{i1} y_i^{m_i - 1} + \cdots + C_{im_i},$$

in which

$$C_{ij} \in K[u_1, \cdots, u_d, y_1, \cdots, y_{i-1}], \quad i = 1, \cdots, n; j = 0, 1, \cdots, m_i.$$

The initials I_i of A_i are then just the polyaomials $I_i = C_{i0} \in K[u_1, \cdots, u_d, y_1, \cdots, y_{i-1}]$. We call each inequation

$$I_i \neq 0$$

a *non-degeneracy condition*.

Let a polynomial $G \in K[u_1, \cdots, u_d, y_1, \cdots, y_n]$ be given. Construct the remainder R of G with respect to Φ. Then by the remainder formula we have

$$I_1^{s_1} \cdots I_n^{s_n} G = Q_1 A_1 + \cdots + Q_n A_n + R,$$

for certain non-negative integers $s_i \geqslant 0$, with each $Q_i \in K[u_1, \cdots, u_d, y_1, \cdots, y_n]$.

We shall investigate the necessary and sufficient conditions such that

$$G = 0$$

may be deduced as a consequence of the equations $A_i = 0$, $i = 1, \cdots, n$. We shall prove that, under the subsidiary non-degeneracy conditions $I_i \neq 0$ and under the hypothesis that Φ is irreducible, the necessary and sufficient condition is just $R = 0$. Whether the set Φ is irreducible or not, the sufficiency of the condition is quite evident from the above remainder formula. So we have the following

Theorem 1. *Let Φ, A_i, I_i, G be as above and $R = 0$; then under the non-degeneracy conditions*

$$I_i \neq 0, i = 1, \cdots, n,$$

$G = 0$ is a consequence of $A_i = 0$, $i = 1, \cdots, n$, whether Φ is reducible or not.

If Φ is irreducible, under the non-degeneracy conditions for $G=0$ to be a consequence of $A_i=0$, $i=1, \cdots, n$, the condition $R=0$ is not only necessary but also sufficient, as in the following theorem which follows directly from Lemma 3 in Section 3.

Theorem 2. *Let Φ, A_i, I_i, G be as above and Φ be irreducible. If under the non-degeneracy conditions $I_i \neq 0$ the equation $G=0$ is a consequence of the equations $A_i=0$, $i=1, \cdots, n$ (for a certain extension field of K), then the remainder R of G with respect to Φ is 0.*

Remark. The proofs of these theorems depend very much on the theory developed in Section 3 and are rather involved. If we restrict ourselves to real field as is the case of ordinary Euclidean geometry and pay no attention to the constructive aspects, then the proofs will be much simpler.

We now give the proof of the Mechanization Theorem of unordered geometries in its algebraic form.

Given a geometrical statement (S) in a certain unordered geometry, our object is to give a mechanical method to decide whether (S) is true or not. For this purpose we choose first a coordinate system, express the points involved by coordinates, denote these coordinates by x_i, and arrange them in a certain definite order:

$$x_1 \prec x_2 \prec \cdots \prec x_N.$$

Next we translate the various geometrical relations in the statement (S) into algebraic relations of these coordinates. Then the hypothesis in the statement (S) will be translated into a system of equations

$$F_1=0, \ \cdots, \ F_s=0,$$

in which F_i are polynomials in $K[x_1, \cdots, x_N]$, with K the basic field of characteristic 0 associated to the geometry in question. Actually all these polynomials are with rational or even integer coefficients. The conclusions of the statement (S) will then be turned into another system of equations

$$G_1=0, \ \cdots, \ G_t=0,$$

with all G_j being polynomials in $K[x_1, \cdots, x_N]$, also with rational or integer coefficients. Without loss of generality we may suppose that there is only one such polynomial G_j, denoted simply by G henceforward. The polynomials F_i are then called *hypothesis polynomials* of the statement (S), and the G_j's or G the *conclusion polynomial(s)* of (S).

