Randomized Zero Testing of Radical Expressions and Elementary Geometry Theorem Proving

Daniela Tulone[1]*, Chee Yap[2]**, and Chen Li[2]

[1] Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974
daniela@research.bell-labs.com
[2] Department of Computer Science, Courant Institute, New York University,
251 Mercer Street, New York, NY 10012
{yap,chenli}@cs.nyu.edu

Abstract. We develop a probabilistic test for the vanishing of *radical expressions*, that is, expressions involving the four rational operations $(+, -, \times, \div)$ and square root extraction. This extends the well-known Schwartz's probabilistic test for the vanishing of polynomials. The probabilistic test forms the basis of a new theorem prover for conjectures about ruler & compass constructions. Our implementation uses the `Core Library` which can perform exact comparison for radical expressions. Some experimental results are presented.

1 Introduction

Several approaches to proving theorems in Elementary Geometry using constructive methods in Computer Algebra were proposed in the 1980s [7]. These were much more successful than earlier approaches based on purely logical or axiomatic approaches. Thus, Kutzler, Stifter [14] and Kapur [12] proposed methods based on Gröbner Bases. Carrà and Gallo [1, 8] devised a method using the dimension underlying the algebraic variety. Hong [11] introduced semi-numerical methods ("proof by example" techniques) based on gap theorems. An acclaimed approach in this area is due to Wu [21, 23, 22] who applied the concept of characteristic sets to geometric theorem proving. Extensive experimentation with Wu's method were reported by Chou [3, 5].

All these algebraic approaches begin by translating the geometric statements into algebraic ones. A proposed geometry theorem (also called a *conjecture*) is translated algebraically into two parts: a system H of multivariate polynomials called the *hypothesis*, and a single polynomial T called the *thesis*. The conjecture is true if the vanishing of the hypothesis system implies the vanishing of the thesis polynomial. From the viewpoint of algebraic geometry, proving the conjecture amounts to showing that $Var(H) \subseteq Var(T)$ where $Var(S)$ is the algebraic variety defined by a set S of polynomials. This basic formulation must be refined in order to handle degeneracy conditions.

* This work was performed while Daniela Tulone was at NYU.
** This work is supported in part by NSF Grant #CCR 9402464

Wu's "basic method" computes the pseudo-remainder of the polynomial thesis with respect to the Wu-Ritt extended characteristic set of the hypotheses system. If the pseudo-remainder vanishes, then the conjecture is true provided the initials of the extended characteristic set do not vanish. Wu's basic method has been successfully used to prove many classical and some new theorems in plane analytic geometry. The basic method fails if the variety $Var(H)$ is reducible. To handle this, Wu's "complete method" begins by decomposing $Var(H)$ into irreducible components and applying the basic method to each component. A drawback in Wu's method is that it works with an algebraically closed field. In particular, it is not a complete method for the real algebraic varieties. The present paper addresses a special case of real algebraic varieties.

Gröbner bases methods can be doubly exponential in the worst case [17, 24]. The complexity for Wu's method is somewhat better but remains an issue. To circumvent the high complexity, we investigate probabilistic methods [20] combined with "proof by example" techniques [11]. In probabilistic theorem proving, we do not prove the validity of a conjecture in the classical sense. Instead, we either prove the invalidity of a conjecture (by showing a counter example) or else classify the conjecture as "true with the high probability $1-\varepsilon$". This latter classification must be properly understood since, classically, it is nonsense to say that a theorem is true with some probability. What is meant is that, relative to a set of experiments we conduct, the probability that the conjecture is false *and* we failed to discover this, is less than ε.

An interesting approach along these lines was given by Carrà, Gallo and Gennaro [2]. They applied the Schwartz-Zippel [20, 27] probabilistic test for the vanishing of pseudo-remainders in Wu's method. They considered conjectures in the classical setting of *ruler & compass constructions*. Such conjectures are examined by testing the vanishing of Wu's pseudo-remainder for randomly chosen examples. Each example is specified by a random choice of values for its parameters. The random choices come from some suitable *test set* whose cardinality depends on the degree of the pseudo-remainder. The extended characteristic set as well as the pseudo-remainder are computed. If the pseudo-remainder is zero, then the example is successful; otherwise, as in Wu's method, further investigation is called for. While implementing their method, one of us (D.T.) discovered a serious efficiency issue. The degree of the pseudo-remainder is very high: if the conjecture involves C ruler & compass construction steps, then, the degree of the pseudo-remainder in [2] (following [9, 10]) has the following bound:

$$D = 2^{O(C^3)} C^{O(C^2)} \quad .$$

The cardinality of the test set is $2D$, which is too large in practice. This bound applies to the test for "generic truth". For "universal truth", D can be improved to $2P \cdot 3^{C+1}$ where P is the number of points in the construction. Unfortunately, practically no classical theorems are universal truths.

Summary of New Results. (1) We develop an extension of the Schwartz-Zippel probabilistic zero test. While the Schwartz-Zippel test is applicable to polynomials, we treat radical expressions by admitting the additional operations of division and square-roots. This adds considerable complexity to the proofs. Furthermore, for efficiency considerations, we use straight line programs to represent radical expressions. The asymptotic time complexity of our probabilistic test is a low-order polynomial. Since radical expressions are common in many applications, we expect this new test to be generally useful.

(2) We address the problem of computer proofs of geometric conjectures about ruler & compass constructions. The zero test of radical expressions is tailor fitted for this problem. Moreover, we combine randomness with the numerical approach of Hong to give additional efficiency. Thus, our approach appears to be intrinsically more efficient than previous general approaches (e.g., Wu's or Gröbner bases).

(3) Our prover is implemented using the `Core library` [15, 13, 19]. This is an unexpected application of our library, which was designed as a general `C++`-package to support the Exact Geometric Computation [26, 25] approach to robust algorithms. Preliminary experimental results are quite promising. We expect further improvements by fine-tuning our library for this specific application. Our prover is currently distributed with version 1.3 of the `Core library` (Aug. 15, 2000) and available from `http://cs.nyu.edu/exact/core/`.

Overview. The paper is organized as follows: Section 2 gives an overview of geometric conjectures about ruler & compass constructions. Section 3 gives our extension of Schwartz's probabilistic test to radical expressions. Section 4 addresses the application of our new probabilistic test to theorem proving. We conclude in Section 5.

2 Theorem Proving for Ruler & Compass Constructions

We follow the algebraic approach which has been well-summarized by Chou [5]. Ruler & compass operations may be seen as constructing lengths, points, lines and circles, collectively called *geometric objects*. A collection of such geometric objects will be called a *geometric scene*. We consider geometric scenes that are constructed incrementally using ruler & compass operations. The algebraic analogue of constructing a geometric object O amounts to introducing a pair of variables (x, y) and corresponding polynomial equations $h_i(x, y, z, \ldots)$ $(i = 1, 2, \ldots)$ that must be satisfied if (x, y) lies on O. Here, h_i may involve other variables $z, \ldots$, from previously constructed objects. We shall classify the *variables* introduced by our constructions into two sorts: *independent* and *dependent* variables. For short, the independent variables will be called *parameters*. It is instructive to give a concrete example (Figure 1 from [5]).

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

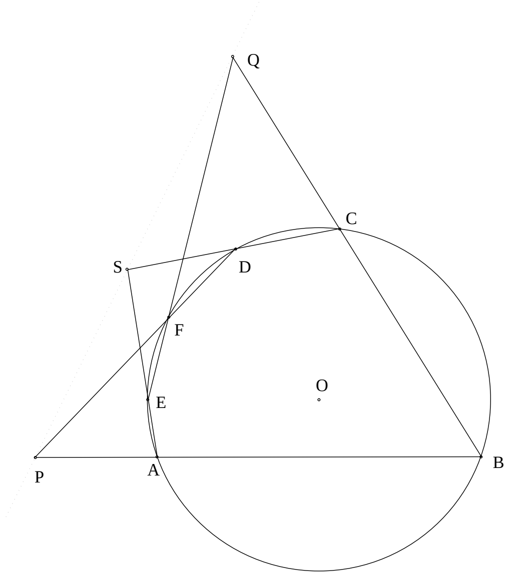

Fig. 1. Pascal's Theorem.

Let $A = (0,0)$, $O = (u_1, 0)$, $B = (x_1, u_2)$, $C = (x_2, u_3)$, $D = (x_3, u_4)$, $F = (x_4, u_5)$, $E = (x_5, u_6)$, $P = (x_7, x_6)$, $Q = (x_9, x_8)$, and $S = (x_{11}, x_{10})$. This gives the following equations for the hypotheses.

