# 北京大学

# 博士研究生学位论文

## 题目：柱形代数分解的几何理论


姓　　名：陈日增

学　　号：2101110063

院　　系：数学科学学院

专　　业：应用数学

研究方向：实代数几何

导师姓名：夏壁灿


☑ 学术学位　　☐ 专业学位

二〇二六 年 五 月

# 版权声明

# 摘要

柱形代数分解是实代数几何研究中的基本核心工具之一. 它不仅为实闭域理论, 量词消去算法等实代数几何的理论基石提供了关键构造, 还在自动定理证明, 生化反应网络分析, 可满足性检验乃至机器人运动规划等诸多领域中得到广泛应用. 虽然文献中已有多种计算柱形代数分解的算法, 但是其实际效率在面对复杂系统时仍面临显著瓶颈. 其中一个最突出的挑战是, 算法计算复杂度会在针对大型多项式系统构建分解时急剧上升. 导致这一现象的核心原因在于现有算法未能充分和系统地利用问题本身所蕴含的丰富代数结构 (方程信息), 使得投影阶段产生大量冗余胞腔并形成性能瓶颈.

本论文旨在从更本质的几何观点出发, 借助现代代数几何的理论框架来重新审视并革新柱形代数分解的构造问题. 我们在几何纤维基数不变性 (一个纯粹的几何性质) 和半代数连续截面的存在性 (一个半代数性质) 间建立了一个基础性的对应关系. 这一发现不仅在实代数几何与其复对应理论之间建立了深刻的联系, 也为分解的构造提供了清晰的几何解释. 基于这一核心结论, 我们证明了一般的柱形投影总可被有效地分层为一系列具有半代数截面的规则片段. 这直接引出了一个全新的算法设计理念: 在投影的每一步, 系统且完整地利用所有已知的等式信息来引导分层过程, 从而避免不必要的分支. 实验结果表明, 这一新算法在多个标准测试集上表现出了极高的效率, 显著优于传统方法.

受上述柱形代数分解几何方法的启发, 本文进一步将相关思想推广至更一般的映射层面, 研究了实代数簇间覆叠映射的可计算判定判据. 我们证明了: 对于具有局部常值几何纤维的实簇拟有限平坦态射, 其在实点上的映射必构成覆叠映射. 因它能将覆叠性质的验证归结为纯代数条件, 这一判定准则具有重要的理论价值; 尤为值得注意的是, 该判据对奇异簇情形同样适用, 从而扩展了现有理论的适用范围. 此外, 判据中所依赖的 “平坦性” 与 “局部常值几何纤维” 条件, 均可通过本文设计的配套算法进行有效验证, 确保了其在实际中的可计算性.

作为柱形代数分解理论能力的应用拓展, 我们界定并研究了一类被称为 “三角扩张” 的特殊超越问题. 简而言之, 三角扩张是指通过向一个仅包含有限个实零点的函数环 (作为基环) 上添加三角函数而生成的一类单变量解析函数环. 这类形式化体系刻画了实践中一大类同时包含多项式约束与基本指数以及三角函数的混合系统. 我们证明, 在此类问题中, 三角函数的固有周期性可以通过柱形分解的框架得到系统性利用. 其核心机制为利用分解过程识别并压缩周期结构, 从而将解空间的搜索范围从整个实轴安全地缩减至一个可精确计算的有界区间. 这为有效求解一类长期被视为困难的混合代

数超越约束问题提供了可行牢靠的方法.

关键词：柱形代数分解，实代数几何，有限平展态射，超越约束求解。

# Inside the Geometry of Cylindrical Algebraic Decomposition

Rizeng Chen (Applied Mathematics)
Directed by Prof. Bican Xia

## ABSTRACT

Cylindrical algebraic decomposition is a fundamental tool in the study of real algebraic geometry. It has not only provided a crucial construction for the theory of real closed fields and quantifier elimination, which are the theoretical foundation of real algebraic geometry, but has also been applied in various fields such as automatic theorem proving, biochemical network analysis, satisfiability checking and robotic motion planning. Although there are many algorithms to compute a cylindrical algebraic decomposition, their practical performance is still very limited when facing a complicated polynomial system. A key challenge is the surge of computational complexity when constructing a cylindrical algebraic decomposition for a large polynomial system. The main reason for this is that the existing algorithms fail to exploit the equations systematically, resulting in too many cells during the projection and leading to a bottleneck.

In this dissertation, we revisit this problem from a more geometric perspective, where the construction of cylindrical algebraic decomposition is studied through the lens of modern algebraic geometry. It is showed that the geometric fiber cardinality (geometric property) decides the existence of semi-algebraic continuous sections (semi-algebraic property). This has not only bridged real algebraic geometry and its complex counterpart, but also provides a clear geometric insight for the construction of the decomposition. Based on this, we proved that a general cylindrical projection can always be effectively stratified into pieces admitting semi-algebraic sections. This directly led to a completely new algorithmic concept: at each stage of the projection, all known equation information is systematically and completely exploited to guide the stratification process, thereby avoiding unnecessary cells. Experimental results showed that this new algorithm exhibits extremely high efficiency on multiple benchmarks, significantly outperforming traditional methods.

Inspired by the geometric approach to cylindrical algebraic decomposition, we further lift this idea to more general morphisms, and investigate an effective criterion for covering maps between real algebraic varieties. We showed that for a quasi-finite flat morphism between real varieties with locally constant geometric fibers, the map on the real points is a covering. This criterion is important, as it reduces the verification of covering properties to purely algebraic conditions. Remarkably, this criterion is even applicable to singular varieties, thereby extending the scope of existing theory. Furthermore, the conditions of the criterion (“flatness” and “locally constant geometric fibers”) can be effectively verified by the algorithms developed in this dissertation, ensuring its practical computability.

As an application that extends the theoretical capability of cylindrical algebraic decomposition, we identify and study a special class of transcendental problems termed “trigonometric extension”. Formally, a trigonometric extension is a ring of univariate analytic functions generated by adjoining trigonometric functions to a base ring of functions possessing only finitely many real zeros. This formalism captures a broad and practical class of mixed systems involving both polynomial constraints and basic exponential/trigonometric functions. We demonstrate that for this significant class of problems, cylindrical algebraic decomposition becomes applicable to utilize the periodicity of trigonometric functions. Specifically, the decomposition process can identify and compress these periodic structures, thereby safely reducing the solution search space from the entire real axis to a single, explicitly computable bounded interval. This provides a concrete and feasible pathway for solving a category of mixed algebraic-transcendental problems that have long been considered challenging.

KEY WORDS: Cylindrical Algebraic Decomposition, Real Algebraic Geometry, Finite Étale Morphism, Transcdental Constraints Solving.

# Contents

# Chapter 1 Introduction

## 1.1 Background

Human has been dreaming of carrying out all the mathematical proofs automatically for centuries. Leibniz, now commonly known as one of the inventor of calculus, was one of first people dared to dream so. He developed a machine, *Staffelwalze*, allowing all four basic arithmetic operations for the first time in the history and he was then led to comparing logical reasoning to mechanism [Cou01, Page 115]. Since then, people realized that it might be possible to build a machine that performs calculations to determine the truth value of a mathematical statement.

The enthusiasm was put to the highest peak in the first half of the twentieth century after Hilbert proposed the program named after him [HA28]. The program called for

1. a formalization of all mathematics in a proof system (the first-order logic), and
2. a method that would determine in a finite number of well-defined effective steps whether a first-order statement is true or not.

The call of the method is known as "Hilbert's *Entscheidungsproblem*", where the German word stands for "decision problem", a terminology still being used today. Hilbert and his collaborators were very optimistic towards an answer for Entscheidungsproblem.

Shortly after, an answer was found by a young logician named Kurt Gödel [Göd31]. However, it was not the answer that Hilbert really anticipated but quite the opposite. Gödel showed that from comparatively weak systems like Peano Arithmetic to strong systems like Whitehead and Russell's *Principia Mathematica*, there are always true propositions expressible in that system that cannot be proved. Later Church and Turing independently proved that there is no algorithm to determine the truth value of a first-order statement [Chu36; Tur36]. These results negate the expectations put forward in Hilbert's program.

But not all hopes are lost. Near the same time as Gödel published his incompleteness theorem, Tarski found that the theory of real closed fields $\langle R, +, \times, <, 0, 1\rangle$ admits quantifier elimination, thus allowing effective algorithms to decide statements in elementary algebra and geometry [Tar51]. However, the publication was postponed because of the outbreak of the World War II. By the time Tarski's paper was published, digital computers had emerged and Tarski's method drew attentions from researchers in the industry. McKinsey from RAND corporation was one of them.

McKinsey's goal was to build a machine that employs Tarski's method to solve problems from games. He wrote to von Neumann to ask if it is possible to build a special-purpose machine to undertake the task. In the response letter, von Neumann pointed out that it is not the type of machine that matters but the running time, and he showed skepticism towards a practical implemention of Tarski's method because of its combinatorial nature [Neu05, Page 178-180]. Two years later, McKinsey left RAND and it seems that none of his plans were carried out.

In fact, the complexity of Tarski's method cannot be bounded by any finite tower of exponential functions. Even for very small instances, Tarski's method becomes unmanageable by modern computers. So any effort trying to implement Tarski's method is doomed to fail and it is important to find a more feasible alternative approach. Cylindrical algebraic decomposition (c.a.d.), proposed by Collins, is the first algorithm that is feasible in practical computations. Roughly speaking, the goal of it is to decompose semi-algebraic sets into cylindrically arranged cells that are homeomorphic to open balls of various dimensions. The study of cylindrical algebraic decomposition began with Collins's landmark paper [Col75]. Though Collins's original motivation was to develop an algorithm that improves Tarski's method, it is now proved that the underlying idea also plays an important role in studying the geometry of semi-algebraic sets and cylindrical algebraic decomposition has become a fundamental tool in real algebraic geometry.

Collins's algorithm takes a family of $n$-variate polynomials $\mathcal{P}_n$ as input (e.g. all polynomials that occur in a first-order formula), and it returns a cell decomposition of the $n$-dimensional space where the cells are arranged cylindrically and each input polynomial is sign-invariant on every cell. Collins's strategy for cylindrical algebraic decomposition consists of two phases: repeated projection and lifting. In the projection phase, the algorithm repeatedly "projects" polynomials. To be more precise, it computes

- a family of $(n-1)$-variate polynomials $\mathcal{P}_{n-1}$ from the input $\mathcal{P}_n$,
- then a family of $(n-2)$-variate polynomials $\mathcal{P}_{n-2}$ from the previous result $\mathcal{P}_{n-1}$,
- ...
- finally a family of univariate polynomial $\mathcal{P}_1$ from $\mathcal{P}_2$.

Then the algorithm moves on to the lifting phase. In the lifting phase, the algorithm successively constructs sign-invariant cells for $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_n$. See Figure 1.1.

Often the projection phase takes a long time, leading to a bottleneck. Hence there have been exhaustive efforts on improving the projection phase by selecting a smaller family of

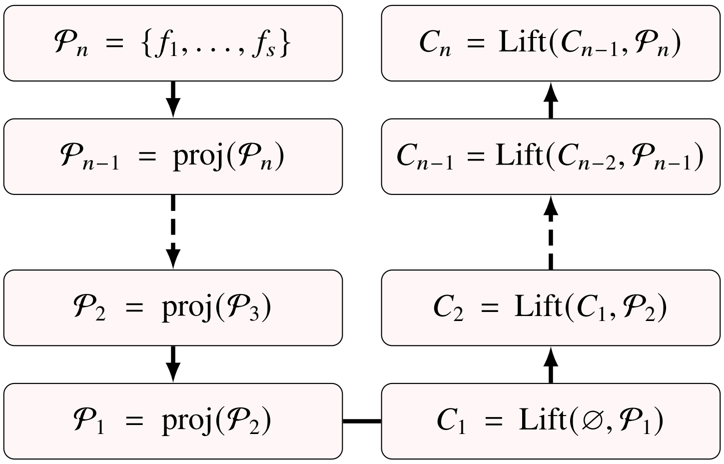



Figure 1.1 The workflow of a classical c.a.d. algorithm

"projection" polynomials, e.g. [Hon90], [Laz94], [McC98] and [Bro01].

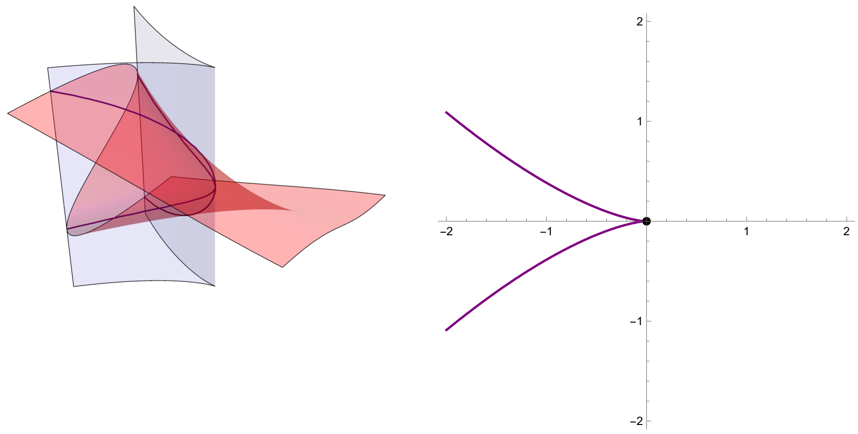
## 1.2 Challenge

It was soon noticed that some information can be used to accelerate the computation of a cylindrical algebraic decomposition, especially the equations. For some applications, there are already equations in the input (e.g. Real Root Classification, Automatic Theorem Proving). Also, equations are generated during the projection phase.

To illustrate this observation, let us consider a toy polynomial system $x^3 + px + q = 4p^3 + 27q^2 = 0$ in parameters $p, q$ and variable $x$. It defines a curve in the three-dimensional space with two irreducible components intersecting at the origin. The curve is depicted in Figure 1.2(a) in purple, as the intersection of two surfaces defined by $x^3 + px + q = 0$ in red and $4p^3 + 27q^2 = 0$ in blue, respectively. One wants to study the number of real roots of the system. This is a classical application of cylindrical algebraic decomposition. Certainly there are equations in the input and extra equations are generated during the projection ($p = q = 0$, reflecting the intersection of two irreducible components). See Figure 1.2(b).

For larger non-trivial inputs, this phenomena become much more frequent: many equations are already in the input and many additional equations are systematically generated during the projection phase. Therefore it is an important challenge to exploit these equations.

Another challenge comes from non-algebraic constraints. For example, exponential functions and trigonometric functions naturally arise in solutions of linear ordinary differential equations, which model the trajectory of some simple dynamic systems. Tarski himself has considered the decidability of the theory of the real numbers together with the exponential




(a) The curve as intersection of two surfaces

(b) The projection image

Figure 1.2 Equations naturally arise

function (known as Tarski's exponential function problem). However, due to the lack of tools and the hardness of transcendental number theory, results are rare in this direction.

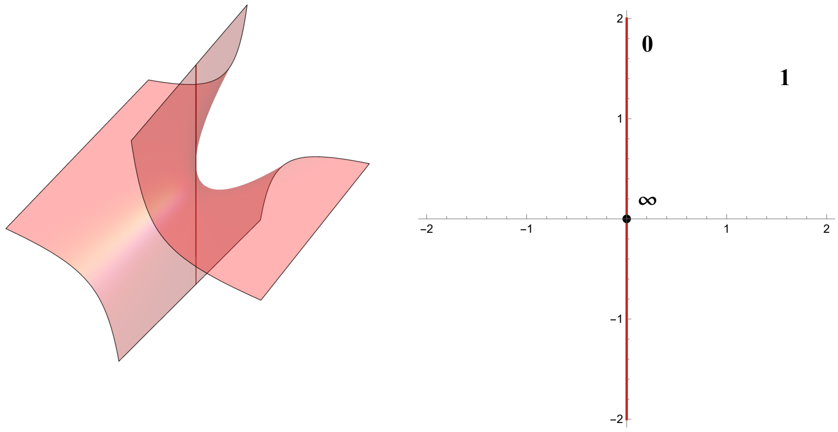
## 1.3 Previous Results

There have been intensive efforts regarding the issue of exploiting equations in c.a.d. computation and a significant process is made to utilize the equational constraints in the input in [McC99; McC01], based on McCallum's projection operator [McC98]. The main idea is to select an equation (if there are any) at each stage of projection and decompose the hypersurface defined by the equation only and implicit equations are discovered by resultants of known equations. Later McCallum and Brown generalized this to two equations [BM05; MB09]. It is also found that savings can be made during the lifting phase [EBD15]. For a recent summary on utilizing equational constraints in cylindrical algebraic decomposition, we refer readers to [EBD20].

However, there is still more to be done.

1. The previous results did not make full use of the equations. Only one or two equations are used as "pivot" in each stage of projection. And they ignored many equations emerging in the middle of projection (e.g. the origin in Figure 1.2(b)).
   To further demonstrate this, let us consider another toy equation $px + q = 0$, where $p, q$ are parameters and $x$ is the unknown. Its locus is depicted in Figure 1.3(a). Clearly, the equation has exactly one solution when $p \neq 0$, zero solution when $p = 0$ but $q \neq 0$, and infinite solutions when $p = q = 0$. Therefore during the projection, equations $p = 0$

and $p = q = 0$ naturally arise. However, the latter is neglected in classical algorithms. Instead, they compute a $p, q$-sign-invariant c.a.d. of the $pq$-plane. As a result, the $pq$-plane is divided into 9 cylindrical cells (four quadrants, four half-axes and the origin), although 5 is enough ($p > 0$, $p < 0$, the positive and negative half-axes of $q$, and the origin).




(a) The surface $V(px + q)$, note that the $x$-axis is contained in the surface

(b) Number of solutions in $x$

Figure 1.3 Some equations are ignored

2. The result in [McC01] works with irreducible polynomials. A complicated preprocessing needs to be done to apply their result.
3. McCallum's projection operator [McC98] could fail in rare cases. A key property "order-invariance" is not maintained during the lifting phase, if the projection of a hypersurface is not quasi-finite. As a result, the correctness is not guaranteed in this case. This is called the "Nullification Problem" in the literature. For such an example, we refer readers to Example 4.

Although our goal is to find the truth value of a logic statement, the difficulty of utilizing equations comes from the geometry side. Indeed, the conjunction of several equations correspond to an algebraic variety. However, a classical algorithm only handles some families of hypersurfaces (i.e. solution set of a single equation) and it is not suitable for a general variety. Readers may recall Figure 1.1 for an overview of a classical algorithm. Therefore, in order to exploit all the equations systematically, we must develop new results that enable cylindrical algebraic decomposition to directly handle arbitrary varieties. Here, algebraic geometry naturally comes into play and we are led to study effective criteria for covering maps between real varieties.

There have been studies about covering maps between real varieties from the theoretical side. Delfs and Manfred discussed covering maps in the category of locally semi-algebraic spaces [DK84, Section 5], generalizing the notion of semi-algebraic spaces to study the topology of semi-algebraic sets, especially their homology and homotopy [DK81a; DK81b; DK85; Kne89]. Schwartz characterized a covering map $f$ of locally semi-algebraic spaces by the associated morphism of the corresponding real closed spaces being finite and flat [Sch88]. The theory of real closed spaces is introduced by Schwartz as a generalization of locally semi-algebraic spaces [Sch89]. Baro, Fernando and Gamboa studied the relationship between branched coverings of semi-algebraic sets and its induced map on the spectrum of the ring of continuous semi-algebraic functions in [BFG22]. A common feature of these works is the use of (various generalizations of) semi-algebraic spaces. Hence, the sheaf equipped is the sheaf of continuous semi-algebraic functions. This is not really convenient for our purpose, because the family of continuous semi-algebraic functions is not of finite type over $R$ (it is not even noetherian) and fails to be a computable object. We shall seek some criterion that is expressible from the scheme structure. Lazard and Rouillier provided a sufficient condition for $f : X \to Y$ to be a covering in [LR07], but the assumption is too strong (both $X$ and $Y$ are non-singular, $X$ is equi-dimensional, $f$ is finite, all fibers are reduced).

As for non-algebraic problems, the literature is relatively small and the decidability of a large part of transcendental theory remains unknown. Many efforts have been put into showing the decidability or undecidability of certain theories of transcendental functions. Assuming Schanuel's Conjecture (see Conjecture 1), Macintyre and Wilkie showed that the real exponential field $\mathbb{R}_{\exp}$ is decidable [MW96], providing a partial answer to Tarski's exponential function problem. Achatz, McCallum and Weispfenning proposed an algorithm to isolate real roots of polynomial-exponential functions and showed the unconditional decidability of a special class of polynomial-exponential problems [AMW08]. Later McCallum and Weispfenning generalized their work to the logarithmic function and the inverse tangent function [MW12]. Strzeboński studied the real root isolation of exp-log functions [Str08], tame elementary functions [Str09] and exp-log-arctan functions [Str12], the termination of his algorithms relies on Schanuel's Conjecture.

When trigonometric functions are taken into account, the corresponding problem becomes very complicated. Succeeding Richardson [Ric68], Matiyasevich [Mat73], Caviness [Cav70] and Wang [Wan74], Laczkovich proved that there is no algorithm to test whether a polynomial in $x$, $\sin(x^n)$ and $\sin(x\sin(x^n))$ $(n = 1, 2, \ldots)$ has a real root or not [Lac03]. The undecidability of a general theory involving trigonometric function is mainly due to the unsolvability of Diophantine equations (there is no algorithm to find integer solutions of polynomial equations, Davis-Putnam-Robinson-Matiyasevich Theorem) and an unlimited use of trigonometric functions would allow us to encode integers in the theory. Still, there are hopes in some special theories. Anai and Weispfenning gave an algorithm to decide linear-trigonometric problems [AW00]. Macintyre showed that the theory $\mathbb{R}_{\exp,\cos\restriction[0,n],\sin\restriction[0,n]}$, the reals with the exponential function and the restricted trigonometric functions (i.e. the trigonometric functions are only interpreted in a bounded interval $[0, n]$ and are set to 0 elsewhere.) is decidable, provided Schanuel's Conjecture is true [Mac16]. Chen and Liu proposed an algorithm to prove mixed trigonometric-polynomial inequalities on a bounded interval [CL20].

## 1.4 Contribution

In this dissertation, we develop a new simple algorithm that exploits all the equations, based on the study of finite free morphisms of affine real varieties.

1. All the equations (from both the input and the intermediate stages) are used.
2. It does not rely on radical (square-free) computation nor primary decomposition (factorization), so it is conceptually simple.
3. It is a complete algorithm that never fail (no nullification problem).
4. Most importantly, it is generally faster than the existing methods. A benchmark is reported in Section 3.4.2.

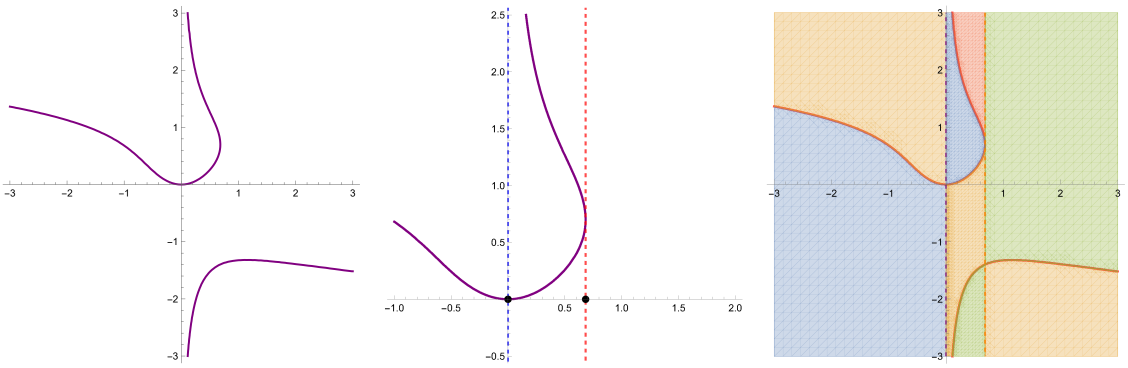
The key result enabling all these advances is the following:

**Main Theorem A** (Theorem 3.3.7)**.** *Suppose* $\pi : X \underset{\text{closed}}{\hookrightarrow} \mathbb{A}^1_Y \to Y$ *is a finite free morphism between affine* $R$*-varieties over a real closed field* $R$*. Let* $S \subseteq Y(R)$ *be a semi-algebraic connected semi-algebraic set on which* $\#X_{\overline{y}}$ *is a constant, then there are semi-algebraic continuous functions* $\xi_1, \ldots, \xi_l : S \to \pi_R^{-1}(S)$ *such that* $\pi \circ \xi_i = \mathrm{id}_S$ *and* $\{\xi_i(y)\}_{i=1}^l$ *is the real fiber of* $y$ *in* $X(R)$*.*

It shows that a finite free morphism $f : X \to Y$ of affine real varieties admits semi-algebraic continuous sections on regions having a constant geometric fiber size, provided that $X$ is a closed subscheme of the cylinder $\mathbb{A}^1_Y$. This is a non-trivial generalization of [Col75, Theorem 1] in the algebro-geometric context, connecting a geometric invariant (cardinality of geometric fiber) and a semi-algebraic property (existence of semi-algebraic continuous sections). Thus the theorem itself provides a link between real and complex algebraic geometry.

The result is turned into an algorithm by an effective version of Grothendieck's Generic Freeness and a geometric fiber counting technique (Hermite quadratic form). Very roughly speaking, our algorithm starts with some varieties. Then we stratify their projection images into "flat" (fibers changes continuously) and "unramified" (fibers do not break into different branches) pieces.

We illustrate the above idea with a c.a.d. of the plane. The same idea applies to high-dimensional varieties. Let $f = xy^3 - y + x^2$, the goal is to compute an $f$-sign-invariant c.a.d. of the plane.




(a) Locus of $f$ (b) Points where "flatness" or "unramifiedness" fails (c) The final c.a.d.

Figure 1.4 An example of c.a.d.

We study the projection of the curve $V(f)$. It can be seen that the fibers change continuously when $x \neq 0$, so the affine line is split into two parts where the projection is "flat". Notice that the fiber ramifies when $27x^5 - 4 = 0$ (e.g., $x = \sqrt[5]{4/27}$ on the real axis). Therefore the $x \neq 0$ part is further divided according to whether $27x^5 - 4$ is zero or not.

So the projection is "flat" and "unramified" over

- $x \neq 0 \wedge 27x^5 - 4 \neq 0$,
- $x \neq 0 \wedge 27x^5 - 4 = 0$,
- $x = 0$.

By Main Theorem A, the projection of $V(f)$ admits sections over each semi-algebraically connected components of these regions. As a result, a c.a.d. is constructed over them, see Figure 1.4(c).

The crucial difference of our approach and the existing methods is the systematic use of equations, which requires a deeper understanding of morphisms of real varieties instead of just hypersurfaces. The algebro-geometric result Main Theorem A is the essential ingredient to the new algorithm.

Next, we further seek criteria for covering maps between real varieties and generalize Main Theorem A.

**Main Theorem B** (Theorem 4.2.1)**.** *Suppose $\pi : X \to Y$ is a quasi-finite flat $R$-morphism with locally constant geometric fibers of $R$-varieties over a real closed field $R$. Then $\pi_R : X(R) \to Y(R)$ is a covering map in the Euclidean topology.*

Notice that no assumptions (e.g. reduced, equi-dimensional, non-singular, etc) are posed on the source $X$ and base $Y$. The result is very sharp, as we will see in Example 9 that even a slight weakening of the assumptions on $\pi$ leads to counter-examples. Also, it includes several known results.

During the development of Main Theorem B, we also prove a result that might be of independent interests.

**Main Theorem C** (Theorem 4.1.4)**.** *Let $k$ be a field of characteristic 0. If $\pi : X \to Y$ is a quasi-finite flat $k$-morphism of $k$-varieties with locally constant geometric fibers, then $\pi_{\text{red}} : X_{\text{red}} \to Y_{\text{red}}$ is a finite étale morphism.*

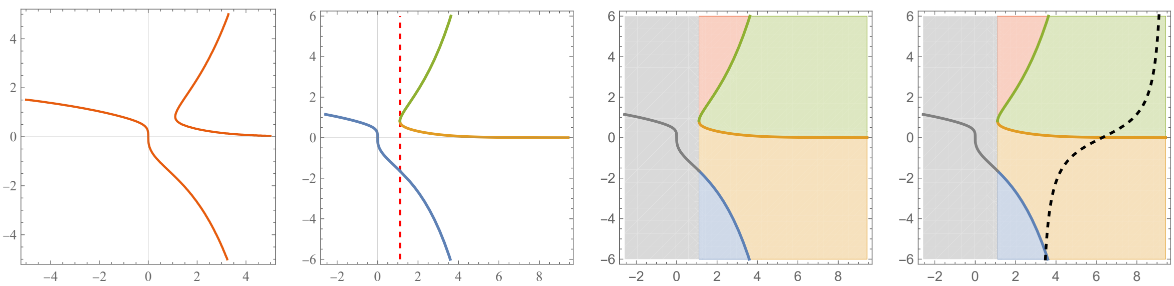
So for a quasi-finite flat morphism of $k$-varieties ($\operatorname{char} k = 0$), if the cardinality of geometric fiber is a constant, then the ramification is purely due to the non-reduced structures of $X, Y$ and can be resolved by a reduction. Moreover, the flatness is also preserved (notice that the reduction of a flat map is not necessarily flat in general). Proving that $\pi_{\text{red}}$ is flat is actually the main difficulty in establishing Main Theorem C. Our strategy is to use a dévissage-type argument to reduce to showing that $X_{\text{red}}$ is locally isomorphic to the spectrum of a finite standard étale algebra over an affine open of $Y_{\text{red}}$.

Finally, we apply ideas from cylindrical algebraic decomposition to study a special family of theories that contains trigonometric function. It is observed in [Mac16] that, when trigonometric functions are restricted to a bounded interval, the "bounded" theory is significantly simpler, while the "unbounded" theory $\mathbb{R}_{\exp,\cos,\sin}$ is much more complicated, linked to an undecidable theory $\mathbb{C}_{\exp}$. This observation is also supported by other results like [Str09], [MW12], [CL20]. Therefore, a reduction from an unbounded problem to a bounded problem could be favorable: if there is a decision algorithm for the bounded problem, then the reduction provides a decision algorithm for the unbounded problem. Similarly, if the unbounded problem is undecidable, then researchers should not waste their time on the bounded version either.

We identify a case called "trigonometric extension", where every unbounded problem can be effectively reduced to an equivalent bounded problem.

**Main Theorem D** (Theorem 5.3.1)**.** *Let $S$ be a subring of univariate analytic functions in $x$ that satisfies certain finiteness and computability conditions (see Section 5.1 for a detailed setting). The language $\mathcal{L}(S[\sin x, \cos x])$ is decidable if and only if there exists an algorithm $A(\varphi, N, M)$ that computes the truth value of $(\forall x)((N \leq x \leq M) \to \varphi(x))$, where $\varphi \in \mathcal{L}(S[\sin x, \cos x])$ is quantifier-free and $N, M \in \mathbb{R}$.*

The core idea of the above results is to exploit the periodicity of trigonometric function, so the root distribution of a ring element follows a similar pattern for all $x$ with sufficiently large absolute value. See Figure 1.5 for a demonstration. So instead of looking for all $x \in \mathbb{R}$, it suffices to look for $x$ in a bounded interval. This reduces an unbounded problem to a bounded problem, which is much easier.



(a) The curve defined by $y^3-(e^x-1)y+x=0$.

(b) For sufficiently large $x$, the curve breaks into three branches.

(c) The right half-plane is then partitioned into cylindrically stacked cells.

(d) The trigonometric function $y = \tan\frac{x}{2}$ will run through each cell in each period.

Figure 1.5 The root distribution of $\tan^3\left(\frac{x}{2}\right)-(e^x-1)\tan\left(\frac{x}{2}\right)+x=0$ is related to the periodicity of trigonometric functions

The reduction theorem (Main Theorem D) implies several decidability results. For example, if Schanuel's Conjecture is true, then the trigonometric extension of the ring of analytic exp-log-arctan functions is decidable (a precise definition can be found in Definition 5.3.3.) As a corollary, we prove that a special class of reachability problem is decidable when assuming Schanuel's Conjecture. These results are remarkable because they provide a (conditional) algorithm to decide certain theories that involves trigonometric functions, while the general theory is undoubtedly undecidable. Moreover, the multivariate case is also discussed. We show that a fragment of the multivariate theory can be reduced to the univariate case, but the general theory is undecidable.

# Chapter 2 Preliminary

Since this dissertation utilize tools from several different branches of mathematics, a brief summary of the required theories for readers' convenience is provided in this chapter. We discuss about some basic scheme theory in the modern algebraic geometry, a few definitions needed to study real algebraic geometry and some constructive methods. If the reader is well acquainted with them, then this chapter can be skipped and the reader can come back later when necessary. If the reader is not familiar with some of the concepts listed above, then this chapter should serve as a quick summary of the only basic notion and results that will be used in this dissertation.

In this dissertation, all rings are commutative with a multiplicative identity. Unless otherwise specified, all fields are of characteristic 0.

## 2.1 Schemes

We will state our results using Grothendieck's language of schemes. While it is true that most of our results can be formulated in the classical language of varieties, they will become very complicated and less intuitive. Since we are doing algebraic geometry over a non-algebraically closed field ($\mathbb{R}$ for instance) and prime/radical ideals are rare in computation, schemes should serve as the right settings (we do not require varieties to be integral for the same reason).

Here we provide some basic notions and results about schemes. Reader more familiar with the classical language of varieties can also take a look at [EH00] or [GP08, Appendix A.3-A.5] for a comparison of classical and modern algebraic geometry.

Our main reference will be Hartshorne's classical textbook *Algebraic Geometry* [Har77], supplemented by Liu's *Algebraic Geometry and Arithmetic Curves* [Liu02]. In case that a terminology is not explained here, please refer to [Har77] or [Liu02].

### 2.1.1 Affine Varieties

Let $S$ be a ring. The **spectrum** of $S$, denoted by $\operatorname{Spec} S$, is a topological space together with a sheaf of regular functions on it. The topological space of $\operatorname{Spec} S$ is the set of prime ideals in $S$, equipped with the topology that takes closed sets as

$$V(I) = \{p \in \operatorname{Spec} S | I \trianglelefteq S,\ I \subseteq p\}.$$

This topology is named after Zariski. The **distinguished open sets** in $\operatorname{Spec} S$ are the sets of the form

$$D(h) = \{p \in \operatorname{Spec} S | h \in S,\ h \notin p\},$$

which is the complement of $V(\langle h \rangle)$. Distinguished open sets form a base for the Zariski topology [Har77, Section II.2]. Then $\mathbb{A}^n_S$, the **affine $n$-space** over $S$ is the spectrum of $S[x_1, \ldots, x_n]$, i.e. $\mathbb{A}^n_S = \operatorname{Spec} S[x_1, \ldots, x_n]$.

An **affine scheme** is the spectrum of some ring $S$. An affine $k$-scheme is the spectrum of a $k$-algebra. Therefore an **affine $k$-variety** is just the spectrum of a finitely generated $k$-algebra [Liu02, Definition 2.3.47]. Because we are primarily interested in affine schemes in this paper, and the category of affine schemes is equivalent to the category of commutative rings, with arrows reversed [EH00, Prop. I-41], we can often omit the structure sheaf.

An (abstract) $k$-**variety** is a scheme that locally looks like an affine variety with a separatedness condition. More formally, a $k$-variety is separated, of finite type scheme over a field $k$.

A **morphism** $f : \operatorname{Spec} B \to \operatorname{Spec} A$ of affine schemes is given by the associated ring map $\varphi : A \to B$: suppose $p \in \operatorname{Spec} B$, then $f(p) = \varphi^{-1}(p)$, which is a prime ideal in $A$.

A **fiber product** $X \times_Z Y$ of two affine schemes $X = \operatorname{Spec} A$ and $Y = \operatorname{Spec} B$ over a base scheme $Z = \operatorname{Spec} S$ (i.e. there are morphisms $X \to Z$ and $Y \to Z$), is the spectrum of the tensor product $A \otimes_S B$.