The proof of the Mechanization Theorem consists in exhibiting a mechanical procedure which permits to determine first in a finite number of steps a set of polynomials $D_1, \cdots, D_r$ for *non-degeneracy conditions*, with all D_k in $K[x_1, \cdots, x_N]$, which will actually be all with rational or even integer coefficients. Secondly the same mechanical procedure will also permit to decide in a finite number of steps whether under the non-degeneracy conditions

$$D_1 \neq 0, \ \cdots, \ D_r \neq 0,$$

the equation $G=0$ will be a consequence of $F_1=0, \cdots, F_r=0$.

With the language of algebraic geometry, the proof of Mechanization Theorem can also be restated in an alternative form in the following manner:

Denote the set of hypothesis polynomials F_i by $\Sigma=\{F_i\}$. The set Σ defines an algebraic variety $|\Sigma|$, with dimension d, viz., the dimension of any characteristic set of Σ. The proof of the Mechanization Theorem consists then in exhibiting a mechanical procedure which permits to determine a set of polynomials $D_1, \cdots, D_r$ such that in adjoining each D_i to Σ, the resulting polynomial set $\Sigma+D_i$ will define an algebraic variety $|\Sigma+D_i|$ of dimension $< d$. Furthermore, the same procedure will permit to decide, under the non-degeneracy conditions $D_1 \neq 0, \cdots, D_r \neq 0$, whether $G=0$ or not; in other words, whether G will be 0 or not on the remaining part of the algebraic variety $|\Sigma|$ after removal of the true subvarieties $|\Sigma+D_i|$.

As briefly indicated in Section 3, we can decompose the algebraic variety into irreducible components, each of which has an irreducible basic set Φ_i which determines in turn that irreducible component in question, denoted by $|\Omega_{\Phi_i}|$. Furthermore, in the case the dimension d_i of $|\Omega_{\Phi_i}|$ is less than the dimension d of $|\Sigma|$, then this true subvariety is got from a certain previous $|\Omega_{\Phi_j}|$ by adjoining to Φ_j some polynomial D_i which is either an initial I_k or some G_i in the previous notations and $|\Omega_{\Phi_i}|$ is a subvariety of $|\Phi_j+D_i|$. We take each such D_i as a non-degeneracy polynomial. Suppose after removal of all these true subvarieties, the remaining irreducible components of dimension d are

$$|\Omega_{\Phi_1}|, \cdots, |\Omega_{\Phi_t}|.$$

Denote the initials of each Φ_j by $I_{j1}, \cdots, I_{jk}$ and consider them also as non-degeneracy polynomials D_{jk}. Now whether $G=0$ is a consequence of $F_1=0, \cdots, F_r=0$ under the non-degeneracy conditions $D_i \neq 0$, $D_{jk} \neq 0$, is just the same as whether $G=0$ on the remaining parts of $|\Omega_{\Phi_1}|, \cdots, |\Omega_{\Phi_t}|$ after removal of the components $|\Phi_j+D_i|$ and those defined by $D_{jk}=0$. By Theorems 1, 2 above this can be decided by whether the remainders of G with respect to Φ_j are all 0. It furnishes the mechanical procedure required and thus gives the proof of the Mechanization Theorem in question.

The above mechanical procedure of theorem-proving is theoretically quite simple in appearance. However it would be quite difficult to apply this method to the proof of concrete theorems. The reason is that the irreducible decomposition of algebraic varieties depends on factorization of polynomials which, though theoretically almost self-evident, is a rather difficult problem in practice for which no method of high efficiency is available even up to now. Consequently, the above method is entirely non-feasible in practice. Fortunately, for the theorem-proving in geometries, we usually hope that the theorem in question is really a true theorem and we hope to prove it true in an affirmative manner. For this purpose it is enough to prove, by Theorem 1, that the remainder of the conclusion polynomial G is 0 with respect to some ascending set,whether irreducible or not. Therefore, to each concrete theorem whose truth is to be tested and to be proved in the case it is really true, we may apply Theorem 1 directly. If by computation we know that G has its remainder 0 with respect to the ascending set, then the theorem in question is true and the computation furnishes actually a proof of this theorem. In this case everything is done. Only in the case that the remainder is not 0 should we ask further whether the corresponding ascending set is irreducible or not. For this reason we shall modify the above mechanical procedure of proof to the following form which has been proved to be very efficient in practice (some examples will be given in the next section).

The modified mechanical procedure runs somewhat as follows.