Equation	*Geometry*	*Remark*
$h_1 : x_1^2 - 2u_1x_1 + u_2^2 = 0$	$[OA \equiv OB]$	Introduces x_1, u_2
$h_2 : x_2^2 - 2u_1x_2 + u_3^2 = 0$	$[OA \equiv OC]$	Introduces x_2, u_3
$h_3 : x_3^2 - 2u_1x_3 + u_4^2 = 0$	$[OA \equiv OD]$	Introduces x_3, u_4
$h_4 : x_4^2 - 2u_1x_4 + u_5^2 = 0$	$[OA \equiv OF]$	Introduces x_4, u_5
$h_5 : x_5^2 - 2u_1x_5 + u_6^2 = 0$	$[OA \equiv OE]$	Introduces x_5, u_6
$h_6 :$ $(u_5 - u_4)x_7 + (-x_4 + x_3)x_6 +$ $u_4x_4 - u_5x_3 = 0$	$[P \in DF]$	Introduces x_6, x_7
$h_7 : u_2x_7 - x_1x_6 = 0$	$[P \in AB]$	Constrains x_6, x_7
$h_8 :$ $(u_6 - u_5)x_9 + (-x_5 + x_4)x_8 +$ $u_5x_5 - u_6x_4 = 0$	$[Q \in FE]$	Introduces x_8, x_9
$h_9 :$ $(u_3 - u_2)x_9 + (-x_2 + x_1)x_8 +$ $u_2x_2 - u_3x_1 = 0$	$[Q \in BC]$	Constrains x_8, x_9
$h_{10} : u_6x_{11} - x_5x_{10} = 0$	$[S \in AE]$	Introduces x_{10}, x_{11}
$h_{11} :$ $(u_4 - u_3)x_{11} + (-x_3 + x_2)x_{10} +$ $u_3x_3 - u_4x_2 = 0$	$[S \in CD]$	Constrains x_{10}, x_{11}

The conclusion that P, Q, S are collinear can be translated into the following polynomial:

$$g = (x_8 - x_6)x_{11} + (-x_9 + x_7)x_{10} + x_6x_9 - x_7x_8 = 0.$$ ■

In general, we get a system of polynomial equations, $h_1 = h_2 = \cdots = h_\ell = 0$ where $h_i \in \mathbb{R}[u_1, \ldots, u_m, x_1, \ldots, x_n]$ ($\mathbb{R}$ is the field of real numbers), the $u_1, \ldots, u_m$ are parameters, and the $x_1, \ldots, x_n$ are dependent variables. The conjecture has the form:

$$(\forall \mathbf{u}, \mathbf{x})[h_1 = h_2 = \cdots = h_\ell = 0 \;\;\Rightarrow\;\; g = 0] \tag{1}$$

where $\mathbf{u} = (u_1, \ldots, u_m)$, $\mathbf{x} = (x_1, \ldots, x_n)$ and $g = g(\mathbf{u}, \mathbf{x}) \in \mathbb{R}[\mathbf{u}, \mathbf{x}]$.

Degeneracy and Generic Truth. A theorem of the form (1) is called a *universal truth.* It turns out that the classical notion of theoremhood is more subtle, and this led Wu to formulate the notion of *generic truth.* We formalize it as follows: let $\Delta_1, \ldots, \Delta_k$ be predicates on the variables $\mathbf{u}, \mathbf{x}$. We call each Δ_i a *non-degeneracy condition.* The conjecture (1) is *generically true* relative to $\{\Delta_1, \ldots, \Delta_k\}$ if

$$(\forall \mathbf{u}, \mathbf{x})[\Delta_1, \Delta_2, \ldots, \Delta_k, h_1 = h_2 = \cdots = h_\ell = 0 \;\;\Rightarrow\;\; g = 0]. \tag{2}$$

Classical ruler-and-compass theorems are indeterminate in that they do not explicitly specify the degenerate conditions. Hence part of “proving a classical theorem” involves discovering a suitable set of non-degeneracy conditions. Hopefully the set is minimal is some sense (but not necessarily unique). The simplest kind of non-degeneracy condition has the form

$$\Delta : d \neq 0$$

where d is a polynomial. Call this the *first kind* of non-degeneracy condition. The *degree* of the Δ is equal to the total degree of d. If each Δ_i has degree d_i, then the *degree* of $\{\Delta_1, \ldots, \Delta_k\}$ is $\sum_{i=1}^{k} d_i$. Typical examples of the first kind of non-degeneracy may require two points to be distinct or two lines to be non-parallel. It is easy to see that both have degree 2.

Example 1 (continued). The non-degeneracy conditions require the intersection points P, S and Q be not at infinity. Equivalently, the following pairs of lines are not parallel: $\{AB, DF\}$, $\{BC, FE\}$, $\{CD, EA\}$. So the degree of these non-degeneracy conditions is 6.

Second Kind of Degeneracy. The *second kind* of non-degeneracy condition arises for theorems in the real field. For example, when we define a point by the intersection of two circles, we require that these two circles intersect. Or, when we define three collinear points A, B and C, we may require B to lie between the other two points. Such non-degeneracy conditions have the form

$$\Delta : d \geq 0$$

where d is a polynomial. We can modify this condition using a well-known trick:

$$\Delta' : \;\; \exists z, \; d - z^2 = 0$$

where z is a new variable. The existential quantifier on z can be pulled out as a prenex universal quantifier. Thus, we can formulate the conjecture as

$$(\forall \mathbf{u}, \mathbf{x}, \mathbf{z}) \ \ (\Delta', H \ \ \Rightarrow \ \ T).$$

In practice, there may be other ways to handle this: in the Pascal example, such non-degeneracies demand that the parameters u_j (for $j = 2, 3, 4, 5, 6$) satisfy $|u_j| \leq |u_1|$. Our prover can handle non-degeneracy conditions of the second kind when put in this form. Indeed, in all the examples we looked at in [5], such a formulation is possible.

Reduction to Radical Expressions. In a ruler & compass construction, each dependent variable is a radical function of the previously introduced variables. As exemplified by Pascal's Theorem, all the dependent variables are introduced either (i) singly by a single equation (e.g., x_1 is introduced by $h_1 = 0$) or (ii) in pairs by two equations (e.g., x_6, x_7 are introduced by $h_6 = h_7 = 0$). As all equations are at most quadratic, the x_i's can be replaced by radical expressions involving the u_j's. Let $G = G(\mathbf{u})$ be the radical expression after such a substitution into $g(\mathbf{u}, \mathbf{x})$. The universal truth conjecture (1) now says

$$(\forall \mathbf{u})[G = 0],$$

with an analogous statement for generic truth. Another issue arises: each radical is determined only up to a ± sign. Hence, if there are r radicals in G, we must replace $G = 0$ by the system of 2^r radical expressions, $G_1 = G_2 = \cdots = G_{2^r} = 0$, in which each of the 2^r possible sign combinations are used. If a single function $G^*(\mathbf{u})$ is desired, we can use $G^* = \sum_{i=1}^{2^r} G_i^2$. The appearance of "$2^r$" in this expression may be disturbing from a complexity viewpoint. Several observations suggest that this is not serious in practice. First, r is typically small ($r = 5$ in Pascal's theorem). Next, we can reduce the number of summands in G^* from the worst case of 2^r terms. There are two ways this can happen: (A) Symmetries in the problem may arise so that many of the G_i's can be omitted. (B) Certain sign combinations may be excluded by the nature of the construction and/or theorem so that G^* may represent a sum of less than 2^r radical expressions. In particular, using (A) and (B), we can always omit half of the summands in standard geometric theorems. Thus, 2^{r-1} terms suffice in G^*.

Example 2 (Butterfly Theorem). We illustrate the reduction in the number of terms in G^* using the Butterfly Theorem in [5, Example 2.4, p. 9]. The theorem concerns 4 co-circular points A, B, C and D. Let O be the center of this circle and E be the intersection of AC and BD. The points A, B, C, D, E form a "butterfly". If the line perpendicular to OE and passing through E intersects the lines AD and BC at G and F (respectively), then the theorem says that segments EF and EG have the same length. There are 3 quadratic equations in formulating this theorem (so $r = 3$). In the construction described by Chou, the point E is placed at the origin $(0, 0)$ and O is placed at $(u_1, 0)$. A is freely placed at (u_2, u_3). The point C is now completely determined, and has two possible solutions. In one solution, C and A coincide, and the nature of the theorem excludes this case. Next, the points B is freely chosen on the circle (and this introduces one parameter). Again there are two possible solutions. But it is clear by symmetry that we can arbitrarily choose one of them without loss of generality. Therefore, G^* only needs two terms (corresponding to choosing the 2 solutions for D). ■

The fact that our prover can address theorems about real geometry is illustrated by the following simple example.

Example 3 (Triangle Bisectors). Let A, B, C be three non-linear points, and D be the intersection point of the angle bisectors of $\angle A$ and $\angle B$ in the triangle $\triangle ABC$. We want to prove that D must be on the bisector of $\angle C$ in $\triangle ABC$.

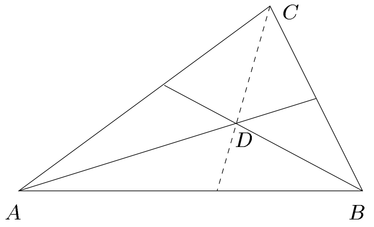

Fig. 2. Coincidence of three angle bisectors.

Let $A = (0,0)$, $B = (u_1,0)$, $C = (u_2,u_3)$, $D = (x_4,x_5)$. This gives the following equations for the hypotheses.