### 2.1.2 Points in a Scheme

If $X = \operatorname{Spec} A$ is an affine scheme and $p \in X$, then the **residue field** of $p$ on $X$ is the residue field of the local ring at $p$: $\kappa_p = A_p/pA_p = \operatorname{Frac}(A/p)$ (see [Har77, Exercise II.2.7]). Suppose $X$ is an affine $k$-scheme, then $p \in X$ is said to be a **$k$-rational point** if the residue field of $p$ is $k$ (see [Har77, Exercise II.2.8]). The set of all the $k$-rational points in $X$ is denoted by $X(k)$. Equivalently, a $k$-rational point can be characterized by a $k$-morphism $\operatorname{Spec} k \to X$. A **geometric point** $\bar{x}$ of a scheme $X$ is a morphism $\bar{x} : \operatorname{Spec} \Omega \to X$, where $\Omega$ is an algebraically closed field. We say $\bar{x}$ lies over $x$ to indicate that $x \in X$ is the image of $\bar{x}$. Point $p \in X$ is said to be **regular** or **non-singular**, if the local ring $A_p$ is a regular local ring (i.e. its maximal ideal $pA_p$ can be generated by exactly $\dim A_p$ elements).

Let $k$ be a field. We use the notation $k^n$ to denote the **classical affine $n$-space** (the $n$-fold cartesian product of $k$). Please be aware that $k^n \subsetneq \mathbb{A}_k^n$ and it inherits the Zariski topology from the affine $n$-spaces. In fact, if $X$ is an affine $k$-variety with a closed immersion $X \cong V(I) \subseteq \mathbb{A}_k^n$, then the set of the $k$-rational points $X(k)$ can be identified with the common solutions of $I$ in $k^n$ [Liu02, Proposition 3.2.18].

$$X(k) \stackrel{\text{1-1}}{\longleftrightarrow} \{(x_1, \ldots, x_n) \in k^n | f(x_1, \ldots, x_n) = 0, \forall f \in I\}.$$

This bridges the gap between the classical algebraic geometry and the modern algebraic geometry. Now an algebraic set in $k^n$ is just the set of $k$-rational points of a $k$-variety. In particular, $k^n = \mathbb{A}_k^n(k)$.

Suppose $f : X \to Y$ is a $k$-morphism of $k$-schemes $X, Y$. We add the subscript $k$ to $f$ to denote the **restriction of $f$ on the $k$-rational points**:

$$\begin{aligned} f_k : X(k) &\to Y(k) \\ x &\mapsto f(x). \end{aligned}$$

If $k$ is algebraically closed and $X$, $Y$ are $k$-varieties, then $f_k$ is just the classical morphism of classical varieties.

### 2.1.3 Properties of Morphisms

If $f : X \to Y$ is a morphism of schemes, $y \in Y$, the **(scheme-theoretic) fiber** of $f$ over $y$ is defined to be the $\kappa_y$-scheme

$$X_y = X \times_Y \operatorname{Spec} \kappa_y.$$

Similarly, we define the **geometric fiber** of $f$ over $y$ to be

$$X_{\overline{y}} = X_y \times_Y \operatorname{Spec} \Omega = X \times_Y \operatorname{Spec} \kappa_y \times_{\kappa_y} \operatorname{Spec} \Omega = X \times_Y \operatorname{Spec} \Omega,$$

where $\overline{y} : \operatorname{Spec} \Omega \to Y$ is a geometric point of $Y$ lying over $y$. One can show that the underlying topological space of $X_y$ is homeomorphic to the subset $f^{-1}(y)$ of $X$ (see [Har77, Exercise II.3.10]). When $X = \operatorname{Spec} B$ and $Y = \operatorname{Spec} A$ are affine, then $y$ is a prime ideal in $A$ so $X_y = \operatorname{Spec} B \otimes_A \operatorname{Frac}(A/y)$ and $X_{\overline{y}} = \operatorname{Spec} B \otimes_A \overline{\operatorname{Frac}(A/y)}$ by the definition of fiber product.

A morphism $f : \operatorname{Spec} B \to \operatorname{Spec} A$ is a **finite** morphism, if $B$ is a finite $A$-module via the associated ring map. A morphism $f : X \to Y$ is **quasi-finite**, if $X_y = f^{-1}(y)$ is a finite set for every $y \in Y$. A finite morphism is always quasi-finite [Har77, Exercise II.3.5].

Suppose $X = \operatorname{Spec} B$ and $Y = \operatorname{Spec} A$ are two affine schemes. Morphism $f : X \to Y$ is said to be **free**, if $B$ is a free $A$-module via the associated ring map $f^\# : A \to B$ (see [Har77, Page 109]). Let $p \in X$ and $q = f(p) \in Y$, then $f$ is said to be **flat at** $p$ if $B_p$ is a flat $A_q$-module (i.e. tensor product with $B_p$ in the $A_q$-module category preserves left exactness). The morphism $f$ is said to be **flat** if it is flat at all points of $X$. The notion of freeness is closely related to flatness, because a finite module over a noetherian local ring is flat if and only if it is a free module [Eis95, Exercise 6.2].

Keep notations as in the previous paragraph. The map $f$ is said to be **smooth**, if it is flat and all points of all geometric fibers $X_{\overline{y}}$ are regular. It is an **étale** morphism if it is both quasi-finite and smooth.

### 2.1.4 Chow Form

Let $X \subsetneq \mathbb{A}_k^n$ be a reduced zero-dimensional variety embedded in $n$-dimensional space. Then we may consider all the hyperplanes in $\mathbb{A}_k^n$ that pass through $X$. It is known that the family of such hyperplanes is a hypersurface in the dual space $\left(\mathbb{P}_k^n\right)^*$ defined by a single polynomial (see e.g. [GKZ94, Section 3.2B, Proposition 2.2]). This polynomial $\mathrm{Ch}_X$ is called the Chow form of $X$. When $k$ is algebraically closed, it is easy to write down the explicit formula for Chow form:

$$\mathrm{Ch}_X = \prod_{\xi \in X} (\lambda_0 + \xi_1\lambda_1 + \cdots + \xi_n\lambda_n),$$

where $(\xi_1, \ldots, \xi_n)$ is the coordinate of $\xi$ in $\mathbb{A}_n^k$ and $(\lambda_0 : \lambda_1 : \cdots : \lambda_n)$ is the homogeneous coordinate of the dual space.

## 2.2 Real Closed Fields and Semi-algebraic Sets

Now we move to real algebraic geometry, which is a rich theory in its own right. The literature for real algebraic geometry includes [BPR06], [BCR98], [Man20] and a more recent book [Sch24].

If $k$ is a field, then $k^{(2)}$ is the set of **squares** of elements in $k$ and $\sum k^{(2)}$ is the set of **sums of squares** of elements in $k$.

A field $k$ is said to be **real**, if one of the following conditions holds [BPR06, Theorem 2.11].

1. $-1 \notin \sum k^{(2)}$.
2. For every $x_1, \ldots, x_n \in k$, $\sum_{i=1}^n x_i^2 = 0 \implies x_1 = \cdots = x_n = 0$.
3. $k$ can be **ordered**, that is, there exists a total order $\leq$ on $k$ satisfying $x \leq y \implies x + z \leq y + z$ and $0 \leq x \wedge 0 \leq y \implies 0 \leq xy$. The elements of $k$ greater than zero are said to

be **positive**.

A field $R$ is **real closed**, if one of the following conditions holds [BPR06, Theorem 2.14].

1. The field $R$ is an ordered field whose positive elements have a square root and every polynomial of odd degree has a root in $R$.
2. The field $R$ is not algebraically closed, but $R[i] = R[x]/(x^2 + 1)$ is an algebraically closed field.
3. The field $R$ has no non-trivial real algebraic extension.
4. The field $R$ is an ordered field that has the **intermediate value property** (IVP). IVP means, for any $P \in R[x]$, if $P(a)P(b) < 0$ for some $a, b \in R(a < b)$, then there exists some $x \in (a, b)$ such that $P(x) = 0$.

If $k$ is an ordered field, then there is a unique smallest (necessarily algebraic) real closed field containing $k$ (up to isomorphisms), extending the order on $k$. This is the **real closure** of $k$ [BCR98, Section 1.3].

From now on, we fix $R$ to denote a real closed field, and $C = R[i] = R[x]/\langle x^2 + 1\rangle$ is the algebraic closure of $R$.

A **basic semi-algebraic set** in $R^n$ is a set of the form

$$\left\{ x \in R^n \middle| \bigwedge_{1 \le i \le s} f_i(x) = 0 \wedge \bigwedge_{1 \le i \le s'} h_i(x) > 0 \right\}.$$

A **semi-algebraic set** is a finite union of basic semi-algebraic sets. The collection of semi-algebraic sets is the smallest family of sets in $R^n$ definable by polynomial equations and inequalities that is closed under the boolean operations [BPR06, Section 2.3]. One of the most important properties about semi-algebraic sets is that the projection of a semi-algebraic set is still semi-algebraic (Tarski-Seidenberg Principle, see [BPR06, Theorem 2.92]). A **semi-algebraic function** is a map between semi-algebraic sets whose graph is also a semi-algebraic set.

There is a more refined topology on semi-algebraic sets. Since $R$ is ordered, we can define the **Euclidean topology** on $R^n$ by taking open balls

$$B(\boldsymbol{x}, r) = \{\boldsymbol{x'} \in R^n \mid \|\boldsymbol{x} - \boldsymbol{x'}\| < r\}$$

as a basis for a topology space, where $\|\boldsymbol{x}\|$ is the usual **Euclidean norm** $\sqrt{x_1^2 + \cdots + x_n^2}$. Semi-algebraic sets inherit the topology from the embedding [BPR06, Section 3.1]. When we talk about the continuity of a semi-algebraic function, we always use the Euclidean topology. A semi-algebraic set is said to be **semi-algebraically connected**, if it is not a disjoint union of two non-empty closed semi-algebraic subsets [BPR06, Section 3.2].

## 2.3 Effective Methods in Algebraic Geometry

### 2.3.1 Gröbner Basis

In 1965, Buchberger proposed the concept of Gröbner basis and found the Buchberger's Criterion on S-polynomials together with an algorithm to compute a Gröbner basis in his PhD thesis [Buc65], and he added a complete proof for the termination in a journal article [Buc70] later. The theory of Gröbner basis is now one of the most important tools in computational algebraic geometry and symbolic computation. Most definitions in this subsection can be found in [CLO15] or [Kem11, Section 9.1].

To introduce the theory of Gröbner basis, we need to define monomial orders. Let $S$ be a noetherian ring and $S[x_1, \ldots, x_n]$ is a polynomial ring over $S$. A monomial is a product of the variables. A **monomial order** $T$ is a total order on the set of monomials that satisfies

1. (Well-ordering) Every non-empty set of monomials has a smallest element, and
2. (Multiplicative) if $\boldsymbol{x}^\alpha, \boldsymbol{x}^\beta$ and $\boldsymbol{x}^\gamma$ are three monomials and $\boldsymbol{x}^\alpha > \boldsymbol{x}^\beta$, then $\boldsymbol{x}^\alpha \cdot \boldsymbol{x}^\gamma > \boldsymbol{x}^\beta \cdot \boldsymbol{x}^\gamma$.

A term is a product $s \cdot \boldsymbol{x}^\alpha$, where $s \in S \backslash \{0\}$ and $\boldsymbol{x}^\alpha$ is a monomial. Obviously every non-zero polynomial $f$ in $S[x_1, \ldots, x_n]$ can be uniquely written as a sum of terms such that no two terms share the same monomial

$$f = \sum_{\alpha} s_\alpha \boldsymbol{x}^\alpha.$$

Without loss of generality, we can assume that the terms $s_\alpha \boldsymbol{x}^\alpha$ are ordered in the descending order with respect to $T$. Then the largest term on the right hand side is said to be the **leading term** of $f$ (with respect to $T$). The notation $\mathtt{LT}_T(f)$ stands for the leading term. Similarly the **leading coefficient** $\mathtt{LC}_T(f)$ is the coefficient of the leading term and the **leading monomial** $\mathtt{LM}_T(f)$ is the monomial of the leading term.

Here are some examples of monomial orders. Let $\boldsymbol{x}^\alpha = x_1^{a_1} \cdots x_n^{a_n}$ and $\boldsymbol{x}^\beta = x_1^{b_1} \cdots x_n^{b_n}$ be monomials.

The **lexicographic order** is given by saying $\boldsymbol{x}^\alpha > \boldsymbol{x}^\beta$ if and only if there is some $i$ such that $a_1 = b_1, \ldots, a_{i-1} = b_{i-1}$ but $a_i > b_i$.

The **graded reverse lexicographic order** (abbreviated as **grevlex**) first compares the total degree of monomials, the larger total degree is, the larger monomial is. If the total degrees are the same, then $\boldsymbol{x}^\alpha > \boldsymbol{x}^\beta$ if and only if there is some $i$ such that $a_n = b_n, a_{n-1} = b_{n-1}, \ldots, a_{i+1} = b_{i+1}$ but $a_i < b_i$. It is observed that in practice, this is the monomial order that leads to the fastest computation.

A **block order** is a monomial order assembling two smaller orders. Suppose variables are assigned to two groups $\boldsymbol{x}$ and $\boldsymbol{y}$, and $T_x$ ($T_y$ respectively) is an order for monomials in $\boldsymbol{x}$ ($\boldsymbol{y}$ respectively). Then the block order $T = T_x > T_y$ is the order for monomials in $\boldsymbol{x}$ and $\boldsymbol{y}$ that first compares $\boldsymbol{x}$ by $T_x$, breaking ties with comparing $\boldsymbol{y}$ with $T_y$. A block order can be used to compute elimination ideals.

Fix a monomial order $T$. If $I$ is an ideal in the polynomial ring $S[x_1, \dots, x_n]$, the **ideal of leading terms** of $I$ is the ideal generated by the leading terms of elements of $I$,

$$\langle \mathtt{LT}_T(I) \rangle = \langle \mathtt{LT}_T(f) | f \in I \rangle.$$

A **Gröbner basis** $\mathcal{G}$ of the ideal $I$ with respect to $T$ is a set of generators $\{g_1. \dots, g_s\}$ such that the ideal of leading terms can also be generated by the leading terms of $g_i$,

$$\langle \mathtt{LT}_T(I) \rangle = \langle \mathtt{LT}_T(g_1), \dots, \mathtt{LT}_T(g_s) \rangle.$$

Notice: although most literature defines Gröbner basis for ideals in a polynomial ring over a field $k$, the same definition extends to arbitrary noetherian rings without any difficulty, see [Möl88].

### 2.3.2 Subresultants

Subresultant characterizes the degree of the greatest common divisor of two univariate polynomials.

**Definition 2.3.1.** Let $A$ be a commutative ring with identity. Given two polynomials $f = \sum_{i=0}^{p} a_i x^i, g = \sum_{i=0}^{q} b_i x^i \in A[x]$ ($\deg f = p > q = \deg g$). Define $A$-linear map

$$\begin{aligned} \varphi_j : A[x]_{q-j} \oplus A[x]_{p-j} &\to A[x]_{p+q-j} \\ (u, v) &\mapsto uf + vg. \end{aligned}$$

between free $A$-modules.

The **$j$-th subresultant polynomial** $\mathtt{sResP}_{x,j}(f, g)$ is the polynomial $\sum_{i=0}^{j} d_{j,i} x^i$ where $d_{j,i}$ is the determinant of the composition $\pi_i \circ \varphi_j$. Here $\pi_i : A[x]_{p+q-j} \to A^{p+q-2j}$ the projection map forgetting coefficients of $1, \dots, x^j$ except $x^i$.

The **$j$-th subresultant**, denoted by $\mathtt{sRes}_{x,j}(f, g)$, is the leading coefficient $d_{j,j}$ of the $j$-th subresultant polynomial $\mathtt{sResP}_{x,j}(f, g)$. Especially, the 0-th subresultant $\mathtt{sRes}_{x,0}(f, g)$ is the resultant of $f$ and $g$.

**Theorem 2.3.2** ([BPR06, Proposition 4.25, Corollary 8.58])**.** *Let $k$ be a field. The greatest common divisor of $f$ and $g$ in $k[x]$ is of degree $j$, if and only if*

$$\mathtt{sRes}_{x,0}(f,g) = \cdots = \mathtt{sRes}_{x,j-1}(f,g) = 0, \mathtt{sRes}_{x,j}(f,g) \neq 0.$$

*Moreover, the $j$-th subresultant polynomial $\mathtt{sResP}_{x,j}$ coincides with the greatest common divisor of $f$ and $g$ in this case.*

From the determinantal definition, it is immediate that subresultants and subresultant polynomials have the specialization property, provided that the leading coefficients are not mapped to zero.

**Lemma 2.3.3.** *Let $\varphi : A \to B$ be a ring homomorphism and $f, g \in A[x]$. Suppose that $\mathtt{LC}_x(f), \mathtt{LC}_x(g) \notin \ker\varphi$, then*

$$\varphi(\mathtt{sRes}_{x,j}(f,g)) = \mathtt{sRes}_{x,j}(\varphi(f), \varphi(g))$$

*and*

$$\varphi(\mathtt{sResP}_{x,j}(f,g)) = \mathtt{sResP}_{x,j}(\varphi(f), \varphi(g)).$$

### 2.3.3 Separating Elements

Suppose $k$ is a field of characteristic 0. Let $A$ be a reduced $k$-algebra of finite type, and let $B = A[x_1, \ldots, x_n]/I$ be an $A$-algebra that is also a free $A$-module of finite rank $d$. Every element $\sigma \in B$ defines an $A$-linear multiplication map $L_\sigma : \begin{array}{ccc} B & \to & B \\ b & \mapsto & \sigma b \end{array}$. Denote the characteristic polynomial $\det(\lambda I - L_\sigma) \in A[\lambda]$ by $\chi_\sigma$.

**Definition 2.3.4.** An element $\sigma \in B$ is said to be **separating** over $p \in \operatorname{Spec} A$, if

$$\#(\operatorname{Spec} B)_{\overline{p}} = d - \deg_\lambda \gcd(\chi_\sigma(p), \chi'_\sigma(p)).$$

This is equivalent to say the degree of square-free part of $\chi_\sigma(p)$ is equal to cardinality of geometric fiber over $p$. If $\sigma$ is separating over every $p \in \operatorname{Spec} A$, then we say $\sigma$ is a **separating element** of $B$.

It is a well-known fact that there is always a separating element for a zero-dimensional $k$-varieties when $\operatorname{char} k = 0$.

**Lemma 2.3.5** ([BPR06, Lemma 4.90])**.** *Suppose $k$ is a field of characteristic 0. Let $A = k[x_1, \ldots, x_n]/J$ be a zero-dimensional $k$-algebra such that $\dim_k A = d$. Among the elements* *of*

$$\left\{x_1 + ix_2 + \cdots + i^{n-1}x_n\right\}_{0\le i\le (n-1)\binom{d}{2}},$$

*at least one of them is a separating element for A.*

Stickelberger's Eigenvalue Theorem justifies the name "separating element". See [CLO05, Theorem 4.2.7] for a proof.

**Theorem 2.3.6** (Stickelberger)**.** *Let $k$ be an algebraically closed field. If $J$ is a zero-dimensional ideal in $k[x_1,\dots,x_n]$ and $\sigma \in k[x_1,\dots,x_n]$, then*

$$\det(\lambda I - L_\sigma) = \prod_{p\in V(J)} (\lambda - \sigma(p))^{\mu(p)},$$

*where $\mu(p) = \dim_k(k[x_1,\dots,x_n]/J)_p$ is the multiplicity of $p$ in $V(J)$ and $L_\sigma$ is the multiplication map by $\sigma$ in $k[x_1,\dots,x_n]/J$. Moreover we have*

$$\operatorname{tr} L_\sigma = \sum_{p\in V(J)} \mu(p)\sigma(p) \text{ and } \det L_\sigma = \prod_{p\in V(J)} \sigma(p)^{\mu(p)}.$$

By the Stickelberger's Eigenvalue Theorem, the degree of the square-free part of $\chi_\sigma(p)$ is less than or equal to the number of points in the geometric fiber $(\operatorname{Spec} B)_{\overline{p}}$, and the equality holds if and only if $\sigma(p)$ takes distinct values on each point of the geometric fiber. In this case, $\sigma$ separates different points in $(\operatorname{Spec} B)_{\overline{p}}$.

### 2.3.4 Effective Generic Freeness

Suppose $A$ is a noetherian domain and $B$ is a finitely generated $A$-algebra. If $M$ is a finitely generated $B$-module, then there exists an element $0 \neq a \in A$ such that $M_a$ is a free $A_a$-module. This is a classical result due to Grothendieck, called Grothendieck's Generic Freeness Lemma [EGA $\text{IV}_2$, Lemme 6.9.2], which is more commonly known as "Generic Flatness" in algebraic geometry.

The theory of Gröbner basis gives it a constructive proof, which is known to many people as a folklore [Vas97]. We collect the results here.

**Lemma 2.3.7** (Effective Generic Freeness)**.** *Let $I \trianglelefteq k[\boldsymbol{y},\boldsymbol{x}]$ and $J = I \cap k[\boldsymbol{y}] \trianglelefteq k[\boldsymbol{y}]$, $T_x$ (resp. $T_y$) is a monomial order in $\boldsymbol{x}$ (resp. $\boldsymbol{y}$). $T = T_x > T_y$ is a block order. Suppose $\mathcal{G}$ is a reduced Gröbner basis of $I$ with respect to $T$, $\mathcal{G}_y = \mathcal{G} \cap k[\boldsymbol{y}]$ and $\mathcal{G}_x = \mathcal{G}\backslash\mathcal{G}_y$. Then*

*(a) The set $\mathcal{G}_y$ is a Gröbner basis of $J$ with respect to $T_y$;*

*(b) For all $f \in \mathcal{G}_x$, the leading coefficient $\mathtt{LC}_{T_x}(f) \notin J$ ;*

*(c)* *Set* $w = \prod_{f \in \mathcal{G}_x} \mathrm{LC}_{T_x}(f)$*, then for all prime ideals* $p \trianglelefteq k[\boldsymbol{y}]$ *containing* $J$*: if* $w \notin p$*, then the specialization of* $\mathcal{G}$ *at* $\kappa_p = \mathrm{Frac}(k[\boldsymbol{y}]/p)$ *is a Gröbner basis for* $I \otimes_k \kappa_p$ *with respect to* $T_x$*.*

*(d)* *Moreover,* $k[\boldsymbol{y}, \boldsymbol{x}]_w / I_w$ *is a free* $(k[\boldsymbol{y}]_w / J_w)$*-module. A basis of this free module can be chosen to be the monomials in* $\boldsymbol{x}$ *not divisible by any leading monomial of elements of* $\mathcal{G}_x$ *with respect to* $T_x$*.*

**Proof.**

(a) Elimination property of Gröbner basis [CLO15, Section 3.1, Exercise 5].

(b) Since $\mathcal{G}$ is a reduced Gröbner basis, no leading monomial can divide another leading monomial. If there is some $f \in \mathcal{G}_x$ such that $l = \mathrm{LC}_{T_x}(f) \in J$, then by the previous result, there is some $g \in \mathcal{G}_y$ such that $\mathrm{LT}_{T_y}(g) | \mathrm{LT}_{T_y}(l)$. Therefore $\mathrm{LT}_T(g) | \mathrm{LT}_T(f)$, which is a contradiction.

(c) Kalkbrener's stability criterion [Kal97, Theorem 3.1].

(d) This is [Kem11, Theorem 10.1].

□

# Chapter 3 A Geometric Approach to Cylindrical Algebraic Decomposition

## 3.1 Problem Statement

In this section, we give the precise statement of the problem. For this, we will first present here the definition of cylindrical algebraic decomposition, which is essentially the same as [BPR06, Definition 5.1]. A minor difference is that here a c.a.d. does not necessarily cover the whole affine space $R^n$ (a natural generalization for applications like real root classification).

**Definition 3.1.1.** Let $R$ be a real closed field. A **cylindrical algebraic decomposition** (c.a.d.) $\mathfrak{D}$ in $R^n$, is a sequence $S_1, \ldots, S_n$, where for each $1 \le i \le n$:

1. The set $S_i$ consists of disjoint semi-algebraic sets in $R^i$, which are called the **cells** of level $i$.
2. Each cell of level 1 is either a point or an interval.
3. For every $1 < i \le n$, and every $S \in S_i$, the **silhouette**
$$\pi_{i-1}(S) = \{(y_1, \ldots, y_{i-1}) \in R^{i-1} | \exists x \in R \text{ s.t. } (y_1, \ldots, y_{i-1}, x) \in S\}$$
is a cell of level $i-1$.
4. For every $1 \le i < n$, and every $S \in S_i$, there are finitely many semi-algebraic continuous functions $\xi_{S,1}, \ldots, \xi_{S,l_S} : S \to R$ (the **sections**) such that any cell in $S_{i+1}$ lying over $S$ is either the graph of some $\xi_{S,j}$
$$\left\{(y, \xi_{S,j}(y)) \in S \times R \middle| y \in S\right\},$$
or the **band**
$$\left\{(y, x) \in S \times R \middle| y \in S,\ \xi_{S,j}(y) < x < \xi_{S,j+1}(y)\right\},$$
where we take $\xi_{S,0}(x) = -\infty$ and $\xi_{S,l_S+1}(x) = +\infty$.

We say $\mathfrak{D}$ is **$\mathcal{F}$-sign-invariant** for a finite family of $n$-variate polynomials $\mathcal{F}$, if for every $f \in \mathcal{F}$ and every cell $C$ of level $n$, $f$ is sign-invariant on $C$.

We say $\mathfrak{D}$ is **adapted to** a finite family of semi-algebraic sets $\{L_1, \ldots, L_s\}$ in $R^n$, if each $L_i$ is a union of level $n$ cells. Notice that the disjointness of cells implies that for each cell $S \in S_n$ of level $n$ and each semi-algebraic set $L_i$, either $S \subseteq L_i$ or $S \cap L_i = \varnothing$.

**Problem.** Design an algorithm that

**Inputs:** A list of varieties $V(I_1), \ldots, V(I_s)$ and

**Outputs:** A cylindrical algebraic decomposition adapted to $\{V(I_1), \ldots, V(I_s)\}$.

Note that a c.a.d. adapted to $\{V(0), V(f_1), \ldots, V(f_s)\}$ is an $\{f_1, \ldots, f_s\}$-sign-invariant c.a.d.. So the classical problem of computing a sign-invariant c.a.d. is a special case here.

Note also that it is easy to convert such an algorithm to accept semi-algebraic sets input.

## 3.2 Counting Geometry Fiber Cardinality

Our goal in this section is to count the cardinality of geometric fiber of a finite free morphism. Throughout this section, $k$ is a field of characteristic 0.

A function $\varphi : X \to \mathbb{Z}$ is said to be **lower semi-continuous**, if $\{x \in X | \varphi(x) \le n\}$ is closed for all $n \in \mathbb{Z}$. The main result in this section is the lower semi-continuity of geometric fiber cardinality:

**Theorem 3.2.1** (Lower Semi-Continuity of Geometric Fiber Cardinality)**.** *If $f : X \to Y$ is a finite, free morphism between affine $k$-varieties $X$ and $Y$, then the cardinality of geometric fiber $y \mapsto \# X_{\overline{y}}$ is a lower semi-continuous function on $Y$.*

*If $U = D(h)$ is a distinguished open subset of $X$, then $y \mapsto \# U_{\overline{y}}$ is also lower semi-continuous on $Y$.*

Now, we will introduce multivariate Hermite quadratic form, which is a trace pairing technique to count the number of geometric points of a zero-dimensional $k$-algebra. It will play a central role in our proof. The tool is proposed by Becker and Wörmann [Bec91; BW94], and independently by Pedersen, Roy and Szpirglas [Ped91; PRS93] in its current form. Scheja and Storch obtained essentially the same result in a different form [SS88, Theorem 94.7]. See [Cox21] for a discussion on the history. This technique is widely used in symbolic computation (for example, [Wei98; PR17; LSED22]). For our convenience, we will rephrase this in the language of schemes.

We need a notation here. Suppose $S$ is a ring and $A$ is an $S$-algebra that is also a free $S$-module of finite rank. Let $f \in A$, then the **multiplication map** $L_f$ is the $S$-module endomorphism given by

$$\begin{aligned} L_f : A &\to A \\ g &\mapsto fg. \end{aligned}$$

Because $A$ is a free $S$-module of finite rank, the trace of $L_f$ is well-defined as the sum of diagonal entries of a matrix representing $L_f$.

**Lemma 3.2.2** (Multivariate Hermite Quadratic Form)**.** *Suppose $k$ is a field. Let $A$ be a zero-dimensional $k$-algebra. Let $\overline{A} = A \otimes_k \overline{k}$ and $X = \operatorname{Spec} \overline{A}$. Define Hermite quadratic form on $A$:*

$$\begin{aligned} H : A \times A &\rightarrow k \\ (f, g) &\mapsto \operatorname{tr} L_{fg}, \end{aligned}$$

*then* $\operatorname{rank} H = \# X$.

*Moreover, if $h \in A$, define*

$$\begin{aligned} H_h : A \times A &\rightarrow k \\ (f, g) &\mapsto \operatorname{tr} L_{fhg}, \end{aligned}$$

*then* $\operatorname{rank} H_h = \# X_h$ *($X_h = \operatorname{Spec} \overline{A_h} = \operatorname{Spec}(A_h \otimes_k \overline{k})$).*

*If $k$ is in addition ordered and the real closure of $k$ is $R$, let $A' = A \otimes_k R$ be the base change of $A$ to $R$ and $X' = \operatorname{Spec} A'$, then* $\operatorname{sign} H = \# X'(R)$*, i.e., the signature of $H$ is exactly the number of $R$-rational points in $X' = \operatorname{Spec} A'$.*

*Similarly,* $\operatorname{sign} H_h = \#\{p \in X'(R) | h(p) > 0\} - \#\{p \in X'(R) | h(p) < 0\}$.

**Proof.** This is just [BPR06, Theorem 4.102], rephrased in the language of schemes. □

**Corollary 3.2.3** (Relative Version)**.** *Let $k$ be a field of characteristic 0. Suppose $X = \operatorname{Spec} B$, $Y = \operatorname{Spec} A$ are two affine $k$-varieties, $p \in Y$ and $f : X \rightarrow Y$ is a $k$-morphism. Let $\kappa_p =$ $\operatorname{Frac} A/p$ be the residue field of $p$. If the fiber $X_p$ or the geometric fiber $X_{\overline{p}}$ is finite, then the $\kappa_p$-bilinear form on $B \otimes_A \kappa_p$ sending $(f, g)$ to $\operatorname{tr} L_{fg}$ has its rank equal to $\# X_{\overline{p}}$.*

*Moreover, if $h \in B$ and $U = D(h) \subseteq \operatorname{Spec} B$ is a distinguished open subset, then the $\kappa_p$-bilinear form on $B \otimes_A \kappa_p$ sending $(f, g)$ to $\operatorname{tr} L_{fhg}$ has its rank equal to $\# U_{\overline{p}}$.*

**Proof.** Notice that the fiber $X_p$ is the spectrum of $B \otimes_A \kappa_p$ and the geometric fiber $X_{\overline{p}}$ is just the spectrum of $B \otimes_A \kappa_p \otimes_{\kappa_p} \overline{\kappa_p}$. Therefore everything follows from Lemma 3.2.2. □

In order to prove Theorem 3.2.1, we extend the classical trace pairing technique to the free module of finite rank case now.

**Definition 3.2.4.** If $A$ is a finitely generated $k$-algebra and $B$ is an $A$-algebra that is also a free $A$-module of finite rank, then we define the **relative Hermite quadratic form** of $B$ to be:

$$\begin{aligned} H : B \times B &\rightarrow A \\ (f, g) &\mapsto \operatorname{tr} L_{fg}. \end{aligned}$$

This is well-defined because $B$ is a free $A$-module of finite rank.

Similarly, if $h \in B$, we define the **relative Hermite quadratic form associated with** $h$ of $B$ to be:

$$\begin{array}{rcl} H_h : B \times B & \to & A \\ (f, g) & \mapsto & \operatorname{tr} L_{fhg}. \end{array}$$

**Lemma 3.2.5** (Specialization Property of relative Hermite quadratic form)**.** *Notations as above and $p \in \operatorname{Spec} A$, then the relative Hermite quadratic forms $H$ and $H_h$ evaluated at $p$ coincides with the classical Hermite quadratic form.*

**Proof.** Choose a basis $\{b_1, \ldots, b_r\}$ of $B$ as free $A$-module, then $H$ ($H_h$ resp.) can be represented as a matrix $\left(\operatorname{tr} L_{b_i b_j}\right)_{i,j}$ ($\left(\operatorname{tr} L_{b_i h b_j}\right)_{i,j}$ resp.). Let $p$ be a prime ideal of $A$ and $\varphi$ is the canonical map $\varphi : B \to B \otimes_A \operatorname{Frac}(A/p)$, the images of $b_i$ in $B \otimes_A A_p \otimes_{A_p} A_p/pA_p \cong B \otimes_A \operatorname{Frac}(A/p)$ is a basis for $\operatorname{Frac}(A/p)$-vector space $B \otimes_A \operatorname{Frac}(A/p)$.

Now the matrix of the classical Hermite quadratic form is given by the traces of $L_{\varphi(b_i b_j)}$, which are exactly $\operatorname{tr} L_{b_i b_j}$ evaluated at $p$. The same argument works for $H_h$. The proof is completed. $\square$

*Proof of Theorem 3.2.1.* Let $X = \operatorname{Spec} B$ and $Y = \operatorname{Spec} A$, then $B$ is a free $A$-module of finite rank via the morphism, choose a basis $\{b_i\}$ for $B$.

By Definition 3.2.4 and Lemma 3.2.5, the relative Hermite quadratic form specializes to all $p \in Y$, and Corollary 3.2.3 shows that the rank of Hermite quadratic form specialized at $p$ is equal to $\#X_{\overline{p}}$. Let $H = (\operatorname{tr} L_{b_i b_j})_{b_i, b_j \in \mathscr{B}}$, then $\operatorname{rank} H(p) < r$ if and only if all the $r$-minors of $H(p)$ vanish. Let $I_r$ be the ideal in $A$ generated by all $r$-minors of $H$, then $\{p \in Y | \operatorname{rank} H(p) \leq r\} = V(I_{r+1})$ is closed, i.e. $y \mapsto \#X_{\overline{y}}$ is lower semi-continuous on $Y$. By replacing $H$ with $H_h$, we can show the same for $y \mapsto \#U_{\overline{y}}$. This completes the proof. $\square$

*Remark* 1. Actually, Theorem 3.2.1 is a special case of a Grothendieck theorem [EGA $\mathrm{IV}_3$, Proposition 15.5.9], which states for a quasi-finite separated flat morphism locally of finite presentation between schemes, using non-constructive techniques like Stein factorization.

## 3.3 The Geometry of Cylindrical Algebraic Decomposition

Throughout this section, $k$ is an ordered field and $R$ is the real closure of $k$. Let $C = R(i) = R[x]/\langle x^2 + 1\rangle$ be the algebraic closure of $R$.

We recall the following two basic facts in real algebraic geometry, which are about semi-algebraic connectedness and continuity of roots, respectively.