Consider a set $\mathscr{P}$ of polynomial sets and a set Δ of polynomials, where Δ is called the *degeneracy set*. In the outset, $\mathscr{P}$ will consist of a single polynomial set, viz. the set of hypothesis polynomials

$$\Sigma = \{F_1, \cdots, F_s\},$$

and the degeneracy set will be an empty one, viz.

$$\Delta = \emptyset.$$

During the procedure we shall increase or decrease the number of polynomial sets in $\mathscr{P}$ and also adjoin non-degeneracy polynomials into Δ to get the final

$$\Delta = \{D_1, \cdots, D_r\}$$

as required.

Step 1. Let $\mathscr{P}$ be non-empty. Then take arbitrarily a polynomial set Σ from $\mathscr{P}$, and remove it from $\mathscr{P}$ to get a new $\mathscr{P}$. Using the well-ordering theorem in Section 2 to enlarge Σ to successive polynomial sets as shown below:

$$\Sigma = \Sigma_1 \subset \Sigma_2 \subset \cdots \subset \Sigma_q = A.$$

If A has an element which is a non-zero number in K, then A is a contradictory set. In this case the hypothesis in the statement (S) is contradictory in itself and the procedure will be stopped. In the contrary case let the basic set of A be

$$\Phi: A_1, A_2, \cdots, A_n.$$

The initials of A_i will be denoted by I_i. By construction, any polynomial in A except A_i will have its remainder 0 with respect to Φ. In that case we have also

$$\dim|\Sigma| = \dim\Phi = N - n = d.$$

If Step 1 is just the first step from the very beginning of the whole procedure, then the dimension d will be recorded for future reference.

If Step 1 is the successive step from the other ones during the procedure, then we compare the new dimension d with the previous d recorded in the beginning.

If this new $d =$ the previously recorded d, then we adjoin the initials I_i to Δ to get some enlarged new degeneracy set Δ, and proceed to Step 2.

If this new $d <$ the previously recorded d, and the present Σ is obtained as some $A + I_i$ or $A + G_j$ during Step 3 below, then we adjoin this I_i or G_j to Δ to get a new Δ. We then return to Step 1 and proceed as before.

Step 2. Find the remainder R of G with respect to Φ.

Suppose $R = 0$. If in $\mathscr{P}$ there is not any more polynomial set, then the statement (S) is true under the non-degeneracy conditions

$$D_k \neq 0 \quad (D_k \in \Delta),$$

and the procedure will be stopped. In this case the theorem is true and is proved under the non-degeneracy conditions. Otherwise we return to Step 1 and proceed again as before.

Suppose $R \neq 0$. Then we proceed to Step 3.

Step 3. Check the irreducibility of the basic set Φ.

Suppose that Φ is irreducible. Then as G has its remainder $\neq 0$ with respect to Φ, by Theorem 2 under the non-degeneracy conditions

$$D_k \neq 0 \qquad (D_k \in \Delta)$$

statement (S) is not true; the procedure will then be stopped. In this case the theorem is not true under the above non-degeneracy conditions.

Suppose that Φ is reducible. Then there will be some decomposition

$$|\Lambda| = |\Lambda + I_1| \cup \cdots \cup |\Lambda + I_{l-1}| \cup |\Lambda + G_1| \cup \cdots \cup |\Lambda + G_k|.$$

Consider such $\Lambda + I_i$ and $\Lambda + G_j$ as new polynomial sets Σ, and adjoin all these to $\mathscr{P}$ to get a new enlarged set $\mathscr{P}$. Then return to Step 1 and proceed again as before.

According to the previous sections, the above procedure should stop in a finite number of steps. In this way we get a final degeneracy set

$$\Delta = \{D_k\}$$

and one of the following three conclusions should be true:

1) Under the non-degeneracy conditions

$$D_k \neq 0 \qquad (D_k \in \Delta)$$

the hypotheses in the statement (S) are contradictory in themselves.

2) Under the above non-degeneracy conditions, or under the additional hypothesis $D_k \neq 0$, the statement (S) is true, or, what is the same, the theorem in question is true.

3) Under the above non-degeneracy conditions, or under the additional hypothesis $D_k \neq 0$, the statement (S) is not true, or, what is the same, the theorem is not true.