Equation	*Geometry*	*Remark*
$h_1 : x_1^2 - u_1^2 = 0, x_1 \geq 0$	$[x_1 \equiv \|AB\|]$	Introduces x_1
$h_2 : x_2^2 - u_2^2 - u_3^2 = 0, x_2 \geq 0$	$[x_2 \equiv \|AC\|]$	Introduces x_2
$h_3 : x_3^2 - (u_1 - u_2)^2 - u_3^2 = 0, x_3 \geq 0$	$[x_3 \equiv \|BC\|]$	Introduces x_3
$h_4 : (x_1u_2 - x_2u_1)x_4 + x_1u_3x_5 = 0$	$[D \in \mathit{bisector}(\angle A)]$	Constrains x_4, x_5
$h_5 : [(u_2 - u_1)x_1 + u_1x_3](x_4 - u_1) + x_1u_3x_5 = 0$	$[D \in \mathit{bisector}(\angle B)]$	Constrains x_4, x_5

The conclusion that D is on the bisector of angle $\angle C$ can be formulated as the following thesis:

$$g = (x_4 - x_2)(u_1x_2 - u_2x_2 + u_2x_3) - (x_5 - x_3)(x_3 - x_2)u_3 = 0$$

The formulation explicitly introduces inequalities for x_1, x_2, x_3 to pick the internal angle bisectors. When regarded as a complex theorem, no such inequalities are allowed. In this case, each "bisector" can refer to either the internal or external bisector of an angle, so there are a total of $8 = 2^3$ choices for these bisectors. The "thesis" is true for exactly four of these choices, which also means that the theorem is false in complex geometry. Let $G(\mathbf{u})$ be the radical expression after eliminating the dependent variables from g. The 8 choices of bisectors correspond to different assignment of signs to the three radicals in $G(\mathbf{u})$. Our prover can be used to test the validity of each choice. ■

3 Randomized Zero Testing for Radical Expressions

3.1 Straight Line Programs

We need to generalize expressions to *straight line programs* (SLP). A SLP π is a sequence of *steps* where each step is an assignment to a new *programming variable.* The ith step of a SLP has one of the forms

$$z_i \leftarrow x_i \circ y_i, \qquad (\circ \in \{+, -, \times, \div\}) \tag{3}$$

$$z_i \leftarrow \sqrt{x_i} \tag{4}$$

where z_i is a newly introduced programming variable, x_i and y_i are either real constants, *input variables* or programming variables introduced in some earlier steps. Alternatively, we call an input variable an *independent variable* (or, *parameter*) and a programming variable a *dependent variable.* These x_i and y_i are said to be *used* in the ith step. The last introduced variable is called the *main variable* and it is never used. In general, a SLP can have branching steps. But this possibility is not considered in this paper.

An *expression* is a SLP where, with the exception of the main variable, each programming variable is used exactly once. Underlying each SLP is a labeled and ordered *dag* (directed acyclic graph) defined in the obvious way: each node corresponds to a constant or variable in the SLP. We often use the terms "nodes" and "variables" interchangeably. For the steps in (4) (resp., (3)), we introduce edges that are directed from x_i (resp., x_i and y_i) to z_i. We use standard graph-theoretic terminology to talk about this dag: sinks, sources, predecessor/successor nodes, etc. If (u, v) is an edge of the dag, we call u the *predecessor* of v, and call v the *successor* of u. The nodes labeled by input variables or constants are *source nodes* while the non-source are labeled by programming variables. The sources may be called *leaves* in case the dag is a tree. The non-source nodes are associated with an operation $(\pm, \times, \div, \sqrt{\cdot})$ – so we may speak of "radical nodes", "multiplication nodes", etc. Variables that are not used correspond to *sink nodes* in the dag. The main variable corresponds to a sink node which we call *root.* The *radical depth* of a node u is the maximum number of radical nodes in a path from u to any root node, inclusive of the end points. Thus, if u is a radical node, then the radical depth of u is at least 1. For each node u, its *induced dag* is the subdag comprising all the nodes that can reach u along a path. A SLP is said to be *rooted* if the root is the unique sink. The dags corresponding to expressions are ordered trees (hence rooted). Our SLP's are assumed rooted unless otherwise noted.

Values. Let $\mathbf{u} = (u_1, \ldots, u_m)$ be the input variables. For each variable u in a SLP π, we inductively define its *value* to be an appropriate element $val_\pi(u)$

in an algebraic extension of $\mathbb{Q}(\mathbf{u})$. The extension is obtained by adjunction of square roots. The *value* of π is the value of its main variable. More precisely, let $Q_0 = \mathbb{Q}(\mathbf{u})$ and define the tower of extensions defined by π to be

$$Q_0 \subseteq Q_1 \subseteq Q_2 \subseteq \cdots \subseteq Q_r \tag{5}$$

where $Q_i := Q_{i-1}(\sqrt{\alpha_i})$ and the ith square-root in π has operand $\alpha_i \in Q_{i-1}$. A SLP π is also said to *compute* a collection $V \subseteq Q_r$ of values provided each $v \in V$ is the value of some variable in π.

Rational Degrees. Let x be a node in a SLP π. We define the *rational degree* $\mathrm{rdeg}_\pi(x)$ of x (the subscript π is usually dropped). We need some auxiliary notions. For any node or variable x, let $\mathrm{RAD}(x)$ denote the set of radical nodes in the subdag of π rooted at x. Write $\mathrm{RAD}(x, y)$ for $\mathrm{RAD}(x) \setminus \mathrm{RAD}(y)$ (set difference). Also let $\rho(x) := |\,\mathrm{RAD}(x)|$ and $\rho(x, y) := |\,\mathrm{RAD}(x, y)|$. We will inductively define $\mathrm{rdeg}(x)$ to be a pair of natural numbers $(a, b) \in \mathbb{N}^2$, but usually write it as "$a : b$". These two numbers are the "upper" and "lower" degrees of x and denoted $\mathrm{udeg}(x)$ and $\mathrm{ldeg}(x)$. Thus,

$$\mathrm{rdeg}(x) = \mathrm{udeg}(x) : \mathrm{ldeg}(x).$$

Assuming $\mathrm{rdeg}(x) = a_x : b_x$ and $\mathrm{rdeg}(y) = a_y : b_y$, we inductively define $\mathrm{rdeg}(z)$ using the table:

z	$\mathrm{udeg}(z)$	$\mathrm{ldeg}(z)$
constant	0	0
parameter	1	0
$x \times y$	$a_x 2^{\rho(y,x)} + a_y 2^{\rho(x,y)}$	$b_x 2^{\rho(y,x)} + b_y 2^{\rho(x,y)}$
$x \div y$	$a_x 2^{\rho(y,x)} + b_y 2^{\rho(x,y)}$	$b_x 2^{\rho(y,x)} + a_y 2^{\rho(x,y)}$
$x \pm y$	$\max(a_x 2^{\rho(y,x)} + b_y 2^{\rho(x,y)}, b_x 2^{\rho(y,x)} + a_y 2^{\rho(x,y)})$	$b_x 2^{\rho(y,x)} + b_y 2^{\rho(x,y)}$
$\sqrt{x}$	a_x	b_x

The *rational degree* of the SLP π is defined to be $a : b$ where $a = \max_x \mathrm{udeg}(x)$, $b = \max_x \mathrm{ldeg}(x)$, and x ranges over the nodes in π. Note that if π is division-free, then $\mathrm{ldeg}(x) = 0$ for all x.

Alternative Approach. It is useful to have an alternative approach to rdeg which does not involve $\rho(x, y)$ or $\rho(y, x)$. In particular, we define $\mathrm{rdeg}_2(z) = \mathrm{udeg}_2(z) : \mathrm{ldeg}_2(z)$ inductively using the following table: as before, we assume $\mathrm{rdeg}_2(x) = a_x : b_x$ and $\mathrm{rdeg}_2(y) = a_y : b_y$.

z	$\mathrm{udeg}_2(z)$	$\mathrm{ldeg}_2(z)$
constant	0	0
parameter	1	0
$x \times y$	$a_x + a_y$	$b_x + b_y$
$x \div y$	$a_x + b_y$	$b_x + a_y$
$x \pm y$	$\max\{a_x + b_y, b_x + a_y\}$	$b_x + b_y$
$\sqrt{x}$	$\frac{a_x}{2}$	$\frac{b_x}{2}$

Notice that these degrees are no longer natural numbers but binary fractions. The following lemma gives the connection between the two definitions of rdeg.