**Lemma 3.3.1** ([BPR06, Proposition 3.12]). *Let $S$ be a semi-algebraically connected semi-algebraic set and let $f : S \to R$ be a locally constant semi-algebraic function (i.e. given $x \in S$, there is an open $U \subset S$ such that for all $y \in U$, $f(y) = f(x)$), then $f$ is a constant.*

**Theorem 3.3.2** ([BPR06, Theorem 5.12], Continuity of Roots). *Let $P \in R[x_1, \ldots, x_n]$ and let $S$ be a semi-algebraic subset of $R^{n-1}$. Assume that $\deg_{x_n}(P)$ is constant on $S$ and that for some $a' \in S$, $z_1, \ldots, z_j$ are the distinct roots of $P(a'; x_n)$ in $C = R[i]$, with multiplicities $\mu_1, \ldots, \mu_j$, respectively.*

*If the open disks $B(z_i, r) \subset C$ are disjoint, then there is an open neighborhood $V$ of $a'$ such that for every $x' \in V \cap S$, the polynomial $P(x'; x_n)$ has exactly $\mu_i$ roots, counted with multiplicities, in the disk $B(z_i, r)$, for $i = 1, \ldots, j$.*

### 3.3.1 Geometric Delineability for a Variety

In this subsection, we will extend Collins's notion of delineability to a variety. The goal is to show that a finite free morphism of $R$-varieties admits semi-algebraic sections on the regions that have a constant geometric fiber cardinality if it factors through the cylinder over the base. We will see that counting the geometric fiber lies at the heart of geometry of cylindrical algebraic decomposition.

**Lemma 3.3.3** (Rank-Signature Lemma). *If $S$ is a semi-algebraically connected semi-algebraic subset of $\mathrm{Sym}(R, n) := \{M \in R^{n\times n} | M_{ij} = M_{ji}\}$ (with the Euclidean topology inherited from $R^{n\times n}$) and all the matrices in $S$ are of the same rank, then all matrices in $S$ have the same signature.*

**Proof.** Consider the map

$$\begin{aligned} \chi : S &\to R^n \\ M &\mapsto (a_{n-1}, a_{n-2}, \ldots, a_0), \end{aligned}$$

where $\lambda^n + \sum_{i=0}^{n-1} a_i \lambda^i$ is the characteristic polynomial of $M$. By some abuse of notations, we also use $\chi_M$ to denote the characteristic polynomial of $M$, since this would not cause any confusion. Because $M$ is symmetric and all the entries are in $R$, it is diagonalizable and all the eigenvalues are in $R$ [BPR06, Theorem 4.43]. Suppose all the matrices in $S$ are of the same rank $r$, then for all $M \in S$, there are exactly $r$ non-zero eigenvalues. To show the signature is invariant, it suffices to show the number of positive eigenvalues (counted with multiplicity) is invariant now.

By Theorem 3.3.2, the continuity of roots in the coefficients, for all $M \in S$, there is a neighborhood $U \subseteq S$ of $M$ such that for all $M' \in U$, $\chi_M$ and $\chi_{M'}$ share the same number of positive roots, counted with multiplicity. Consider the function

$$\begin{array}{rcl} \psi : S & \to & R \\ M & \mapsto & \sum_{\lambda>0} \nu_\lambda(\chi_M) \end{array}$$

that counts the number of positive roots with multiplicity, where $\nu_\lambda(\chi_M)$ is the order of vanishing of $\chi_M$ at $\lambda$. Then $\psi$ is a locally constant function. Because a locally constant semi-algebraic function defined on a semi-algebraically connected semi-algebraic set is a constant (see Lemma 3.3.1), it remains to show $\psi$ is a semi-algebraic function now.

A polynomial $f$ has exactly $p$ distinct positive roots with multiplicity $d_1, \dots, d_p$ is equivalent to

$$\begin{aligned} (\exists x_1) \cdots (\exists x_p) \left( \left( \bigwedge_{i=1}^{p} x_i > 0 \right) \wedge \bigwedge_{i=1}^{p} \left( \bigwedge_{j=0}^{d_i-1} f^{(j)}(x_i) = 0 \wedge f^{(d_i)}(x_i) \neq 0 \right) \right. \\ \left. \wedge (\forall y) \left( \left( (y > 0) \wedge (\bigwedge_{i=1}^{p} y \neq x_i) \right) \implies (f(y) \neq 0) \right) \right). \end{aligned}$$

So by the Tarski-Seidenberg principle [BPR06, Theorem 2.92],

$$\begin{aligned} C_{d_1,\dots,d_p} := \{(a_{n-1}, \dots, a_0) | \lambda^n + \sum_{i=0}^{n-1} a_i \lambda^i = 0 \text{ has exactly } p \text{ positive roots} \\ \text{with multiplicity } d_1, \dots, d_p\} \end{aligned}$$

is a semi-algebraic set. Then

$$\begin{aligned} B_l := \{(a_{n-1}, \dots, a_0) | \lambda^n + \sum_{i=0}^{n-1} a_i \lambda^i = 0 \text{ has } l \text{ positive roots}\} \\ = \bigcup_{d_1+\cdots+d_p=l} C_{d_1,\dots,d_p} \end{aligned}$$

is also semi-algebraic. So the graph of $\psi$ is just $\bigcup_{l=0}^{n} (\chi^{-1}(B_l) \times \{l\})$, which is clearly semi-algebraic. Therefore $\psi$ is a semi-algebraic function by definition. Hence $\psi$ is a constant function on $S$ and the signature of matrices in $S$ is invariant.

□

By counting geometric fiber, we also learn information about the real fiber, as shown in the theorem below.

**Theorem 3.3.4** (Geometric Fiber v.s. Real Fiber)**.** *Suppose $X = \operatorname{Spec} B, Y = \operatorname{Spec} A$ are two $R$-varieties and $f : X \to Y$ is a finite free $R$-morphism. Let $h \in B$. Consider the induced morphism of $R$-rational points $f_R : X(R) \to Y(R)$. Let $S \subseteq Y(R)$ be a semi-algebraically connected semi-algebraic set such that $\#(D(h))_{\overline{y}}$ is a constant for all $y \in S$, then $\#(f_R^{-1}(y) \cap D(h))$ is a constant over $S$.*

**Proof.** Because we can define $H_{h^2}$, the relative Hermite quadratic form associated with $h^2$ of $B$ (Definition 3.2.4) and it specializes to all points in $\operatorname{Spec} A$ (Lemma 3.2.5). By Lemma 3.2.2, the rank of $H_{h^2}$ is invariant over $S$, since $D(h) = D(h^2)$. Then Lemma 3.3.3 tells us the signature of $H_{h^2}$ is also invariant over $S$. Therefore using Lemma 3.2.2 again, we see that the cardinality of real fibers $\#(f_R^{-1}(y) \cap D(h))$ is invariant over $S$ (notice that $f_R^{-1}(y) = X_y(R)$). $\square$

To show that the real fibers can be glued to sections, we need a stronger condition in morphism.

**Definition 3.3.5.** Let $X, Y$ be affine $k$-varieties and $\pi : X \to Y$ is a $k$-morphism. Suppose $\pi$ factors through the cylinder $Y \times_k \mathbb{A}^1_k$:

$$\begin{tikzcd} X \arrow[rr, "\pi"] \arrow[dr, hook] & & Y \\ & Y \times_k \mathbb{A}^1_k \arrow[ur] & \end{tikzcd}$$

and $X \to Y \times_k \mathbb{A}^1_k$ is a closed immersion. Then we say $f : X \to Y$ is a **Cylindrical Projection** (or a corank 1 projection).

The following lemma says a cylindrical projection is finite and free, if and only if $X$ can be cut out by a monic polynomial in the cylinder $Y \times_k \mathbb{A}^1_k = \mathbb{A}^1_Y$.

**Lemma 3.3.6** ([Eis95, Prop. 4.1])**.** *Let $A$ be a ring and let $J \lhd A[x]$ be an ideal in the polynomial ring in one variable over $A$. Let $B = A[x]/J$, and let $s$ be the image of $x$ in $B$.*

a. *$B$ is generated by $\leq n$ elements as an $A$-module if and only if $J$ contains a monic polynomial of degree $\leq n$. In this case $B$ is generated by $1, s, \ldots, s^{n-1}$. In particular, $B$ is a finitely generated $A$-module if and only if $J$ contains a monic polynomial.*
b. *$B$ is a finitely generated free $A$-module if and only if $J$ can be generated by a monic polynomial. In this case $B$ has a basis of the form $1, s, \ldots, s^{n-1}$.*

The next theorem shows that, if $X \to Y$ is a finite free cylindrical projection of $R$-varieties, then the real fibers are continuous functions on the base.

**Theorem 3.3.7.** *Suppose $\pi : X \to Y$ is finite free cylindrical projection between affine $R$-varieties. Let $S \subseteq Y(R)$ be a semi-algebraic connected semi-algebraic set on which $\#X_{\overline{y}}$ is a constant, then there are semi-algebraic continuous functions $\xi_1, \dots, \xi_l : S \to \pi_R^{-1}(S)$ such that $\pi \circ \xi_i = \mathrm{id}_S$ and $\{\xi_i(y)\}_{i=1}^{l}$ is the real fiber of $y$ in $X(R)$.*

**Proof.** We may embed $Y$ in $\mathbb{A}_R^n = \operatorname{Spec} R[y_1, \dots, y_n]$ as a closed subscheme. Then by $\pi$ is a cylindrical projection, $X$ can be embedded in $\mathbb{A}_R^{n+1} = \operatorname{Spec} R[y_1, \dots, y_n, x]$ too. Here is the diagram showing the morphisms among all the schemes we just mentioned:

$$\begin{tikzcd}
X \arrow[r, hook, "i_X", "closed"'] \arrow[d, "\pi"] & \mathbb{A}_R^{n+1} \arrow[d, "\pi"] \\
Y \arrow[r, hook, "i_Y", "closed"'] & \mathbb{A}_R^{n}.
\end{tikzcd}$$

And what we are proving is that there are sections $\xi_i$ making the induced diagram commute:

$$\begin{tikzcd}
\pi_R^{-1}(S) \arrow[d, "\pi_R"] & \subseteq & X(R) \arrow[r, hook, "(i_X)_R"] \arrow[d, "\pi_R"] & R^{n+1} \arrow[d, "\pi_R"] \\
S \arrow[u, dashed, bend left, "\xi_i"] & \subseteq & Y(R) \arrow[r, "(i_Y)_R"] & R^n.
\end{tikzcd}$$

Since $\pi$ is a cylindrical projection, we can choose a monic polynomial $f \in R[y_1, \dots, y_n, x]$ such that $X$ is cut out by the locus of $f$ over the cylinder $Y \times_R \mathbb{A}_R^1$, using Lemma 3.3.6. By Theorem 3.3.4, the number of real fibers is a constant $m$ on $S$. Define

$$\begin{aligned}
\xi_i : S &\to \pi_R^{-1}(S) \\
\boldsymbol{y} &\mapsto (\boldsymbol{y}, x_i),
\end{aligned}$$

where $x_i$ is the $i$-th largest root of $f(\boldsymbol{y}; x) = 0$ ($i = 1, \dots, m$), then $\xi_i$ are semi-algebraic functions by definition.

It remains to show the continuity. Choose some particular $\boldsymbol{y} \in S$ and let $x_1, \dots, x_l$ be the distinct solutions of $f(\boldsymbol{y}; x) = 0$ in $C$ with multiplicities $\mu_1, \dots, \mu_l$. By renumbering the roots if necessary, we may assume $x_1 < \cdots < x_m$ are the real roots of $f(\boldsymbol{y}; x) = 0$. Suppose $U_1, \dots, U_l$ are disjoint open Euclidean neighborhoods of $x_1, \dots, x_l$ such that $U_{m+1} \cap R = \cdots = U_l \cap R = \varnothing$.

By Theorem 3.3.2, the continuity of roots in the coefficients, there is a neighborhood $U$ in $S$ of $\boldsymbol{y}$ such that for all $\boldsymbol{y}' \in U$, the equation $f(\boldsymbol{y}'; x) = 0$ has exactly $\mu_i$ solutions in $U_i$ for all $1 \le i \le l$ (counted with multiplicity). But $\#X_{\overline{y}} = l$ is invariant on $S$, forcing that there is exactly 1 solution of multiplicity $\mu_i$ to $f(\boldsymbol{y}'; x) = 0$ in $U_i$ for $1 \le i \le l$. Also, for $m + 1 \le i \le l$, the root of $f(\boldsymbol{y}'; x) = 0$ in $U_i$ must be non-real because $U_i \cap R = \varnothing$. So for $1 \le i \le m$, the root of $f(\boldsymbol{y}'; x) = 0$ in $U_i$ is real, hence equal to the last coordinate of $\xi_i(\boldsymbol{y}')$. This shows the continuity for $\xi_i$.

□

To show our theory does generalize Collins's original cylindrical algebraic decomposition, we will show that [Col75, Theorem 1] is a special case of Theorem 3.3.7.

**Definition 3.3.8** (Collins's Delineability)**.** Let $f(x_1, \dots, x_r) \in R[x_1, \dots, x_r]$ $(r \geq 2)$, and let $S \subseteq R^{r-1}$ be a semi-algebraic set. The roots of $f$ are delineable over $S$ and that $\xi_1, \dots, \xi_k$ $(k \geq 0)$ delineate the real roots of $f$ over $S$ if the following conditions are satisfied:

1. There are $m \geq k$ positive integers $\mu_i$ such that if $(a_1, \dots, a_{r-1}) \in S$ then
$$f(a_1, \dots, a_{r-1}, x) = 0$$
has exactly $m$ distinct roots (in $C$), with multiplicities $\mu_1, \dots, \mu_m$.
2. $\xi_1 < \xi_2 < \dots < \xi_k$ are semi-algebraic continuous functions from $S$ to $R$.
3. If $(a_1, \dots, a_{r-1}) \in S$ then $\xi_i(a_1, \dots, a_{r-1})$ is a root of $f(a_1, \dots, a_{r-1}, x) = 0$ of multiplicity $\mu_i$ for all $1 \leq i \leq k$ and $(a_1, \dots, a_{r-1}) \in S$.
4. If $(a_1, \dots, a_{r-1}) \in S$, $b \in R$ and $f(a_1, \dots, a_{r-1}, b) = 0$ then $b = \xi_i(a_1, \dots, a_{r-1})$ for some $i$.

**Corollary 3.3.9** (Collins cf. [Col75, Theorem 1])**.** *Let $f(x_1, \dots, x_r)$ be an $R$-polynomial, $r \geq 2$. Let $S$ be a semi-algebraically connected semi-algebraic subset of $R^{r-1}$. If $\mathtt{LC}_{x_r}(f)$ has no zero on $S$ and the number of distinct roots of $f$ in $x_r$ is invariant on $S$, then the roots of $f$ are delineable on $S$.*

**Proof.** Set $A = R[x_1, \dots, x_{r-1}]$ and $B = R[x_1, \dots, x_r]/\langle f \rangle$. Let $X = \operatorname{Spec} B$ and $Y = \operatorname{Spec} A$. There is a natural morphism $\pi : X \longrightarrow Y$ induced by the ring map $R[x_1, \dots, x_{r-1}] \longrightarrow R[x_1, \dots, x_r]/\langle f \rangle$. If $\deg_{x_r}(f) = 0$ then we are already done. So let us suppose $\deg_{x_r}(f) > 0$ and set $w = \mathtt{LC}_{x_r}(f)$, then $B_w$ is a free $A_w$-module of finite rank (generated by monomials $1, x_r, \dots, x_r^{\deg_{x_r}(f)-1}$). In other words, $\operatorname{Spec} B_w \longrightarrow \operatorname{Spec} A_w$ is a finite free morphism. The rest immediately follows from Theorem 3.3.7 (note that the constant multiplicity is implied in the proof of Theorem 3.3.7.) □

Motivated by Collins's Delineability, we will define Geometric Delineability for a variety.

**Definition 3.3.10** (Geometric Delineability I)**.** Let $T = V(f_1, \dots, f_s)$ be a variety in $\mathbb{A}_R^{n+1}$, where $f_i \in R[x_1, \dots, x_{n+1}]$. Let $S$ be a semi-algebraically connected semi-algebraic set in $R^n$. We say $T$ is **(strictly) geometrically delineable** over $S$, if the following conditions holds:

1. For all $a = (a_1, \ldots, a_n) \in S$, the number of distinct solutions (in $C$) of

$$f_1(a; x_{n+1}) = \cdots = f_s(a; x_{n+1}) = 0$$

is finite and invariant.

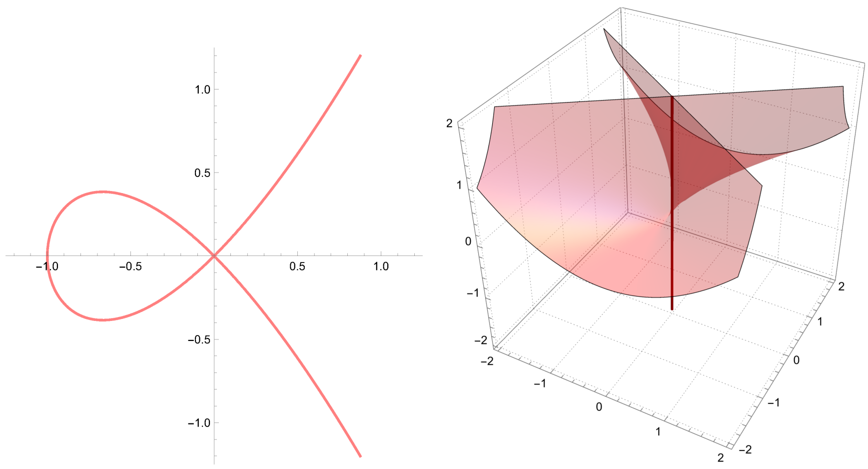
2. There are semi-algebraic continuous functions $\xi_1, \ldots, \xi_k : S \to T(R)$ such that $\pi \circ \xi_i = \mathrm{id}_S$, where $\pi : \mathbb{A}_R^{n+1} \to \mathbb{A}_R^n$ is the projection map. Additionally, we can require $\xi_i$ to be arranged in strictly ascending order in the last coordinate.

3. If $t \in T(R) \cap (S \times R^1)$, then there is some unique $1 \leq i \leq k$ such that $t = \xi_i(\pi(t))$.

The semi-algebraic continuous functions $\xi_i$ are called the **sections** of $T$ over $S$.

We make the convention that if $T \cap (S \times R^1) = S \times R^1$ (i.e., $T$ is a cylinder over $S$), then we say $T$ is **(vacuously) geometrically delineable** over $S$.

**Example 1** (Geometric Delineability Examples)**.**




(a) Nodal Curve (b) Whitney Umbrella

Figure 3.1 Several Geometric Objects to Demonstrate Geometric Delineability

Figure 3.1(a) is the nodal curve $E : y^2 - x^2(x+1) = 0$ in $R^2$. Let

$$I_1 = (-\infty, -1), p_1 = \{-1\}, I_2 = (-1, 0), p_2 = \{0\} \text{ and } I_3 = (0, +\infty)$$

be intervals or points in $R^1$. It is easy to see that $E$ is geometrically delineable over these subsets of $R^1$. For example, $E$ is geometrically delineable over $I_2$ because there are two different roots of $y^2 - x^2(x+1) = 0$ in $y$ when $x \in I_2$, given by $y = \pm\sqrt{x^2(x+1)}$.

Figure 3.1(b) is the Whitney's umbrella $U : x^2 - y^2 z = 0$ in $R^3$. Clearly $U$ is geometrically delineable over $\{(x, y) \in R^2 | y \neq 0\}$ because $z = \frac{x^2}{y^2}$ over this region. Also $U$ is geometrically delineable over $\{(x, y) \in R^2 | x \neq 0, y = 0\}$ because there is no solution in $z$ when $y = 0$

but $x \neq 0$. By the convention we made in Definition 3.3.10, $U$ is (vacuously) geometrically delineable over $\{(0,0)\}$ because the whole $z$-axis is contained in $U$.

We can rephrase Theorem 3.3.7 in our new terminology by just saying $X$ is geometrically delineable over $S$.

**Theorem 3.3.7** (Reformulation in Geometric Delineability)**.** *If $\pi : X \to Y$ is a finite free cylindrical projection between affine $R$-varieties. Let $S \subseteq Y(R)$ be a semi-algebraic connected semi-algebraic set such that $\#X_{\overline{y}}$ is invariant over $S$, then $X$ is geometrically delineable over $S$.*

*Remark* 2. While Collins's delineability in [Col75] requires constant multiplicity in each section, this requirement is dropped in our definition of geometric delineability. However, these two notions are equivalent for a finite free cylindrical projection from a hypersurface (recall the proof of Theorem 3.3.7). The constant multiplicity is never used in the c.a.d. construction, so we do not include it in the definition. There is an alternative definition of delineability from Arnon, Collins and McCallum that do not require a constant geometric fiber cardinality or a constant multiplicity [ACM84].

We propose Algorithm 3.1 to compute the regions over which a variety $V(I)$ is geometrically delineable. The extra output will be useful later (Lemma 3.3.17).

*Remark* 3. The reason for this name "`UnramifiedStratification`" is that the cardinality of geometric fiber remains the same over the regions defined by each pair of ideals in Output 1, so the fiber does not "ramify" into different branches. It should be also noted that, the regions defined by Output 1 are not necessarily disjoint.

**Theorem 3.3.11.** *Algorithm 3.1 terminates. For each tuples $(J_i, J'_i)$ in the Output 1, $V(I)$ is geometrically delineable over any semi-algebraic connected semi-algebraic set contained in $V(J_i)\backslash V(J'_i)$.*

*Moreover, for each tuples $(J_i, J'_i)$ in the Output 1, the locally closed set $V(J_i)\backslash V(J'_i)$ can be obtained by finite intersections, unions and differences of varieties defined by the ideals $\{\mathfrak{a}_i\}$ in the Output 2.*

**Proof.** *Termination*. It suffices to show that there is no infinite recursion. By Lemma 2.3.7(b), for every $f \in \mathcal{G}_x$, the leading coefficient $w_0$ is not in $I$. So $I + \langle w_0 \rangle$ is a strictly larger ideal. Since $k[\boldsymbol{y}, x]$ is noetherian, the algorithm certainly terminates.

*Correctness*. The correctness can be proved by a standard technique called "Noetherian Induction" [Har77, Exercise II.3.16]. In other words, it suffices to show that if our algorithm works for closed subschemes of $V(I)$, then it works for $V(I)$ too.

**Algorithm 3.1:** UnramifiedStratification

**Input:** An ideal $I \trianglelefteq k[\boldsymbol{y}, x]$; Here, $k$ is a subfield of $R$.

**Output 1:** A list of pairs of ideals in $k[\boldsymbol{y}]$: $\{(J_i, J_i')\}$ such that $V(I)$ is geometrically delineable over any semi-algebraically connected semi-algebraic set contained in $V(J_i)\backslash V(J_i')$.

**Output 2:** An extra list of ideals $\{\mathfrak{a}_i\}$ in $k[\boldsymbol{y}]$. For each tuples $(J_i, J_i')$ in the Output 1, the locally closed set $V(J_i)\backslash V(J_i')$ is a finite Boolean combination of the sets $\{V(\mathfrak{a}_i)\}$ (using $\cap, \cup, \backslash$).

Compute a reduced Gröbner basis $\mathcal{G}$ of $I$ w.r.t some block order $x > \boldsymbol{y}$;
$\mathcal{G}_y := \mathcal{G} \cap k[\boldsymbol{y}], \mathcal{G}_x := \mathcal{G}\backslash\mathcal{G}_y$;
**if** $\mathcal{G}_x = \varnothing$ **then** // $V(I)$ is a cylinder
  **return** $\{(\langle\mathcal{G}_y\rangle, \langle 1\rangle)\}, \{\langle\mathcal{G}_y\rangle\}$ ; // Vacuously geometrically delineable
$w := \prod_{f\in\mathcal{G}_x} \mathrm{LC}_x(f)$ ; // Effective Generic Freeness
$H := \texttt{HermiteQuadraticForm}(\mathcal{G}), s := \texttt{sizeof}(H)$;
$o_1 := \{(\langle\mathcal{G}_y\rangle + \texttt{minors}(i+1, H), \langle\mathcal{G}_y\rangle + \texttt{minors}(i, H)\cdot\langle w\rangle)\}_{i=0}^{s}$;
/* $\texttt{minors}(i, H)$ computes the ideal generated by all $i$-minors in $H$. It returns $\langle 1\rangle$ when $i = 0$, and $\langle 0\rangle$ when $i = s+1$. */
$o_2 := \{\langle\mathcal{G}_y\rangle + \texttt{minors}(i+1, H)\}_{i=0}^{s}$;
**foreach** $f \in \mathcal{G}_x$ **do** // Recursively look at the non-free locus
  $w_0 := \mathrm{LC}_x(f)$;
  **if** $\langle\mathcal{G}_y\rangle + \langle w_0\rangle \neq (I + \langle w_0\rangle) \cap k[\boldsymbol{y}]$ **then**
    $o_2 := o_2 \cup \{\langle\mathcal{G}_y\rangle + \langle w_0\rangle\}$;
  Merge $o_1$ and $o_2$ with $\texttt{UnramifiedStratification}(I + \langle w_0\rangle)$;
**return** $o_1$, $o_2$;

**Algorithm 3.2:** HermiteQuadraticForm

**Input:** A reduced Gröbner basis $\mathcal{G}$ of an ideal $I \trianglelefteq k[\boldsymbol{y}, x]$ with respect to some block order $x > \boldsymbol{y}$.

**Output:** The relative Hermite quadratic Ffrm $H$.

$\mathcal{G}_y := \mathcal{G} \cap k[\boldsymbol{y}], \mathcal{G}_x := \mathcal{G}\backslash\mathcal{G}_y$;
Choose $f \in \mathcal{G}_x$ of the minimal degree, set $w = \mathrm{LC}_x(f)$;
$s := \deg f$ ; // The monomial basis $\mathscr{B} = \{1, x, \ldots, x^{s-1}\}$
**for** $i := 0$ **to** $s-1$ **do**
  **for** $j := 0$ **to** $s-1$ **do** // Compute the trace of $L_{x^{i+j}}$
    $H_{i,j} := 0$;
    **for** $k := 0$ **to** $s-1$ **do** // Compute the image of $x^k$ under $L_{x^{i+j}}$
      $r := \texttt{PolynomialDivisionRemainder}(x^{i+j+k}, f)$ ; // Polynomial Division over $k(\boldsymbol{y})[x]$. This does the NormalForm computation
      $H_{i,j} := H_{i,j} + \texttt{coef}(r, k)$ ; // the $k$-th coefficient of $r$
Multiply $H$ by a suitable power of $w$ so entries of $H$ are in $k[\boldsymbol{y}]$;
/* The entries of $H$ are in $k[y]_w$, so there is some power $w^\ell$ that cancels all the denominators. This does not affect the rank of $H$. */
**return** $H$;

In Line 4, the algorithm finds that $I$ is fully contained in $k[\boldsymbol{y}]$, so $V(I)$ is vacuously geometrically delineable over $V(\langle \mathcal{G}_y \rangle)$ and our algorithm is correct in this case.

Now suppose $\mathcal{G}_x \neq \varnothing$, then by Effective Generic Freeness (Lemma 2.3.7(d)), $k[\boldsymbol{y}, x]_w / I_w$ is a free $(k[\boldsymbol{y}]_w / J_w)$-module, where $J = \langle \mathcal{G}_y \rangle = I \cap k[\boldsymbol{y}]$. Also, this module is necessary of finite rank, otherwise $\mathcal{G}_x$ cannot be non-empty. Therefore define $X = (\operatorname{Spec} k[\boldsymbol{y}, x]_w / I_w) \times_k R$ and $Y = (\operatorname{Spec} k[\boldsymbol{y}]_w / J_w) \times_k R$, the morphism

$$\pi : X \to Y$$

is a finite free cylindrical projection of $R$-varieties.

Next, the algorithm computes relative Hermite quadratic form $H$ by calling Algorithm 3.2. The correctness of Algorithm 3.2 is quite direct, as it simply computes the matrix of left multiplication by $x^{i+j}$ using the monomial basis $\mathscr{B}$ and takes the trace.

Then, the algorithm calculates ideals of minors. By the specialization property of relative Hermite quadratic form (Lemma 3.2.5) and Corollary 3.2.3, if all $(i+1)$-minors vanish at $p \in Y$, but some $i$-minor does not vanish at $p$, then the rank of $H$ at $p$ is $i$ and $\# X_{\overline{p}} = i$.

As a consequence, the geometric fiber cardinality is a constant $i$ on

$$\begin{aligned} & D(w) \cap \left( V\left( \langle \mathcal{G}_y \rangle + \mathtt{minors}(i+1, H) \right) \setminus V\left( \langle \mathcal{G}_y \rangle + \mathtt{minors}(i, H) \right) \right) \\ = \; & V\left( \langle \mathcal{G}_y \rangle + \mathtt{minors}(i+1, H) \right) \setminus V\left( \langle \mathcal{G}_y \rangle + \mathtt{minors}(i, H) \cdot \langle w \rangle \right) \\ = \; & V\left( \langle \mathcal{G}_y \rangle + \mathtt{minors}(i+1, H) \right) \setminus \left( V\left( \langle \mathcal{G}_y \rangle + \mathtt{minors}(i, H) \right) \cup V(\langle \mathcal{G}_y \rangle + \langle w \rangle) \right). \end{aligned} \tag{3.1}$$

Suppose $S$ is a semi-algebraically connected semi-algebraic set contained in it, then by Theorem 3.3.7, we can conclude that $V(I)$ is geometrically delineable over $S$. And the set in Equation (3.1) can be obtained from finite intersections, unions and differences of the varieties defined by Output 2, if $V(\langle \mathcal{G}_y \rangle + \langle w \rangle)$ is so.

Finally, Algorithm 3.1 takes care of the non-free locus $V(\langle \mathcal{G}_y \rangle + \langle w \rangle)$, which is the union $\bigcup_{f \in \mathcal{G}_x} V(\langle \mathcal{G}_y \rangle + \langle \mathtt{LC}_x(f) \rangle)$. For any $f \in \mathcal{G}_x$, let $w_0 = \mathtt{LC}_x(f)$, $I' = I + \langle w_0 \rangle$ and $J' = I' \cap k[\boldsymbol{y}]$. Clearly, if $V(I')$ is geometrically delineable over some $S \subseteq V(J')$ then $V(I)$ is also geometrically delineable over $S$ because $V(I') = V(I) \cap (V(w_0) \times \mathbb{A}^1_R)$. So the recursion is correct. Also, Line 12 guarantees that $V(\langle \mathcal{G}_y \rangle + \langle w \rangle)$ is a union of varieties defined by Output 2.

□

**Example 2.** Let us work with a concrete example. Consider the cylindrical projection

$$X = \operatorname{Spec} k[a, b, c, x] / \langle a x^2 + b x + c \rangle \to Y = \operatorname{Spec} k[a, b, c].$$

Algorithm 3.1 (`UnramifiedStratification`) shall recursively stratify $Y$ into subschemes on which $X$ is geometrically delineable.

The effective generic freeness gives $w = a$, so $D(a) \subsetneq Y$ is a subscheme of $Y$ where the projection is finite and free. On this subscheme, the determinant of Hermite quadratic form is the discriminant $b^2-4ac$. Therefore $X$ is geometrically delineable over $D(b^2-4ac)\cap D(a)$ and $V(b^2-4ac)\cap D(a)$. These two sets are added to $o_1$. At the same time, the closed subschemes $V(0)$ and $V(b^2-4ac)$ are added to $o_2$.

Next, the algorithm stratifies the non-free locus $V(a)$. Apply the generic freeness to $\langle a, ax^2+bx+c\rangle$ yields $w = b$. Therefore the $V(a)\cap D(b)$ is a subscheme of $Y$ where the projection is finite and free, of degree 1. Thus $X$ is geometrically delineable over $V(a)\cap D(b)$, which is added to $o_1$. The closed subscheme $V(a)$ is appended to $o_2$. By Line 12, $V(a,b)$ is also appended to $o_2$.

Finally, the algorithm stratifies $V(a,b)$. The pullback of $V(a,b)$ to $X$ is

$$\operatorname{Spec} k[a,b,c,x]/\langle a,b,ax^2+bx+c\rangle = \operatorname{Spec} k[a,b,c,x]/\langle a,b,c\rangle.$$

This is vacuously geometrically delineable. Therefore $V(a,b,c)$ is appended to both $o_1$ and $o_2$.

To sum up, $X$ is geometrically delineable over

$$o_1 = \{D(a)\cap D(b^2-4ac), D(a)\cap V(b^2-4ac), V(a)\cap D(b), V(a,b,c)\},$$

which are finite Boolean combinations of

$$o_2 = \{V(0), V(b^2-4ac), V(a), V(a,b), V(a,b,c)\}.$$

### 3.3.2 Geometric Delineability for Several Varieties

We now extend Geometric Delineability to multiple varieties, just as Collins extended his delineability to multiple polynomials.

**Definition 3.3.12** (Geometric Delineability II)**.** A family of varieties $\{V_i\}_{i=1}^{l}$ in $\mathbb{A}_R^{n+1}$ is said to be **(jointly) geometrically delineable** over a semi-algebraically connected semi-algebraic set $S \subseteq R^n$, if each $V_i$ is geometrically delineable over $S$, and all the sections of $V_i$ $(i = 1,\dots,l)$ are either identical or disjoint. As a consequence, the set of sections $\{\xi_{ij} : S \to R^{n+1}\}$ can be sorted in strictly ascending order in the last coordinate.

The next lemma concludes the joint geometric delineability from the geometric delineability of the intersection.

**Lemma 3.3.13.** *Let $X$, $Y$ be two varieties in $\mathbb{A}_R^{n+1}$, and $Z = X \cap Y$. Suppose $X$, $Y$ and $Z$ are all geometrically delineable over a semi-algebraically connected semi-algebraic set $S \subseteq R^n$. Then $\{X, Y\}$ is jointly geometrically delineable over $S$.*

**Proof.** If either $X$ or $Y$ is vacuously geometrically delineable over $S$ then the lemma is trivially true and we can simply rule out this case.

Now suppose $\xi$ ($\eta$ resp.) is a section of $X$ ($Y$ resp.) over $S$ and $\xi(s_0) = \eta(s_0)$ for some $s_0 \in S$. Since $Z = X \cap Y$ is geometrically delineable over $S$, there is a section $\zeta$ of $Z$ such that $\xi(s_0) = \eta(s_0) = \zeta(s_0)$. Because $Z \subseteq X$, for every $s \in S$, $\zeta(s)$ has to agree with some section of $X$. Then by the continuity, $\zeta$ has to agree with the same section in an Euclidean neighborhood of $s$. Therefore

$$S_\xi = \{s \in S | \zeta(s) = \xi(s)\}$$

is an Euclidean open non-empty semi-algebraic subset of $S$. By an identical argument, we can show that for every section $\xi^*$ of $X$: $\{s \in S | \zeta(s) = \xi^*(s)\}$ is an Euclidean open semi-algebraic subset of $S$. So these open subsets constitute a finite disjoint open cover for $S$. But $S$ is semi-algebraically connected, forcing all but one are them is empty. This shows $S_\xi = S$, i.e. $\zeta$ and $\xi$ agree everywhere. Similarly $\zeta$ and $\eta$ must agree on $S$. Therefore all sections of $X$ and $Y$ are either identical or disjoint. □

Therefore, to show the joint geometric delineability of a family of varieties $\{V_i\}$, it suffices to show the geometric delineability of each variety $V_i$ and intersection $V_i \cap V_j$ for each pair of varieties. The next theorem shows that we can conclude the geometric delineability of a closed subscheme $Z \subseteq X$ (an intersection for example) from a weaker condition, if we have the geometric delineability of $X$.

**Theorem 3.3.14.** *Suppose $\pi : X \to Y$ is a finite free cylindrical projection of $R$-varieties. Let $Z$ be a closed subscheme of $X$. If $\pi|_Z : Z \to Y$ is also finite free and $S \subseteq Y(R)$ is a semi-algebraically connected semi-algebraic set satisfying that $\#X_{\overline{y}}$ is a constant for all $y \in S$, then $\#Z_{\overline{y}}$ is also a constant for all $y \in S$.*

*As a corollary, $Z$ is geometrically delineable over $S$.*

**Proof.** Let $Y = \operatorname{Spec} A$. Since $\pi$ is a finite free morphism, $X$ can be identified as $\operatorname{Spec} A[x]/\langle f \rangle$ for some monic $f \in A[x]$ by Lemma 3.3.6. Because $\pi|_Z : Z \to Y$ is finite and free, $Z \cong \operatorname{Spec} A[x]/\langle g \rangle$ for some monic $g \in A[x]$ by Lemma 3.3.6 again.