Generally speaking, the degeneracy conditions

$$D_k = 0$$

are not worth any more consideration. If there is some necessity to consider such a degeneracy condition $D_k = 0$, we may simply take it as a new hypothesis to be adjunct to the statement (S), i.e., we consider $\{F_1, \cdots, F_r, D_k\}$ instead of $\{F_1, \cdots, F_r\}$ and then proceed as above.

The above mechanical procedure is very feasible. We have implemented it on small computers, proving and thus also discovering quite non-trivial theorems in this way. The next section will describe a few illustrative examples.

§5. Programming and Examples.

It is clear how to program according to the procedure described in the preceding sections. In fact, programming has been done and various theorems have been proved on rather small computers. Before we explain certain theorems proved in this way, let us first add some remarks.

First, we may lessen the labour of computation by modifying slightly the definition of the basic set and characteristic set. Thus, we shall define an ascending set

$$\mathscr{A}: A_1, A_2, \cdots, A_r$$

to be one *in loose sense* or *in weak sense* in requiring only that each A_i in the set be reduced merely with respect to the variables occuring in the leading term of A_i alone. The notions of basic set, etc. derived in this way are then also said to be in *loose sense or weak sense.* This will not affect the final conclusions but will greatly simplify the programming and the computation Thus, the polynomial set corresponding to the hypothesis of a theorem in the ordinary geometry is usually already in the form of an ascending set and hence also a basic set in the above loose or weak sense. In the worse case a few strokes of simple hand computations may be required. The procedure of well-ordering is not necessary in general because it is quite laboursome.

Secondly, we are only interested in arriving at *true* theorems so that only the sufficiency part of our criterion will be considered in the programming. Thus, if the remainder of the conclusion polynomial with respect to the hypothesis polynomial set, supposed already a basic set in loose sense, is zero, then the theorem is *true* generically under the non-degeneracy conditions furnished by the initials of the hypothesis polynomials and we have achieved our aim. Only in the case of non-zero remainders is the truth of theorem doubtful, and further investigations about the reducibility of the polynomials may then be required.

Finally, we remark that though the hypotheses as well as the conclusion polynomials usually have only a few terms, the polynomials got successively during the reduction in the determination of the remainder may rise up quickly to hundreds and thousands of terms. To avoid the appearance of this phenomenon the following *branching* device has been adopted in our programming. Thus, let some polynomial g of the form (m_p = degree in y_p of A_i of class p in $\mathscr{A}$)

$$g = g_0 y_p^{m_p-1} + g_1 y_p^{m_p-2} + \cdots + g_{m_p-1},$$

in which each g_i is of class $< p$, appear during the successive reduction of the conclusion polynomial. Then, instead of verifying further whether the remainder of g with respect to $\mathscr{A}$ is zero, we may verify this for each g_i in turn. Furthermore, we shall use an *index set* $[TCD]$ to indicate the complexity of a polynomial, where T is the number of terms, C the class, and D the degree in the leading variable y_C of the polynomial. The successive reduction of the conclusion polynomial up to the final remainder which constitutes in fact a *proof* of the theorem in the case of zero remainder may then be clearly shown by a *flowing chart* of the index sets. As a simple example, with suitable coordinates the well-known Pappus Theorem will have 6 hypothesis polynomials already in the form of a basic set in the loose sense whose index sets are:

$$[4\ \ 7\ \ 1],\ [3\ \ 8\ \ 1],\ [4\ \ 9\ \ 1],\ [3\ \ 10\ \ 1],\ [4\ \ 11\ \ 1],\ [4\ \ 12\ \ 1].$$

The conclusion polynomial has an index set [6 12 1] and the flowing chart of the reductions, as done on a computer, runs as follows:

$$[6\ \ 12\ \ 1] \longrightarrow [8\ \ 11\ \ 1] \longrightarrow [12\ \ 10\ \ 1] \longrightarrow [16\ \ 9\ \ 1] \longrightarrow$$
$$[18\ \ 8\ \ 1] \longrightarrow [16\ \ 7\ \ 1] \longrightarrow 0.$$

The final zero means that the theorem is true (of course generically only) and is proved with the above running chart as a proof. Remark that different choices of coordinates will give rise to different running charts which correspond to different proofs.