Lemma 1. *For any variable z in a SLP, we have*

$$\mathrm{udeg}(z) = 2^{\rho(z)}\,\mathrm{udeg}_2(z), \qquad \mathrm{ldeg}(z) = 2^{\rho(z)}\,\mathrm{ldeg}_2(z).$$

3.2 Equivalent Transformations

Two variables (resp. SLP's) are said to be *equivalent* if they have the same value. Transformations of an SLP that do not change its value are called *equivalent transformations* (but the set of computed values may change). Equivalent transformations may change the rational degree, as when applying the distributive law:

$$z(x+y) \Rightarrow zx + zy. \tag{6}$$

It is easy to verify that the rational degree of the left-hand side is at most that of the right-hand side. We next show that the rational degree is preserved in the absence of division (but allowing radicals):

Lemma 2. *If π is division-free, then the transformation (6) preserves* rdeg *of π. In particular,*

$$\mathrm{rdeg}(z(x+y)) = \mathrm{rdeg}(zx+zy).$$

Proof. We only need to consider the upper degrees. With $\mathrm{udeg}(x) = a_x$, etc, as before, we have

$$\mathrm{udeg}(z(x+y)) = 2^{\rho(xy,z)}a_z + 2^{\rho(z,xy)}\max\{a_x 2^{\rho(y,x)}, a_y 2^{\rho(x,y)}\}$$

while

$$\begin{aligned}\mathrm{udeg}(zx+zy) &= \max\{a_{zx}2^{\rho(zy,zx)}, a_{zy}2^{\rho(zx,zy)}\}\\ &= \max\{(a_z 2^{\rho(x,z)} + a_x 2^{\rho(z,x)})2^{\rho(zy,zx)}, (a_z 2^{\rho(y,z)} + a_y 2^{\rho(z,y)})2^{\rho(zx,zy)}\}.\end{aligned}$$

The lemma follows if we now verify the following:

$$\begin{aligned}\mathrm{RAD}(xy,z) &= \mathrm{RAD}(x,z) \uplus \mathrm{RAD}(zy,zx),\\ \mathrm{RAD}(xy,z) &= \mathrm{RAD}(y,z) \uplus \mathrm{RAD}(zx,zy),\\ \mathrm{RAD}(z,xy) \uplus \mathrm{RAD}(y,x) &= \mathrm{RAD}(z,x) \uplus \mathrm{RAD}(zy,zx),\\ \mathrm{RAD}(z,xy) \uplus \mathrm{RAD}(x,y) &= \mathrm{RAD}(z,y) \uplus \mathrm{RAD}(zx,zy).\end{aligned}$$

Our notation here, $A \uplus B$, refers to disjoint union of the sets A and B. Let us only prove the first equation: the RHS is equivalent to $\mathrm{RAD}(x,z) \uplus \mathrm{RAD}(y,zx)$. We may verify that the union is indeed disjoint, and equal to $\mathrm{RAD}(xy,z)$. The other equations can be proved similarly. We omit the details here. ■

Next, we show that applying the associative laws for multiplication and addition does not affect rational degree. This follows from the following general result:

Lemma 3. *Let x_i be variables in π and $r_i = |\mathrm{RAD}(x_1,\ldots,x_k) \setminus \mathrm{RAD}(x_i)|$. Then*

$$\begin{aligned}
\mathrm{rdeg}(\prod_{i=1}^{k} x_i) &= \sum_{i=1}^{k} \mathrm{rdeg}(x_i)2^{r_i} \\
\mathrm{udeg}(\sum_{i=1}^{k} x_i) &= \max_{i=1}^{k}\{\mathrm{udeg}(x_i)2^{r_i} + \sum_{j=1, j\neq i}^{k} \mathrm{ldeg}(x_j)2^{r_j}\} \\
\mathrm{ldeg}(\sum_{i=1}^{k} x_i) &= \sum_{i=1}^{k} \mathrm{ldeg}(x_i)2^{r_i}
\end{aligned}$$

The above lemma justifies a generalization of SLP's in which we allow addition nodes and multiplication nodes to take an arbitrary number of arguments. These are called "sum" or $\sum$-nodes, and "product" or $\prod$-nodes, respectively. Such an SLP is called a *generalized SLP*. A path in a generalized SLP dag is said to be *alternating* if along the path, no two consecutive nodes are $\sum$-nodes and no two consecutive nodes are $\prod$-nodes. The SLP is *alternating* if every path is alternating. Clearly, any SLP can be made alternating without changing its rational degree. We can eliminating any non-alternating path in the SLP by aggregating the consecutive additions (or multiplications) using the $\sum$ (or $\prod$) operations. This process will terminate because each elimination reduces the number if nodes in a SLP.

3.3 Preparation

A SLP in which the last three steps has the form

$$\begin{aligned}
&\cdots \\
&x \leftarrow \sqrt{w_C} \\
&y \leftarrow x \times w_B \\
&z \leftarrow y + w_A
\end{aligned}$$

is said to be *prepared* (or in prepared form). Here w_A, w_B, w_C are variables or constants. Thus z is the main variable, and x is the last radical variable to be introduced. Intuitively, the radical x has been brought up as close to the root as possible, in preparation for a transformation (to be introduced) to remove the radical. We also call x the *prepared variable.* If the values of w_A, w_B, w_C are given by the expressions A, B, C (resp.) then the value of z is given by the expression

$$A + B\sqrt{C}.$$

Note special forms of this expression when $A = 0$ or $B = 1$, or both. If the SLP has no square roots, it is considered prepared already. Our goal is to prepare a given SLP, and to bound the resulting rational degree.

Let us now prepare a radical node A_0 with radical depth 1. Assume the SLP is division-free. Let A_n, B_n be expressions ($n \geq 0$). The expressions E_n, for $n \geq 0$ is defined inductively as follows: $E_0 = A_0 \times B_0$, and for $n \geq 1$,

$$E_n = (E_{n-1} + A_n)B_n = ((E_{n-2} + A_{n-1})B_{n-1} + A_n)B_n = \cdots .$$

To show the dependence of E_n on the A_n's and B_n's, we may also write $E_n = E_n(A_0, B_0, A_1, B_1, \ldots, A_n, B_n)$. Viewed as a tree, E_n is essentially a single alternating path from the root down to A_0. This path is left-branching only and the root is a ×-node. Also write: $B_{(n)} := \prod_{j=0}^{n} B_j$.

Lemma 4. *For $n \geq 1$, the expression $E_n(A_0, B_0, \ldots, A_n, B_n)$ is equivalent to the expression*

$$E'_n := (A_0 \times B_{(n)}) + E_{n-1}(A_1, B_1, \ldots, A_n, B_n)$$

Moreover, if E_n is division-free, then $\text{rdeg}(E_n) = \text{rdeg}(E'_n)$.

Proof. Proof by induction. When $n = 1$,

$$\begin{aligned} E_1 &= (A_0 \times B_0) + A_1) \times B_1 \\ &= (A_0 \times B_0 \times B_1) + A_1 \times B_1. \end{aligned}$$

Assume that this lemma is held for $n \leq k$, then for $n = k + 1$,

$$\begin{aligned} E_{k+1} &= (E_k + A_{k+1}) \times B_{k+1} \\ &= ((A_0 \times B_{(k)}) + E_{k-1}(A_1, B_1, \ldots, A_k, B_k) + A_{k+1}) \times B_{k+1} \\ &= (A_0 \times B_{(k+1)}) + E_k(A_1, B_1, \ldots, A_{k+1}, B_{k+1}). \end{aligned}$$

Thus we know the equivalence of this transformation is held for any $n \in \mathbb{N}$.

In both cases, we only apply the distributive and associative laws, which do not change the rational degree when E_n is division free. ■

This is illustrated in the case $n = 2$ by Figure 3. Note that the variable A_0 is prepared in E'_n. Actually, E_n in this lemma can be a generalized SLP so that the A_i, B_i's need not be distinct and the nodes can be $\sum$- and $\prod$-nodes. Then there is a corresponding equivalent SLP E'_n; this is the version that we will use in the next theorem.

We address the problem of multiple uses of a node. A node u is *used* k times if there are k distinct paths from the root to u. If a radical node u of radical depth 1 is used k times, then if we judiciously apply the previous lemma k times, each time eliminating one "use" of u, we obtain:

Theorem 1. *Suppose π is a division-free SLP and u is a radical node in π with radical depth of* 1. *Then we can transform π into an equivalent SLP π' such that* $\text{udeg}(\pi) = \text{udeg}(\pi')$. *Moreover, either no node in π' has the value $val_\pi(u)$ or else, there is a node u' in π' with the following properties:*

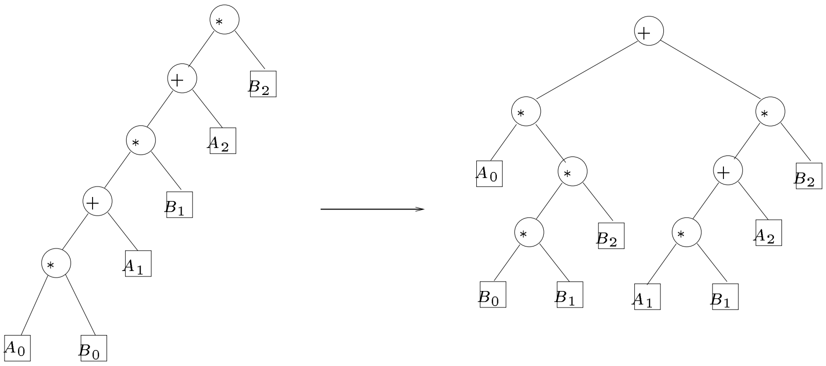

Fig. 3. The Transformation $E_2 \mapsto E'_2$.