Now embed $Y$ in some affine space $\mathbb{A}^n_R = \operatorname{Spec} R[y_1, \dots, y_n]$ as a closed subscheme, so the monic polynomials $f, g$ can be taken from $R[y_1, \dots, y_n, x]$ by a slight abuse of notations. Let $(c_1, \dots, c_n) \in S$ and let $x_1, \dots, x_l \in C$ be the distinct solutions of $f(c_1, \dots, c_n; x) = 0$ in $C$. Denote the multiplicity of $x_i$ as a root of $f(c_1, \dots, c_n; x) = 0$ by $\mu_i$. Notice that $Z$ is a subscheme of $X$, so by renumbering the roots if necessary, we can assume that $x_1, \dots, x_m$ are the all distinct solutions of $g(c_1, \dots, c_n; x) = 0$ in $C$ $(0 \leq m \leq l)$. Denote the multiplicity of $x_i$ as a root of $g(c_1, \dots, c_n; x) = 0$ by $\lambda_i$.

Choose disjoint open Euclidean neighborhoods $U_1, \dots, U_l$ for $x_1, \dots, x_l$. By the continuity of roots (Theorem 3.3.2), there is an open Euclidean neighborhood $U$ of $(c_1, \dots, c_n)$ such that for all $(c'_1, \dots, c'_n) \in U \cap S$, the equation $f(c'_1, \dots, c'_n; x) = 0$ has exactly $\mu_i$ solutions, counting multiplicity in $U_i$ for all $1 \leq i \leq l$, and the equation $g(c'_1, \dots, c'_n; x) = 0$ has exactly $\lambda_i$ solutions, counting multiplicity in $U_i$ for all $1 \leq i \leq m$.

As $\# X_{\overline{y}}$ is a constant over $S$, for all $(c'_1, \dots, c'_n) \in U \cap S$, the equation $f(c'_1, \dots, c'_n; x) = 0$ should have only one solution of multiplicity $\mu_i$ in $U_i$ for all $1 \leq i \leq l$. Notice that the roots of $g(c'_1, \dots, c'_n; x) = 0$ are also roots of $f(c'_1, \dots, c'_n; x) = 0$, hence the equation $g(c'_1, \dots, c'_n; x) = 0$ have only one solution of multiplicity $\lambda_i$ in $U_i$ for all $1 \leq i \leq m$, and these are all the roots of $g$ by counting degree. Therefore for all $(c'_1, \dots, c'_n) \in U \cap S$, the number of distinct complex solutions to the equation $g(c'_1, \dots, c'_n; x) = 0$ is a constant $m$. In other words, $\# Z_{\overline{y}}$ is a locally constant function on $S$. It is easy to show that this is a semi-algebraic function using an argument similar to Lemma 3.3.3. So by Lemma 3.3.1, $\# Z_{\overline{y}}$ is a constant on $S$.

The last assertion follows immediately from Theorem 3.3.7. $\square$

**Corollary 3.3.15.** *Suppose $X, Z$ are closed subschemes of $\mathbb{A}^{n+1}_R$ and $Z$ is a closed subscheme of $X$. Let $Y, W$ be affine subschemes of $\mathbb{A}^n_R$ and let $\pi : \mathbb{A}^{n+1}_R \to \mathbb{A}^n_R$ be the projection map (a subscheme is an open subscheme of a closed subscheme). If $X \times_{\mathbb{A}^n_R} Y \to Y$ and $Z \times_{\mathbb{A}^n_R} W \to W$ are finite free morphisms, and $S \subseteq (Y \cap W)(R)$ is a semi-algebraically connected semi-algebraic set on which $X_{\overline{y}}$ is a constant, then $Z$ is geometrically delineable over $S$.*

**Proof.** We have the following commutative diagram.

$$\begin{tikzcd}
 & Z \times_{\mathbb{A}^n_R} (Y \cap W) \arrow[r] \arrow[d] \arrow[dl, hook, "closed"'] \arrow[dr, phantom, "\square"] & Z \times_{\mathbb{A}^n_R} W \arrow[d] \\
X \times_{\mathbb{A}^n_R} (Y \cap W) \arrow[r] \arrow[d] \arrow[dr, phantom, "\square"] & Y \cap W \arrow[r] \arrow[d] \arrow[dr, phantom, "\square"] & W \arrow[d] \\
X \times_{\mathbb{A}^n_R} Y \arrow[r] & Y \arrow[r] & \mathbb{A}^n_R
\end{tikzcd}$$

Notice that finite free morphisms are stable under base change, so $Z \times_{\mathbb{A}^n_R} (Y \cap W) \to Y \cap W$ and $X \times_{\mathbb{A}^n_R} (Y \cap W) \to Y \cap W$ are also finite free morphisms. Applying Theorem 3.3.14 to the upper-left triangle in the diagram completes the proof. ☐

We present Algorithm 3.3, which stratifies the image of a projection into free-loci.

**Algorithm 3.3:** FreeStratification

**Input:** An ideal $I \trianglelefteq k[\boldsymbol{y}, x]$; Here, $k$ is a subfield of $R$.

**Output 1:** A list of pairs of ideals in $k[\boldsymbol{y}]$: $\{(J_i, J_i')\}$ such that the projection is free on $V(J_i) \backslash V(J_i')$

**Output 2:** An extra list of ideals $\{\mathfrak{a}_i\}$ in $k[\boldsymbol{y}]$. For each tuples $(J_i, J_i')$ in the Output 1, the locally closed set $V(J_i) \backslash V(J_i')$ is a finite Boolean combination of the sets $\{V(\mathfrak{a}_i)\}$ (using $\cap, \cup, \backslash$)

Compute a reduced Gröbner basis $\mathcal{G}$ of $I$ w.r.t some block order $x > \boldsymbol{y}$;
$\mathcal{G}_y := \mathcal{G} \cap k[\boldsymbol{y}], \mathcal{G}_x := \mathcal{G} \backslash \mathcal{G}_y$;
**if** $\mathcal{G}_x = \varnothing$ **then** // $V(I)$ is a cylinder
  **return** $\{(\langle \mathcal{G}_y \rangle, \langle 1 \rangle)\}, \{\langle \mathcal{G}_y \rangle\}$ ; // Clearly a free morphism
$w := \prod_{f \in \mathcal{G}_x} \mathrm{LC}_x(f)$ ; // Effective Generic Freeness
$o_1 := \{(\langle \mathcal{G}_y \rangle, \langle \mathcal{G}_y \rangle + \langle w \rangle)\}$;
/* Notice that $V(\langle \mathcal{G}_y \rangle) \backslash V(\langle \mathcal{G}_y \rangle + \langle w \rangle) = \operatorname{Spec} k[\boldsymbol{y}]_w / \langle \mathcal{G}_y \rangle_w$ is affine. Thus it makes sense to say the projection is free on this set. */
$o_2 := \{\langle \mathcal{G}_y \rangle\}$;
**foreach** $f \in \mathcal{G}_x$ **do** // Recursively look at the non-free locus
  $w_0 := \mathrm{LC}_x(f)$;
  **if** $\langle \mathcal{G}_y \rangle + \langle w_0 \rangle \neq (I + \langle w_0 \rangle) \cap k[\boldsymbol{y}]$ **then**
    $o_2 := o_2 \cup \{\langle \mathcal{G}_y \rangle + \langle w_0 \rangle\}$;
  Merge $o_1$ and $o_2$ with `FreeStratification`$(I + \langle w_0 \rangle)$;
**return** $o_1$, $o_2$;

**Theorem 3.3.16.** *Algorithm 3.3 terminates. For each tuples $(J_i, J_i')$ in the Output 1, the projection is free over $V(J_i) \backslash V(J_i')$.*

*Moreover, for each tuples $(J_i, J_i')$ in the Output 1, the locally closed set $V(J_i) \backslash V(J_i')$ can be obtained by finite intersections, unions and differences of varieties defined by the ideals $\{\mathfrak{a}_i\}$ in the Output 2.*

**Proof.** The proof is identical to Theorem 3.3.11. ☐

### 3.3.3 A New Algorithm for Cylindrical Algebraic Decomposition

We assume that field operations in $R$ $(+, -, \times, /, <, =)$ can be effectively performed, so the real roots of a univariate equation can be effectively represented and computed (e.g. $R = \mathbb{R}_{\text{alg}}$).

Solving equations can be done by the real root isolation algorithm when $R$ is archimedean (see [BPR06, Algorithm 10.55]) and Thom Encoding if $R$ is a general real closed field (see [BPR06, Algorithm 10.105, 10.106]).

The following simple lemma is crucial in our algorithm for cylindrical algebraic decomposition.

**Lemma 3.3.17.** *Let $\mathcal{T} = \{T_i\}$ be a finite family of semi-algebraic sets in $R^n$, and $\mathfrak{D}$ is a cylindrical algebraic decomposition adapted to $\mathcal{T}$. Suppose $\mathcal{T}' = \{T'_j\}$ is another finite family of semi-algebraic sets in $R^n$ such that each $T'_j$ can be obtained by a finite number of intersections, unions, and differences of elements in $\mathcal{T}$. Then $\mathfrak{D}$ is also a cylindrical algebraic decomposition adapted to $\mathcal{T}'$.*

**Proof.** It suffices to prove that for any $T_i, T_j \in \mathcal{T}$, $\mathfrak{D}$ is also adapted to $\mathcal{T} \cup \{T_i \cap T_j, T_i \cup T_j, T_i \backslash T_j\}$.

To see this, write $T_i = \cup_{\alpha\in\Lambda} S_\alpha$ and $T_j = \cup_{\alpha\in\Lambda'} S_\alpha$ as unions of disjoint cells of level $n$. Then $T_i \cup T_j = \cup_{\alpha\in\Lambda\cup\Lambda'} S_\alpha$ is a union of cells of level $n$. Similarly $T_i \cap T_j = \cup_{\alpha\in\Lambda\cap\Lambda'} S_\alpha$ and $T_i \backslash T_j = \cup_{\alpha\in\Lambda\backslash\Lambda'} S_\alpha$. □

**Algorithm 3.4:** CADProjection

**Input:** A list of ideals $\{I_k\}_{k=1}^s$ in $k[\boldsymbol{y}, x]$.
**Output 1:** A list of pairs of ideals in $k[\boldsymbol{y}]$: $\{(J_i, J'_i)\}$.
**Output 2:** An extra list of ideals $\{\mathfrak{a}_i\}$ in $k[\boldsymbol{y}]$.

```
o_1 = ∅, o_2 = ∅;
for k := 1 to s do
    Merge o_1, o_2 with UnramifiedStratification(I_k)
for k := 1 to s do
    for j := 1 to k − 1 do
        Merge o_1, o_2 with FreeStratification(I_j + I_k);
return o_1, o_2;
```

Algorithm 3.4 reduces a higher-dimensional c.a.d. computation to a lower-dimensional c.a.d. computation. It can be thought as an adaptation of Projection Operator to the geometric context.

**Theorem 3.3.18.** *Suppose $\{V(I_k)\}$ is a finite family of varieties in $\mathbb{A}_R^n$. A cylindrical algebraic decomposition $\mathfrak{D}'$ adapted to varieties defined by Output 2 of* `CADProjection({I_k})` *(Algorithm 3.4) can be lifted to a cylindrical algebraic decomposition $\mathfrak{D}$ adapted to $\{V(I_k)\}$.*

**Proof.** Observe that cells in a cylindrical algebraic decomposition are always semi-algebraically connected, because a cell of level $> 1$ is either a graph of a semi-algebraic continuous function defined over its silhouette, or a band between two graphs, and a cell of level 1 is either an interval or a point. So our assertion follows from an easy induction.

Using Lemma 3.3.17, it suffices to prove that a cylindrical algebraic decomposition $\mathfrak{D}'$ adapted to the family of constructible sets $\{V(J_i)\backslash V(J_i')\}$ can be lifted to a cylindrical algebraic decomposition $\mathfrak{D}$ of $\{V(I_k)\}$, where $\{(J_i, J_i')\}$ is the Output 1 (recall Theorem 3.3.11 and Theorem 3.3.16).

Let $S$ be a level $n-1$ cell of $\mathfrak{D}'$. For each $I_k$, since `CADProjection` contains the Output 1 of `UnramifiedStratification`$(I_k)$, $V(I_k)$ is geometrically delineable over $S$ by Theorem 3.3.11.

Next, we will show the joint geometric delineability of every pair of varieties $(V(I_j), V(I_k))$ on $S$. By Lemma 3.3.13, it suffices to show the intersection $V(I_j) \cap V(I_k) = V(I_j + I_k)$ is geometrically delineable over $S$. We may assume that $V(I_j)$ is not a cylinder over $S$ and $V(I_j+I_k)$ is non-empty over $S$ (otherwise the geometric delineability is trivial). Because `CADProjection` contains the Output 1 of `FreeStratification`$(I_j + I_k)$, the projection of $V(I_j + I_k)$ is free over $S$. By applying Corollary 3.3.15, we see that $V(I_j+I_k)$ is indeed geometrically delineable over $S$.

It remains to write down a cylindrical algebraic decomposition $\mathfrak{D}$ adapted to $\{V(I_k)\}$. We start with $\mathfrak{D}'$ and construct the cells of level $n$ to complete $\mathfrak{D}$. Let $S$ be cell of level $n-1$ in $\mathfrak{D}'$, each $V(I_k)$ is geometrically delineable over $S$ and the sections are either disjoint or identical. So we let the image of sections to be the cells of level $n$ over $S$. If there is some $V(I_k)$ vacuously geometrically delineable over $S$ then we also let the bands between sections/above the highest section/below the lowest section to be the cells of level $n$ over $S$. Clearly the cells are disjoint semi-algebraic sets. Therefore $\mathfrak{D}$ is indeed a cylindrical algebraic decomposition adapted to $\{V(I_k)\}$. □

It is time to introduce our cylindrical algebraic decomposition algorithm based on the geometric tools we developed (see Algorithm 3.5). The correctness of Algorithm 3.5 is obvious from Theorem 3.3.18.

**Example 3.** We continue to study Example 2 and compute a cylindrical algebraic decomposition adapted to $V(ax^2 + bx + c) \subsetneq \operatorname{Spec} R[a, b, c, x]$.

By Theorem 3.3.18, the task is reduced to computing a cylindrical algebraic decomposition adapted to $\{V(0), V(b^2 - 4ac), V(a), V(a, b), V(a, b, c)\}$ in $\operatorname{Spec} R[a, b, c]$.

**Algorithm 3.5:** GeometricCAD

**Input:** A list of varieties $\{V(I_k)\}_{k=1}^{s}$ in $\mathbb{A}_R^n$.

**Output:** A cylindrical algebraic decomposition adapted to $\{V(I_1), \ldots, V(I_s)\}$.

**if** $n > 1$ **then** // The Projection Phase
    $o_1, o_2$ := `CADProjection`$(\{I_k\}_{k=1}^{s})$;
    $\mathfrak{D}$ := `GeometricCAD`$(o_2)$ ; // Compute a c.a.d. adapted to $o_2$
**else**
    $\mathfrak{D} := \{\bullet\}$ ; // A c.a.d. of $R^0$
`cells` := $\varnothing$ ; // The cells of level $n$, initialized to be an empty list
**foreach** *cell $S$ of level $n-1$ in $\mathfrak{D}$* **do** // The Lifting Phase
    Choose a sample point $s \in S \subseteq R^{n-1}$;
    $b$ := `False`;
    $\xi := \varnothing$ ; // The sections over $S$, initialized to be an empty list
    **foreach** *ideal $I_k$* **do**
        $\langle f \rangle$:=Specialize $I_k$ with $s \in S$ ; // After specialization, the result is a principal ideal in $R[x_n]$
        **if** $f \neq 0$ **then**
            $\{\rho_i\}$ := The real roots of $f$ in ascending order;
            **foreach** $\rho_i$ **do**
                Append $\{(s, x) | s \in S,$
                 $x$ is the i-th least real root of $I_k$ specialized at $s\}$ to $\xi$;
        **else** // $V(I_k)$ is vacuously geometrically delineable over $S$
            $b$ := `True`;
    Sort $\xi$ in ascending order, `cells` := `cells` $\cup\, \xi$;
    **if** $b$ **then**
        Append the bands over $S$ to `cells`;
Append `cells` to $\mathfrak{D}$, **return** $\mathfrak{D}$;

Applying Algorithm 3.1 (`UnramifiedStratification`) to each variety, we see that each of them is geometrically delineable over cells of a cylindrical algebraic decomposition adapted to $\{V(0), V(a), V(a,b)\}$ in Spec $R[a,b]$. Using Algorithm 3.3 (`FreeStratification`), the joint geometric delineability can be established from free stratifications of pairwise intersections of varieties at Level 3. In this case, the results of `FreeStratification` are already contained in the results of `UnramifiedStratification`. Therefore, the task is further reduced to computing a cylindrical algebraic decomposition adapted to $\{V(0), V(a), V(a,b)\}$ in Spec $R[a,b]$, which can be furthermore reduced to dividing the real axis by the origin.

We start the lifting phase from the decomposition of $R^1$ consisting of three cells $\{a < 0\}$, $\{a = 0\}$, $\{a > 0\}$. This lifts to a decomposition of $R^2$ consisting five cells

$$\{\{a < 0\}, \{a = 0 \wedge b < 0\}, \{a = 0 \wedge b = 0\}, \{a = 0 \wedge b > 0\}, \{a > 0\}\},$$

adapting to $\{V(0), V(a), V(a,b)\}$. Continuing the lifting, we obtain a decomposition of $R^3$ consisting of eleven cells:

$$\begin{aligned}\Bigg\{\left\{a < 0 \wedge c < \frac{b^2}{4a}\right\}, \left\{a < 0 \wedge c = \frac{b^2}{4a}\right\}, \left\{a < 0 \wedge c > \frac{b^2}{4a}\right\},\ \{a = 0 \wedge b < 0\},\\ \{a = 0 \wedge b = 0 \wedge c < 0\}, \{a = 0 \wedge b = 0 \wedge c = 0\}, \{a = 0 \wedge b = 0 \wedge c > 0\},\\ \{a = 0 \wedge b > 0\}, \left\{a > 0 \wedge c < \frac{b^2}{4a}\right\}, \left\{a > 0 \wedge c = \frac{b^2}{4a}\right\}, \left\{a > 0 \wedge c > \frac{b^2}{4a}\right\}\Bigg\};\end{aligned}$$

and finally, a decomposition of real points of $V(ax^2 + bx + c)$ consisting of nine cells:

$$\begin{aligned}\Bigg\{\left\{a < 0 \wedge c = \frac{b^2}{4a} \wedge x = \frac{-b}{2a}\right\},\\ \left\{a < 0 \wedge c > \frac{b^2}{4a} \wedge x = \frac{-b + \sqrt{b^2 - 4ac}}{2a}\right\}, \left\{a < 0 \wedge c > \frac{b^2}{4a} \wedge x = \frac{-b - \sqrt{b^2 - 4ac}}{2a}\right\},\\ \left\{a = 0 \wedge b < 0 \wedge x = \frac{c}{b}\right\}, \{a = 0 \wedge b = 0 \wedge c = 0\}, \left\{a = 0 \wedge b > 0 \wedge x = \frac{c}{b}\right\},\\ \left\{a > 0 \wedge c < \frac{b^2}{4a} \wedge x = \frac{-b - \sqrt{b^2 - 4ac}}{2a}\right\}, \left\{a > 0 \wedge c < \frac{b^2}{4a} \wedge x = \frac{-b + \sqrt{b^2 - 4ac}}{2a}\right\},\\ \left\{a > 0 \wedge c = \frac{b^2}{4a} \wedge x = \frac{b}{2a}\right\}\Bigg\}.\end{aligned}$$

## 3.4 Analysis

### 3.4.1 Connection to the Prior Work

We compare our theory to some other existing methods. We hope the following discussion can be inspiring and illustrate the similarities and differences between our results and previous theories.

#### 3.4.1.1 Collins's Original CAD

It is very clear that our theory is deeply influenced by Collins's fundamental work [Col75]. Our algorithm follows the Projection-Lifting paradigm developed by Collins. Also, Algorithm 3.1 and 3.3 are analogues to Collins's Projection Operator `PROJ1` and `PROJ2`.

While Collins's goal is to compute an $\mathcal{F}$-sign-invariant cylindrical algebraic decomposition of the whole real affine space, our result is a c.a.d. adapted to a family of varieties. In order to compute an $\mathcal{F}$-sign-invariant c.a.d., Collins used his Projection Operator `PROJ` to reduce the task to computing a $\mathtt{PROJ}(\mathcal{F})$-sign-invariant c.a.d. in a lower-dimensional space. This idea is evolved in our geometric theory: a c.a.d. adapted to a family of varieties can be lifted from another lower-dimensional c.a.d. adapted to varieties.

To exploit all equations, one has to study the geometry more carefully, so there is a significant difference in techniques involved. The mathematics behind Collins's work is the classical Elimination Theory based on resultants and subresultants, and our theory has more recent ingredients from modern algebraic geometry, which are translated into algorithms by Gröbner basis via the standard algebra-geometry dictionary. These new ingredients allow us to manipulate varieties defined by ideals, rather than hypersurfaces defined by polynomials and reveal the geometric nature behind cylindrical algebraic decomposition.

In fact, a sign-invariant c.a.d. can be thought as a c.a.d. adapted to the hypersurfaces (variety of a single polynomial). Therefore, Collins's Projection can be explained in our new theory as a special case that tackles hypersurfaces only. The Collins-Hong Projection Operator `PROJ` consists of three types of polynomials: leading coefficients, (sub-)discriminants and subresultants. By Corollary 3.3.9, the leading coefficient of $f$ is used to construct the generic freeness of the map $\operatorname{Spec} R[x_1, \ldots, x_{n+1}]/\langle f\rangle \to \operatorname{Spec} R[x_1, \ldots, x_n]$, the sub-discriminants of $f$ are used to count the geometric fiber of the same map. Similarly, the subresultants of $f$ and $g$ are used to construct the generic freeness of the map $\operatorname{Spec} R[x_1, \ldots, x_{n+1}]/\langle f, g\rangle \to \operatorname{Spec} R[x_1, \ldots, x_n]$! So there are sufficiently many polynomials in the Collins-Hong Projection Operators to define the regions returned by our Algorithms 3.1 and 3.3.

The advantage of the new idea is the ability to exploit all equations in the projection phase, avoiding unnecessarily refined decomposition. Actually, a $\mathtt{PROJ}(\mathcal{F})$-sign-invariant c.a.d. is not necessarily needed. Since we can conclude the geometric delineability over a region, it should be more economic if we work directly with regions instead of polynomials. The reason is simple: a $\mathtt{PROJ}(\mathcal{F})$-sign-invariant c.a.d. is adapted to all semi-algebraic sets that can be defined by $\mathtt{PROJ}(\mathcal{F})$, but a c.a.d. adapted to some semi-algebraic sets defined by $\mathtt{PROJ}(\mathcal{F})$ is not necessarily $\mathtt{PROJ}(\mathcal{F})$-sign-invariant. Therefore, a $\mathtt{PROJ}(\mathcal{F})$-sign-invariant c.a.d. might decompose the real space into more pieces than what is really needed. The following example we learned from Christopher Brown can be used to show the advantage of geometric theory.

**Example 4.** Let $f = w^2 + ((x+y)^2u - yz) \in R[x, y, z, u, w]$. We will build an $f$-sign-invariant cylindrical algebraic decomposition of $R^5$, using the classical method and our method.

| | $\mathcal{F}$-sign-inv. c.a.d., where $\mathcal{F}$ = | c.a.d. adapted to |
|---|---|---|
| Level 5 | $w^2 + (x+y)^2u - yz$ | $V(0), V(f)$ |
| Level 4 | $(x+y)^2u - yz$ | $V(0), V((x+y)^2u - yz)$ |
| Level 3 | $(x+y)^2, yz$ | $V(0), V(x+y), V(x+y, yz)$ |
| Level 2 | $x+y, y$ | $V(0), V(x+y), V(x, y)$ |
| Level 1 | $x$ | $V(0), V(x)$ |

Table 3.1 A comparison of computation of $f$-sign-invariant c.a.d.

Using the Collins-Hong Projection Operator [Col75; Hon90], the problem is reduced to computing lower-dimensional sign-invariant c.a.d.s of $R^4, \dots, R^1$ and the projection polynomials can be found in the left column of Table 3.1. The final result consists of 297 cells in $R^5$.

When applying McCallum-Brown Projection Operator [McC98; Bro01], the same family of projection polynomials is generated. However, the nullification problem occurs, as the projection of $V((x+y)^2u - yz)$ downto $\mathbb{A}^3_R$ is not quasi-finite (it has 1-dimensional fibers over $V(x+y, yz)$). In this case, the use of McCallum-Brown projection is not guaranteed to be correct by their theory (although it generates the same c.a.d. as Collins-Hong Projection does).

Notice that a c.a.d. adapted to $\{V(f), V(0)\}$ is an $f$-sign-invariant c.a.d. of $R^5$, so it enables the use of our `GeometricCAD` algorithm. In our theory, this problem is reduced to compute a lower-dimensional c.a.d. adapted to another family of varieties recursively. The right column in Table 3.1 shows the varieties needed to compute the c.a.d.. The final result consists of 75 cells in $R^5$.

Comparing 75 cells with 297 cells, we see a huge saving in cell numbers. Why does `GeometricCAD` generate a more compact c.a.d.? The reason lies in the geometric nature of cylindrical algebraic decomposition: sign invariance is more than necessary to conclude delineability (or geometric delineability). To see this, consider the c.a.d. constructed at level 3: a $\{(x+y)^2, yz\}$-sign-invariant c.a.d. $\mathfrak{D}_3$ versus a c.a.d. $\mathfrak{D}'_3$ adapted to $\{V(0), V(x+y), V(x+y, yz)\}$. Clearly $\mathfrak{D}_3$ is also adapted to these varieties. But $(x+y)^2u - yz$ (or its locus) is obviously (geometrically) delineable over $V(0)\backslash V(x+y)$, $V(x+y)\backslash V(x+y, yz)$ or $V(x+y, yz)$ and the sign invariance of $\{(x+y)^2, yz\}$ is an overkill – we do not really care what happens over $V(yz)$, as long as $x+y \neq 0$. So $\mathfrak{D}_3$ partitions $R^3$ into more cells than $\mathfrak{D}'_3$, resulting a larger decomposition in the end. As a consequence, the classical algorithm yields 3, 13, 39, 99 and 297 cells at each level, while our algorithm `GeometricCAD` generates 3, 9, 13, 25 and 75 cells at each level.

#### 3.4.1.2 CAD with Equational Constraints

Usually, it is not necessary to decompose the whole affine real space but only a variety. For many problems from the real world, there are equational constraints that must be met.

Collins pointed out that the construction of cylindrical algebraic decomposition can be remarkably accelerated if taking equational constraints into consideration in [Col98]. Later McCallum made precise Collins's idea in [McC99] by reducing the projection polynomials in the first stage of projection phase. In [McC01], he further showed that repeated applications of "semi-restricted projection" in the projection phase is possible. The lower-dimensional equational constraints are obtained by a technique called "EC Propagation". To be more precise, Collins wrote in [Col98]:

> If $f_1$ and $f_2$ are both equational constraints polynomials then so is their resultant, $\mathtt{Res}(f_1, f_2)$, since
>
> $$f_1 = 0 \wedge f_2 = 0 \implies \mathtt{Res}(f_1, f_2) = 0.$$

And the propagation can be repeated many times.

Our strategy to exploit equational constraints here is significantly different from Collins's idea. Since our theory is based on varieties, the ability to exploit equational constraints is intrinsic. If we need to decompose $V(I)$ into cells on which $g_1, \ldots, g_t$ are sign-invariant, then it suffices to compute a cylindrical algebraic decomposition adapted to

$$\{V(I), V(I + \langle g_1 \rangle), \ldots, V(I + \langle g_t \rangle)\}.$$

We believe that our algorithm exploits equational constraints in a better way, as the following example demonstrates.

**Example 5** (Example 7 in [FIS15])**.** The goal is to compute a cylindrical algebraic decomposition adapted to a one-dimensional variety (curve) $C \subseteq \mathbb{A}_R^5 = \operatorname{Spec} R[x, y, z, u, w]$, and $C$ is defined by the locus of polynomials

$$\begin{aligned}
f_1 &= wu - 1,\\
f_2 &= w^2 - 2wy + u^2 - 2uz - x,\\
f_3 &= 32y^2 + 168yz + 40yx - 270y + 8z^2 + 20zx - 390z + 4x^2 - 105x + 450\\
f_4 &= 240yz - 16yx^2 - 532yx - 4480y - 800z^3 - 1240z^2x - 17720z^2 - 408zx^2\\
&\quad -6214zx + 25240z - 40x^3 - 550x^2 + 5695x + 1050 \text{ and}\\
f_5 &= 320yzx + 8320yz + 32yx^2 + 264yx - 14840y + 240z^2 + 16zx^2 - 372zx\\
&\quad -23380z - 140x^2 - 2575x + 36750.
\end{aligned}$$

Our algorithm `GeometricCAD` finishes the computation in 10.67 seconds, generating 200 cells at level 5.

We tried to compute a c.a.d. adapted to $C$ with the aid of QEPCAD [Bro03], which applies McCallum's equational constraint operator [McC99; McC01] after manually declaring $f_1, \ldots, f_5$ as equational constraints. However, QEPCAD quit after 14 minutes because of a software failure.

The use of resultants is completely abandoned in our geometric theory, and we believe this is beneficial because resultants do not always reflect the geometry faithfully. The phenomenon becomes apparent in EC Propagation as explained below.

1. The locus of $\mathtt{Res}_x(f, g)$ is not necessarily the closure of the projection of $V(f, g)$. In fact, we have

$$\begin{aligned}
V(\mathtt{Res}_x(f, g)) &= V(\mathtt{LC}_x(f), \mathtt{LC}_x(g)) \cup \pi(V(f, g))\\
&= V(\mathtt{LC}_x(f), \mathtt{LC}_x(g)) \cup \overline{\pi(V(f, g))}, \qquad (\star)
\end{aligned}$$

   which follows from Proposition 6 and Corollary 7 in [CLO15, Chapter 3, Section 6] quite easily. So there is a superfluous set $V(\mathtt{LC}_x(f), \mathtt{LC}_x(g))$, which does not necessarily lie in the silhouette of $V(f, g)$.
2. Successively applying resultants may distort the geometry, making things even worse. Suppose $f, g, h$ are three polynomials in $x, y, z$. Set $r_1 = \mathtt{Res}_x(f, g)$, $r_2 = \mathtt{Res}_x(f, h)$

and $r = \mathtt{Res}_y(r_1, r_2)$. Then

$$\begin{aligned}\pi_1(\pi_2(V(f,g,h))) &= \pi_1(\pi_2(V(f,g)\cap V(f,h)))\\ &\subseteq \pi_1(\pi_2(V(f,g))\cap \pi_2(V(f,h)))\\ &\subseteq \pi_1(V(r_1)\cap V(r_2))\\ &= \pi_1(V(r_1,r_2))\\ &\subseteq V(r).\end{aligned}$$

The first containment is by $\varphi(S\cap T)\subseteq \varphi(S)\cap\varphi(T)$ and the others follow from Equation ($\star$). Each containment can be strict, which leads to an unnecessarily large estimate of the silhouette of $V(f,g,h)$ (usually reflected in superfluous factors of successive resultants).

The failure of QEPCAD in Example 5 is possibly due to these reasons. The problem is overcome when using Gröbner basis, because the vanishing set of the elimination ideal is always the closure of the projection.

#### 3.4.1.3 CAD by Regular Chain

Apart from Collins's work, there is another approach to cylindrical algebraic decomposition, proposed by Chen, Moreno Maza, Xia and Yang in [Che+09] and [CMM14a], using regular chains. Our work was firstly motivated by [Che+09] in the following way: since regular chains can be obtained from a lexicographic Gröbner basis computation [WDM20], it is natural to ask if it is possible to build c.a.d. using only Gröbner basis?

Soon it turns out that [Che+09] reflects some geometric nature of cylindrical algebraic decomposition and one of the highlights in [Che+09] is that instead of directly building a cylindrical algebraic decomposition in $R^n$, they first build a cylindrical algebraic decomposition in $C^n$ then refine it to a cylindrical algebraic decomposition in $R^n$. From our perspective, a regular chain computation immediately provides generic freeness by inverting the leading coefficients and another discriminant computation counts the geometric fiber.

But the dissimilarity is also clear. Thanks to the standard algebra-geometry dictionary, our theory can be directly translated into an algorithm, while the algorithm in [Che+09] is sophisticated, using many high-level routines as black boxes due to the obscure relation between regular chains and varieties. Also, we can develop our theory much further without cylindrical projection, we do not use it until Theorem 3.3.7, which is about geometric delineability. It seems that recursively projecting the largest variable is an intrinsic nature of the theory of regular chains. This feature of regular chains makes it similar to the lexicographic Gröbner basis, which is usually hard to compute. As the readers can see, the monomial order does not play an important role in our work, and we can use cheaper monomial orders like graded reverse lexicographic order (grevlex) or block-grevlex (for elimination). Faugère's `FGb` package supports these two monomial orders [Fau10], and is believed to be the state-of-the-art implementation of Gröbner basis.

#### 3.4.1.4 Preprocessing with Gröbner Basis

The idea of combining Gröbner basis and cylindrical algebraic decomposition is actually not new. Buchberger and Hong has considered replacing equational constraints with their lexicographic Gröbner basis in [BH91], which usually speeds up the computation. It is also observed that, however sometimes computing a Gröbner basis would slow down the computation. Then, the experiments in [BH91] were revisited by Wilson, Bradford and Davenport in [WBD12], with more advanced implementations of the algorithms. Huang *et al.* applied machine learning to decide when to use a Gröbner basis in [Hua+16].

In short, the role that Gröbner basis plays in these works is a preprocessor, which is not the case in our theory. We use Gröbner basis as a fundamental algorithm to deal with ideals and varieties effectively.

#### 3.4.1.5 Comprehensive Gröbner System

Suppose $I \trianglelefteq k[\boldsymbol{y}, \boldsymbol{x}] = k[y_1, \ldots, y_m, x_1, \ldots, x_n]$. A Comprehensive Gröbner System for $I$ is a stratification of $\overline{k}^m$ into constructible sets $S_i$ such that each set is tagged with a finite subset $\mathcal{G}_i = \{f_{i,1}, \ldots, f_{i,n_i}\}$ of $k[\boldsymbol{y}, \boldsymbol{x}]$, satisfying for every $Y \in S_i$, the specialization of $\mathcal{G}_i$ at $Y$ is a Gröbner basis of the specialization of $I$ at $Y$. And a comprehensive Gröbner basis for $I$ is a finite subset of $k[\boldsymbol{y}, \boldsymbol{x}]$ that is a Gröbner basis under all specializations. These concepts are proposed by Weispfenning in [Wei92]. The theory of Comprehensive Gröbner System/Basis is very rich and well suited to study parametric problems, readers may refer to [LSW19] for a beautiful survey on this subject.

In [Wei98], Weispfenning used Comprehensive Gröbner Bases and Hermite quadratic forms to develop a real quantifier elimination procedure. His approach roughly goes as:

1. Convert the input to the prenex disjunctive normal form (DNF), which reduces the problem to producing an equivalent formula of

$$(\exists x_1) \cdots (\exists x_k)(\bigwedge_i f_i(x_1, \ldots, x_n) = 0 \wedge \bigwedge_i h_i(x_1, \ldots, x_n) > 0),$$

   which is the projection of a basic semi-algebraic set.

2. Reduce to the generically finite case using Comprehensive Gröbner Bases, then compute the Hermite quadratic forms associated with $1, h_i, h_i^2$.
3. Then the number of real solutions can be read off from the sign variations of the characteristic polynomials of the quadratic forms (especially when the number is zero).

This idea was further developed in [FIS15], where they used a more advanced Comprehensive Gröbner System algorithm by Suzuki and Sato [SS06].