We have applied our program to the proof of various famous theorems in the ordinary geometry: theorems of Kcukou, Pappus, Pascal, Simson, Feuerbach, Morley, etc. Perhaps the proof of the theorem of Morley is the most difficult and is quite instructive in itself. So let us state the theorem in full below.

Theorem of Morley. *For a triangle $A_1A_2A_3$ the neighbouring trisectors of the three angles of the triangle will intersect to form 27 triangles in all, of which 18 are equilateral.*

In appearance this theorem is out of the reach of our method which works only for unordered geometries without notion of order or only for theorems not involving order relations in an ordered geometry. Thus, in an unordered geometry, there is no notion of rays and an angle cannot be defined in the usual way as two rays emanating from a common point. However, we can define an angle $\angle(l_1, l_2)$ simply as an ordered pair of lines l_1, l_2, and attribute a magnitude $T(l_1, l_2)$ to it corresponding to the tangent function of the angle in the case of ordinary geometry.

We may now define a bisector of the angle $\angle(l_1, l_2)$ in the unordered geometry as a line such that the *reflection* (well-defined in the geometry) of l_1 with respect to t is just l_2. If l_1, l_2 intersect, then t is a line through the intersecting point such that $T(t, l_1) = T(l_2, t)$ corresponding to the ordinary formula $\angle(t, l_1) \equiv \angle(l_2, t)$ or $2\angle(t, l_1) \equiv \angle(l_2, l_1) \operatorname{mod} \pi$. However, in the unordered geometry there may exist two such bisectors for an angle and there is no means to distinguish these two bisectors.

Similar ambiguity occurs for trisectors of an angle. To fix the ideas, let us call a line t a *primary trisector* of an angle $\angle(l_1, l_2)$ if a formula in T holds which corresponds to the ordinary formula $3\angle(t, l_1) \equiv \angle(l_2, l_1) \operatorname{mod} \pi$. There are 3 such primary trisectors which there is no means to distinguish from each other. To each such primary trisector t however is uniquely associated a *secondary trisector* t' such that $T(l_2, t') = T(t, l_1)$.

Consider now a triangle $A_1A_2A_3$. Let t_1 be any one of the primary trisectors of the angle $\angle(A_1A_2, A_1A_3)$ at vertex A_1 with associated secondary trisector t'_1. Similarly let t_2, t'_2 be a primary and an associated secondary trisector of the angle $\angle(A_2A_3, A_2A_1)$ and t_3, t'_3 be those of the angle $\angle(A_3A_1, A_3A_2)$. Let t_1, t'_2 intersect at a point A_4, in notation $A_4 = t_1 \wedge t'_2$. Similarly let $A_6 = t_2 \wedge t'_3$, $A_5 = t_3 \wedge t'_1$. The triangles $A_4A_5A_6$ are clearly 27 in all. The Morley theorem asserts that 18 among them are equilateral.

First of all we have to settle the problem how the 18 triangles should be chosen. For this let us denote by θ an angle for which the T-value has square $= 3$. In ordinary geometry this means $\theta = \pm\frac{\pi}{3} \operatorname{mod} 2\pi$. Remark in passing that in an unordered geometry it is not legitimate to speak about $+\sqrt{3}$ or $-\sqrt{3}$. Now we choose the primary trisectors t_1, t_2, t_3 such that some relation in the T-values corresponding to the ordinary formula

$$\angle(t_1, A_1A_2) + \angle(t_2, A_2A_3) + \angle(t_3, A_3A_1) \equiv \theta \operatorname{mod} 2\pi$$

holds true. Under this condition the number of triangles $A_4A_5A_6$ is then reduced to 18 which will be proved to be all equilateral.