1. u' is the prepared variable in π'
2. u' is the unique node in π' with value $val_\pi(u)$.

Proof. We may assume that π is a generalized, alternating SLP. Fix any path p from u to the root and we may assume that this alternating sum-product path has the same form as the path from A_0 to the root of E_n in lemma 4. We then apply the previous lemma in which u now plays the role of the node A_0 in E_n. This collapses the path p to length 2, as in the lemma and the resulting SLP is in a prepared form $E' = u \times A + B$. If the variable u is used in A and/or B, then we can repeat this process for another path p' (if any) in A or B. We can repeat this process for the subexpressions A and/or B, if they contain references to the node u as well. There are two cases:

1. u is used in A, then A is transformed to $A' = u \times A_1 + B_1$ and $E' = u \times B_1 + (A_1 u^2 + B)$. Remember that u is a square root and thus the expression u^2 effectively eliminates the square root operation here;
2. u is used in B, then B is transformed to $B' = u \times A_2 + B_2$ and $E' = u \times (A + A_2) + B_2$.

In both cases, we can see that E' is still in a prepared form. We keep this process until there is no use of u except the one that is in the prepared position and has a unique path to the root with length 2. Since there must be a finite number of uses of u, this iterative process will eventually terminate. At that point, the resulting SLP π' has the desired form: π' is prepared and u is the main prepared variable. It is also clear that if there are other nodes with the same value as u, they can also be merged with u by the same process. Hence, u will be the unique node with value $val_\pi(u)$.

Note that we apply the commutative, associative and distributive laws in these transformations. The commutative and associative transformations do not change the rational degree. Since π is division free, Lemma 2 tells us that the distributive transformation preserves the rational degree too. Therefore, the preparation transformation does not change the rational degree of π. ■

We say that π' is obtained by the process of "preparing" u in π.

3.4 Main Result

Let π be a SLP whose value is $V = V(\mathbf{u}) \in \mathbb{Q}_r$ (see (5)). We define the real function $f_\pi : \mathbb{R}^m \to \mathbb{R}$ where $f_\pi(a_1, \ldots, a_m)$ is the value of the main variable in π when we evaluate each dependent variable at $\mathbf{a} = (a_1, \ldots, a_m) \in \mathbb{R}^m$, following π in a step-by-step fashion. The *domain* of f_π comprises those $\mathbf{a} \in \mathbb{R}^m$ where $f_\pi(\mathbf{a})$ is defined. Similarly, we define an associated real function $f_V : \mathbb{R}^m \to \mathbb{R}$. Note that the domain of f_π is always a subset of f_V. The following example shows that it may be a proper subset: let π compute the value $V = \sum_{i=0}^{n-1} x^i$ using Horner's rule, and let π' compute the same V using the formula $V = \frac{x^n-1}{x-1}$. Then π and π' are equivalent, but $\pi(1) = n$ while $\pi'(1)$ is undefined. The domain of π (and V) is $\mathbb{R}$ but the domain of π' is $\mathbb{R} - \{1\}$.

Theorem 2. *Suppose $V = V(\mathbf{u})$ is the non-zero value of a rooted division-free SLP π. Then there exists a non-zero polynomial $P(\mathbf{u})$ such that $\mathtt{Zero}(V) \subseteq \mathtt{Zero}(P)$ with $\deg P(\mathbf{u}) \leq \mathrm{udeg}(\pi)$.*

Proof. We show the existence of the polynomial $P(\mathbf{u})$ by induction on the number r of square roots in π. For $r = 0$, the result holds because V is already a polynomial of degree $\mathrm{udeg}(\pi)$.

Assume $r > 0$ and let u be a radical node of radical depth 1 in π. We prepare u, leading to an equivalent SLP (which we still call π). The udeg of π is unchanged by this transformation. If C is the value of u, then the value of π can be written as

$$V = A + B\sqrt{C}$$

where A, B, C belongs to Q_{r-1} (recall that values of programming variable introduced before the rth root extraction belongs to the field Q_{r-1}, by definition of Q_{r-1}). If $B = 0$ then $V = A$ and the result is true by the inductive hypothesis applied to A (which has $\leq r - 1$ square roots). Otherwise, by applying some further (obvious) transformations, we transform π to some π' whose value is

$$V' = A^2 - B^2C. \tag{7}$$

Note that π' has $\leq r-1$ square-roots. If $V' = 0$ then $0 = V' = (A + B\sqrt{C})(A - B\sqrt{C})$. Since Q_r is a UFD and $V = A + B\sqrt{C} \neq 0$ (by assumption), we conclude that $A - B\sqrt{C} = 0$, i.e., $\sqrt{C} = A/B \in Q_{r-1}$. Thus $V = A + B\sqrt{C} = 2A$. Then V can be computed by some SLP with $\leq r - 1$ square-roots, and the result follows by inductive hypothesis.

So assume $V' \neq 0$. By induction, $\mathtt{Zero}(V') = \mathtt{Zero}(A^2 - B^2C) \subseteq \mathtt{Zero}(P)$ for some P with $\deg(P) \leq \mathrm{udeg}(V')$. Since $\mathtt{Zero}(V) \subseteq \mathtt{Zero}(V')$, it remains to show that $\mathrm{udeg}(V') \leq \mathrm{udeg}(V)$. We have

$$\begin{aligned}&\mathrm{udeg}(V) = \mathrm{udeg}(A + B\sqrt{C})\\ &= \max\{\mathrm{udeg}(A)2^{\rho(B\sqrt{C},A)}, \mathrm{udeg}(B\sqrt{C})2^{\rho(A,B\sqrt{C})}\}\end{aligned}$$

$$
\begin{aligned}
&\geq \max\{\mathrm{udeg}(A)2^{1+\rho(B^2C,A)}, \left[\mathrm{udeg}(B)2^{\rho(\sqrt{C},B)} + \mathrm{udeg}(C)2^{\rho(B,\sqrt{C})}\right] 2^{\rho(A,B^2C)}\} \\
&= \max\{2\,\mathrm{udeg}(A)2^{\rho(B^2C,A)}, \left[\frac{\mathrm{udeg}(B^2)}{2}2^{1+\rho(C,B^2)}+\mathrm{udeg}(C)2^{\rho(B^2,C)}\right]2^{\rho(A,B^2C)}\} \\
&\geq \max\{\mathrm{udeg}(A^2)2^{\rho(B^2C,A)}, \left[\mathrm{udeg}(B^2)2^{\rho(C,B^2)} + \mathrm{udeg}(C)2^{\rho(B^2,C)}\right] 2^{\rho(A^2,B^2C)}\} \\
&= \mathrm{udeg}(A^2 - B^2C) = \mathrm{udeg}(V').
\end{aligned}
$$

■

3.5 Presence of Division

What if the SLP is not division-free? Note that the presence of division is very common. For instance, when we intersect two lines in the construction, it gives rise to an expression with division. There is a well-known transformation to move all divisions towards the root, merging them as we go. An instance of this transformation is

$$\frac{A}{B} + \frac{A'}{B'} \Rightarrow \frac{AB' + A'B}{BB'}.$$

Unfortunately, the number of radical nodes may be doubled because if we move a division node past a radical node, we obtain two radical nodes:

$$\sqrt{\frac{A}{B}} \Rightarrow \frac{\sqrt{A}}{\sqrt{B}}. \tag{8}$$

Hence we give two versions of this transformation in the following lemma: in version (i) we do not move any division node past a radical node, and in version (ii) we remove all but at most one division node.

Lemma 5 (Elimination of Division). *Let π be a rooted SLP.*
(i) There is an equivalent SLP π' in which each division node is either the root of π or the child of a radical node. Moreover, $\mathrm{rdeg}(\pi') = \mathrm{rdeg}(\pi)$ *and π' has the same number of radical nodes as π.*
(ii) There is an equivalent SLP π'' with only one division node which is also the root. In this case $\mathrm{rdeg}(\pi'') \leq 2^r\,\mathrm{rdeg}(\pi)$.

The proof of (ii) exploits the alternative definition of $\mathrm{udeg}(u)$. Because the justification of the alternative definition is long, we only refer to the details in [15].

The value of the SLP π'' has the form A/B where A, B are division-free. Intuitively, to check if $A/B = 0$, we check if $A = 0$ subject to $B \neq 0$. Since A is division-free, we may apply main theorem (see next Section). This effectively amounts to doubling the number of square roots to prove a theorem involving division.