We find that our Projection algorithms have a similar structure to the Comprehensive Gröbner System construction in [SS06]: compute a reduced Gröbner basis with respect to a block monomial order and pick out the leading coefficients. Moreover, we also use Hermite quadratic forms. These common features make our work comparable to [Wei98; FIS15]. The major difference lies in the ways using Hermite quadratic forms. Weispfenning and his successors directly use the signature of the matrices to count the real roots, while we use the rank to count the complex roots (as Collins uses subresultants). We believe this can result in a significant difference in the number of basic semi-algebraic sets and the number of varieties, because the number of sign variance of the characteristic polynomial is far more difficult to describe than the rank (see [FIS15, Section 5.1]). Moreover, converting to disjunctive normal form will also produce many basic semi-algebraic sets, and many of them are actually empty. These problem are avoided in our theory because cylindrical algebraic decomposition is able to postpone the construction of semi-algebraic sets in the Lifting Phase.

### 3.4.2 Experimental Results

We report an implementation in this subsection and compare it to existing tools. The source codes and all examples in the experiment are available at

https://github.com/xiaxueqaq/GeometricCADv2.

Our algorithm is implemented with Mathematica 13.3 [MMA]. Our experiment runs on a Windows PC with Intel i7-11700 (2.50GHz) CPU and 32GB memory.

We compare our implementation with

1. QEPCAD B 1.74 [Bro03], which applies McCallum-Brown Projection Operator [McC98; Bro01] by default. The command to launch QEPCAD is

   ```
   qepcad +N500000000 +L2000000.
   ```

   Also, we declare all the equational constraints (whenever there are) in QEPCAD, so it can utilize the McCallum Equational Constraint Operator [McC99; McC01].

2. `CylindricalAlgebraicDecompose` [CMM14b] in MAPLE 2021, which uses the algorithm in [CMM14a] by Regular Chains. This tool is denoted by RC-CAD later.

The testing examples are taken from [BF91], [EBD20], [FIS15] and [Hon91], except 5-VAR, which is Example 4. We summarize information about the benchmark in Table 3.2, including the number of variables, polynomials, equations and the (Krull) dimension. The number of variables varies from 3 to 10 and the benchmark has both low-dimensional and high-dimensional varieties. Also, there are two families of scalable examples (Hong-$n$ and cyclic-$n$). Hence our benchmark is a comprehensive one.

The result is recorded in Table 3.3. “>1h” stands for a killed computation that consumes more than 1 hour. “Error” means the computation halts due to a bug. Overall, it is evident that our algorithm solves the most problems. Also, our algorithm is of the best performance in 15 instances, which are 75% of the benchmark.

In the first 6 examples, there is no equational constraints, and their goal is to build a sign-invariant c.a.d. to the input polynomials. The input of Hong-$n$ consists of two $n$-variate polynomials: $\sum_{i=1}^{n} x_i^2 - 1$ and $\prod_{i=1}^{n} x_i - 1$. Although QEPCAD is the fastest tool on small instances like Hong-5 and Hong-6, it fails to finish the computation on larger problems. RC-CAD computes one more c.a.d. than QEPCAD, but it is slower than our algorithm. It crashes in Hong-8 for an unknown reason. Hence it is reasonable to say that our method outperforms the others in this category of data. While QEPCAD and GeometricCAD generate same numbers of cells on Hoon-$n$, it is interesting to see that RC-CAD uses fewer cells in c.a.d. construction.

Moving on to the other 14 examples, there are now equational constraints. The cyclic-$n$ family is a famous system in the symbolic computation community, which consists of cyclic sums of products of variables [BF91]. QEPCAD succeeds in 7 out of 14, but requires longer time in general. The failure of QEPCAD in FIS-7 and FIS-8 is due to an exhausted prime list in the `saclib` library. RC-CAD finishes two more instances and it is the most efficient tool on 3 examples, while our algorithm solves all the problems and uses the least time on 11 examples. Furthermore, neither QEPCAD nor RC-CAD is able to build a c.a.d. for cyclic-6, FIS-5, FIS-6 or FIS-7. As for the cell number, our algorithm always use less cells than QEPCAD, but RC-CAD's result is comparable to ours. A possible explanation for this is that regular chain and block-order Gröbner basis may choose different affine opens in the free locus of projection.

We have not included any complexity analysis because,

1. The decision problem over the real closed field is in EXPSPACE and it is also conjectured that this is EXPSPACE-complete [BOKR84]. Moreover, it is shown in [DH88;

| Examples | #Variables | #Polynomials | #Equations | Dimension |
|---|---|---|---|---|
| 5-VAR | 5 | 1 | 0 | 5 |
| Hong-4 | 4 | 2 | 0 | 4 |
| Hong-5 | 5 | 2 | 0 | 5 |
| Hong-6 | 6 | 2 | 0 | 6 |
| Hong-7 | 7 | 2 | 0 | 7 |
| Hong-8 | 8 | 2 | 0 | 8 |
| cyclic-4 | 4 | 4 | 4 | 0 |
| cyclic-5 | 5 | 5 | 5 | 0 |
| cyclic-6 | 6 | 6 | 6 | 0 |
| EBD-2 | 3 | 3 | 2 | 1 |
| EBD-5 | 5 | 6 | 4 | 1 |
| EBD-8 | 4 | 4 | 3 | 1 |
| FIS-1 | 5 | 2 | 1 | 4 |
| FIS-2 | 4 | 3 | 3 | 1 |
| FIS-3 | 7 | 4 | 4 | 3 |
| FIS-4 | 3 | 3 | 2 | 1 |
| FIS-5 | 5 | 5 | 5 | 0 |
| FIS-6 | 5 | 3 | 3 | 2 |
| FIS-7 | 5 | 6 | 5 | 1 |
| FIS-8 | 10 | 15 | 12 | 0 |

Table 3.2 Information about the Benchmark

| Tools | QEPCAD B | | RC-CAD | | GeometricCAD | |
|---|---|---|---|---|---|---|
| | Time (s) | #Cells | Time (s) | #Cells | Time (s) | #Cells |
| 5-VAR | 1.50 | 297 | 0.09 | 75 | **0.08** | 75 |
| Hong-4 | 1.35 | 575 | 1.19 | 491 | **0.56** | 575 |
| Hong-5 | **1.35** | 2241 | 5.16 | 1731 | 2.53 | 2241 |
| Hong-6 | **1.70** | 7967 | 20.00 | 5631 | 9.77 | 7967 |
| Hong-7 | >1h | N/A | 133.86 | 19743 | **49.52** | 29841 |
| Hong-8 | >1h | N/A | Error | N/A | **270.33** | 103343 |
| cyclic-4 | 1.56 | 129 | 0.06 | 4 | **0.03** | 4 |
| cyclic-5 | >1h | N/A | 4.58 | 10 | **0.58** | 10 |
| cyclic-6 | >1h | N/A | >1h | N/A | **5.48** | 24 |
| EBD-2 | 1.17 | 41 | **0.02** | 4 | **0.02** | 5 |
| EBD-5 | 1.33 | 107 | 0.27 | 13 | **0.05** | 10 |
| EBD-8 | 107.89 | 257 | >1h | N/A | **0.23** | 4 |
| FIS-1 | **3.68** | 114541 | 40.25 | 5333 | 16.11 | 4063 |
| FIS-2 | 1101.60 | 5467 | **1.59** | 79 | 2.66 | 299 |
| FIS-3 | >1h | N/A | **644.91** | 16928 | 1654.84 | 66704 |
| FIS-4 | 4.76 | 227 | 1.83 | 24 | **0.19** | 16 |
| FIS-5 | >1h | N/A | >1h | N/A | **0.20** | 3 |
| FIS-6 | >1h | N/A | >1h | N/A | **2834.00** | 32875 |
| FIS-7 | Error | N/A | >1h | N/A | **10.67** | 200 |
| FIS-8 | Error | N/A | 122.58 | 3 | **12.75** | 3 |

Table 3.3 Computation Time and Number of Cells

Wei88; BD07] that a cylindrical algebraic decomposition needs doubly-exponential number of cells in the worst case. So our algorithm probably does not improve the worst-case doubly-exponential bound of a cylindrical algebraic decomposition construction.

2. Our algorithm heavily relies on Gröbner basis. But the complexity of Buchberger's Algorithm is still not well understood in the geometric context. While the ideal membership problem is shown to be EXPSPACE-complete [MM82; May89], Buchberger's algorithm and its variants are often able to tackle problems of large size. In fact, there are many works to study why Gröbner basis computation is feasible in the cases of real interest in algebraic geometry, and sharper upper bounds are obtained by analyzing zero-dimensional systems [Lak91; LL91], low-dimensional systems [MR10; MR13], regular sequences (Faugère's F5) [Fau02; BFS15], unmixed ideals [Dic+91].

Still, the experimental data shows that our method goes beyond the reach of classical algorithms.

# Chapter 4 An Effective Criterion for Real Algebraic Coverings

In this chapter, we are going to show that a quasi-finite, flat morphism of real varieties with locally constant geometric fibers induces a covering map on the real points. This generalizes the core result in the previous chapter (Theorem 3.3.7) and serve as a criterion to identify covering maps from families of polynomial systems. Moreover, we present algorithms to effectively check this criterion.

We fix $R$ to be a real closed field and $C = R[\sqrt{-1}]$ is the algebraic closure of $R$. We will use a slightly different definition of covering map for arithmetic reasons.

**Definition 4.0.1.** By a **covering map**, we mean a continuous map $p : E \to B$ such that every $b \in B$ has an open neighborhood $U$ whose preimage is a disjoint (possibly empty) union of opens homeomorphic to $U$.

We do not assume $p$ to be surjective here because $E$ and $B$ are usually taken as the rational points of varieties and they sometimes can be empty.

For a quasi-finite morphism of schemes $f : X \to Y$, the function $\begin{array}{rcl} n(y) : Y & \to & \mathbb{N} \\ y & \mapsto & \#X_{\overline{y}} \end{array}$ counts the cardinality of the geometric fiber.

## 4.1 Flat Morphism with Locally Constant Finite Geometric Fibers

In this section, we will show the following: a quasi-finite flat morphism with locally constant geometric fibers between $k$-varieties (char $k = 0$) becomes finite étale after reduction. We shall establish some basic properties first.

**Proposition 4.1.1.** *A quasi-finite flat $k$-morphism with locally constant geometric fibers between $k$-varieties is finite.*

**Proof.** Denote the morphism by $f : X \to Y$. Since $n(y)$ is locally constant, for any $y \in Y$ there is a neighborhood $U$ of $y$ such that $f^{-1}(U) \to U$ is proper, hence finite by [EGA $\text{IV}_3$, Proposition 15.5.9] and [EGA $\text{IV}_3$, Théorème 8.11.1]. □

**Proposition 4.1.2.** *A finite étale $k$-morphism $f : X \to Y$ of $k$-varieties is quasi-finite, flat and has locally constant geometric fibers.*

**Proof.** By [EGA $\text{IV}_4$, Proposition 18.2.8], the function $n(y)$ is locally constant. □

**Lemma 4.1.3.** *Let $A$ be a reduced ring. Polynomials $f, g \in A[x]$ are monic. Suppose for every $p \in \operatorname{Spec} A$, the image of $g$ in $\kappa_p[x]$ divides the image of $f$ in $\kappa_p[x]$, then $g$ divides $f$ in $A[x]$.*

**Proof.** Since $g$ is monic, we can perform polynomial division on $A[x]$: there are $q, r \in A[x]$ such that

$$f = qg + r \ (\deg r < \deg g).$$

We are going to show that $r$ is zero.

Now consider this relation in $\kappa_p[x]$ for any $p \in \operatorname{Spec} A$. Denote the images of $f, g, q, r$ by $\overline{f}, \overline{g}, \overline{q}, \overline{r}$. Then we have

$$\overline{f} = \overline{qg} + \overline{r} \ (\deg \overline{r} < \deg \overline{g}).$$

Therefore $\overline{r} = 0$. In other words, all the coefficients of $r$ vanish at $p$. Because $\bigcap_{p \in \operatorname{Spec} A} p = \operatorname{nil} A = 0$, all the coefficients of $r$ must be zero. The proof is completed.

□

Now we are ready for the main result in this section.

**Theorem 4.1.4.** *Let $k$ be a field of characteristic 0. Suppose $X, Y$ are $k$-varieties and $\varphi : X \to Y$ is a quasi-finite flat $k$-morphism with locally constant geometric fibers, then the reduced map $\varphi_{\mathrm{red}} : X_{\mathrm{red}} \to Y_{\mathrm{red}}$ is (finite) étale.*

*Moreover, the reduced map $\varphi_{\mathrm{red}}$ is locally of the form $\operatorname{Spec} A[\lambda]/\langle u \rangle \to \operatorname{Spec} A$, where $u \in A[\lambda]$ is a monic polynomial and $u'$ is invertible modulo $u$.*

**Proof.** The proof is a standard dévissage argument.

*Step 1. Reduce to the affine case.* First notice that $\varphi$ is actually already finite by Proposition 4.1.1 (so is $\varphi_{\mathrm{red}}$). The question is local, so we may assume $X = \operatorname{Spec} B$, $Y = \operatorname{Spec} A$ are affine varieties and $B = A[x_1, \ldots, x_n]/I$ is a free $A$-module of finite rank $r$ because $\varphi$ is finite flat by [AM69, Exercise 7.15]. Since $\# X_{\overline{y}}$ is locally constant, we may further assume that the geometric fiber cardinality is a constant $s$. Replacing $A$, $B$ with $A_{\mathrm{red}} = A/\operatorname{nil} A$ and $B \otimes_A A_{\mathrm{red}}$ respectively, we may even assume that $A$ is reduced.

*Step 2. Chow form of all fibers.*

Let $p \in Y$, by the existence of separating linear forms ([BPR06, Lemma 4.90]), there is a linear form $\sigma^* = u_1^* x_1 + \cdots + u_n^* x_n \in B$ such that the geometric points of $X_{\overline{p}}$ takes different values. Geometrically speaking, this means for any hyperplane with the normal vector $(u_1^*, \ldots, u_n^*)$, there is at most one point in $X_{\overline{p}}$ on it. Later we shall see that $\sum_{i=1}^n u_i^* x_i$ separates

$X_{\overline{y}}$ for all $y$ in a neighborhood of $p$. Therefore there is a bijection between the geometric fiber $X_{\overline{y}}$ and the family of hyperplanes whose normal vector is $(u_1^*, \ldots, u_n^*)$, passing through $X_{\overline{y}}$.

To study this family of hyperplanes, it is more convenient to consider all the hyperplanes that pass through $X_{\overline{y}}$ at once. This family of hyperplanes is characterized by the Chow form of the reduced geometric fiber $X_{\overline{y},\mathrm{red}} := \left(X_{\overline{y}}\right)_{\mathrm{red}}$:

$$\mathrm{Ch}_{X_{\overline{y},\mathrm{red}}} = \prod_{\xi \in X_{\overline{y}}} (\lambda_0 + \xi_1\lambda_1 + \cdots + \xi_n\lambda_n).$$

We are going to show that there is a nice polynomial that parameterizes Chow forms of the reduced geometric fiber $X_{\overline{y},\mathrm{red}}$ for all geometric points $\overline{y} : \operatorname{Spec}\Omega \to Y$.

Introduce fresh variables $U_1, \ldots, U_n$ and let $A[U] = A[U_1, \ldots, U_n]$, $B[U] = B[U_1, \ldots, U_n]$. It is clear that $B[U]$ is also a free $A[U]$-module of finite rank $r$.

Consider the general linear form

$$\sigma = x_1U_1 + x_2U_2 + \cdots + x_nU_n \in B[U],$$

then it induces an $A[U]$-linear map by left multiplication:

$$\begin{array}{rcl} L_\sigma : B[U] & \to & B[U] \\ b & \mapsto & \sigma b \end{array}.$$

Define

$$\chi_\sigma = \det(\lambda I - L_\sigma) \in A[U][\lambda]$$

to be the characteristic polynomial of $L_\sigma$ and set $\chi'_\sigma = \frac{\partial}{\partial\lambda}\chi_\sigma$. Let

$$\Delta_i = \mathtt{sRes}_{\lambda,i}(\chi_\sigma, \chi'_\sigma) \in A[U]$$

be the subdiscriminants for $i = 0, \ldots, r-1$.

From Stickelberger's Eigenvalue Theorem [CLO05, Theorem 4.2.7], we see that for any geometric point $\overline{y} : \operatorname{Spec}\Omega \to Y$,

$$\chi_\sigma(\overline{y}; U_1, \ldots, U_n, \lambda) = \prod_{\xi \in X_{\overline{y}}} (\lambda - (\xi_1U_1 + \cdots + \xi_nU_n))^{\mu_\xi}.$$

Therefore we must have $\Delta_0 = \Delta_1 = \ldots = \Delta_{r-s-1} = 0$ (there are only $s$ points in $X_{\overline{y}}$). However, $\Delta_{r-s} \neq 0$ (recall that $\Delta_{r-s}(p, u_1^*, \ldots, u_n^*) \neq 0$).

Moreover, for a geometric point $\overline{y} : \operatorname{Spec}\Omega \to Y$ and $u_1, \ldots, u_n \in \Omega$:

$$\Delta_{r-s}(\overline{y}, u_1, \ldots, u_n) \neq 0 \Leftrightarrow u_1x_n + \cdots + u_nx_n \text{ separates } X_{\overline{y}}$$

Now let

$$\gamma_\sigma = \frac{1}{\Delta_{r-s}} \mathtt{sResP}_{\lambda,r-s}(\chi_\sigma, \chi'_\sigma) = \lambda^{r-s} + \cdots \in A[U]_{\Delta_{r-s}}[\lambda].$$

By our construction, for any $\overline{y} : \operatorname{Spec}\Omega \to Y$ and $\boldsymbol{u} = (u_1, \ldots, u_n) \in \Omega^n$:

$$\Delta_{r-s}(\overline{y}, \boldsymbol{u}) \neq 0 \implies \gamma_\sigma(\overline{y}, \boldsymbol{u}; \lambda) \text{ is the gcd of } \chi_\sigma(\overline{y}, \boldsymbol{u}; \lambda) \text{ and } \chi'_\sigma(\overline{y}, \boldsymbol{u}; \lambda).$$

Then there is an

$$\omega = \lambda^s + \cdots \in A[U]_{\Delta_{r-s}}[\lambda]$$

such that $\chi_\sigma = \gamma_\sigma \omega$ by Lemma 4.1.3. Therefore,

$$\Delta_{r-s}(\overline{y}, \boldsymbol{u}) \neq 0 \implies \omega(\overline{y}, \boldsymbol{u}; \lambda) \text{ is the squarefree part of } \chi_\sigma(\overline{y}, \boldsymbol{u}; \lambda).$$

In other words, if $\Delta_{r-s}(\overline{y}, \boldsymbol{u}) \neq 0$, then we have

$$\omega(\overline{y}, \boldsymbol{u}; \lambda) = \prod_{\xi \in X_{\overline{y}}} (\lambda - (\xi_1 u_1 + \cdots + \xi_n u_n)) = \mathrm{Ch}_{X_{\overline{y},\mathrm{red}}}(\lambda, -u_1, \ldots, -u_n).$$

Therefore $\omega(\overline{y}; -U, \lambda)$ parameterizes the Chow form of $X_{\overline{y},\mathrm{red}}$ for every geometric point $\overline{y}$ of $Y$.

*Step 3. Bijection between geometric fibers and the hyperplanes*

Because $\Delta_{r-s}(p, u_1^*, \ldots, u_n^*) \neq 0$, there is an open neighborhood $U$ of $p$ such that for every $y \in U$, we have $\Delta_{r-s}(y, u_1^*, \ldots, u_n^*) \neq 0$. By shrinking $Y = \operatorname{Spec} A$ and $X = \operatorname{Spec} B$ if necessary, we now may assume that $\sigma^* = \sum_{i=1}^n u_i^* x_i$ is a separating element for every geometric fiber.

First we define the morphism sending geometric fibers to the hyperplanes with normal vector $(u_1^*, \ldots, u_n^*)$ passing through the fibers.

Specializing $U_1, \ldots, U_n$ to $u_1^*, \ldots, u_n^*$, then we have $\chi_{\sigma^*} = \det(\lambda I - L_{\sigma^*})$, where $L_{\sigma^*}$ is the multiplication map $b \mapsto \sigma^* b$ on $B$. By Cayley-Hamilton Theorem, there is a well-defined $A$-algebra map

$$\begin{array}{rcl} \psi^\# : A[\lambda]/\langle \chi_{\sigma^*} \rangle & \to & B \\ \lambda & \mapsto & \sigma^* \end{array}.$$

This yields a $Y$-scheme morphism $\psi : X \to Z = \operatorname{Spec} A[\lambda]/\langle \chi_{\sigma^*} \rangle$ that sends every point $\xi \in X_{\overline{y}}$ to the unique hyperplane with normal vector $(u_1^*, \ldots, u_n^*)$ that passes through $\xi$. Here each hyperplane is identified by (the negation of) its constant term $\lambda = \sum_{i=1}^n u_i^* x_i$. By construction, this is a bijection between $X_{\overline{y}}$ and $Z_{\overline{y}}$ for every geometric point $\overline{y}$. Here

$$Z_{\overline{y}} \cong \{H : \text{hyperplanes that pass through } X_{\overline{y}} \text{ with normal vector } (u_1^*, \ldots, u_n^*)\}.$$

*Step 4. Isomorphism $X_{\mathrm{red}} \cong Z_{\mathrm{red}}$.*

Now we are going to show that $\psi_{\mathrm{red}} : X_{\mathrm{red}} \to Z_{\mathrm{red}}$ is actually an isomorphism. This is done by recovering a variety by its Chow form. We are going to show that given a hyperplane that meets $X_{\overline{y}}$ in exactly one point, the Chow form allows us to find the point from the hyperplane.

Define $\omega_0 = \frac{\partial \omega}{\partial \lambda} \in A[U]_{\Delta_{r-s}}[\lambda]$ and $\omega_1 = -\frac{\partial \omega}{\partial U_1}, \ldots, \omega_n = -\frac{\partial \omega}{\partial U_n} \in A[U]_{\Delta_{r-s}}[\lambda]$. Set

$$\omega_i^* = \omega_i|_{U_1 = u_1^*, \ldots, U_n = u_n^*} \in A[\lambda]$$

for $0 \le i \le n$.

We claim that

$$\begin{array}{rcl} \tau^\# : A[x_1, \ldots, x_n] & \to & A[\lambda]/\langle \chi_{\sigma^*} \rangle \\ x_i & \mapsto & \frac{\omega_i^*}{\omega_0^*} \end{array}$$

is a well-defined $A$-algebra map and this induces a $Y$-morphism $\tau : Z \to \mathbb{A}_Y^n$.

To show the claim, we need to prove that $\omega_0^*$ is never zero on $Z$. We will check that $\omega_0^*(\overline{z})$ is non-zero for all geometric points $\overline{z} : \operatorname{Spec} \Omega \to Z$. Let $\overline{y} : \operatorname{Spec} \Omega \to Z \to Y$ be the corresponding geometric point in $Y$. Since there is a bijection between $X_{\overline{y}}$ and $Z_{\overline{y}}$, we see that there is a geometric point $\overline{x} : \operatorname{Spec} \Omega \to X$ making the following diagram commute.

$$\begin{tikzcd}
 & \operatorname{Spec} \Omega \arrow[dl, "\overline{x}"'] \arrow[dr, "\overline{z}"] \arrow[dd, bend left, "\overline{y}"] & \\
X \arrow[rr, bend left, "\psi"] \arrow[dr, "\phi"'] & & Z \arrow[dl] \\
 & Y &
\end{tikzcd}$$

A direct computation shows that $\omega_0^*$ is never vanishing on geometric points of $Z$, so $\omega_0^*$ is a unit modulo $\chi_{\sigma^*}$ and the $A$-algebra map $\tau^\#$ is well-defined by Hilbert's Nullstellensatz:

$$\begin{aligned}
&\omega_0^*(\overline{z})\\
=&\left.\left(\frac{\partial}{\partial\lambda}\prod_{\xi\in X_{\overline{y}}}\left(\lambda-\sum_{i=1}^n\xi_iU_i\right)\right)\right|_{U_1=u_1^*,\ldots,U_n=u_n^*,\lambda=\sigma^*(\overline{x})}\\
=&\left.\left(\sum_{\xi\in X_{\overline{y}}}\prod_{\substack{\xi'\in X_{\overline{y}}\\ \xi'\neq\xi}}\left(\lambda-\sum_{i=1}^n\xi'_iU_i\right)\right)\right|_{U_1=u_1^*,\ldots,U_n=u_n^*,\lambda=\sigma^*(\overline{x})}\\
=&\sum_{\xi\in X_{\overline{y}}}\prod_{\substack{\xi'\in X_{\overline{y}}\\ \xi'\neq\xi}}(\sigma^*(\overline{x})-\sigma^*(\xi'))\\
=&\prod_{\substack{\xi'\in X_{\overline{y}}\\ \xi'\neq\overline{x}}}(\sigma^*(\overline{x})-\sigma^*(\xi'))\\
\neq&0.
\end{aligned}$$

Next, we check the composition $\tau\circ\psi$ on geometric points. Similarly, for $1\leq i\leq n$, we have

$$\begin{aligned}
&\omega_i^*(\overline{z})\\
=&\left.\left(-\frac{\partial}{\partial U_i}\prod_{\xi\in X_{\overline{y}}}\left(\lambda-\sum_{i=1}^n\xi_iU_i\right)\right)\right|_{U_1=u_1^*,\ldots,U_n=u_n^*,\lambda=\sigma^*(\overline{x})}\\
=&\left.\left(\sum_{\xi\in X_{\overline{y}}}\xi_i\prod_{\substack{\xi'\in X_{\overline{y}}\\ \xi'\neq\xi}}\left(\lambda-\sum_{i=1}^n\xi'_iU_i\right)\right)\right|_{U_1=u_1^*,\ldots,U_n=u_n^*,\lambda=\sigma^*(\overline{x})}\\
=&\sum_{\xi\in X_{\overline{y}}}\xi_i\prod_{\substack{\xi'\in X_{\overline{y}}\\ \xi'\neq\xi}}(\sigma^*(\overline{x})-\sigma^*(\xi'))\\
=&(\overline{x})_i\prod_{\substack{\xi'\in X_{\overline{y}}\\ \xi'\neq\overline{x}}}(\sigma^*(\overline{x})-\sigma^*(\xi'))\,.
\end{aligned}$$

Therefore $\frac{\omega_i^*(\overline{z})}{\omega_0^*(\overline{z})}=(\overline{x})_i$. As a consequence, $\tau\circ\psi$ agrees with the closed immersion $\iota: X\to\mathbb{A}_Y^n$ on all geometric points at the set-theoretic level.

Consider the reduced map $\tau_{\mathrm{red}}\circ\psi_{\mathrm{red}}: X_{\mathrm{red}}\to Z_{\mathrm{red}}\to\mathbb{A}_Y^n$ and $\iota_{\mathrm{red}}: X_{\mathrm{red}}\to\mathbb{A}_Y^n$. They are the same on the geometric points, and the source $X_{\mathrm{red}}$ is geometrically reduced (we are in characteristic 0). By [GW20, Exercise 5.17], we actually have that $\tau_{\mathrm{red}}\circ\psi_{\mathrm{red}}=\iota_{\mathrm{red}}$. Therefore $\tau_{\mathrm{red}}$ factors through $X_{\mathrm{red}}$ and it is the inverse of $\psi_{\mathrm{red}}$. This proves the required isomorphism

$X_{\mathrm{red}} \cong Z_{\mathrm{red}}$.

*Step 5. The étaleness.*

It suffices to show that $Z_{\mathrm{red}} \to Y$ is étale, because $X_{\mathrm{red}}$ is isomorphic to $Z_{\mathrm{red}}$.

We will show that $Z_{\mathrm{red}} = \operatorname{Spec} A[\lambda]/\langle\omega^*\rangle$, where $\omega^* = \omega|_{U_1=u_1^*,\dots,U_n=u_n^*} \in A[\lambda]$. In other words, $\sqrt{\langle\chi_{\sigma^*}\rangle} = \langle\omega^*\rangle$.

On the one hand, $\omega^* \in \sqrt{\langle\chi_{\sigma^*}\rangle}$. This is a direct application of Lemma 4.1.3 on $(\omega^*)^r$ and $\chi_{\sigma^*}$.

On the other hand, $\omega^*$ divides every element $f \in \sqrt{\langle\chi_{\sigma^*}\rangle}$. Suppose $f^t \in \langle\chi_{\sigma^*}\rangle$. By polynomial division in $A[\lambda]$, there are $q, \delta \in A[\lambda]$ such that $f = q\omega^* + \delta$ ($\deg_\lambda \delta < \deg_\lambda \omega^*$). Then

$$f^t = (q\omega^* + \delta)^t = \omega^*((\omega^*)^{t-1}q^t + \cdots + tq\delta^{t-1}) + \delta^t \in \langle\chi_{\sigma^*}\rangle \subseteq \langle\omega^*\rangle.$$

This shows that $\delta^t \in \langle\omega^*\rangle$. If $\delta \neq 0$, then there is some geometric point $y : \operatorname{Spec}\Omega \to Y$ such that $\delta(y;\lambda) \neq 0$ in $\Omega[\lambda]$. But $\omega^*(y;\lambda)$ is square-free, $\omega^*(y;\lambda)|\delta^t(y;\lambda)$ implies $\omega^*(y;\lambda)|\delta(y;\lambda)$. So

$$\deg_\lambda \delta \geq \deg_\lambda \delta(y;\lambda) \geq \deg_\lambda \omega^*(y;\lambda) = \deg_\lambda \omega^*,$$

which is a contradiction.

Because $\omega^*$ is a monic polynomial of degree $s$ in $A[\lambda]$, clearly $\operatorname{Spec} A[\lambda]/\langle\omega^*\rangle \to Y$ is flat. Also, we have already seen that $\frac{\partial\omega^*}{\partial\lambda} = \omega_0^*$ is invertible in the previous step. So every geometric fiber has exactly $s$ point, hence regular. By [Har77, Theorem III.10.2], the morphism $Z_{\mathrm{red}} \to Y$ is étale. The proof is completed now.

□

The last assertion of Theorem 4.1.4 is a slightly different version of the fact that étale morphisms are locally standard étale.

**Example 6.** Suppose $k$ is a field of characteristic 0. Consider

$$X = \operatorname{Spec} k[p,q,x]/\langle 4p^3 + 27q^2, x^3 + px + q\rangle, Y = \operatorname{Spec} k[p,q]/\langle 4p^3 + 27q^2\rangle$$

and the canonical projection $\pi : X \to Y$. Clearly $\pi$ is finite and flat. See Figure 4.1.

Let $U = Y_{\mathrm{reg}} = Y \backslash V(p,q)$, then $\pi|_{\pi^{-1}(U)} : X \times_Y U \to U$ is finite, flat and $\#X_{\overline{y}} = 2$ for all $y \in U$. So by Theorem 4.1.4, the reduced map $(X \times_Y U)_{\mathrm{red}} \to U_{\mathrm{red}}$ is finite étale.

Notice that in general, even the flatness is not preserved under reduction. In this example, $\pi_{\mathrm{red}} : X_{\mathrm{red}} \to Y_{\mathrm{red}}$ is not flat! In fact, by a direct computation, one can show that $X_{\mathrm{red}} =$

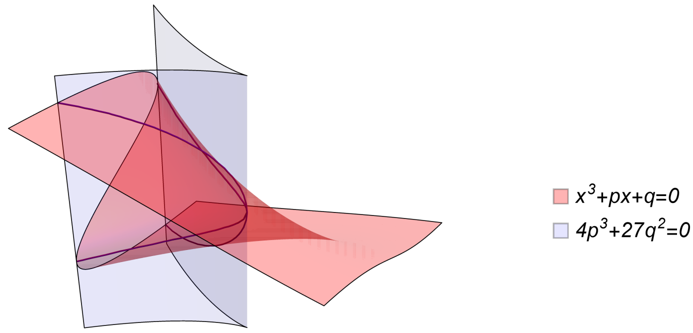



Figure 4.1 $X$ as the intersection of zeros of defining equations

Spec $k[p,q,x]/\langle 6px^2-9qx+4p^2, x^3+px+q\rangle$ and fibers of $\pi_{\mathrm{red}}$ are of degree 2 except the origin's, which is of degree 3.

**Example 7.** When char $k = p > 0$, the same conclusion does not hold anymore. Let $X = \operatorname{Spec}\mathbb{F}_2[t][x]/\langle x^2+t\rangle$ and $Y = \operatorname{Spec}\mathbb{F}_2[t]$, then $X$ and $Y$ are both reduced, the canonical projection $\pi : X \longrightarrow Y$ is flat and the cardinality of geometric fiber is always 1. But $\pi$ is not smooth.

The following corollary is immediate from Proposition 4.1.2 and Theorem 4.1.4.

**Corollary 4.1.5.** *Let $k$ be a field of characteristic 0. Suppose $X, Y$ are $k$-varieties and the $k$-morphism $\varphi : X \longrightarrow Y$ is quasi-finite and flat. Then the followings are equivalent:*

1. *$\varphi_{\mathrm{red}} : X_{\mathrm{red}} \longrightarrow Y_{\mathrm{red}}$ is finite étale.*
2. *The number of geometric fibers $n(y)$ is a locally constant function on $Y$.*

We end this section with a corollary that concerns the geometric fiber cardinality of a subvariety.

**Corollary 4.1.6.** *Let $k$ be a field of characteristic 0. Suppose $X, Y$ are $k$-varieties and $\varphi : X \longrightarrow Y$ is a quasi-finite flat $k$-morphism with locally constant geometric fibers. Let $i : Z \hookrightarrow X$ be a closed subscheme of $X$ and $j : U = X\backslash Z \hookrightarrow X$ be the complement with the open subscheme structure.*

*Suppose that $\psi = \varphi \circ i : Z \hookrightarrow X \longrightarrow Y$ is flat. Then*

1. *The composition $\psi$ is also a finite flat morphism with locally constant geometric fibers, and $\psi_{\mathrm{red}} : Z_{\mathrm{red}} \longrightarrow Y_{\mathrm{red}}$ is finite étale.*
2. *Similarly, $\varphi \circ j$ is finite flat with locally constant geometric fibers, and $\varphi_{\mathrm{red}} \circ j_{\mathrm{red}} : U_{\mathrm{red}} \longrightarrow Y_{\mathrm{red}}$ is finite étale.*

3. *The closed immersion* $i_{\mathrm{red}} : Z_{\mathrm{red}} \hookrightarrow X_{\mathrm{red}}$ *and the open immersion* $j_{\mathrm{red}} : U_{\mathrm{red}} \hookrightarrow X_{\mathrm{red}}$ *are also finite étale.*

**Proof.**

1. We can compute $\#Z_{\overline{y}}$ in two different ways. Notice that $\varphi : X \to Y$ is actually finite by Proposition 4.1.1.
   On the one hand, $\psi : Z \to Y$ is finite and flat, so $\#Z_{\overline{y}}$ is lower semi-continuous by [EGA IV$_3$, Proposition 15.5.9].
   On the other hand, $\#Z_{\overline{y}} = \#(Z \times_X X_{\mathrm{red}})_{\overline{y}}$. By Theorem 4.1.4, $\varphi_{\mathrm{red}} : X_{\mathrm{red}} \to Y_{\mathrm{red}}$ is finite étale. So the fiber $(X_{\mathrm{red}})_y$ is zero-dimensional and regular, thus reduced. This implies that the subscheme $(Z \times_X X_{\mathrm{red}})_y$ is also zero-dimensional and reduced. So $\#(Z \times_X X_{\mathrm{red}})_{\overline{y}} = \deg(Z \times_X X_{\mathrm{red}})_y$ and we have
   $$\deg(Z \times_X X_{\mathrm{red}})_y = \dim_{\kappa_y} \left( \varphi_* \left( i_* O_Z \otimes_{O_X} (O_X)_{\mathrm{red}} \right)_y \otimes_{O_{Y,y}} \kappa_y \right)$$
   is upper semi-continuous [Har77, Exercise II.5.8].
   Therefore $\#Z_{\overline{y}}$ is an integer-valued continuous function, thus a locally constant function. By Theorem 4.1.4 again, $\psi_{\mathrm{red}}$ is finite étale.
2. Observe that $\#U_{\overline{y}} = \#X_{\overline{y}} - \#Z_{\overline{y}}$. Therefore $\#U_{\overline{y}}$ is also locally constant. Notice that the open immersion $j$ is quasi-finite and flat. So $\varphi \circ j$ is a quasi-finite flat morphism with locally constant geometric fibers, hence finite by Proposition 4.1.1. Once more, Theorem 4.1.4 implies that $\varphi_{\mathrm{red}} \circ j_{\mathrm{red}} : U_{\mathrm{red}} \to Y_{\mathrm{red}}$ is finite étale.
3. Combining [Mil80, Corollary I.3.6] and [Har77, Exercise II.4.8], we see that there is a cancellation law of finite étale morphisms. Namely, if $g \circ f$ and $g$ are finite étale morphisms between $k$-varieties, then so is $f$. Therefore the closed immersion $i_{\mathrm{red}} : Z_{\mathrm{red}} \to X_{\mathrm{red}}$ is finite étale, as $\psi_{\mathrm{red}} = \varphi_{\mathrm{red}} \circ i_{\mathrm{red}}$ and $\varphi_{\mathrm{red}}$ are both finite étale. Similarly, $j_{\mathrm{red}}$ is also finite étale.