Adopting now a certain coordinate system with coordinates of various points and the T-

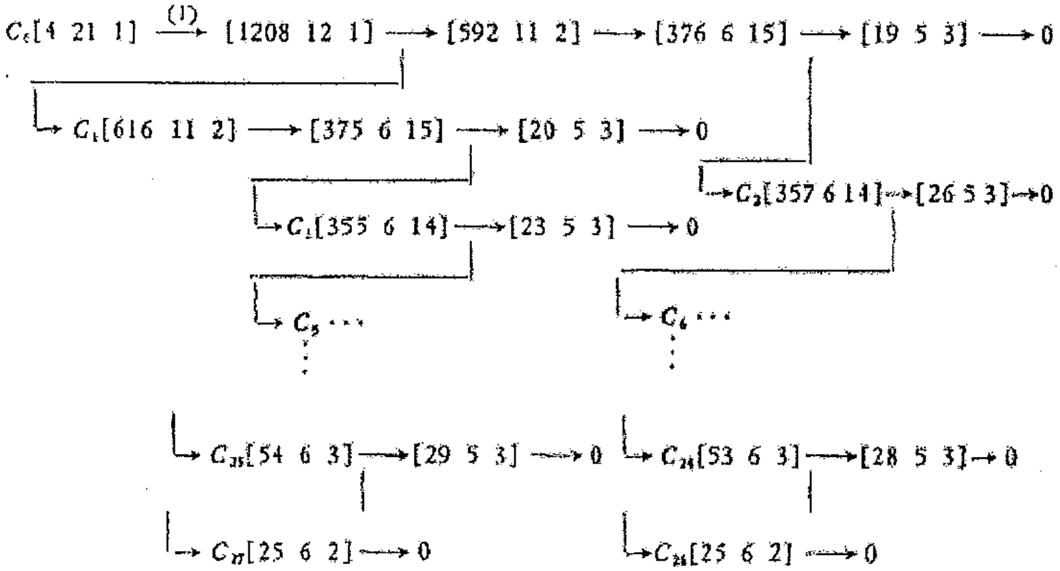
values of various angles involved in the theorem as $x_i's$ arranged in a certain definite order, we shall get a set of hypothesis polynomials H_i, 18 in number, and a certain conclusion polynomial g. Without entering the details we merely list the index sets of various polynomials below:

For hypothesis-polynomials:

[2 3 1], [3 4 1], [4 5 1], [3 7 1], [3 8 1], [4 9 1], [3 10 1],
[2 11 1], [2 12 2], [8 13 1], [4 14 1], [4 15 1], [4 16 1],
[2 17 1], [5 18 1], [3 19 1], [4 20 1], [4 21 1].

For conclusion-polynomial: [4 21 1].

To verify the theorem by means of our program we remark that separation will occur when we come to the point after the reductions with respect to H_9 and H_4. The following is a rough scheme about the successive reductions with index set of successive polynomials indicated.


$C_0[4\ 21\ 1] \xrightarrow{(1)} [1208\ 12\ 1] \longrightarrow [592\ 11\ 2] \longrightarrow [376\ 6\ 15] \longrightarrow [19\ 5\ 3] \longrightarrow 0$
$C_1[616\ 11\ 2] \longrightarrow [375\ 6\ 15] \longrightarrow [20\ 5\ 3] \longrightarrow 0$
$C_3[357\ 6\ 14] \longrightarrow [26\ 5\ 3] \longrightarrow 0$
$C_2[355\ 6\ 14] \longrightarrow [23\ 5\ 3] \longrightarrow 0$
$C_5 \cdots$
$C_6 \cdots$
$C_{25}[54\ 6\ 3] \longrightarrow [29\ 5\ 3] \longrightarrow 0$
$C_{24}[53\ 6\ 3] \longrightarrow [28\ 5\ 3] \longrightarrow 0$
$C_{27}[25\ 6\ 2] \longrightarrow 0$
$C_{26}[25\ 6\ 2] \longrightarrow 0$

Remark that each arrow in the above scheme consists of a number of successive reductions. For example, the arrow marked (1) is detailed as follows.

$$C_0[4\ 21\ 1] \longrightarrow [8\ 20\ 1] \longrightarrow [4\ 19\ 1] \longrightarrow [18\ 18\ 1] \longrightarrow [36\ 17\ 1]$$
$$\longrightarrow [36\ 16\ 1] \longrightarrow [66\ 15\ 1] \longrightarrow [132\ 14\ 1] \longrightarrow [236\ 13\ 2] \longrightarrow [832\ 13\ 1]$$
$$\longrightarrow [1960\ 12\ 3] \longrightarrow [1208\ 12\ 1].$$

Thus a certain polynomial of 1960 terms occurs in the whole procedure of reductions. If we do not adopt separation devices at convenient places in selecting suitable coordinate systems and coordinates of points, the polynomials during the procedure may quickly grow too large to be admitted even by a big computer. For the present case as all remainders (28 in all) are zero, the Morley theorem is true and the above scheme furnishes such a proof of the theorem.