3.6 Improved Square Root Transformation

It turns out that we can exploit another trick motivated by [18] in order to avoid the doubling of the number of square roots. Instead of (8), we use the following transformation to extract division out of square roots:

$$\sqrt{\frac{A}{B}} \Rightarrow \begin{cases} \frac{\sqrt{AB}}{B} & \text{if } \operatorname{udeg}(A) \geq \operatorname{udeg}(B), \\ \\ \frac{A}{\sqrt{AB}} & \text{if } \operatorname{udeg}(A) < \operatorname{udeg}(B). \end{cases} \tag{9}$$

Suppose our transformations for eliminating divisions, using the new rule (9), transform an arbitrary expression z into $U(z)/L(z)$ where $U(z), L(z)$ are division free. Let u_z and ℓ_z denote the $\operatorname{udeg}(U(z))$ and $\operatorname{udeg}(L(z))$. To exploit the advantages of this new rule, we now give an explicit set of inductive rules for computing u_z and ℓ_z:

z	u_z	l_z
constant	0	0
parameter	1	0
$x \times y$	$u_x + u_y$	$l_x + l_y$
$x \div y$	$u_x + l_y$	$l_x + u_y$
$x \pm y$	$\max\{u_x + l_y, l_x + u_y\}$	$l_x + l_y$
$\sqrt{x}$	$\frac{1}{2}(u_x + l_x)$, $(u_x \geq l_x)$; u_x, $(u_x < l_x)$.	l_x, $(u_x \geq l_x)$; $\frac{1}{2}(u_x + l_x)$, $(u_x < l_x)$.

Note that [18] only uses one of two clauses in (9) unconditionally. But the effect of using the two conditional clauses is that the resulting bound u_z is never worse than $2^r \operatorname{udeg}(z)$, which is the bound in Lemma 5. The proofs may be found in [15].

4 Proving by Random Examples

We show how to use our main result to prove theorems about ruler & compass constructions. According to Section 2, this amounts to verifying if a radical expression $G^*(\mathbf{u})$ is identically zero (subject to non-degeneracy conditions). Let $\pi(\mathbf{u})$ be the natural SLP which computes the values of all the dependent variables in a ruler & compass construction, and whose value is the polynomial thesis $G^*(\mathbf{u})$. We give a simple upper estimate on the rdeg of each node in π.

Each "stage" of our construction introduces new points, lines or circles. Let us now be more precise: assume that our system maintains three kinds of geometric objects: points, lines and circles. These are constructed as follows:

- Points: There are three cases. **Case 0**: We can introduce an arbitrary point, P. Then $P.x$ and $P.y$ are free variables (i.e., parameters). **Case 1**: We can introduce an arbitrary point, P on an existing line L or circle C. We may specify either $P.x$ or $P.y$ to be a parameter. The other coordinate is therefore a dependent variable, constrained by an equation. **Case 2**: We can introduce a point P that arises from the intersection of a line/circle with another line/circle. In this case, $P.x$ and $P.y$ are both dependent variables constrained by a pair of simultaneous equations. There is a variation of Case 2, which arises when at least one of the two intersecting objects is a circle. In this case, we allow the user to obtain both the points of intersection[1].

- Lines: Given two existing points, we can construct the line through them.
- Circles: Given three points P, Q, R, we can construct the circle centered at P of radius equal to the distance between Q and R. As a special case, if P is equal to Q or R, we can just use two arguments for this construction.

Lemma 6. *If the dependent variable x is introduced at stage i , then* $\mathrm{rdeg}_2(x) \leq 85^i$*, i.e.,* $\mathrm{udeg}_2(x) \leq 85^i$*,* $\mathrm{ldeg}_2(x) \leq 85^i$*.*

Proof. Proof by induction. Let S_k be the set of objects (points, lines, etc.) available after k construction stages. This lemma is trivially true when $k = 0$ because S_0 is empty.

Let $r_k = 85^k$. By the induction hypothesis, we assume that the coordinate (e.g., for points) or coefficient (e.g., in a line or circle equation) variables for all the objects in S_k have rational degrees at most r_k.

Let us first consider the construction of lines and circles. Recall that in our system, a line refers to one that is constructed by linking two points in S_k; while a circle means one that is constructed with the center in S_k and the radius being the length of some segment between two points in S_k. We represent a line by a linear equation $ax + by + c = 0$. It is easily verified that the rational degrees of a, b and c are at most $2r_k, 2r_k$ and $6r_k$, respectively. Similarly, we represent a circle by an equation in the form of $(x - a)^2 + (y - b)^2 = c^2$ where the rational degrees of a, b and c are at most r_k, r_k and $4r_k$, respectively.

Next, we consider the construction of points. As discussed above, we can have one of the three types of construction (Cases 0, 1, 2) in stage $(k + 1)$. Case 0 is trivial because all the parameters have the rational degree $1 : 0$. Case 1 can be viewed as a simplified Case 2. In the following, we focus on the more interesting Case 2 constructions.

There are three possible constructions in a Case 2 stage.

First, we consider the intersection of two lines $L_1 : a_1x + b_1y + c_1 = 0$ and $L_2 : a_2x + b_2y + c_2 = 0$ where a's, b's and c's can be at most r_k. We obtain the intersection point (x, y) of these two lines as follows,

$$(\frac{c_1b_2 - c_2b_1}{a_1b_2 - a_2b_1}, \frac{c_1a_2 - c_2a_1}{a_2b_1 - a_1b_2}).$$

From the definition (see Section 3.1), the rational degrees for x and y are at most $8r_k$.

[1] It should be possible to allow the user to pick one of the two points using some criteria, but we defer this to a future paper on implementation. This additional power is sometimes needed in ruler-and-compass theorems.

Next, let us consider the intersection of a line $L: a_1x + b_1y + c_1 = 0$ and a circle $C: (x-a_2)^2 + (y-b_2)^2 = c_2^2$. We eliminate y and get a quadratic equation for x as follows:

$$(1+\frac{a_1^2}{b_1^2})x^2 + (-2a_2 + 2\frac{a_1}{b_1}(\frac{c_1}{b_1}+b_2))x + ((\frac{c_1}{b_1}+b_2)^2 + a_2^2 - c_2^2) = 0.$$

Let A, B and C be the three coefficients in the above equation. It can be shown that the rational degrees of them can at most be $4r_k$, $6r_k$ and $10r_k$ respectively. From the above equations, we get $x = \frac{-B\pm\sqrt{B^2-4AC}}{2A}$ and $y = -\frac{a_1x+c_1}{b_1}$. Thus, $\mathrm{rdeg}_2(x) \le 23r_k$ and $\mathrm{rdeg}_2(y) \le 26r_k$.

Thirdly, we consider the intersection of two circles: $C_1: (x-a_1)^2+(y-b_1)^2 = c_1^2$ and $C_2: (x-a_2)^2+(y-b_2)^2 = c_2^2$. We subtract them first to obtain a linear equation first. Then by arguments similar to those used for the intersection of a line and a circle, we can show that the rational degrees for x and y are at most $69r_k$ and $85r_k$, respectively.

Therefore, we know that $\mathrm{rdeg}_2(x) \le 85^i$ for all the nodes at the stage i. ■

REMARK: The constant 85 in the above lemma is clearly very conservative. This bound can be refined, for example, by classifying the stages into the various types of construction.

Corollary 1. *Let the thesis polynomial be $g(\mathbf{u},\mathbf{x})$ with $\deg(g) = d$, and $G(\mathbf{u})$ be any of the 2^r radical expressions derived from $g(\mathbf{u},\mathbf{x})$ by eliminating dependent variables. Then $\mathrm{rdeg}_\pi(G) \le td2^r85^k$ where $g(\mathbf{u},\mathbf{x})$ has t terms and k is the number of construction stages.*

Proof. For Lemma 6, we know that the rational degrees for all the dependent and independent variables are at most 85^k. The thesis G has t terms with total degree at most d. By the inductive definitions of rational degrees, we know that $\mathrm{rdeg}_\pi(G) \le td2^r85^k$. ■

Assume an incremental construction with m parameters, n dependent variables, k stages, and r quadratic equations. Note that t is at most $\binom{m+n+d}{d}$. Moreover, $d \le 2$ in most classical geometric theorems. In our implementation, instead of relying on this crude upper bound, we actually compute the actual bounds on rdeg to achieve better performance. By applying Lemma 5(ii) to π, we obtain π'' with one division at the root, and $\mathrm{rdeg}(\pi'') \le 2^r\,\mathrm{rdeg}(\pi)$. Now the value of π'' (which is G^*) has the form A/B where A, B are division-free. Moreover, $\mathrm{rdeg}_{\pi''}(G^*) \le td2^{2r}85^k$. Clearly, $\mathtt{Zero}(A/B) \subseteq \mathtt{Zero}(A)$. Without loss of generality, assume $A \ne 0$. By our main theorem, $\mathtt{Zero}(A) \subseteq \mathtt{Zero}(P)$ for some polynomial P of degree $\le td2^{2r}85^k$. Then we invoke a simple form of Schwartz's lemma:

Fact 1. *Let $P(\mathbf{u})$ be a non-zero polynomial of degree at most D. If each a_i ($i = 1,\ldots,m$) is randomly chosen from a finite set $S \subseteq \mathbb{R}$. Then the probability that $P(a_1,\ldots,a_m) = 0$ is at most $D/|S|$.*

If we randomly pick the values $\mathbf{a} = (a_1, \ldots, a_m) \in S^m$, and $|S| = td2^{c+2r}85^k$ (for any $c \geq 1$) then the "error probability" of our procedure is given by $\mathtt{Pr}\{A(\mathbf{a}) = 0\} \leq \mathtt{Pr}\{P(\mathbf{a}) = 0\} \leq 2^{-c}$. This constitutes our probabilistic verification of the universal truth of "$G^*(\mathbf{u}) = 0$".