□

## 4.2 Semi-algebraic Covering Map

In this section, we will consider the case where $k = R$ is a real closed field and we will show that a quasi-finite flat morphism of $R$-varieties with locally constant geometric fibers induces a covering map on the $R$-rational points in the Euclidean topology.

**Theorem 4.2.1.** *Let $X$, $Y$ be two $R$-varieties. Suppose that $\pi : X \to Y$ is a quasi-finite flat $R$-morphism with locally constant geometric fibers. Let $y \in Y(R)$. Then there exists $r \in \mathbb{N}$,* *semi-algebraically connected Euclidean neighborhood $U \subseteq Y(R)$ of $y$ and semi-algebraic continuous functions $\xi_1, \ldots, \xi_r : U \to X(R)$ such that*

1. *The number of real fibers over any $y \in U$, is always $r$.*
2. *$\xi_i$ are sections of $\pi$, that is $\pi \circ \xi_i = \mathrm{id}_U$ for $i = 1, \ldots, r$.*
3. *For every $y \in U$: $i \neq j \implies \xi_i(y) \neq \xi_j(y)$.*

*Therefore $\pi_R$ is a covering map in the Euclidean topology.*

**Proof.** We may replace $\pi : X \to Y$ with $\pi_{\mathrm{red}} : X_{\mathrm{red}} \to Y_{\mathrm{red}}$ and assume that $\pi$ is a finite étale morphism by Theorem 4.1.4. Then locally $\pi$ is of the form $\operatorname{Spec} A[\lambda]/\langle u\rangle \to \operatorname{Spec} A$, where $u \in A[\lambda]$ is a monic polynomial and $u'$ is invertible modulo $u$. Now $\operatorname{Spec} A[\lambda]/\langle u\rangle \to \operatorname{Spec} A$ is a finite free morphism of affine $R$-varieties with constant geometric fibers, so we can define $\xi_i$ to be the $i$-th largest real root of $u$ in an Euclidean neighborhood of $y$ and they are continuous semi-algebraic functions by Theorem 3.3.7. □

*Remark* 4. For readers familiar with locally semi-algebraic spaces, it can be immediately concluded that $\pi_R$ is a covering map after applying Theorem 4.1.4, see [DK84, Example 5.5].

*Remark* 5. One possible way to understand Theorem 4.2.1 is to compare it with the Ehresmann's fibration theorem [Ehr50]. Actually, Theorem 4.2.1 implies a semi-algebraic weaker variant of Ehresmann's fibration theorem. Suppose that $\pi : X \to Y$ is a proper morphism of non-singular $R$-varieties with surjective tangent maps $T_\pi : T_x X \to T_{\pi(x)} Y \otimes_{\kappa_y} \kappa_x$, and suppose further that all fibers of $\pi$ are finite, then $\pi_R$ is a covering map in the Euclidean topology. Indeed, such a morphism is quasi-finite, proper and étale. Therefore locally there are semi-algebraic sections to $\pi_R$ by Theorem 4.2.1.

The geometric picture of the criteria is clear. In algebraic geometry, the flatness serves as a notion of continuity in fibers. The local constancy of geometric fibers can be interpreted as the invariance of the number of distinct complex roots. Therefore these two condition together shall imply that the fibers (solutions) are locally continuous functions on the base (in the parameters).

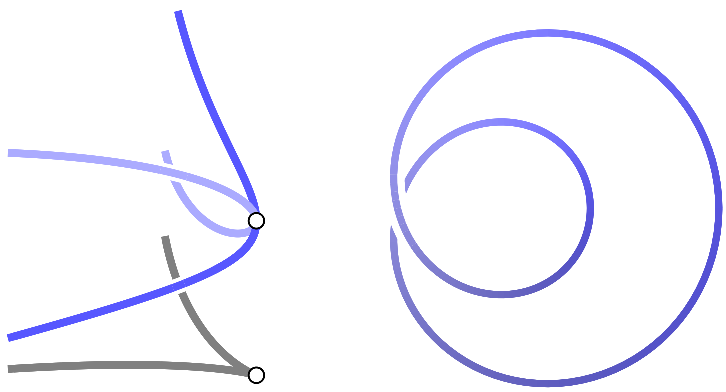
**Example 8.** Here are some examples fulfilling the assumption of Theorem 4.2.1. We shall see that they are indeed covering maps.

(a) Set $Y$ to be the cuspidal curve with singularity removed $V(4p^3 + 27q^2)\backslash V(p, q)$ in the affine plane and set $X$ to be $V(4p^3 + 27q^2, x^3 + px + q)\backslash V(p, q, x)$ in the three dimensional space. Then the projection from $X$ to $Y$ is quasi-finite flat with constant geometric fibers. Therefore the projection is a covering on the real points. See Figure 4.2(a).

(b) Consider $\mathbb{R}$-varieties $X = \operatorname{Spec} \mathbb{R}[x, y, z, w]/\langle x^2+y^2-1, z^2+w^2-1, x-2z^2+1, x+2w^2-1, y-2zw\rangle$ and $Y = \operatorname{Spec} \mathbb{R}[x, y]/\langle x^2 + y^2 - 1\rangle$. The canonical projection $\pi : X \to Y$ is flat with constant geometric fiber size 2. The $\mathbb{R}$-rational points of $Y$ can be modeled by the unit circle $S^1$ and $X(\mathbb{R})$ can be identified with a curve parameterized by $\varphi \mapsto (\cos 2\varphi, \sin 2\varphi, \cos\varphi, \sin\varphi)$ in the torus $S^1 \times S^1 = \{(\cos\theta, \sin\theta, \cos\varphi, \sin\varphi) | \theta, \varphi \in \mathbb{R}\}$. So $\pi_{\mathbb{R}}$ is given by

$$(\cos 2\varphi, \sin 2\varphi, \cos\varphi, \sin\varphi) \mapsto (\cos 2\varphi, \sin 2\varphi).$$

Therefore $X(\mathbb{R})$ is a double cover of $S^1 = Y(\mathbb{R})$. See Figure 4.2(b).
In this example, one can only defined the semi-algebraic sections locally and $X(\mathbb{R})$ is not a disjoint union of copies of $Y(\mathbb{R})$. It is easy to see that the source is connected, but two different global sections will give two disjoint closed subsets. In the corank 1 case, sections can always be defined globally because of the order structure of $R$. In the general case, a covering map is the best we can get.



(a) Double cover of cuspidal curve's smooth locus (b) Non-trivial double cover of $S^1$

Figure 4.2 Examples

The condition here is more general than Collins's cylindrical algebraic decomposition, which states for a corank 1 projection from a hypersurface [Col75, Theorem 1], and "discriminant variety" in the sense of Lazard and Rouillier's that works with equidimensional smooth varieties [LR07, Theorem 1]. Therefore, these previous results can be thought as special cases of our new criterion.

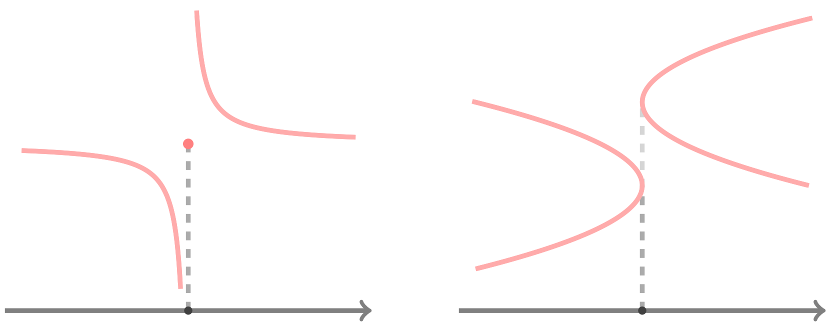
Actually, our new criterion is very sharp, as either the flatness assumption or the constancy assumption on the geometric fibers is dropped, then the conclusion does not hold anymore.

**Example 9.** Theorem 4.2.1 is quite sharp. We present some non-examples here.

(a) The flatness assumption on the morphism is essential and natural. Let $X$ be the union of hyperbola $V(xy-1)$ and the origin $V(x,y)$ on the plane and consider its projection to the affine line. See Figure 4.3(a). Obviously the projection is not a covering. Notice that the geometric fiber size is always 1. In this example, the projection is neither finite nor flat.

For a finite but not flat example, one can modify Example 8(a) by setting $Y = V(4p^3 + 27q^2)$ to be the whole cuspidal curve and adding a point above the singularity, e.g. $X = V(4p^3+27q^2, x^3+px+q) \cup V(x,y,z-1)$. It might be instructive to recall Figure 4.2(a).

(b) Also, the constancy assumption on the geometric fiber cardinality is crucial. Let $X = V((y-1)^2-x) \cup V((y+1)^2+x)$ be the union of two parabolas and let $Y = \operatorname{Spec} R[x]$ be the affine line, then the canonical projection from $X$ to $Y$ is obviously finite and flat. See Figure 4.3(b). Clearly this is not a covering either. Generically, the geometric fiber consists of four distinct complex roots $y = 1 \pm \sqrt{x}, -1 \pm \sqrt{x}$, but over $x = 0$ there are only two distinct pairs of double roots. Notice that the number of real fibers over $Y(R)$ is always 2 in this example. So even a constancy condition on the real fiber is not enough.



(a) Without flatness

(b) Geometric fiber not locally constant

Figure 4.3 Non-examples

The assumption in Theorem 4.2.1 is purely algebro-geometric and independent of the order structure of $R$, yet it is strong enough to conclude the local existence of semi-algebraic sections. It might be instructive to recall Example 8 and Example 9. Of course, such a simple condition is not always necessary, as the map $x \mapsto x^3$ is a semi-algebraic homeomorphism but the geometric fiber cardinality drops when $x = 0$. Also, one may add embedded points to destroy the flatness without changing the map on the topological level.

Still, Theorem 4.2.1 is quite strong among the results of the same kind in the sense that there is no assumption on $X$ or $Y$, so non-reduced, reducible, singular varieties with irreducible components of various dimensions are all allowed. Also, we can work over arbitrary real closed field instead of just real numbers $\mathbb{R}$. Most previous results producing covering maps are from a differential view, relying on smooth structures. Therefore $X, Y$ are required to be equidimensional smooth $\mathbb{R}$-varieties.

The notion of "discriminant variety" due to Lazard and Rouillier in [LR07] is one of the previous results. It can be thought as a special case of Theorem 4.2.1.

**Corollary 4.2.2** (Lazard-Rouillier cf. [LR07, Theorem 1])**.** *Let $X = \operatorname{Spec} B, Y = \operatorname{Spec} A$ be two $R$-varieties. Suppose $X$ is a distinguished open $D(f)$ of a closed subscheme $V(I) \subseteq \mathbb{A}^n_Y = \operatorname{Spec} A[x_1, \ldots, x_n]$, and the projection $\pi : X \hookrightarrow \mathbb{A}^n_Y \to Y$ is dominant.*

*Let $\overline{V(I)} \subseteq \mathbb{P}^n_Y$ be the projective closure of $V(I)$, and let $H = \mathbb{P}^n_Y \backslash \mathbb{A}^n_Y$ be the hyperplane at the infinity. Define*

- *(a) $W_\infty \subseteq Y$ to be the scheme-theoretical image of $\overline{V(I)} \cap H$ in $Y$.*
- *(b) $W_{\mathcal{F}}$ to be the scheme-theoretical image of $V(\langle f \rangle + I)$,*
- *(c) $W_{\text{sing}} \subseteq Y$ to be the singular locus of $Y$,*
- *(d) $W_{\text{sd}}$ to be the scheme-theoretical image of irreducible components of $X$ of dimension $< \dim X$,*
- *(e) $W_{c_1}$ to be the scheme-theoretical image of the singular locus of $X$,*
- *(f) $W_{c_2}$ to be the scheme-theoretical image of the ramification locus $\Delta(X/Y) = \operatorname{Supp} \Omega_{X/Y}$, the support of the sheaf of relative differential on $X$.*

*Then $\pi_R : X(R) \to Y(R)$ is a covering map outside the union $W = W_\infty \cup W_{\text{sing}} \cup W_{\text{sd}} \cup W_{\mathcal{F}} \cup W_{c_1} \cup W_{c_2}$. Moreover, if $R = \mathbb{R}$, then $\pi_R$ is an analytic covering outside $W$.*

**Proof.** First, we retreat to the finite case. Clearly, the morphism $\overline{V(I)} \hookrightarrow \mathbb{P}^n_Y \to Y$ is projective, hence proper [Har77, Theorem II.4.9]. Then the base change to $Y \backslash W_\infty$ is also proper. Notice that $\overline{V(I)} \times_Y (Y \backslash W_\infty) = V(I) \times_Y (Y \backslash W_\infty)$. Therefore, the projection $V(I) \times_Y (Y \backslash W_\infty) \to Y \backslash W_\infty$ is affine and proper, hence finite [Har77, Exercise II.4.6]. After another fiber product by the open subscheme $Y \backslash W_{\mathcal{F}}$, we may now assume that $\pi : X \to Y$ is a finite surjective morphism.

Next, we are going to show that $\pi$ becomes flat after removing $W_{\text{sing}}$, $W_{\text{sd}}$ and $W_{c_1}$. Recall the miracle flatness theorem [Mat89, Theorem 23.1]: for a morphism $\pi : X \to Y$ with a Cohen-Macaulay source $X$ and a regular base $Y$, $\pi$ is flat if and only if for all $x \in X$ and

$y = \pi(x) \in Y$,

$$\dim \mathcal{O}_{X,x} = \dim \mathcal{O}_{Y,y} + \dim \mathcal{O}_{X_y,x}.$$

Now after removing $W_{\mathrm{sing}}$, $W_{\mathrm{sd}}$ and $W_{\mathrm{c}_1}$, our source $X$ is regular (hence Cohen-Macaulay [Mat89, Theorem 17.8]) and of pure dimension $d$, our base $Y$ is regular. Therefore to show the flatness, it suffices to show the equation in dimensions holds everywhere. Notice that $Y$ is also of pure dimension $d$ (choose an irreducible component $Y_i$ of $Y$, there must be an irreducible component $X_i$ of $X$, mapped onto $Y_i$ as $\pi$ is finite surjective, so $\dim X = \dim X_i = \dim Y_i$). Then we have

$$\begin{aligned} \dim \mathcal{O}_{X,x} &= \dim X - \dim \overline{\{x\}} && X \text{ equi-dimensional} \\ &= \dim Y - \dim \overline{\{y\}} && \pi \text{ finite surjective} \\ &= \dim \mathcal{O}_{Y,y} && Y \text{ equi-dimensional} \\ &= \dim \mathcal{O}_{Y,y} + \dim \mathcal{O}_{X_y,x} && X_y \text{ finite} \end{aligned}$$

At last, $W_{\mathrm{c}_2}$ is removed so $\pi$ is finite, flat, unramified. Therefore $\pi|_{\pi^{-1}(Y\backslash W)} : X \times_Y (Y\backslash W) \to Y\backslash W$ is finite étale . Our conclusion follows from Theorem 4.2.1.

When $R = \mathbb{R}$, we can talk about analytic properties of morphisms. Since $\pi|_{\pi^{-1}(Y\backslash W)}$ : $X \times_Y (Y\backslash W) \to Y\backslash W$ is a finite étale morphism between smooth varieties, its tangent map is surjective. It follows immediately from inverse function theorem that $\pi_R$ is a local diffeomorphism. □

*Remark* 6. The corollary can be sharpened. After removing $W_\infty$ and $W_{\mathcal{F}}$, we may assume that $\pi : X \to Y$ is a finite morphism. In this case, the flat locus $F \subseteq Y$ of $\pi$ is an open subset given explicitly by the Fitting ideals [GP08, Corollary 7.3.13]. Recall that $n(y)$ is lower semi-continuous on $F$ [EGA IV$_3$, Proposition 15.5.9]. Therefore there is a biggest open subset $F' \subseteq F$ such that the $n(y)$ is locally constant on $F'$. On $F'$, $\pi$ induces a covering map (but not necessarily an analytic cover).

The complex analogue of Theorem 4.2.1 is also true. It can be proved using Weil restriction (restriction of scalars). Roughly speaking, the Weil restriction $\mathrm{Res}_{C/R}$ of a $C$-variety is the $R$-variety obtained by separating the real and imaginary parts of the coordinates.

**Theorem 4.2.3.** *Let $X$, $Y$ be two $C$-varieties. Suppose that $\pi : X \to Y$ is a quasi-finite flat $C$-morphism with locally constant geometric fibers. Then $\pi_C$ is a covering map in the Euclidean topology.*

**Proof.** Replacing $\pi : X \to Y$ with the reduced map, we may assume that $\pi$ is finite and étale. Because the question is local, we may also assume that $X$, $Y$ are affine varieties. Since $C/R$ is a finite field extension, the Weil restriction $\mathrm{Res}_{C/R}$ exists for quasi-projective varieties [Sch94, Corollary 4.8.1]. Let $Z = \mathrm{Res}_{C/R} X$ and $W = \mathrm{Res}_{C/R} Y$ be the Weil restrictions of $X$ and $Y$ respectively, then $Z$ and $W$ are $R$-varieties whose $R$-rational points can be naturally identified with closed points of $X$ and $Y$. Also, the induced map $\mathrm{Res}_{C/R}(\pi) : Z \to W$ is finite and étale [Sch94, Proposition 4.9, Proposition 4.10.1]. The proof is completed by applying Theorem 4.2.1 to $\mathrm{Res}_{C/R}(\pi) : Z \to W$. □

For subvarieties, the condition for being a covering is easier. This allows us to study inequalities as well.

**Theorem 4.2.4.** *Let $X$, $Y$ be two $R$-varieties. Suppose that $\pi : X \to Y$ is a quasi-finite flat $R$-morphism with locally constant geometric fibers. Let $Z_1, \dots, Z_m$ be closed subschemes of $X$ and set $U_i = X \backslash X_i$ for $i = 1, \dots, m$. If $\pi|_{Z_i} : Z_i \hookrightarrow X \to Y$ is flat for $i = 1, \dots, m$, then for arbitrary intersection $W = \bigcap_{\alpha \in \Lambda} U_\alpha \cap \bigcap_{\beta \in \Lambda'} Z_\beta$, the restriction $\pi|_W$ induces a covering map between $W(R)$ and $Y(R)$ in the Euclidean topology.*

**Proof.** By Theorem 4.1.4 and Corollary 4.1.6, we may assume that $X, Y, Z_1, \dots, Z_m$ are all reduced, $\pi$ is finite étale, and the closed (open respectively) immersions $Z_i \hookrightarrow X$ ($U_i \hookrightarrow X$ respectively) are finite étale. Then the immersion $W \hookrightarrow X$ is finite étale, as the product of finite étale morphisms is still finite étale. Then the composition $\pi|_W : W \hookrightarrow X \to Y$ is also a finite étale morphism and the conclusion follows from Theorem 4.2.1. □

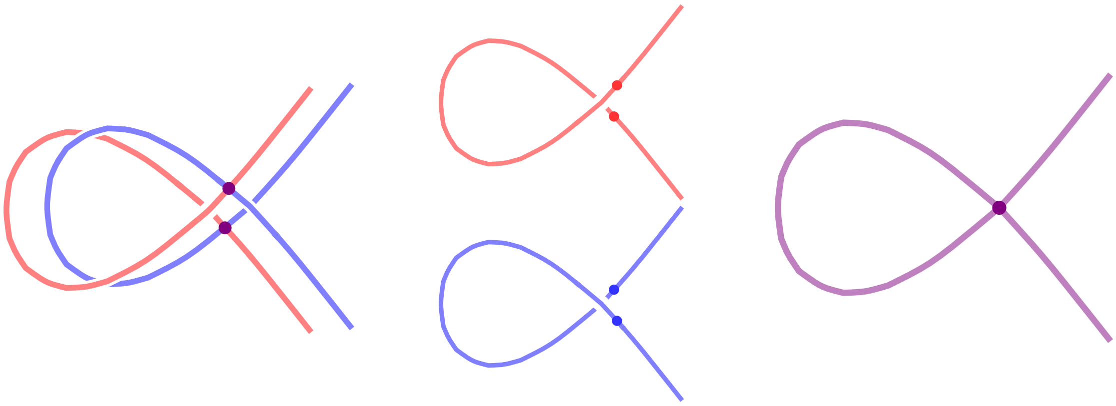
*Remark* 7. This generalizes Theorem 3.3.14 (which itself is a generalization of [Col75, Theorem 5] and [ACM84, Theorem 3.3]), using a different strategy (Corollary 4.1.6).

**Example 10.** This example is taken from [Har77, Exercise III.10.6]. Consider

$$I = \langle y_1 - x_1^2 + 1, y_2 - x_1 x_2, x_2^2 - (x_1^2 - 1)^2 \rangle, J = I \cap k[y_1, y_2] = \langle y_1^2(y_1 + 1) - y_2^2 \rangle.$$

Set $X = \mathrm{Spec}\, k[y_1, y_2, x_1, x_2]/I$ and $Y = \mathrm{Spec}\, k[y_1, y_2]/J$. Then $Y$ is a nodal curve. The normalization of $Y$ coincides with the blow-up of $Y$ at its singular locus $(0, 0)$. $X$ is the variety gluing two normalizations of $Y$ along the exceptional divisors. See Figure 4.4.

Later we shall see that the projection from $X$ to $Y$ is indeed a quasi-finite, flat morphism with locally constant geometric fibers in Example 11. As a result, the projection induces a covering map on the real points by Theorem 4.2.1. Note $X$ and $Y$ are both singular, so results based on the theory of smooth manifolds like Lazard-Rouillier (Corollary 4.2.2) do not apply here.



(a) Variety $X$ (b) Two copies of normalization of $Y$ (c) The nodal curve $Y$

Figure 4.4 A finite étale double cover of the nodal curve

The last generator of $I$, $x_2^2 - (x_1^2 - 1)^2$ can be factorized as $(x_2 - x_1^2 + 1)(x_2 + x_1^2 - 1)$. Each factor defines an irreducible component of $X$. Choose $x_2 - x_1^2 + 1$ and let $Z$ be the corresponding irreducible component of $X$. Then the projection from $Z$ to $Y$ is the blow-up map. It is easy to see that the restriction of this projection on the regular points $Y_{\text{reg}}$ is flat. Therefore, $Z \times_Y Y_{\text{reg}} \longrightarrow Y_{\text{reg}}$ is again a finite flat morphism with locally constant geometric fibers and it induces covering maps on the real points (Corollary 4.1.6 and Theorem 4.2.4).

Because $Z$ is defined by $x_2 - x_1^2 + 1 = 0$ on $X$ and $Z \times_Y Y_{\text{reg}} \longrightarrow Y_{\text{reg}}$ is flat, the complement $U = X \backslash Z$ yields another covering over $Y_{\text{reg}}$. In particular, the semi-algebraic set defined by

$$\left(y_1 - x_1^2 + 1 = 0 \wedge y_2 - x_1 x_2 = 0 \wedge x_2^2 - (x_1^2 - 1)^2 = 0 \wedge x_2 - x_1^2 + 1 > 0\right)$$
$$\bigwedge \neg \left(y_1 = 0 \wedge y_2 = 0\right)$$

is a covering over the real points of $Y_{\text{reg}}$.

## 4.3 Effective Aspect of the Criterion

In this section, we are going to show that the condition in Theorem 4.2.1 can be effectively verified, at least in the affine case. As there are algorithms to compute the radical ideal [KL91; EHV92; Lap06], it suffices to check finiteness and étaleness for the reduced map by Corollary 4.1.5. The setting in this section is as follows:

Let $I \trianglelefteq k[y_1, \ldots, y_m, x_1, \ldots, x_n]$ and $J = I \cap k[y_1, \ldots, y_m]$. Then

$$B = k[y_1, \ldots, y_m, x_1, \ldots, x_n]/I \text{ and } A = k[y_1, \ldots, y_m]/J$$

are $k$-algebras of finite type and $B$ is also an $A$-algebra. Let $T_x$ ($T_y$ resp.) be a monomial order in $x_i$'s ($y_i$'s resp.). Let $\mathcal{G}$ be a Gröbner basis of $I$ in $k[y_1, \ldots, y_m, x_1, \ldots, x_n]$, the global polynomial ring with respect to the block order $T = T_x > T_y$. Additionally, let $\mathcal{G}_y = \mathcal{G} \cap k[y_1, \ldots, y_m]$ and $\mathcal{G}_x = \mathcal{G} \backslash \mathcal{G}_y$, so $\mathcal{G}_y$ is a Gröbner basis for $J$ and $\mathcal{G}_x$ consists of the defining equations for $B$ over $A$.

First, we check the finiteness of the map. We present a test that is slightly more general than the version in [GP08, Proposition 3.1.5] (here we have more flexibility in the choice of monomial order $T$).

**Lemma 4.3.1** (Finiteness Test, c.f. [GP08, Proposition 3.1.5])**.** *The following are equivalent:*

1. *$B$ is a finite $A$-module.*
2. *There are monic polynomials $\varphi_i(\lambda) \in A[\lambda]$ such that $\varphi_i(x_i) = 0$ in $B$.*
3. *For each $x_i$, there is $g_i \in \mathcal{G}$ such that the leading monomial $\mathtt{LM}_T(g_i)$ is a power of $x_i$.*

**Proof.**
(1)⇔(2). We know that the $A$-algebra $B$ is finite if and only if $B$ is finitely generated as an $A$-algebra by finitely many integral elements [Eis95, Corollary 4.5].
(2)⇒(3). From (2), we see that for each $x_i$, there is a polynomial in $I$ whose leading term is $x_i^{s_i}$ for some positive integer $s_i$ (recall that $T$ is a block order). Therefore $x_i^{s_i}$ is in the leading term ideal $\langle \mathtt{LT}_T(I) \rangle$. As $\langle \mathtt{LT}_T(I) \rangle$ is generated by leading terms of Gröbner basis elements, there must be some $g_i \in \mathcal{G}$ whose leading term is $x_i^{t_i}$ ($t_i \leq s_i$).
(3)⇒(1). Suppose $x_i^{t_i}$ is the leading term of $g_i$. We claim that $B$ is generated by monomials

$$\mathcal{M} = \left\{ \prod_{i=1}^{n} x_i^{a_i} \middle| 0 \leq a_1 \leq t_1 - 1, \ldots, 0 \leq a_n \leq t_n - 1 \right\} \tag{$\star$}$$

as an $A$-module. Let $f \in k[y_1, \ldots, y_m, x_1, \ldots, x_n]$ be a representative of an element $b$ of $B$. By the polynomial division algorithm [CLO15, Theorem 2.3.3], there is another representative $r \in k[y_1, \ldots, y_m, x_1, \ldots, x_n]$ such that none of the monomials in $r$ is divisible by any of the leading monomials of $\mathcal{G}$, especially $x_1^{t_1}, \ldots, x_n^{t_n}$. Therefore $r$ provides a way to write $b$ as an $A$-linear combination of monomials in $\mathcal{M}$. Our claim is shown and the proof is completed. □

We remark that the finiteness check has been studied under a different notion called "properness defect" [Jel99; Sta02; SEDS04; Mor11]. Of course, for affine morphisms, being proper is equivalent to being finite [Har77, Exercise II.4.6].

Lemma 4.3.1 directly translates to an algorithm. See Algorithm 4.1.

**Algorithm 4.1:** `IsFinite`

**Input:** A block-order Gröbner basis $\mathcal{G}$ for $I \trianglelefteq k[y_1, \ldots, y_m, x_1, \ldots, x_n]$.

**Output:** `True` if $B$ is a finite $A$-module, otherwise `False`.

```
F_1, ..., F_n := False;
foreach g ∈ G do
    if LM_T(g) is of the form x_i^s then
        F_i := True;
return ⋀_{i=1}^n F_i;
```

Finiteness is a local property on the base. Therefore we can still say something about it even when the morphism is not finite: there is a biggest open subset of the base on which the restriction is finite. Actually, this set can be immediately read from the block order Gröbner basis $\mathcal{G}$ in the affine case. The basic idea is almost identical to Lemma 4.3.1, but it is trickier due to the presence of non-constant leading coefficients.

**Theorem 4.3.2** (Non-finite Locus)**.** *The non-finite locus of* Spec $B \to$ Spec $A$ *can be characterized in the following way:*

1. *The set*

$$J_i = \mathtt{LC}_{x_i}(I \cap k[y_1, \ldots, y_m, x_i]) = \left\{ \mathtt{LC}_{x_i}(f) \middle| f \in I \cap k[y_1, \ldots, y_m, x_i] \right\}$$

*is an ideal in $k[y_1, \ldots, y_m]$.*

2. *Let $W_i = V(J_i)$. Then $W = \bigcup_{i=1}^n W_i$ is the smallest closed subset of* Spec $A$ *such that the restriction*

$$\operatorname{Spec} B \times_{\operatorname{Spec} A} (\operatorname{Spec} A \backslash W) \to (\operatorname{Spec} A \backslash W)$$

*is finite.*

3. *Moreover, $J_i$ is generated by*

$$\left\{ \mathtt{LC}_{T_x}(g) \middle| g \in \mathcal{G}, \mathtt{LM}_{T_x}(g) = x_i^\alpha \text{ for some } \alpha \geq 0 \right\},$$

*the leading coefficients of members of Gröbner basis $\mathcal{G}$ whose leading monomial is of the form $x_i^\alpha$.*

**Proof.**

1. Trivial.
2. Let $D(h) = \operatorname{Spec} A_h \subseteq \operatorname{Spec} A$ be a distinguished open defined by $h \neq 0$. Notice that

$$
\begin{aligned}
& D(h) \cap W = \varnothing \\
\Leftrightarrow\ & V(h) \supseteq W \\
\Leftrightarrow\ & \exists s > 0 \text{ s.t. } h^s \in \textstyle\bigcap_{i=1}^n \mathtt{LC}_{x_i}(I \cap k[y_1, \ldots, y_m, x_i]) \\
\Leftrightarrow\ & \exists s > 0 \text{ s.t. } \textstyle\bigwedge_{i=1}^n \exists f_i \in I \cap k[y_1, \ldots, y_m, x_i] \text{ s.t. } \mathtt{LC}_{x_i}(f_i) = h^s \\
\Leftrightarrow\ & \textstyle\bigwedge_{i=1}^n \exists f_i \in I_h \cap k[y_1, \ldots, y_m, x_i]_h \text{ s.t. } \mathtt{LC}_{x_i}(f_i) = 1 \\
\Leftrightarrow\ & x_1, \ldots, x_n \text{ are integral over } A_h
\end{aligned}
$$

Recall that $B_h$ is finite if and only if $x_1, \ldots, x_n$ are integral over $A_h$. Therefore, $B_h$ is finite over $A_h$ if and only if $D(h) \cap W = \varnothing$. So $W \subseteq \operatorname{Spec} A$ is the smallest closed subset making

$$\operatorname{Spec} B \times_{\operatorname{Spec} A} (\operatorname{Spec} A \backslash W) \to (\operatorname{Spec} A \backslash W)$$

is finite.

3. Let $f \in I \cap k[y_1, \ldots, y_m, x_i]$ and set $x_i^\beta := \mathtt{LM}_{T_x}(f)$. Because $\mathcal{G}$ is a block-order Gröbner of $I$ over $k$, we can rewrite $f = \sum_{g \in \mathcal{G}} c_g g$ such that $\mathtt{LM}_T(f) = \max_{g \in \mathcal{G}} \mathtt{LM}_T(c_g)\mathtt{LM}_T(g)$ by polynomial division algorithm [CLO15, Theorem 2.3.3]. Since $T$ is the block order $T_x > T_y$, we have $x_i^\beta = \max_{g \in \mathcal{G}} \mathtt{LM}_{T_x}(c_g)\mathtt{LM}_{T_x}(g)$.
   Extracting the leading terms in $T_x$, we see that

$$
\begin{aligned}
\mathtt{LC}_{T_x}(f) x_i^\beta &= \mathtt{LT}_{T_x}(f) \\
&= \sum_{\substack{g \in \mathcal{G}, \\ x_i^\beta = \mathtt{LM}_{T_x}(c_g)\mathtt{LM}_{T_x}(g)}} \mathtt{LT}_{T_x}(c_g)\mathtt{LT}_{T_x}(g) \\
&= \sum_{\substack{g \in \mathcal{G}, \\ x_i^\beta = \mathtt{LM}_{T_x}(c_g)\mathtt{LM}_{T_x}(g)}} \mathtt{LC}_{T_x}(c_g)\mathtt{LC}_{T_x}(g)\mathtt{LM}_{T_x}(c_g)\mathtt{LM}_{T_x}(g) \\
&= \sum_{\substack{g \in \mathcal{G}, \\ x_i^\beta = \mathtt{LM}_{T_x}(c_g)\mathtt{LM}_{T_x}(g)}} \mathtt{LC}_{T_x}(c_g)\mathtt{LC}_{T_x}(g) x_i^\beta.
\end{aligned}
$$

   Canceling $x_i^\beta$ from both sides, we conclude that $\mathtt{LC}_{T_x}(f)$ is an element of $\langle \mathtt{LC}_{T_x}(g) | g \in \mathcal{G}, \mathtt{LM}_{T_x}(g) = x_i^\alpha \text{ for some } \alpha \geq 0 \rangle$. This finishes the proof.

□

By Theorem 4.3.2, Algorithm 4.2 can be used to compute the non-finite locus of a morphism.

We compare Algorithm 4.2 with the results presented in [Sta02; Mor11] now. First, we work without passing to the projective space, while both [Sta02] and [Mor11] introduce extra variables to homogenize. Second, only one Gröbner basis computation is needed in our

**Algorithm 4.2:** `NonFiniteLocus`

**Input:** A block-order Gröbner basis $\mathcal{G}$ for $I \trianglelefteq k[y_1, \ldots, y_m, x_1, \ldots, x_n]$.
**Output:** A defining ideal for the smallest closed subset $W$ of Spec $A$ such that the restriction on Spec $A \backslash W$ is finite.

```
1 J_1, ..., J_n := J;
2 foreach g ∈ G_x do
3     if LM_{T_x}(g) is of the form x_i^α then
4         J_i := J_i + ⟨LC_{T_x}(g)⟩;
5 return ⋂_{i=1}^n J_i;
```

method, namely a block-order Gröbner basis, while both [Sta02] and [Mor11, Algorithm 1] require multiple eliminations. Third, our method is more general, while [Sta02] deals with irreducible varieties only and [Mor11] works over $\mathbb{C}$.

We recall that the flatness of a finitely presented module can be checked via the Fitting ideals.