We add finally that the above scheme shows that we have indeed proved a theorem a little more general than the original one. For the same proof holds also in the case of certain unordered geometries like complex geometries, for example. In such geometries isotropic lines may exist. However, if we restrict our theorem so that no isotropic lines are involved in the statement, then the mechanical proof applies still.

As a further example let us consider the problem of determining all triangles ABC with two equal bisectors t_A and t_B of angles A and B. It is well-known, but is quite non-trivial to prove, that the triangle ABC should be isoceles ($AC = BC$) if the two equal bisectors in question are both *internal* ones. Mr. S. C. Chou has raised the question of proving this fact by the mechanical theorem-proving method. Now it is easy to see that $AC = BC$ would not be true (generically) if one of the bisectors t_A, t_B is an *internal* and the other is an *external* one. Chou and I have tried on the computer and found the rather unexpected result that $AC = BC$ is still not true if the equal bisectors are both *external* ones.

In principle the above problem is again out of reach of our method. However, in view of the nature of the problem that the order relations only enter the hypothesis but not the conclusion at all, our method in combination with that of Seidenberg in reducing inequalities to equalities by introducing new auxiliary variables will lead to some information about the final results to be found. Thus, let us denote by AE and BD the two equal bisectors in question and by I their point of intersection. Take coordinates with

$A = (-1, 0)$, $B = (+1, 0)$, $I = (x_2, x_3)$, $C = (x_{12}, x_{13})$, etc.

Denote also the slopes of AE, BD by x_4, x_5, etc. Introduce a further auxiliary variable x_1 by setting

$$x_4x_5 = -x_1^2, \tag{1}$$

or

$$x_4x_5 = +x_1^2. \tag{2}$$

Equation (1) means that AE, BD are either both internal or both external bisectors which will be distinguished by either

$$x_3x_{13} > 0,$$

or

$$x_3x_{13} < 0.$$

On the other hand equation (2) means that one of AE, BD is an internal while the other is an external bisector.

Consider e.g. the case of equation (1). From the hypothesis including the equality of bisectors we get on running the program a set of equations, with extraneous factors corresponding to degenerate cases already removed, as follows:

$$x_1f(x') = 0, \tag{3}$$

with

$$f(x) = (1 - x_1^2)(x_1^2 - 1)^2(x_1^2 - 2) - 4, \tag{4}$$

$$x_3^2 = x_1^2(1 - x_1^2), \tag{5}$$

$$(1 - x_1^2)x_3x_{13} = 2x_3^2, \tag{6}$$

etc.

Equation (5) shows that in the non-degenerate case we have

$$x_2^2 < 1. \tag{7}$$

Equation (6) shows that we have

$$x_1^2 < 1 \text{ or } > 1$$

according as the two bisectors AE, BD are both internal ones or both external ones.

Suppose first $x_1^2 < 1$. Then from (4) we see that $f(x) < 0$. From (3) it follows that we have necessarily

$$x_3 = 0.$$

This just proves the classical theorem that a triangle with two equal *internal* bisectors is isosceles.

Suppose next $x_1^2 > 1$ so that the two bisectors are both *external* ones. Then $f(x) = 0$ will have positive roots of x_1^2 for $x_2^2 < 1$ so that there are an infinity of *non-isosceles* triangles ABC with equal *external* bisectors AE, BD for which the corresponding point $I(x_2, x_3)$ will lie on a certain oval defined by the following equation together with (7):

$$x_3^6 - 4x_3^4(1 - x_2^2) + 5x_3^2(1 - x_2^2)^2 - 2(1 - x_2^2)^3 - 4(1 - x_2^2)^2 = 0.$$

The case of equation (2) or the case of one internal and one external bisector can be treated in entirely the same manner. We find thus infinities of non-isosceles triangles with equal bisectors one internal and one external for which the corresponding points I will lie on two ovals defined by the same equation above with the restriction $x_2^2 > 1$. The problem raised above is thus completely settled.