An alternative to testing $G^*(\mathbf{u}) = 0$ is viewing the problem as testing the simultaneous vanishing of a set of polynomial $\mathcal{G} := \{G_1(\mathbf{u}), \ldots, G_{2^r}(\mathbf{u})\}$. This reduces the complexity in two ways:

- The root bound (which determines the precision necessary to numerically determine the sign of radical expressions in the Core Library) is smaller.
- The size of the test set S is smaller.

We also have a further choice when testing $\mathcal{G}$: we can randomly choose some G_i to test for its vanishing, or we can choose to randomly test all the G_i's for their vanishing. However, the random choice of G_i does not seem to be the most efficient way to test a theorem.

Degeneracies of the First Kind. We now address the generic truth of "$G^*(\mathbf{u}) = 0$". The notion of "error probability" becomes an interesting issue. First consider only non-degeneracy conditions of the first kind, $\Delta : \delta \neq 0$. For simplicity, assume the ith ruler & compass construction step introduces exactly one such condition, $\delta_i \neq 0$, of degree ≤ 2. Since there are k stages of construction, the non-degeneracy condition becomes $\delta^* := \delta_1 \delta_2 \cdots \delta_k \neq 0$. The degree of δ^* is thus at most $2k$.

There are two natural models of what it means to have an "error probability" $\leq 2^{-c}$: (A) The "strict model" says that our sample space is now restricted to $S^m \setminus \{\mathbf{a} : \delta(\mathbf{a}) = 0\}$. (B) Alternatively, we can say that the sample space is still S^m but the theorem is trivially true at $S^m \cap \{\mathbf{a} : \delta(\mathbf{a}) = 0\}$. Given a finite test set S, the possible zeros of δ^* (i.e., degenerate configurations) in S^m is at most $2^{2r}\,\mathrm{udeg}(\delta^*)|S|^{m-1}$. With a large enough test set S, we can make the probability that degenerate cases are chosen in the test (i.e., $2^{2r}\,\mathrm{udeg}(\delta^*)/|S|$) arbitrarily small. We adopt the model A in the next theorem:

Theorem 3. *Conjectures about ruler & compass constructions with s non-degenerate conditions of the first kind can be verified with error probability $\leq 2^{-c}$ in time polynomial in the parameters $2^r, 2^s, k, c, \lg(t)$ and $\lg(d)$, where r is the number of square roots in the thesis radical expression $G(\mathbf{u})$, k is the number of construction stages, t is the number of monomials in the thesis polynomial $g(\mathbf{u}, \mathbf{x})$, and d is the total degree of g.*

Proof. Each construction introduces a constant number of new operations into the final radical thesis expression $G^*(\mathbf{a})$. Thus, the cost to construct the thesis expressions $G^*(\mathbf{a})$ is bound by $O(k)$. Next, let us consider the complexity in verifying $G(\mathbf{a})$ for some sample configuration $\mathbf{a} = (a_1, a_2, \ldots a_m)$ randomly chosen from a finite test set S with a cardinality of $2^{2r+c}85^k td$. From the discussion above, we know that the failure probability of this test is at most

2^{-c}. Without loss of generality, we can assume all the elements in S are integers. So the bit length of each instance value is bounded by $L = \lg(|S|) = O(r + c + \lg(t) + \lg(d) + k)$. In our root bound based approach to determine the exact sign of an algebraic expression [16], the number of bits which need to compute in the verification is bounded by $O(pL2^{2r})$, where p is the total number of operations in G^* which is bounded by $O(k)$. It is known that the time complexity of arithmetic operations among multiple precision numbers are no more than $O(\ell^2)$ where ℓ is the bit length of operands. We have a total of 2^r radical thesis expressions to verify. So the complexity to verify the vanishing of G^*, when exact arithmetic is employed, is polynomial in $2^r, k, c, \lg(t)$ and $\lg(d)$.

In presence of s non-degeneracy conditions of the first kind, let $\Delta(\mathbf{u})$ be the product of all of them. It is a radical expression in $\mathbf{u}$. By our main theorems, the number of zeros of Δ in S^m, N, is polynomial in 2^s and 2^r. In the worst case, we may meet at most N degenerate cases before we get the first non-degenerate one. So the worst case complexity for our complete method is polynomial in $2^r, 2^s, k, c, \lg(t)$ and $\lg(d)$. ■

Degeneracies of the Second Kind. As noted, degeneracies of the second kind can often be reduced to simple constraints on the domains of the parameters, possibly depending on the values of other parameters. For instance, we noted that in Pascal's Theorem, the parameters u_i $(i = 2, \ldots, 6)$ must satisfy $|u_i| \le |u_1|$. Our prover can handle such degeneracies by exploiting the following more general form of fact 1: define the *generalized degree* of $p(x_1, \ldots, x_n)$ to be $(d_1, \ldots, d_n)$ where the degree of p is d_1 when viewed as a polynomial in x_1 and its leading coefficient inductively has generalized degree $(d_2, \ldots, d_n)$. Suppose $S_1, \ldots, S_n$ are finite sets of real numbers, then it can be shown that if we choose $(u_1, \ldots, u_n)$ randomly from $S_1 \times S_2 \times \cdots \times S_n$, the probability that p is non-zero and $p(u_1, \ldots, u_n) = 0$ is at most

$$\frac{d_1}{|S_1|} + \cdots + \frac{d_n}{|S_n|}.$$

The main extra complexity caused by this version of our prover is that we need to evaluate the parameters at rational values (instead of just at integer values).

The current implementation does not handle the second kind of degeneracy in the above way, but we plan to rectify this in the future. Instead, it detects when an example $\mathbf{a} \in S^m$ is degenerate, discards it and generates another example, etc. Under probability model (A) above, this means that we do not have an á priori bound on the running time, but the error probability is correct. Of course, under model (B), there is no need to generate another example; but this does not seem like a reasonable model.

Degenerate Ruler-and-Compass Constructions. Certain theorems amount to detecting the validity of construction steps. We give a simple example from

[6] of a theorem true in real geometry but false in the complex geometry. The construction amounts to picking two points $P_1(0,0)$ and $P_2(u,0)$ where u is a free parameter. Also let P_3 be the midpoint of P_1P_2, and P_4 the midpoint of P_1P_3. Let L be the bisector of the segment P_1P_2, and C be the circle centered at P_1 with radius P_1P_4. Let P_5 be the intersection of L and C. The thesis is $P_1 = P_2$ or equivalently $u = 0$. This conjecture is true in real geometry, but it is false in the complex plane because $u = \sqrt{-1}$ is a solution. This is an interesting example because the thesis does not depend on the construction at all. It is an indirect way of asserting the validity of the construction steps. In implementing a prover that takes inputs from the user, we need to guard against being asked to prove such theorems. This amounts to an extreme form of the second kind of degeneracy.

Timing. The following table lists some theorems from Chou [5]. However, the last row (Tri-Bisector theorem) is the real geometry example from Section 2. The timings are for two values of c (this means the probability of error is at most 2^{-c}). We also arbitrarily "perturb" the hypothesis of each theorem by randomly changing one coefficient of one of the input polynomials, and report their timings as well. These are all false theorems, naturally. Our tests were performed on a Sun UltraSPARC-IIi (440 MHz, 512 MB). The times are all in seconds, and represent the average of 6 runs each. The prover uses Core Library, Version 1.3. Actually, the library is directly modified so that we compute the exact rational degrees of the expressions (rather than use the estimates of the Lemma 6). For comparison, we include the timings reported by Chou [5] using the approaches of Wu and of Gröbner Bases. The final column in the table gives the page number in Chou's book [5].

No.	Theorem	$c = 10$	$c = 20$	Perturbed	Char Set	Gröbner	Page
1	Pappus	0.020	0.020	0.007	1.52	33.32	100
2	Pappus Point	0.110	0.113	0.023	4.87	67.62	100
3	Pappus-dual	0.020	0.020	0.013	1.45	25.53	111
4	Nehring	8.300	8.390	0.107	4.15	159.3	115
5	Chou-46	0.070	0.073	0.020	88.13	37.65	124
6	Ceva	0.030	0.033	0.017	1.12	3.47	264
7	Simson	193.22	262.49	0.023	1.22	5.02	240
8	Pascal	1715.8	2991.6	0.037	29.6	>14400	103
9	Tri-Bisector	20.027	38.350	0.010	–	–	–

Let r be the number of square roots in the radical expression representing a theorem. If $r = 0$, we say the theorem is linear. A large part[2] of the 512 theorems in Chou's book are linear. Only the last two theorems (Simson and Pascal) in the above list are non-linear, with $r = 1$ and $r = 5$, respectively. Evidently non-linear theorems represent a challenge for our current system. Recall that there

[2] The theorems in Chou's book include an original list of 366 theorems from [4], of which 219 are reported to be linear [5, p. 12].

are 2^r (or 2^{r-1} by symmetry) possible sign assignments to the radicals in $G(\mathbf{u})$. Our prover has three verification modes: (1) random mode, (2) exhaustive mode, and (3) specified mode. These correspond, respectively, to testing (1) a random sign assignment, (2) all sign assignments and (3) a user-specified assignment. For linear theorems, these modes are irrelevant. In the above table, we test Simson's theorem in the exhaustive mode, Pascal's theorem in the random mode and Tribisector in the specified mode. So our timing for Pascal's theorem should really be multiplied by $2^4 = 16$.