**Lemma 4.3.3** (Flatness Test, [GP08, Corollary 7.3.13])**.** *Suppose $B$ is a finite $A$-module with a finite presentation*

$$A^s \xrightarrow{N} A^r \to B \to 0,$$

*where $N \in A^{r\times s}$. Define the i-th Fitting ideal of $B$, $F_i(B) \trianglelefteq A$ to be ideal generated by all $(r-i)$-minors of $N$. Define*

$$F(B) = \sum_{i=1}^{r} \left((0 : F_{i-1}(B)) \cap F_i(B)\right).$$

*Then,*

1. *$B$ is flat over $A$ if and only if $1 \in F(B)$.*
2. *Let $p \in$ Spec $A$, then $B_p$ is flat over $A_p$ if and only if $p \notin V(F(B))$.*

The idea of flatness test is that flatness can be checked locally. Namely, $B_p$ is a free (equivalently, flat) $A_p$-module of rank $r$, if and only if $F_r(B_p) = A_p$ and $F_{r-1}(B_p) = 0$, i.e. $p$ lies outside the zero locus of $F_r(B)$ and the support of $F_{r-1}(B)$, which is the zero locus of the annihilator $\mathrm{Ann}(F_{r-1}(B)) = (0 : F_{r-1}(B))$.

To check the flatness of a finite $A$-algebra $B$, we still need a finite presentation of it as an $A$-module. Recall from Equation ($\star$) that $B$ can be generated by monomials

$$\mathcal{M} = \left\{ \prod_{i=1}^{n} x_i^{a_i} \,\middle|\, 0 \le a_1 \le t_1 - 1, \ldots, 0 \le a_n \le t_n - 1 \right\}.$$

This gives a surjection $\bigoplus_{x^\alpha \in \mathcal{M}} A x^\alpha \to B \to 0$, whose kernel is $(I/J) \cap \bigoplus_{x^\alpha \in \mathcal{M}} A x^\alpha$. We claim that the kernel is generated by

$$\mathcal{N} := \left\{ g \cdot x^\alpha \middle| g \in \mathcal{G}_x, x^\alpha \in \mathcal{M}, \deg_{x_i}(g \cdot x^\alpha) \le t_i - 1 \text{ for } i = 1, \dots, n \right\}.$$

To see the claim is true, it suffices to observe that if $f \in I/J$ is supported on $\mathcal{M}$, then the polynomial division algorithm [CLO15, Theorem 2.3.3] would rewrite $f$ as a linear combination of elements of $\mathcal{N}$. Now we obtain a finite presentation of $B$ as an $A$-module

$$\bigoplus_{g \cdot x^\alpha \in \mathcal{N}} A g x^\alpha \to \bigoplus_{x^\alpha \in \mathcal{M}} A x^\alpha \to B \to 0$$

Following the above discussion, we propose Algorithm 4.3 to check the flatness of the finite $A$-algebra $B$.

**Algorithm 4.3:** `IsFiniteFlat`

**Input:** A block-order Gröbner basis $\mathcal{G}$ for $I \trianglelefteq k[y_1, \dots, y_m, x_1, \dots, x_n]$.
**Output:** `True` if $B$ is a (finite) flat $A$-module, otherwise `False`.

**if** `IsFinite`$(\mathcal{G})$ = `False` **then**
  **return** `False`;
Choose $g_1, \dots, g_n$ from $\mathcal{G}$ such that $\mathtt{LM}_T(g_i)$ is a power of $x_i$;
Let $t_1, \dots, t_n$ be the degrees of $g_1, \dots, g_n$ in $x_1, \dots, x_n$ respectively;
$\mathcal{M} := \left\{ \prod_{i=1}^n x_i^{a_i} \middle| 0 \le a_1 \le t_1 - 1, \dots, 0 \le a_n \le t_n - 1 \right\}$;
$\mathcal{N} := \left\{ g \cdot x^\alpha \middle| g \in \mathcal{G}_x, x^\alpha \in \mathcal{M}, \deg_{x_i}(g \cdot x^\alpha) \le t_i - 1 \text{ for } i = 1, \dots, n \right\}$;
$r := |\mathcal{M}|, s = |\mathcal{N}|$;
Let $N$ be the $r \times s$ matrix whose $(i, j)$-entry is the $\mathcal{M}_i$-coefficient of $\mathcal{N}_j$;
**for** $i := 1$ **to** $r$ **do**
  $F_i := \langle \mathtt{minors}(N, r - i) \rangle + J$;
$F := \sum_{i=1}^r ((J : F_{i-1}) \cap F_i)$;
**if** $1 \in F$ **then**
  **return** `True`;
**else**
  **return** `False`;

Finally, we show that the étaleness can be effectively verified, provided $B$ is finite and flat over $A$.

**Theorem 4.3.4** (Étaleness Test)**.** *Suppose $B$ is finite and flat over $A$. Write $\mathcal{G}_x = \{g_1, \dots, g_t\}$.*

*Define*

$$\mathcal{J} = \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \cdots & \frac{\partial g_t}{\partial x_1} \\ \vdots & & & \vdots \\ \frac{\partial g_1}{\partial x_n} & \cdots & \cdots & \frac{\partial g_t}{\partial x_n} \end{pmatrix}$$

*and* $\mathrm{Jac}(B)$ *is the ideal generated by all n-minors of* $\mathcal{J}$ *plus* $I$*. Then* $B$ *is étale over* $A$ *if and only if* $1 \in \mathrm{Jac}(B)$.

**Proof.** We know that a smooth morphism is just a flat morphism with regular geometric fibers. Therefore to test whether a finite flat morphism is étale, it suffices to check if there is any singular (geometric) fiber. By Jacobian criterion [Eis95, Corollary 16.20], the singularity of fibers is just the zero set of the ideal $\mathrm{Jac}(B)$. Therefore, all (geometric) fibers are regular if and only if $1 \in \mathrm{Jac}(B)$. □

Algorithm 4.4 checks the finiteness and étaleness of a morphism. Its correctness follows from Lemma 4.3.1, Lemma 4.3.3 and Theorem 4.3.4.

---

**Algorithm 4.4:** `IsFiniteÉtale`

---

**Input:** A block-order Gröbner basis $\mathcal{G}$ for $I \trianglelefteq k[y_1, \ldots, y_m, x_1, \ldots, x_n]$.
**Output:** `True` if $B$ is finite étale over $A$, otherwise `False`.

**if** `IsFiniteFlat`($\mathcal{G}$) = `False` **then**
  **return** `False`;
Let $\mathcal{J}$ be the matrix $\left(\frac{\partial g}{\partial x_i}\right)_{i=1,\ldots,n, g\in\mathcal{G}_x}$;
Jac := $\langle$`minors`$(\mathcal{J}, n)\rangle + I$;
**if** $1 \in$ Jac **then**
  **return** `True`;
**else**
  **return** `False`;

---

We remark that the non-finite-étale locus can be computed from Algorithm 4.2 (finiteness), Fitting ideals in Algorithm 4.3 (flatness) and an elimination of the Jacobian ideal in Algorithm 4.4 (étaleness). This gives the biggest open subset of the base where the projection is finite étale.

**Example 11.** Recall the double cover $X$ of nodal curve $Y$ in Example 10. Their defining ideals are

$$I = \langle y_1 - x_1^2 + 1, y_2 - x_1x_2, x_2^2 - (x_1^2 - 1)^2\rangle, J = I \cap k[y_1, y_2] = \langle y_1^2(y_1 + 1) - y_2^2\rangle.$$

We will verify that the projection from $X$ to $Y$ is finite étale using Algorithm 4.4.

Using a block order $\texttt{grevlex}(x_1, x_2) > \texttt{grevlex}(y_1, y_2)$, we compute a Gröbner basis for $I$:

$$\mathcal{G} = \left\{y_1^3 + y_1^2 - y_2^2, -y_1x_2 + y_2x_1 - x_2, y_1^2x_1 - y_2x_2, x_2^2 - y_1^2, x_1x_2 - y_2, x_1^2 - y_1 - 1\right\}.$$

As $x_1^2 - y_1 - 1, x_2^2 - y_1^2 \in \mathcal{G}$, we conclude that the projection is finite. The monomial basis can be chosen to be $\mathcal{M} = \{1, x_1, x_2, x_1x_2\}$. The relations are given by

$$\mathcal{N} = \left\{-y_1x_2 + y_2x_1 - x_2, y_1^2x_1 - y_2x_2, x_1x_2 - y_2\right\}.$$

Then the presentation matrix is

$$N = \begin{pmatrix} 0 & y_2 & -y_1 - 1 & 0 \\ 0 & y_1^2 & -y_2 & 0 \\ -y_2 & 0 & 0 & 1 \end{pmatrix}$$

Expanding the minors of $N$, we see that the Fitting ideals are

$$F_0 = F_1 = \langle y_1^3 + y_1^2 - y_2^2\rangle = J, F_2 = F_3 = \langle 1\rangle.$$

Therefore the projection is also flat.

Finally we check the smoothness of fibers. The Jacobian matrix is

$$\mathcal{J} = \begin{pmatrix} y_2 & y_1^2 & 0 & x_2 & 2x_1 \\ -y_1 - 1 & -y_2 & 2x_2 & x_1 & 0 \end{pmatrix}.$$

So the Jacobian ideal is $I$ plus the ideal generated by all 2-minors of $\mathcal{J}$. A direct computation shows that the Jacobian ideal is $\langle 1\rangle$. Therefore all fibers are smooth and the projection is finite étale.

## 4.4 Covering Stratification, Sampling and Matrix Completion

We discuss some applications of our criterion in the section.

The first application is to stratify a quasi-finite morphism of $R$-varieties so the map is a covering in each piece.

**Proposition 4.4.1.** *Suppose $f : X \to Y$ is a quasi-finite $R$-morphism of $R$-varieties, then there are finitely many locally closed subschemes $\{Y_i\}$ of $Y$, whose union covers $Y$, such that the restriction map $f_i : X_i = X \times_Y Y_i \to Y_i$ is a covering map on the $R$-rational points.*

**Proof.** By taking reduced subscheme structures and applying generic flatness repeatedly, there exists $t \geq 0$ and closed subschemes

$$Y \supseteq Y_{\mathrm{red}} = V_0 \supsetneq V_1 \supsetneq \cdots \supsetneq V_t = \varnothing$$

such that $X \times_Y (V_i \backslash V_{i+1})$ is flat over $V_i \backslash V_{i+1}$ (the generic flatness stratification, see [Stacks, Lemma 0H3Z]). Set $Y_i = V_i \backslash V_{i+1}$, then the restriction of $f$ on $Y_i$ is quasi-finite and flat. By the lower semi-continuity of geometric fiber cardinality for a quasi-finite flat $R$-morphism [EGA $\mathrm{IV}_3$, Proposition 15.5.9], we can further stratify $Y_i$ into locally closed subschemes $Y_{ij}$ where the cardinality of geometric fibers is a constant. Now apply Theorem 4.2.1 and the proof is completed. $\square$

**Example 12.** This example is from algebraic statistics [HS14, Example 1.17]. A random variable $\xi$ follows an unknown discrete distribution $(p_0, p_1, p_2)$, where the distribution shall satisfy a statistical model:

$$F(p_0, p_1, p_2) = p_2(p_1 - p_2)^2 + (p_0 - p_2)^3 = 0.$$

We can model all possible distributions by $V(F)$ in the projective plane $\mathbb{P}^2_p$. Here the subscript $p$ stands for the probability.

Suppose our observation for $\xi$ is $(u_0, u_1, u_2)$, and we model all observations by another projective plane $\mathbb{P}^2_u$. Our goal is to find the most "reasonable" distribution from the observation. This can be done by the method of maximum likelihood estimation (MLE). The family of all estimated distributions for all observations is the *likelihood correspondence*, a closed subvariety

$$X \hookrightarrow V(F) \times \mathbb{P}^2_u \hookrightarrow \mathbb{P}^2_p \times \mathbb{P}^2_u.$$

By a standard argument, the defining equations for $X$ in $\mathbb{Q}[p_0, p_1, p_2, u_0, u_1, u_2]$ can be found by saturating the ideal $\left\langle \begin{vmatrix} 1 & 1 & 1 \\ u_0p_1p_2 & p_0u_1p_2 & p_0p_1u_2 \\ \partial F/\partial p_0 & \partial F/\partial p_1 & \partial F/\partial p_2 \end{vmatrix}, F \right\rangle$ with respect to $\langle p_0p_1p_2(p_0 + p_1 + p_2)\rangle \cap \langle p_0 - p_2, p_1 - p_2\rangle$.

The fibers of projection downto the observation space $\pi : X \to \mathbb{P}^2_u$ are the MLEs of the observation. The length of the generic fiber is known as the *ML degree* of the statistic model $F$. It is known that the ML degree of $F$ is 5 [HS14], which means that there are 5 sets of "distributions" $(p_0, p_1, p_2)$ over the complex number for a general observation data $(u_0, u_1, u_2)$.

However, a complex "distribution" does not make much sense, as probabilities $p_0, p_1, p_2$ must be real numbers. We will now scrutinize the real fibers of $\pi_{\mathbb{R}}$ more carefully. We invite the readers to verify our arguments below using Algorithm 4.1-4.4 on a computer algebra system.

First, the projection $\pi : X \longrightarrow \mathbb{P}_u^2$ is already a finite one. A brute-force way to see this is to check that all fibers are finite, as a quasi-finite projective morphism is finite. A more elegant solution is to cover $\mathbb{P}_u^2$ by affine opens $D_+(u_0 + u_1 + u_2), D_+(u_0 + 3u_2)$ and $D_+(u_0(u_0 + u_1))$. Their preimages are again affine opens $D_+(p_0 + p_1 + p_2), D_+(p_2)$ and $D_+(p_1)$ respectively, so we are in the affine case where Algorithm 4.1 is applicable.

Next, one can verify by Algorithm 4.3 that the non-flat locus of the finite morphism $\pi$ is a single point $(1 : 1 : 1) \in \mathbb{P}_u^2$, i.e. the restriction of $\pi$ on $\mathbb{P}_u^2\backslash(1 : 1 : 1)$ is flat. Therefore, all fibers of $\pi$ are of the same length except $(1 : 1 : 1)$'s [Har77, Corollary III.9.10]. This length is the ML degree of $F$, which is 5. But the length of fiber over $(1 : 1 : 1)$ is 6. Hence the jump in length is purely due to the non-flatness.

Then, the discriminant locus is a curve $C$ of degree 8 in $\mathbb{P}_u^2$ (this can be found by an elimination on the Jacobian ideal in Algorithm 4.4). This is where the geometric fiber cardinality is different from the expected number 5. The exceptional point $(1 : 1 : 1)$ also lies on $C$. Therefore, the restriction of $\pi$ on $\mathbb{P}_u^2\backslash C$ is a finite flat morphism with constant geometric fibers. By Theorem 4.2.1, $\pi_{\mathbb{R}}$ is a covering map on $\mathbb{P}_u^2(\mathbb{R})\backslash C(\mathbb{R})$, and the number of real fibers is a constant on each connected component of $\mathbb{P}_u^2(\mathbb{R})\backslash C(\mathbb{R})$.

The singular locus of $C$ consists of 21 points, and 7 of them are real. The geometric fiber cardinality $n(y)$ is 4 on the smooth points of $C$, but this number drops to 3 on $\operatorname{Sing} C$. Therefore, the restriction of $\pi$ on $C\backslash(\operatorname{Sing} C \cup (1 : 1 : 1))$ (base change) is again a finite flat morphism with constant geometric fibers and we can apply Theorem 4.2.1 again here.

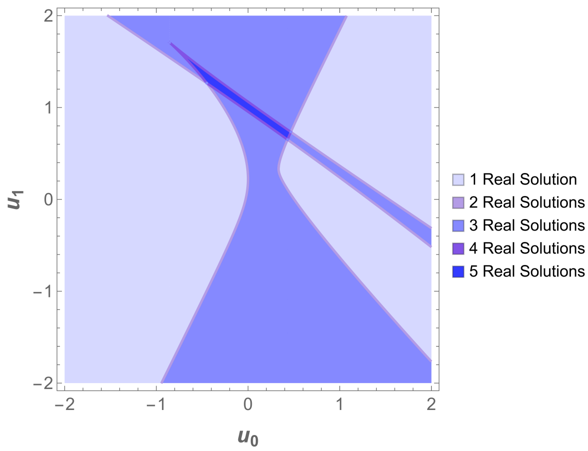



Figure 4.5 The number of real fibers in $D_+(u_0 + u_1 + u_2)$.

We plot the regions where the observation has $1, \ldots, 5$ distinct real MLEs using different colors in Figure 4.5. One can summarize the result as follows:

1. There is a region in the center of the figure where each observation has 5 different real

MLEs.

2. Each observation on the boundaries of the aforementioned region has 4 different real MLEs, except for four corners, which are singularities of the discriminant locus. Each corner point corresponds to three real $(p_0 : p_1 : p_2)$.
3. The region for 3 real MLEs is connected (the upper, lower and middle-right regions in Figure 4.5 are connected at the infinity).
4. For an observation with a unique real MLE, there are two connected components: the left-hand region and the upper-right region actually lie in the same connected component.
5. The remaining points are the border between observations with 1 and 3 real MLEs. They have two distinct real MLEs.

One can even require that all the observation data $(u_0 : u_1 : u_2)$ are positive and look for the number of positive MLEs $(p_0 : p_1 : p_2)$. It can be shown that when the observation is positive, one of the real solutions is never positive, but the remaining real solutions are. See Figure 4.6.

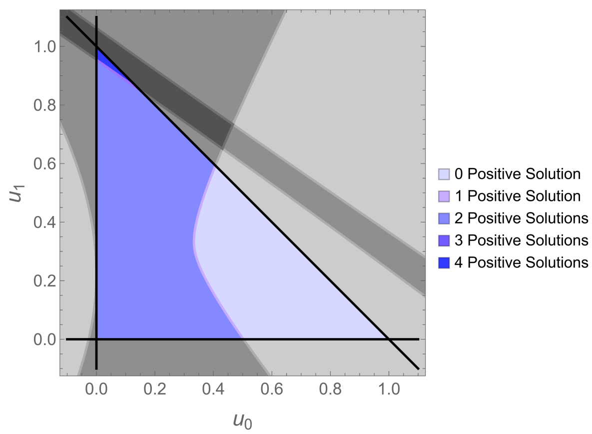



Figure 4.6 The number of positive fibers in the positive quadrant of $D_+(u_0 + u_1 + u_2)$.

Previously, [RT15; RT17] developed a notion call data-discriminant, which corresponds to the curve $C$ in this example. The main result about data-discriminant locus in [RT15; RT17] is that the number of real or positive solutions is uniform over each connected component of the complement of the data-discriminant locus. Using our results, it is even possible to analyze the special data on the data-discriminant locus.

Let us turn to a related problem. Finding at least one point on each semi-algebraically connected component of a given real algebraic set is a classical problem in computational real algebraic geometry. We point out that one can reduce the problem to several smaller similar problems, using the covering stratification.

**Proposition 4.4.2.** *Suppose $f : X \to Y$ is a quasi-finite $R$-morphism of $R$-varieties, then finding at least one sample point on each semi-algebraically connected component of $X(R)$ can be reduced to finding at least one sample point on each semi-algebraically connected component of $Y_i(R)$ for some locally closed subschemes $Y_i \subseteq Y$.*

**Proof.** By Proposition 4.4.1, there are finitely many locally closed subschemes $Y_i \subseteq Y$ such that $X_i = X \times_Y Y_i \to Y_i$ induces a covering map on the rational points. Clearly, if we have found at least one sample point on each semi-algebraically connected component of each $X_i(R)$, then they meet every semi-algebraically connected component of $X(R)$. So it suffices to show that sample points meeting each semi-algebraically connected components of $Y_i(R)$ can be lifted to sample points meeting each semi-algebraically connected components of $X_i(R)$. Let $S = \{y_i\} \subseteq Y_i(R)$ be the sample points of $Y_i(R)$, we claim that $f_R^{-1}(S) = \bigcup_i f_R^{-1}(y_i)$ is a set of sample points meeting each semi-algebraically connected components of $X_i(R)$.

Let $C \subseteq X_i(R)$ be a semi-algebraically connected component of $X_i(R)$. Pick arbitrary $x_0 \in C$ and let $y_0 = f(x_0)$. Then there is some $y_i \in S$ such that $y_0$ and $y_i$ lie in the same semi-algebraically connected component. Notice that for a semi-algebraic set, being semi-algebraically connected is equivalent to being semi-algebraically path-connected [BCR98, Proposition 2.5.13]. So there is a semi-algebraic path $\gamma : [0, 1] \to Y_i(R)$ such that $\gamma(0) = y_0$ and $\gamma(1) = y_i$. By the unique path lifting property [DK84, Proposition 5.7] (or [DK81a, Theorem 3.3]), there is a semi-algebraic path $\tilde{\gamma} : [0, 1] \to X_i(R)$ such that $\tilde{\gamma}(0) = x_0$ and $f_R \circ \tilde{\gamma} = \gamma$. Then $\tilde{\gamma}(1) \in f_R^{-1}(y_i)$ is the required sample point meeting $C$. □

We remark that this process can be repeated by choosing a projection from $Y$.

At last, we study a completion-type matrix problem using all the techniques we have developed so far. Here, a matrix $M$ is partially observed and we need to recover it by an either low-rank or positive semi-definite hypothesis on $M$.

**Example 13** (Matrix Completion)**.** Consider the symmetric matrix

$$
\begin{aligned}
M &= \begin{pmatrix} -v & 0 & u-x & 2v-x \\ 0 & u+v & -v-x & u+2v-x \\ u-x & -v-x & -u+2v+y & -2u-v+y \\ 2v-x & u+2v-x & -2u-v+y & y \end{pmatrix} \\
&= u \cdot \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & -1 & -2 \\ 0 & 1 & -2 & 0 \end{pmatrix} + v \cdot \begin{pmatrix} -1 & 0 & 0 & 2 \\ 0 & 1 & -1 & 2 \\ 0 & -1 & 2 & -1 \\ 2 & 2 & -1 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & -x & -x \\ 0 & 0 & -x & -x \\ -x & -x & y & y \\ -x & -x & y & y \end{pmatrix}.
\end{aligned}
$$

Here, the last summand $\left(\begin{smallmatrix} 0 & 0 & -x & -x \\ 0 & 0 & -x & -x \\ -x & -x & y & y \\ -x & -x & y & y \end{smallmatrix}\right)$ is the observed part of $M$ and $u, v$ are unknown. Question:

(i) (*Low-rank completion*) For which $x, y \in \mathbb{R}$ values can $M$ be completed to a real matrix of rank 2, i.e. there exist $u, v \in \mathbb{R}$ such that $M$ is a real symmetric matrix of rank 2?

(ii) (*PSD completion*) For which $x, y \in \mathbb{R}$ values can $M$ be completed to a p.s.d. of rank 2, i.e. there exist $u, v \in \mathbb{R}$ such that $M$ is a real symmetric positive semi-definite matrix of rank 2?

Let $I$ be the ideal generated by all 3-minors of $M$ and let $J$ be the ideal generated by all 2-minors of $M$. Set $X = \operatorname{Spec} \mathbb{R}[x, y, u, v]/I$, $Z = \operatorname{Spec} \mathbb{R}[x, y, u, v]/J$ and $Y = \operatorname{Spec} \mathbb{R}[x, y]$. Clearly $Z$ is a closed subvariety of $X$, and the real matrix of rank 2 are the $\mathbb{R}$-rational points of $X \backslash Z$. It can be easily checked that the projection $X \to Y$ is quasi-finite.

By recursively applying generic flatness and counting the geometric fiber, we may stratify $Y$ into a union of locally closed subvarieties $Y_i$ such that the projection $X \times_Y Y_i \to Y_i$ is a quasi-finite, flat morphism with constant geometric fibers. For example, the stratification can be taken as

$$
\begin{cases}
Y_1 = D(x(x-y)(8x^4 + 4x^2y^2 - 4xy^3 + y^4)), \\
Y_2 = V(8x^4 + 4x^2y^2 - 4xy^3 + y^4) \cap D(x(x-y)), \\
Y_3 = V(x) \cap D(x-y), \\
Y_4 = V(x-y) \cap D(x), \\
Y_5 = V(x, y).
\end{cases}
$$

Similarly $Y$ can be stratified into a union of locally closed subvarieties $Y'_j$ such that the projection $Z \times_Y Y'_j \to Y'_j$ is flat. A possible stratification is

$$\begin{cases} Y'_1 &=& D(x), \\ Y'_2 &=& V(x) \cap D(y), \\ Y'_3 &=& V(x,y). \end{cases}$$

Then for arbitrary $Y_i$ and $Y'_j$, we have $X \times_Y Y_i \times_Y Y'_j \to Y_i \times_Y Y'_j$ is a quasi-finite, flat morphism with constant geometric fibers and $Z \times_Y Y_i \times_Y Y'_j \to Y_i \times_Y Y'_j$, the restriction of projection on $Z$, is flat. By Theorem 4.2.4, the projection of $X(\mathbb{R})\backslash Z(\mathbb{R})$ on $Y_i(\mathbb{R}) \cap Y'_j(\mathbb{R})$ is a covering map.

The real affine plane $\mathbb{R}^2$ can be decomposed into 9 semi-algebraically connected sets $C_1, \ldots, C_9$, depending on the sign condition of $x$ and $y - x$, such that for all $Y_i$, $Y'_j$, either $C_k \cap Y_i(\mathbb{R}) \cap Y'_j(\mathbb{R}) = \varnothing$ or $C_k \subseteq Y_i(\mathbb{R}) \cap Y'_j(\mathbb{R})$.

Then for each cell $C_k$, the number of real fibers of $X\backslash Z$ is a constant. Therefore we can choose one point from each cell $C_k$ to check whether $M$ can be completed to a real symmetric matrix of rank 2 for all $(x, y) \in C_k$. For example, let us choose $(x, y) = (1, 1)$ from $\{(x, y)|x > 0, y = x\}$ and plug it into $M$. We get

$$\begin{pmatrix} -v & 0 & u-1 & 2v-1 \\ 0 & u+v & -v-1 & u+2v-1 \\ u-1 & -v-1 & -u+2v+1 & -2u-v+1 \\ 2v-1 & u+2v-1 & -2u-v+1 & 1 \end{pmatrix}.$$

By setting all 3-minors to zero, we get a zero-dimensional polynomial system in $u, v$. There are 6 complex solutions and 2 of them are real matrices of rank 2. They are

$$\begin{pmatrix} 0 & 0 & -1 & -1 \\ 0 & 0 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 \end{pmatrix} \text{ and } \begin{pmatrix} -1 & 0 & 0 & 1 \\ 0 & 2 & -2 & 2 \\ 0 & -2 & 2 & -2 \\ 1 & 2 & -2 & 1 \end{pmatrix}.$$

Therefore we can conclude that for all $x = y > 0$, the matrix $M$ can be completed to a real symmetric matrix of rank 2. Repeat this process for all other cells, we can see that the matrix $M$ can be completed to a real symmetric matrix of rank 2 if and only if $x \neq 0$. This solves the first part of the question.

As for the second part of the question, we recall the Rank-Signature lemma (Lemma 3.3.3 or equivalently [BPR06, Proposition 9.14]), which says for a continuous family of real symmetric matrices, the signature does not change if the rank remains the same. Using the unique path lifting property of covering maps, we can see that for every pair of points $p, q$ in the same cell $C_k$, the fibers of $p, q$ in $X(\mathbb{R})\backslash Z(\mathbb{R})$ can be connected using paths in $X(\mathbb{R})\backslash Z(\mathbb{R})$. So the endpoints share the same signature and there is a p.s.d. in the fiber of some $p_0 \in C_k$ if and only if there is a p.s.d. in fibers of all $p \in C_k$. Therefore, again it suffices to check only one point in $C_k$. For example, since all the real fibers of $(x, y) = (1, 1)$ in $X(\mathbb{R})\backslash Z(\mathbb{R})$ are indefinite, $M$ cannot be completed to a p.s.d. for all $x = y > 0$. Continuing this procedure, we see that it is impossible to complete $M$ to be a p.s.d. of rank 2 for any $(x, y)$.

# Chapter 5 Application in Transcendental Problems

This chapter is based on the joint work with Bican Xia.

In this chapter, we will apply the techniques in cylindrical algebraic decomposition to solve problems involving transcendental constraints. We will adapt well-established numerical methods (fixed-point iteration) to a symbolic context which reveals a root distribution pattern for some transcendental functions related to trigonometric functions. The notion of delineability from cylindrical algebraic decomposition plays an important role in our construction.

## 5.1 Settings

Let $R[x]$ be the univariate polynomial ring in $x$ over $R$. Suppose $k$ is a non-negative integer, then $R[x]_k$ is the free $R$-module of all polynomials in $x$ of degree $< k$.

In this chapter, we use $S$ to denote a ring extension of $\mathbb{Q}[x]$ that is

1. A subring of $C^\omega(\mathbb{R}) = \{f(x) : \mathbb{R} \to \mathbb{R} | f \text{ is analytic}\}$.
2. Closed under differentiation. In other words, if $f(x) \in S$, then the derivative $f'(x)$ is also in $S$.
3. Every non-zero element $f \in S$ has only finitely many zeros in $\mathbb{R}$.
4. Computable.

Clearly, $S$ is an integral domain by the identity theorem on analytic functions. The third condition is related to o-minimality. We say $S$ is **computable**, if

(4a) Addition and multiplication can be effectively performed.

(4b) It can be effectively decided whether $s \in S$ is zero.

(4c) Given $x \in \mathbb{Q}$, the sign of $s(x)$ can be effectively computed.

(4d) There exists an algorithm to compute an upper bound and a lower bound of all real roots of $s(x) \in S$.

Let $R$ be a ring of functions from $\mathbb{R}$ to $\mathbb{R}$. We define $\mathcal{L}(R)$, the **language** of $R$ to be the collection of the following expressions.

1. If $f \in R$, then $E(x) = (f(x) \triangleright 0)$ is an (atomic) expression ($\triangleright \in \{\leq, \geq, =, <, >, \neq\}$).
2. If $E_1(x)$ and $E_2(x)$ are expressions in $\mathcal{L}(R)$, then $E_1(x) \wedge E_2(x)$, $E_1(x) \vee E_2(x)$, $E_1(x) \to E_2(x)$, $\neg E_1(x)$ are expressions in $\mathcal{L}(R)$.
3. If $x$ is free in the expression $E(x) \in \mathcal{L}(R)$, then $(\forall x)E(x)$ and $(\exists x)E(x)$ are in $\mathcal{L}(R)$.

We say $R$ or $\mathcal{L}(R)$ is **decidable** over the reals, if there exists an algorithm $A$ that computes the truth value of a fully-quantified expression in $\mathcal{L}(R)$. Otherwise $R$ is said to be **undecidable**.

The **trigonometric extension** of $S$ is the ring $S[\sin x, \cos x]$. Because $\sin^2 x + \cos^2 x = 1$, we can use the Tangent Half Angle Formula to get a more compact description.

$$\begin{array}{rcl} S[\sin x, \cos x] & \to & S[\tan \frac{x}{2}]_{1+\tan^2 \frac{x}{2}} \\ \sum_{i,j} s_{i,j}(x) \sin^i x \cos^j x & \mapsto & \sum_{i,j} s_{i,j}(x) \left(\frac{2\tan\frac{x}{2}}{1+\tan^2\frac{x}{2}}\right)^i \left(\frac{1-\tan^2\frac{x}{2}}{1+\tan^2\frac{x}{2}}\right)^j. \end{array}$$

The image is said to be the **canonical form** of an element of $S[\sin x, \cos x]$.

Cleaning all the denominators, we are reduced to studying $S[\tan \frac{x}{2}]$. The singularity at $x = (2k+1)\pi$ ($k \in \mathbb{Z}$) can be handled by the following lemma.

**Lemma 5.1.1.** *Suppose $f \in S[\sin x, \cos x]$, either for all $k \in \mathbb{Z}$, $f((2k+1)\pi) = 0$, or there are some $\kappa_-, \kappa_+ \in \mathbb{Z}$ such that $f((2k+1)\pi) = 0$ implies $\kappa_- \le k \le \kappa_+$.*

**Proof.** Write $f$ as $\sum_{i,j} s_{i,j}(x) \sin^i x \cos^j x$, $(s_{i,j}(x) \in S)$. Then $f((2k+1)\pi) = \sum_j (-1)^j s_{0,j}((2k+1)\pi)$. Let

$$s(x) = \sum_j (-1)^j s_{0,j}(x),$$

then we have

$$f((2k+1)\pi) = 0 \Leftrightarrow s((2k+1)\pi) = 0.$$

But $s \in S$, either $s = 0$ so $x = (2k+1)\pi$ is a root of $f(x) = 0$ for all $k \in \mathbb{Z}$, or all real roots of $s(x)$ is bounded by $(-B, B)$ for some $B > 0$, so if $|k|$ is too large then $f((2k+1)\pi) = s((2k+1)\pi) \neq 0$.

□

## 5.2 The Reduction Theorem

Our goal is to prove the following result.

**Theorem 5.2.1** (Reduction Theorem)**.** *Let $f_i \in S[\sin x, \cos x]$ for $i = 1, \dots, s$, then there are effective bounds $N, M \in \mathbb{R}$ such that any quantifier-free formula $\varphi(x)$ whose atoms are of the form $f_i \triangleright 0$ is true for all $x \in \mathbb{R}$ if and only if $\varphi(x)$ is true for all $x \in [N, M]$.*

In order to prove Theorem 5.2.1, we need to study the root distribution of members $f(x)$ of $S[\sin x, \cos x]$. We will see that all sufficiently large real roots of $f(x) = 0$ share a similar pattern. Let $\frac{g(x)}{(1+\tan^2\frac{x}{2})^\alpha}$ ($g(x) \in S[\tan\frac{x}{2}]$) be the canonical form of $f(x)$, then $g(x)$ and $f(x)$

share the same roots except the possible roots at the singularity $x = (2k+1)\pi (k \in \mathbb{Z})$, which can be handled by Lemma 5.1.1. So it suffices to study the root distribution of $g(x) = 0$.

**Lemma 5.2.2.** *Suppose $g(x, y) \in S[y]$, then there exists some effective $M_0, N_0 \in \mathbb{R}$ such that the roots of $g(x, y) = 0$ in $y$ are delineable on $(M_0, +\infty)$ and $(-\infty, N_0)$.*

*Moreover, there is some $g_0 \in S[y]$ such that for all $u \in (-\infty, N_0) \cup (M_0, +\infty)$, the specialization $g_0(u; y)$ is the square-free part of $g(u; y)$.*

**Proof.** It suffices to prove the $(M_0, +\infty)$ part, since the argument needed for $(-\infty, N_0)$ is identical.

Let $s_n \in S$ be the leading coefficient of $g(x, y) \in S[y]$. Because $s_n \neq 0$, there is some effective $B \in \mathbb{R}$ such that $s_n(x) \neq 0$ for all $x > B$. Now let $\Delta_k = \mathrm{sRes}_{y,k}(g, \frac{\partial g}{\partial y}) \in S$ for $k = 0, \ldots, \deg_y g - 1$. Then there is some $0 \leq l \leq \deg_y g - 1$ such that $\Delta_k = 0$ for all $0 \leq k < l$ but $\Delta_l \neq 0$. Therefore there is another effective $B' \in \mathbb{R}$ such that for all $x > B'$: $\Delta_0(x) = \cdots = \Delta_{l-1}(x) = 0$, but $\Delta_l(x) \neq 0$. Let $M_0 = \max\{B, B'\}$, suppose $g(x, y) = \sum_{j=0}^{n} s_j(x) y^j$, define the map

$$\begin{aligned} \psi : (M_0, +\infty) &\rightarrow \mathbb{R}^{n+1} \\ x &\mapsto (s_0(x), s_1(x), \ldots, s_n(x)) \end{aligned}$$

and let $T$ be the image of $\psi$. By the construction of $(M_0, +\infty)$ and $T$, the polynomial $\sum_{j=0}^{n} x_j y^j$ is delineable on $T$ and we can pullback along $\psi$ to conclude that $g(x, y)$ is delineable on $(M_0, +\infty)$.