We have also applied our method to the mechanical theorem discovering of "new" theorems in ordinary geometry. Several theorems have been discovered in this way. We shall illustrate below.

Ex. Pascal-Conic Theorem

Suppose we are given 6 points $A_1, \cdots, A_6$ on the same conic. Let us call any point of intersection $A_iA_j \cap A_kA_l$ (for i, j, k, l mutually unequal) a *Pascal point*. Such Pascal points are 45 in all which lie three by three on 60 so-called *Pascal lines*. These points and lines constitute a configuration which has been much studied by numerous geometers including Steiner, Staudt, Cayley, Kirkmann. However, most of the interesting theorems found by them are of a *linear* character: collinearity of certain points and concurrency of certain lines. Now we put the following problem: What theorems of a *quadratic* character can be found about this configuration? In particular, we ask what combinations of 6 among the 45 Pascal points will lie on the same conic (*co-conic* for short). Of course we are only interested in such combinations of 6 Pascal points lying on some conic not degenerated into two Pascal lines.

The problem will be studied with further specialization. Consider for example a permutation $s = (123456)$ which will act on the 45 Pascal points. We now ask for what Pascal points P the six points P, sP, $s^2P, \cdots, s^5P$ will lie on some non-degenerate conic. By trials we see that the only possible points are $A_1A_3 \wedge A_2A_5$ or the equivalent ones. Assuming that the usual Pascal theorem is known, then this amounts to whether the hexagons formed of the six points s^iP, $i = 0, 1, \cdots, 5$, are Pascalian or not, i.e., whether the three points of intersection of the opposite sides of the hexagons are collinear or not. Formulating the theorem to be proved in this way we verified again on the computer that this is really the case. So we get a number of non-degenerate conics on each of which lie 6 Pascal points. We call these conics the *Pascal conics* and the theorem thus discovered the *Pascal-Conic Theorem*. It was first discovered in 1980 and verified on an HP9835A.

Of course it is very likely that the theorem is known already in the last century. Moreover, simple and elegant proof may also be easily found for this theorem. However, these are neither of any interest nor of any importance to us from the point of view of mechanical theorem proving. The example shown may well indicate the powerfulness in discovering really non-trivial new theorems in various kinds of geometries besides the ordinary geometry, e.g. the non-Euclidean geometries, the circle geometries, or geometries of even more modern nature, in which known interesting theorems are rare. Even in the case of Pascal configurations we may put forward some problems to which our method may give some answer: Are there other conics through at least 6 of the Pascal points or touching at least 6 Pascal lines besides those found above? Are there any interesting geometrical relations between these conics and the various Pascal points, Pascal lines and other known points and lines of significance ? Are there also cubic relations between the 45 Pascal points, i.e., are there non-degenerate cubics passing through at least 9 out of the 45 Pascal points, etc. Of course innumerable problems can be set forth in this way.

References

[1] Hilbert, D., Grundlagen der Geometrie, Teubner, 1899.

[2] Ritt, J. F., Differential equations from the algebraic standpoint, Amer Math. Soc. 1932.

[3] Ritt, J. F., Differential algebra, Amer. Math. Soc., 1950.

[4] Wu Wen-tsün, On the decision problem and the mechanization of theorem proving in elementary geometry, *Scientia Sinica*, 21 (1978), 159—172.

初等几何定理机器证明的基本原理

吴文俊

（中国科学院系统科学研究所，北京）

摘　要

1976 与 1977 之交，我发现了一个初等几何定理证明的机械化方法，见文献 [4]. 这一方法适用于各种无序的但满足 Pascal 公理的初等几何，或各种初等几何中不牵涉次序关系的那类定理.本文 § 4 叙述了这一方法所依据的基本原理并给出了详细证明. 在 § 2 与 § 3 中则阐述了基本原理所依赖的关于多项式组的整序理论与代数簇的构造性理论. 二者俱源出 Ritt 的著作，见文献 [2, 3]. 最后在 § 5 中以 Morley 定理与我所发现的 Pascal 锥线定理为例，说明这一方法在计算机上实施的具体情况.