It is interesting to note that we have never observed a single wrong conclusion from our probabilistic tests – all true theorems are reported as true, and all perturbed theorems are reported as false. In some sense, that is not surprising because the probabilistic bounds based on Schwartz's lemma seem overly conservative in all real situations.

The running times for linear theorems are pretty consistent across different runs. However, for the non-linear theorems, the timing can show much more variation. This is not unexpected since the running time depends on the bit size of the random example. A more prominent behavior comes from the clustering of times around certain values. For instance, for Simson ($c = 20$), the times cluster around 10 seconds and around 70 seconds. This "multimodal" behavior of the timings are again seen in Pascal. This can be attributed to the random choice of signs for the radicals in non-linear theorems. This may also account for the curious relative times for Simson $c = 10$ and $c = 20$.

The performance of our library is critically dependent of good root bounds (an area of research that we are actively working on [16]). It should be possible to exploit prover-specific techniques to improve the speed, but this has not been done. There are several issues to bear in mind when comparing our method with Wu's method:

- Chou's timings would look considerably better using hardware available today.
- The actual theorems proved by Wu's method are not strictly comparable to ours in two important aspects: Wu's method proves theorems about complex geometry while ours is about real geometry. On the other hand, Chou's algorithm is deterministic while ours is probabilistic.
- Our method is extremely effective for discarding wrong or perturbed conjectures. It is unclear if Wu's method will be much faster for perturbed theorems, since the algorithm would still have to execute the same basic steps. The ability to quickly reject false theorems is extremely useful in applications where the user has many conjectures to check but most of the conjectures are likely to be false.
- One of the strengths of Wu's methods (as compared to Gröbner bases, say) is its ability to discover non-degeneracy conditions. A similar capability is embedded in our approach – this simply amounts to detecting when a construction step is ill-defined.

5 Final Remarks

In this paper, we have developed a generalization of the Schwartz-Zippel randomized zero test for the class of radical expressions. Such a test is expected to have many applications as radical expressions are quite common. Here, we focus on their use in proving theorems about ruler & compass constructions. Some features of our prover are:

- It proves theorems about real (rather than complex) geometry, under the limitation that there is no inequalities appearing in the thesis.
- It is probabilistic, so that speed can be traded-off against error probability.
- It detects wrong conjectures very quickly.
- It is extremely effective for linear theorems (the majority of the theorems in [5]).
- It exploits the special nature of ruler & compass constructions.

Because of the last feature, our approach may ultimately prove to be more efficient for *this* class of problems than other more general techniques. However, our results so far have not been conclusive in the case of non-linear theorems. The following are some open problems:

- Improve our zero test for straight line programs that involve division.
- Develop techniques to make our approach faster for non-linear theorems.
- Extend our randomized techniques to theorems that have inequalities in the theses. This seems to call for radically new ideas.

Acknowledgments. Daniela Tulone would like to acknowledge Alfredo Ferro and Pina Carrà for their advise and discussions in the initial stages of this project.

References

1. G. Carrà Ferro and G. Gallo. A procedure to prove geometrical statements. In L. Huguet and A. Poli, editors, *Proc. 5th Int. Conf. on Applied Algebra, Algebraic Algorithms and Error-Correcting Codes*, volume 356 of *LNCS*, pages 141–150, Berlin, June 1989. Springer.
2. G. Carrà Ferro, G. Gallo, and R. Gennaro. Probabilistic verification of elementary geometry statements. In D. Wang, editor, *Proceedings of the International Workshop on Automated Deduction in Geometry (ADG-96)*, volume 1360 of *LNAI*, pages 87–101, Berlin, Sept. 27–29 1997. Springer.
3. S.-C. Chou. Proving elementary geometry theorems using Wu's algorithm. *Contemporary Mathematics*, 29:243–286, 1984.
4. S.-C. Chou. Proving geometry theorems using Wu's Method: A collection of geometry theorems proved mechanically. Technical Report 50, Institute for Computing Science, University of Texas, Austin, July 1986.
5. S.-C. Chou. *Mechanical Geomtry Theorem Proving.* D.Reidel Publishing Company, 1988.
6. P. Conti and C. Traverso. Proving real geometry theorems and the computation of the real radical. In *Proceedings of the Third International Workshop on Automated Dedu ction in Geometry (ADG 2000)*, Sept. 2000.

7. A. Ferro and G.Gallo. Automatic theorem proving in elementary geometry. *Le Matematiche*, XLIII(fasc. I):195–224, 1988.
8. G. Gallo. *La Dimostrazione Automatica in Geometria e Questioni di Complessitá Correlate.* Tesi di dottorato, University of Catania, Italy, 1989.
9. G. Gallo and B. Mishra. Efficient algorithms and bounds for Wu-Ritt characteristic sets. In F. Mora and C. Traverso, editors, *Effective Methods in Algebraic Geometry (Proc. MEGA'90)*, volume 94 of *Progress in Mathematics*, pages 119–142. Birkhauser, Boston, 1991.
10. G.Gallo and B.Mishra. Wu-Ritt characteristic sets and their complexity. In *Computational Geometry: Papers from the DIMACS Special Year*, volume 6, pages 111–136. AMS and ACM, 1991.
11. J.-W. Hong. Proving by example and gap theorem. *27th Annual Symposium on Foundations of Computer Science*, pages 107–116, 1986.
12. D. Kapur. Using Groebner bases to reason about geometry problems. *J. Symbolic Comp.*, 2:399–412, 1986.
13. V. Karamcheti, C. Li, I. Pechtchanski, and C. Yap. A core library for robust numeric and geometric computation. In *Proc. 15th ACM Symp. on Computational Geometry*, pages 351–359, June 1999. Miami Beach, Florida.
14. B. Kutzler and S. Stifter. Automated geometry theorem proving using Buchberger's algorithm. *Proc. Symp. on Symbolic and Algebraic Computation*, pages 209–214, 1986.
15. C. Li. *Exact Geometric Computation: Theory and Applications.* PhD thesis, Courant Institute of Mathematical Sciences, New York University, Jan. 2001. URL: http://www.cs.nyu.edu/csweb/Research/theses.html.
16. C. Li and C. Yap. A new constructive root bound for algebraic expressions. In *Proceedings of the Twelfth ACM-SIAM Symposium on Discrete Algorit hms (SODA 2001)*, pages 496–505. ACM and SIAM, Jan. 2001.
17. E. W. Mayr and A. R. Meyer. The complexity of the word problems for commutative semigroups and polynomial ideals. *Advances in Mathematics*, 46:305–329, 1982.
18. K. Mehlhorn and S. Schirra. A generalized and improved constructive separation bound for real algebraic expressions. Technical Report MPI-I-2000-004, Max-Planck-Institut für Informatik, Nov. 2000.
19. K. Ouchi. Real/Expr: Implementation of an exact computation package. Master's thesis, New York University, Department of Computer Science, Courant Institute, January 1997.
20. J. T. Schwartz. Probabilistic verification of polynomial identities. *J. ACM*, 27(4):701–717, Oct. 1980.
21. W.-T. Wu. On decision problem and the mechanization of theorem proving in elementary geometry. *Scientia Sinica*, 21:157–179, 1978.
22. W.-T. Wu. Some recent advances in mechanical theorem proving of geometries. In *Automated Theorem Proving: After 25 Years*, volume 29 of *Contemporary Mathematics*, pages 235–242. American Mathematical Society, Providence, Rhode Island, 1984.
23. W.-T. Wu. Basic principles of mechanical theorem proving in elementary geometries. *Journal of Automated Reasoning*, 2(4):221–252, 1986.
24. C. K. Yap. A new lower bound construction for commutative Thue systems, with applications. *Journal of Symbolic Computation*, 12:1–28, 1991.
25. C. K. Yap. Robust geometric computation. In J. E. Goodman and J. O'Rourke, editors, *Handbook of Discrete and Computational Geometry*, chapter 35, pages 653–668. CRC Press LLC, 1997.

26. C. K. Yap. Towards exact geometric computation. *Computational Geometry: Theory and Applications*, 7:3–23, 1997. Invited talk, Proceed. 5th Canadian Conference on Comp. Geometry, Waterloo, Aug 5–9, 1993.

27. R. Zippel. *Effective Polynomial Computation.* Kluwer Academic Publishers, 1993.