As for the last assertion, let $s = \Delta_l$, and there exists $d(x, y) \in S[y]$ such that for all $u > M_0$, $d(u; y)$ is the greatest common divisor of $g(u; y)$ and $\frac{\partial g}{\partial y}(u; y)$ by Theorem 2.3.2. Notice also that $\mathtt{LC}_y(d) = s$. By performing polynomial division, there exists some unique $g_0, R \in S[y]$ such that $s^{n-l+1} g = g_0 d + R$, $\deg_y R < \deg_y d$. After specializing to any $u \in (M_0, +\infty)$, we see that $s^{n-l+1}(u) g(u; y) = g_0(u; y) d(u; y) + R(u; y)$, But $d(u; y)$ divides $g(u; y)$ and $\deg_y R(u; y) \leq \deg_y R < \deg_y d = \deg_y d(u; y)$, forcing $R(u; y) = 0$. So $g_0(u; y)$ is the square-free part of $g(u; y)$. □

**Definition 5.2.3.** Let $k$ be a positive integer, and let $\xi(x)$ be a root function of $g(x, y) \in S[y]$ on $(M_0, +\infty)$. The **k-th contraction mapping $T_{\xi,k}$ associated with** $\xi$ is defined to be

$$\begin{aligned} T_{\xi,k} : [2k\pi - \pi, 2k\pi + \pi] &\rightarrow [2k\pi - \pi, 2k\pi + \pi] \\ x &\mapsto 2\arctan(\xi(x)) + 2k\pi. \end{aligned}$$

**Lemma 5.2.4.** *Suppose $g(x, y) \in S[y]$ and $\tau \in (0, 1)$, there are effective $M_1, N_1 \in \mathbb{R}$ such that if $\xi(x)$ is a root function of $g(x, y) = 0$ defined on $(M_1, +\infty)$ or $(-\infty, N_1)$, then $\left|\frac{2\xi'(x)}{\xi^2(x)+1}\right| < \tau$.*

**Proof.** Again it suffices to prove the $(M_1, +\infty)$ part.

Applying Lemma 5.2.2, we can see there is an effective $M_0 \in \mathbb{R}$ such that the roots of $g(x, y)$ in $y$ are delineable on $(M_0, +\infty)$, and there exists $g_0 \in S[y]$ such that $g_0(u; y)$ is the square-free part of $g(u; y)$ for all $u \in (M_0, +\infty)$. Clearly $\{(x, y) \in \mathbb{R}^2 | x > M_0, g(x, y) = 0\} = \{(x, y) \in \mathbb{R}^2 | x > M_0, g_0(x, y) = 0\}$, so the root functions of $g$ are also the root functions of $g_0$ on $(M_0, +\infty)$.

Now let

$$h_{\pm} = 2\frac{\partial g_0}{\partial x} \pm \tau \frac{\partial g_0}{\partial y}(y^2 + 1) \in S[y],$$

which are well-defined because $S$ is closed under differentiation. We claim that $g_0$ has no common real solution with $h_{\pm}$ for sufficiently large $x$.

Let $\delta_k^{\pm} = \mathrm{sRes}_{y,k}(h_{\pm}, g_0) \in s$ for $k = 0, \ldots, \deg_y g_0$, then there is some $0 \leq l \leq \deg_y g_0$ such that $\delta_0^+ = \cdots = \delta_{l-1}^+ = 0$ but $\delta_l^+ \neq 0$. Similarly there is some $0 \leq l' \leq \deg_y g_0$ such that $\delta_0^- = \cdots = \delta_{l'-1}^- = 0$ but $\delta_{l'}^- \neq 0$. Let $M_1 \geq M_0$ be an (effective) upper bound of all real roots of $\delta_l^+(x)$ and $\delta_{l'}^-(x)$. By Theorem 2.3.2, there exists $d_+, d_- \in S[y]$ such that for all $u > M_1$, $d_+(u; y)$ ($d_-(u; y)$ respectively) is the greatest common divisor of $g_0(u; y)$ and $h_+(u; y)$ ($h_-(u; y)$ respectively). Now we need to prove $d_{\pm}(u; y)$ have no real root when $u > M_1$.

Because for all $u \in (M_1, +\infty)$, $\deg_y d_{\pm}(u; y) = \deg_y d_{\pm}$ and $d_{\pm}(u; y)$ divide $g_0(u; y)$, which is square-free, it is easy to see that $d_{\pm}$ are delineable on $(M_1, +\infty)$. So the root functions of $d_{\pm}$ on $(M_1, +\infty)$ are also the root functions of $g_0$. Suppose $\zeta(x)$ is a root function of $d_+$, then $g_0(x, \zeta(x)) = h_+(x, \zeta(x)) = 0$. By the implicit function theorem, the derivative of $\zeta(x)$ is $-(\frac{\partial g_0}{\partial x})/(\frac{\partial g_0}{\partial y})|_{(x,\zeta(x))}$ (notice that $\frac{\partial g_0}{\partial y}|_{(x,\zeta(x))}$ is not zero because $g_0$ is square-free after specialization). So,

$$\begin{aligned} h_+(x, \zeta(x)) = 0 \quad &\Longrightarrow \quad -2\zeta'(x) + \tau((\zeta(x))^2 + 1) = 0 \\ &\Longrightarrow \quad \zeta' = \tfrac{1}{2}\tau(\zeta^2 + 1). \end{aligned}$$

The only solution to this ordinary differential equation is $\zeta(x) = \tan(\frac{1}{2}\tau x + c)$ for some constant $c$, whose domain is clearly not $(M_1, +\infty)$. This is a contradiction. Similarly a root function $\zeta(x)$ of $d_-$ must satisfies $\zeta' = -\frac{1}{2}\tau(\zeta^2 + 1)$, whose solution is $-\tan(\frac{1}{2}\tau x - c)$ for some constant $c$, a contradiction again.

Therefore for $u > M_1$, $g_0(u; y)$ has no common real root with $h_{\pm}(u; y)$. Let $\xi(x)$ be a root function of $g_0$, then $h_{\pm}(x, \xi(x)) \neq 0$. So $2\xi'(x) \pm \tau((\xi(x))^2 + 1)$ is sign-invariant on $(M_1, +\infty)$. If $2\xi'(x) - \tau((\xi(x))^2 + 1) > 0$ for all $x \in (M_1, +\infty)$, then

$$(2 \arctan \xi(x))' = \frac{2\xi'(x)}{(\xi(x))^2 + 1} > \tau,$$

for all $x \in (M_1, +\infty)$. This is absurd, because $2\arctan\xi(x)$ is bounded, while $\int \tau \mathrm{d}x = \tau x + c$ is unbounded. So for all $x > M_1$: $2\xi'(x) - \tau((\xi(x))^2 + 1) < 0$. In other words, $\frac{2\xi'(x)}{(\xi(x))^2+1} < \tau$. By the same argument, we can show $\frac{2\xi'(x)}{(\xi(x))^2+1} > -\tau$. Hence $\left|\frac{2\xi'(x)}{\xi^2(x)+1}\right| < \tau$ for all $x \in (M_1, +\infty)$. □

**Theorem 5.2.5.** *Suppose $g(x, y) \in S[y]$, then there are effective integers $k_-, k_+$ such that for all $k > k_+$ and all $k < k_-$: if $\xi(x)$ is a root function of $g(x, y) = 0$ defined on $[2k\pi - \pi, 2k\pi + \pi]$, then $T_{\xi,k}$ has a unique fixed point $r \in (2k\pi - \pi, 2k\pi + \pi)$.*

**Proof.** Choose any $\tau \in (0, 1)$. By Lemma 5.2.4, there is an integer $k_+ = \lfloor \frac{1}{2\pi} M_1 + \frac{1}{2} \rfloor$ and an integer $k_- = \lceil \frac{1}{2\pi} N_1 - \frac{1}{2} \rceil$ such that for all $k > k_+$ and all $k < k_-$:

$$|T'_{\xi,k}(x)| = \left|\frac{2\xi'(x)}{\xi^2(x) + 1}\right| < \tau < 1.$$

By the Lagrange Mean Value Theorem, we obtain that $\forall x,\ y \in [2k\pi - \pi, 2k\pi + \pi]$:

$$|T_{\xi,k}(x) - T_{\xi,k}(y)| < \tau |x - y|.$$

By the Banach fixed point theorem, $T_{\xi,k}$ admits a unique fixed point $r \in (2k\pi - \pi, 2k\pi + \pi)$ (obviously $r$ cannot be the endpoints) and beginning with any initial value $x_0 \in [2k\pi - \pi, 2k\pi + \pi]$, the iteration $x_n = T_{\xi,k}(x_{n-1})$ converges to $r$. □

Clearly, the fixed point of $T_{\xi,k}$ is the solution of $\xi(x) = \tan\frac{x}{2}$ in $(2k\pi - \pi, 2k\pi + \pi)$. So we have proved that for sufficiently large $k$, the equation $\xi(x) = \tan\frac{x}{2}$ has a unique solution in $(2k\pi - \pi, 2k\pi + \pi)$.

**Corollary 5.2.6.** *Suppose $f_1, \ldots, f_s \in S[\sin x, \cos x]$. There are effective $k_+ \in \mathbb{Z}$, $c_+ \in \mathbb{N}$ such that*

1. *For all $k > k_+$, there are exactly $c_+$ roots of $\prod_{j=1}^{s} f_j = 0$, denoted by $r_{k,1} < \cdots < r_{k,c_+}$, in $(2k\pi - \pi, 2k\pi + \pi)$;*
2. *Additionally let $r_{k,0} = 2k\pi - \pi$ and $r_{k,c_++1} = 2k\pi + \pi$. Each $f_j$ is sign-invariant on $\bigcup_{k>k_+}(r_{k,i}, r_{k,i+1})$ and $\{r_{k,i} | k > k_+\}$ for all $i = 0, \ldots, c_+$.*

*There are also effective $k_- \in \mathbb{Z}$, $c_- \in \mathbb{N}$ such that*

3. *For all $k < k_-$, there are exactly $c_-$ roots of $\prod_{j=1}^{s} f_j = 0$, denoted by $r_{k,1} < \cdots < r_{k,c_-}$, in $(2k\pi - \pi, 2k\pi + \pi)$;*
4. *Additionally let $r_{k,0} = 2k\pi - \pi$ and $r_{k,c_-+1} = 2k\pi + \pi$. Each $f_j$ is sign-invariant on $\bigcup_{k<k_-}(r_{k,i}, r_{k,i+1})$ and $\{r_{k,i} | k < k_-\}$ for all $i = 0, \ldots, c_-$.*

**Proof.** Similarly, it suffices to prove (1) and (2).

Let $g_1, \ldots, g_s \in S[y]$ be polynomials that their image in $S[\tan\frac{x}{2}]$ are numerators of canonical forms of $f_1, \ldots, f_s$. By Lemma 5.2.2, there exists $M_0 \in \mathbb{R}$ such that $g = \prod_{i=1}^{s} g_i$ is delineable on $(M_0, +\infty)$, i.e. $\{g_1, \ldots, g_s\}$ is delineable on $(M_0, +\infty)$. Let $\xi_1 < \cdots < \xi_t$ be the root functions of $g$. For the simplicity of notations, we additionally let $\xi_0(x) = -\infty$ and $\xi_{t+1}(x) = +\infty$. Then by Theorem 5.2.5, there exists $k_+ \in \mathbb{Z}$ such that for all $k > k_+$ and $i \in \{1, \ldots, t\}$: there exists unique $r_{k,i} \in (2k\pi - \pi, 2k\pi + \pi)$ such that $\xi_i(r_{k,i}) = \tan\frac{r_{k,i}}{2}$.

By replacing $k_+$ with a larger integer if necessary, we can further assume that for all $j = 1, \ldots, s$, either for all $k > k_+$: $f_j(r_{k,0}) = 0$, or for all $k > k_+$, $f_j(r_{k,0}) \neq 0$ (recall Lemma 5.1.1).

For the other part, let $\Gamma_i = \{(x, y) \in \mathbb{R}^2 | x > 2k_+\pi + \pi, \xi_i(x) < y < \xi_{i+1}(x)\}$ for $i = 0, \ldots, t$, then $g_j(x, y)$ is sign-invariant on $\Gamma_i$ for all $i, j$. We claim that $x \in (r_{k,i}, r_{k,i+1})$ implies $(x, \tan\frac{x}{2}) \in \Gamma_i$. Because $r_{k,i}$ is the unique solution to $\xi_i(x) = \tan\frac{x}{2}$ in $(2k\pi - \pi, 2k\pi + \pi)$ and $r_{k,i} < x$, we must have $\xi_i(x) < \tan\frac{x}{2}$, otherwise by the intermediate value theorem there is another $r' > x$ such that $\xi_i(r') = \tan\frac{r'}{2}$ (recall that $\lim\limits_{u \to 2k\pi+\pi} \tan\frac{u}{2} = +\infty$). Similarly $\tan\frac{x}{2} < \xi_{i+1}(x)$. So by the definition of $\Gamma_i$, the claim $(x, \tan\frac{x}{2}) \in \Gamma_i$ is true. Therefore $g_j(x, \tan\frac{x}{2})$ has a constant sign on $\bigcup_{k>k_+}(r_{k,i}, r_{k,i+1})$. Since $g_j$ only differs from $f_j$ by a power of $(1 + \tan^2\frac{x}{2})$, $f_j$ is sign-invariant on $\bigcup_{k>k_+}(r_{k,i}, r_{k,i+1})$ too.

It remains to prove each $f_j$ is sign-invariant on $\{r_{k,i} | k > k_+\}$ for all $i = 0, \ldots, t$. We discuss by several cases.

1. $i \neq 0$. We have $\tan\frac{r_{k,i}}{2} = \xi_i(r_{k,i})$. There are two sub-cases.
   (1a) $\xi_i(x)$ is a root function of $g_j$, then $f_j(r_{k,i}) = 0$ for all $k > k_+$.
   (1b) $\xi_i(x)$ is not a root function of $g_j$. By the continuity of $g_j(x, \tan\frac{x}{2})$, $\operatorname{sgn} f_j(r_{k,i}) = \operatorname{sgn} g_j(r_{k,i}, \tan\frac{r_{k,i}}{2})$ is the sign of $g_j$ in $\Gamma_i$.
2. $i = 0$. By the choice of $k_+$, either
   (2a) $f_j(2k\pi - \pi) = 0$ $(\forall k \in \mathbb{Z})$, or
   (2b) $f_j(2k\pi - \pi) \neq 0$ for all integers $k > k_+$ and $\operatorname{sgn} f_j(2k\pi - \pi)$ is the sign of $g(x, y)$ in $\Gamma_0$.

After all, $f_j$ is sign-invariant on $\{r_{k,i}\}_{k>k_+}$. □

Finally, we can present the proof of Theorem 5.2.1.

*Proof of Theorem 5.2.1.* Clearly that $\varphi(x)$ is true for all $x \in \mathbb{R}$ implies that $\varphi(x)$ is true for all $x \in [N, M] \subset \mathbb{R}$. It suffices to prove the converse.

Apply Corollary 5.2.6 and set $N = 2(k_- - 1)\pi - \pi$, $M = 2(k_+ + 1)\pi + \pi$. If $x_0 \notin [N, M]$, then by Corollary 5.2.6, there is some $x_0' \in [N, M]$ such that $\operatorname{sgn}(f_i(x_0')) = \operatorname{sgn}(f_i(x_0))$ for all

$i = 1, \ldots, s$. Therefore $\varphi(x_0)$ is true too. □

*Remark* 8. Actually Theorem 5.2.1 can be more directly proved using Lemma 5.2.2. We sketch the proof here.

Indeed, let $g_1, \ldots .g_s$ be the numerators of canonical forms of $f_1, \ldots, f_s$, and let $h_i \in S[y]$ such that $g_i(x) = h_i(x, \tan \frac{x}{2})$. Then by Lemma 5.2.2, $h = \prod_{i=1}^{s}$ is delineable on $I = (2k_+\pi + \pi, +\infty)$ for some computable integer $k_+$ and $\{h_i\}_{i=1}^{s}$ is delineable on $I$, which defines cylindrically stacked cells over $I$ where each $h_i$ is sign-invariant. We may enlarge $k_+$ by Lemma 5.1.1 to handle the singularity of $\tan \frac{x}{2}$, if necessary.

Now suppose $\Phi(a)$ is false for some $a \in (2k\pi - \pi, 2k\pi + \pi)$ $(k > k_+)$. Consider $\Phi^*$ that is the formula replacing all $f_i(x) \triangleright 0$ with $h_i(x, y) \triangleright 0$, then $\Phi^*(a, \tan \frac{a}{2})$ is false too and $\Phi^*$ is false for all points in the cell $C$ containing $(a, \tan \frac{a}{2})$. Notice that the graph of $y = \tan \frac{x}{2}$ must intersect $C$ in each period $(2k\pi - \pi, 2k\pi + \pi)$ $(k > k_+)$, so there is $a' \in (2k_+\pi + \pi, 2k_+\pi + 3\pi)$ such that $(a', \tan \frac{a'}{2}) \in C$ and $\Phi(a') = \Phi^*(a', \tan \frac{a'}{2}) = \texttt{False}$. In particular, $M$ can be chosen to be $2k_+\pi + 3\pi$. The other number $N$ can be found similarly.

This proof is significantly simpler than the presented one. On the other hand, our argument actually proves a more refined result: the graph of $y = \tan \frac{x}{2}$ can only intersect with each section cell $C$ exactly once in every sufficiently large period. Thus, our algorithm can be modified to a real root isolation algorithm easily because of the precise description of the root distribution of $f \in S[\sin x, \cos x]$.

## 5.3 A Framework for Decision Algorithm

The Reduction Theorem (Theorem 5.2.1) provides a framework for decision algorithm. Whenever there is a decision algorithm for bounded problems in $S[\sin x, \cos x]$, we immediately obtain a decision algorithm for unbounded problems too.

**Theorem 5.3.1** (Equivalent Decidability)**.** *The ring $S[\sin x, \cos x]$ is decidable over the reals if and only if there exists an algorithm $A(\varphi, N, M)$ that computes the truth value of $(\forall x)((N \leq x \leq M) \rightarrow \varphi(x))$, where $\varphi \in \mathcal{L}(S[\sin x, \cos x])$ is quantifier-free and $N, M \in \mathbb{R}$.*

**Proof.** By Theorem 5.2.1, there are effective bounds $N, M \in \mathbb{R}$ such that $(\forall x)\varphi(x)$ is true if and only if $(\forall x)((N \leq x \leq M) \rightarrow \varphi(x))$ is true. Therefore the existence of an algorithm $A(\varphi, N, M)$ that decides $(\forall x)((N \leq x \leq M) \rightarrow \varphi(x))$ implies the existence of an algorithm $A'(\varphi)$ that decides $(\forall x)\varphi(x)$. Since $(\exists x)$ is just $\neg(\forall x)\neg$, we can conclude that $S[\sin x, \cos x]$ is decidable, provided the existence of such an algorithm $A$. The converse is trivial. □

Following the previous discussion, it is time to describe a decision algorithm of $S[\sin x, \cos x]$, see Algorithm 5.1 in the next page.

**Algorithm 5.1:** DecideTrigExt

**Input:** A first-order sentence $(Qx)\Phi(x)$ in the language of $S[\sin x, \cos x]$,
$Q \in \{\exists, \forall\}$

**Output:** The truth value of $(Qx)\Phi(x)$

Let $f_1, \ldots, f_s \in S[\sin x, \cos x]$ be all the ring elements that occur in $\Phi$;

Choose $g_1, \ldots, g_s \in S[y]$ such that $g_i(x, \tan\frac{x}{2})$ is the numerator of the canonical form of $f_i$ for each $i = 1, \ldots, s$;

$g \leftarrow \prod_{i=1}^{s} g_i$;

$l \leftarrow \mathtt{LC}_y(g)$;

**for** $i \leftarrow 0$ **to** $\deg_y(g) - 1$ **do**

  $\Delta_i = \mathtt{sRes}_{y,i}(g, \frac{\partial g}{\partial y})$;

$k \leftarrow \min\{k \in \mathbb{N} | \Delta_k \neq 0\}$, $p \leftarrow$ the $k$-th subresultant polynomial;

$g_0 \leftarrow \mathtt{PseudoDivisionQuotient}(g, p)$ ;

Choose a finite interval $(N_0, M_0) \subset \mathbb{R}$ such that all real roots of $l$ and $\Delta_k$ are contained in it by the computability of $S$;

$h_+ \leftarrow 2\frac{\partial g_0}{\partial x} + \tau\frac{\partial g_0}{\partial y}(y^2+1)$, $h_- \leftarrow 2\frac{\partial g_0}{\partial x} - \tau\frac{\partial g_0}{\partial y}(y^2+1)$;

**for** $i \leftarrow 0$ **to** $\deg_y(g_0)$ **do**

  $\delta_i^+ \leftarrow \mathtt{sRes}_{y,i}(h_+, g)$, $\delta_i^- \leftarrow \mathtt{sRes}_{y,i}(h_-, g)$;

$j_+ \leftarrow \min\{j \in \mathbb{N} | \delta_j^+ \neq 0\}$, $j_- \leftarrow \min\{j \in \mathbb{N} | \delta_j^- \neq 0\}$;

Choose a finite interval $(N_1, M_1) \subset \mathbb{R}$, such that $(N_0, M_0) \subseteq (N_1, M_1)$ and all real roots of $\delta_{j_+}^+$ and $\delta_{j_-}^-$ are contained in it, by the computability of $S$ again;

$I \leftarrow (-\pi, \pi)$ ;

**for** $i \leftarrow 1$ **to** $s$ **do**

  $s_i \leftarrow$ Replace all $\sin x$ and $\cos x$ in $f_i$ by 0 and $-1$;

  **if** $s_i \neq 0$ **then**

    Enlarge interval $I$ to contain all real roots of $s_i$;

$(N, M) \leftarrow (N_1, M_1) \cup I$;

$k_- = \lceil \frac{N}{2\pi} - \frac{1}{2} \rceil - 1$, $k_+ = \lfloor \frac{M}{2\pi} + \frac{1}{2} \rfloor + 1$;

**if** $Q = \forall$ **then**

  **return** $(\forall x)((2k_-\pi - \pi \leq x \leq 2k_+\pi + \pi) \to \Phi(x))$

**else**

  **return** $(\exists x)((2k_-\pi - \pi \leq x \leq 2k_+\pi + \pi) \wedge \Phi(x))$

**Corollary 5.3.2.** *The ring* $\mathbb{R}_{\mathrm{alg}}[x, \sin x, \cos x]$ *is decidable.*

**Proof.** This improves our previous result [CX23, Corollary 4.3]. Set $S = \mathbb{R}_{\mathrm{alg}}[x]$, then obviously $S$ satisfies the requirement in section 5.1. So by Theorem 5.3.1, it suffices to have an algorithm that decides the same problem over a bounded interval. Such an algorithm is provided in [MW12] (their result was stated for $S = \mathbb{Z}[x]$, but it generalizes to $\mathbb{R}_{\mathrm{alg}}$ without any difficulty.) □

Schanuel's Conjecture (SC) is a famous open problem in the transcendental number theory, firstly stated in [Lan66]. Our next decidability results rely on the conjecture.

**Conjecture 1** (Schanuel)**.** Suppose $z_1, \ldots, z_n \in \mathbb{C}$ are linear independent over $\mathbb{Q}$, then the transcendental degree of the field extension $\mathbb{Q}(z_1, \ldots, z_n, \mathrm{e}^{z_1}, \ldots, \mathrm{e}^{z_n})$ over $\mathbb{Q}$ is at least $n$.

Schanuel's Conjecture plays a central role in transcendental number theory, because many facts and conjectures are easy consequences of Schanuel's Conjecture (*e.g.*, Gelfond-Schneider Theorem, Lindemann-Weierstrass Theorem, the conjectural algebraic independence of $e$ and $\pi$, etc). It is widely believed that Schanuel's Conjecture is true and there are many existing results relying on it (*e.g.*, [MW96; Ric97; Str08]). However, a proof for it seems to be far beyond contemporary mathematics' reach [Cho99].

We need two more definitions, introduced by Strzeboński.

**Definition 5.3.3** ([Str12, Definition 1])**.** The set of **exp-log-arctan functions** is the smallest set of partial functions $\mathbb{R} \to \mathbb{R}$ containing $\exp(x)$, $\log(x)$ $\arctan(x)$, the identity function and the constant functions that is closed under addition, multiplication and composition of functions.

**Definition 5.3.4** ([Str09, Definition 1.5])**.** The set of **elementary expressions** with coefficients in a computable field $K \subseteq \mathbb{R}$ is the smallest family containing elements of $K$, the identity function, $\exp(x)$, $\log(x)$, trigonometric functions, hyperbolic functions, inverse trigonometric and inverse hyperbolic functions that is closed under addition, subtraction, multiplication, division and composition.

Let $a.b \in \mathbb{R} \cup \{\pm\infty\}$ $(a < b)$. An elementary expression $f$ is **tame** over the interval $(a, b)$ if every subexpression of $f$ of the form $g(h(x))$, where $g$ is a trigonometric function, there exist $c, d \in \mathbb{R}$ such that $c < h(u) < d$ for all $u \in (a, b) \cap \mathrm{Domain}(f)$. An elementary function $f$ is tame over $(a, b)$, if it can be represented by a tame expression.

**Corollary 5.3.5.** *Assuming Schanuel's Conjecture, if ring $R$ consists of analytic exp-log-arctan functions and $R$ is closed under differentiation, then $R[\sin x, \cos x]$ is decidable.*

**Proof.** If Schanuel's Conjecture is true, then the ring $R$ is computable by [Str12, Algorithm 48]. Notice that a non-zero analytic exp-log-arctan function has only finitely many roots (because of the identity theorem of analytic functions and the locus of an exp-log-arctan function is a finite union of intervals and points). So $R$ meets all the requirements in Section 2.1. Let $(Qx)\varphi(x) \in \mathcal{L}(R[\sin x, \cos x])$ and $f_1, \ldots, f_s \in R[\sin x, \cos x]$ are all ring elements occurring in $\varphi(x)$. By Theorem 5.3.1, the variable $x$ can be restrict to be in a bounded interval $[N, M]$ and it suffices to study tame elementary functions $f_i|_{[N,M]}$. Since [Str09, Algorithm 6.3] provides a real root isolation algorithm for tame elementary functions, this bounded interval can be effectively partitioned into finitely many intervals and points, in which each $f_i$ is sign-invariant. Moreover the sign of $f_i$ can be computed by the sign testing algorithm [Str09, Section 3]. Therefore $R[\sin x, \cos x]$ is decidable. □

**Corollary 5.3.6.** *The ring* $\mathbb{R}_{\text{alg}}[x, \mathrm{e}^{\alpha_1 x}, \ldots, \mathrm{e}^{\alpha_s x}, \sin x, \cos x]$ *is decidable* ($\alpha_1, \ldots, \alpha_s \in \mathbb{R}_{\text{alg}}$), *assuming Schanuel's Conjecture.*

**Proof.** Set $R = \mathbb{R}_{\text{alg}}[x, \mathrm{e}^{\alpha_1 x}, \ldots, \mathrm{e}^{\alpha_s x}]$, everything immediately follows from the previous corollary. □

**Corollary 5.3.7** (Reachability)**.** *Assuming Schanuel's Conjecture, if* $\dot{x}(t) = Ax(t), x(0) = \vec{v}$ *is a linear ODE system (*$A \in \mathbb{Q}^{n\times n}, v \in \mathbb{Q}^n$*), let* $V = \operatorname{Spec}(A)$ *be the eigenvalues of* $A$*. Suppose* $\{\operatorname{Im} z | z \in V\}$ *spans a* 1*-dimensional linear space over* $\mathbb{Q}$*, then the reachability problem*

$$(\exists t)(x(t) \in W),$$

*where* $W$ *is a semi-algebraic set, is decidable.*

**Proof.** Because $\{\operatorname{Im} z | z \in V\}$ spans a 1-dimensional linear space over $\mathbb{Q}$, there exists some $\beta \in \mathbb{R}_{\text{alg}}$ such that for all $\omega \in \{\operatorname{Im} z | z \in V\}$: there exists some $m \in \mathbb{Z}$ such that $\omega = m\beta$. Let $\{\operatorname{Re} z | z \in V\} = \{\alpha_1, \ldots, \alpha_s\}$, then each component of the solution $x(t)$ is an element of $\mathbb{R}_{\text{alg}}[t, \mathrm{e}^{\alpha_1 t}, \ldots, \mathrm{e}^{\alpha_s t}, \sin \beta t, \cos \beta t]$ by the solution formula of linear differential systems [Arn06, Corollary 25.2.3].

By definition, a semi-algebraic set can be described by a quantifier-free formula in the first-order language of the reals using elements of $\mathbb{R}_{\text{alg}}[x_1, \ldots, x_n]$. So $(\exists t)(x(t) \in W)$ can be expressed in the first-order language $\mathcal{L}(\mathbb{R}_{\text{alg}}[t, \mathrm{e}^{\alpha_1 t}, \ldots, \mathrm{e}^{\alpha_s t}, \sin \beta t, \cos \beta t])$. Apply Corollary 5.3.6 after the variable substitution $u = \beta t$ and we see that the reachability problem is decidable. □

*Remark* 9. The reachability problem has been intensively studied in [LPY01; AW01; Gan+15; Gan+16]. Their results cover three cases: (i) $A$ is nilpotent, (ii) $A$ is diagonalizable with real eigenvalues, (iii) $A$ is diagonalizable with purely imaginary eigenvalues. It can be easily seen that our corollary is different from theirs. There is also a related problem in theoretical computer science called "Continuous Skölem-Pisot Problem" [Bel+10], asking for the existence of $t$ that satisfies $q^T e^{At} v = 0$ for some given vector $q$. In other words, the goal is to decide whether the trajectory crosses a certain hyperplane. In this direction, [COW16] obtained similar results. In fact, one of their theorems collides with ours by taking $W$ in Corollary 5.3.7 to be a hyperplane.

*Remark* 10. It should be remarked that, although Corollary 5.3.5-5.3.7 rely on Schanuel's Conjecture, we believe that this does not diminish their importance. On the one hand, there is a need for this type of algorithm in the real world, and any new result in this direction will be helpful (even if it relies on some unproven conjecture). On the other hand, the failure of the algorithm will imply the falsity of Schanuel's Conjecture, which will be very surprising and meaningful.

## 5.4 Multivariate Extension

In this section, we will discuss multivariate transcendental problems over the reals. Before showing the undecidability of the general problem, we will prove that a particular extension of the univariate theory is decidable.

### 5.4.1 A Decidable Fragment

The decidability of rings defined in Section 2.1 can be extended to the multivariate case by simply replacing univariate functions $f(x) : \mathbb{R} \to \mathbb{R}$ with multivariate functions $f(x_1, \ldots, x_n) : \mathbb{R}^n \to \mathbb{R}$.

**Theorem 5.4.1.** *Suppose that $\mathbb{R}_{\text{alg}} \subset S[\sin x, \cos x]$ is decidable. Let $\varphi$ be a quantifier-free formula in $\mathcal{L}(S[\sin x, \cos x, y_1, \ldots, y_n])$. The truth value of*

$$\Phi := (Q_0 x)(Q_1 y_1) \cdots (Q_n y_n) \varphi(x, y_1, \ldots, y_n),$$

*where $Q_i \in \{\exists, \forall\}$ $(0 \le i \le n)$, can be effectively computed.*

**Proof.** Because the first order theory of real closed fields admits quantifier elimination [Tar51], let

$$\Phi'(x) := (Q_1 y_1) \cdots (Q_n y_n) \varphi(x, y_1, \ldots, y_n),$$

there is a quantifier-free formula $\psi(x) \in \mathcal{L}(S[\sin x, \cos x])$ that is equivalent to $\Phi'(x)$. Therefore $\Phi$ is equivalent to $(Q_0 x)\psi(x)$, whose truth value can be effectively computed by the assumption. The proof is completed. □

The underlying idea of this theorem dates back to [AMW08], and it was also used and developed in [Str11], [MW12] and [XLY15]. Notice the particular order of quantifiers in $\Phi$ ensures that the quantifier elimination can be applied. This result can be combined with Theorem 5.3.1 and its corollaries to obtain some decidability results.

### 5.4.2 General Undecidability

The following undecidability result is due to Laczkovich.

**Theorem 5.4.2** ([Lac03, Theorem 1])**.** *Let $\mathcal{S}_1$ be the ring generated by the integers and the expressions $x$, $\sin x^n$ and $\sin(x \cdot \sin x^n)$ ($n = 1, 2, \ldots$). For $f \in \mathcal{S}_1$, each of the following statements is recursively undecidable:* (i) *there exists a real number $x$ such that $f(x) > 0$; and* (ii) *there exists a real number $x$ such that $f(x) = 0$.*

We can prove that even the theory of multivariate polynomials and the sine function is undecidable.

**Theorem 5.4.3.** *There does not exist an algorithm $A(n, \Psi)$ that takes an integer $n > 0$ and a closed formula $\Psi$ in $\mathcal{L}(\mathbb{Z}[x_1, \ldots, x_n, \sin x_1, \ldots, \sin x_n])$ as input and returns the truth value of $\Psi$.*

**Proof.** We will show the non-existence by reducing from $\mathcal{S}_1$. Indeed, suppose $f \in \mathcal{S}_1$, then there is an integer $n > 0$ such that $f \in \mathbb{Z}[x, \sin x, \ldots, \sin x^n, \sin(x \cdot \sin x), \ldots, \sin(x \cdot \sin x^n)]$. So there is an integer-coefficient polynomial $g$ such that

$$f(x) = g(x, \sin x, \ldots, \sin x^n, \sin(x \cdot \sin x), \ldots, \sin(x \cdot \sin x^n)).$$

Now introduce new variables $y_1, \ldots, y_n, z_1, \ldots, z_n$. Obviously the formula $(\exists x)(f(x) = 0)$ is equivalent to

$$\begin{aligned} \Psi := (\exists x)(\exists y_1) \cdots (\exists y_n)(\exists z_1) \cdots (\exists z_n) \\ (\bigwedge_{i=1}^{n} (y_i - x^i = 0) \wedge \bigwedge_{i=1}^{n} (z_i - x \cdot \sin y_i = 0) \wedge \\ (g(x, \sin y_1, \ldots, \sin y_n, \sin z_1, \ldots, \sin z_n) = 0)). \end{aligned}$$

Therefore $A(2n + 1, \Psi)$ shall return the truth value of $(\exists x)(f(x) = 0)$. The existence of algorithm $A$ violates the undecidability of $\mathcal{S}_1$ (Theorem 5.4.2). The proof is completed. □

This theorem shows that our decidability results are not very far from being optimal. While it is still possible to have algorithms for small $n$ (for example, $n = 1$ by Corollary 5.3.2), the general problem is clearly beyond the realm of computability.

## 5.5 Experiments

Algorithm 5.1 is implemented with MATHEMATICA 13 and tested with some examples. After the reduction, the bounded problem is solved by MATHEMATICA built-in command `Reduce`, which relies on [Str09; Str12]. The environment is a PC that runs Windows 10 with an Intel Core i7-11700@2.50GHz (8 Cores, 16 Threads) processor and 16 GB of RAM. Our package `TranscendentalProblems` is available at:

https://github.com/xiaxueqaq/TranscendentalProblems.

We use some examples to illustrate Algorithm 5.1.

**Example 14** (Spring-Mass-Bomb)**.** There is a mass $m$ attached to a spring in the water. The position of the mass is denoted by $x(t)$, where $t$ is the time. According to Hooke's law, the force given by the spring is $F = -kx$, where $k$ is the spring constant. The drag force in the water is proportional to the velocity: $f = -bv$ for some constant $b$. Also we have the gravity $G = -mg$ ($g$ is the gravity of Earth). By Newton's second law of motion, the kinetic equation of the mass-spring system is given by:

$$m\ddot{x} = -kx - b\dot{x} - mg.$$

There is also a bomb $m'$ beneath the mass. The bomb will explode when it hits something. Suppose that $y(t)$ stands for the position of the bomb. By the same argument, we can show that the kinetic equation of the bomb is

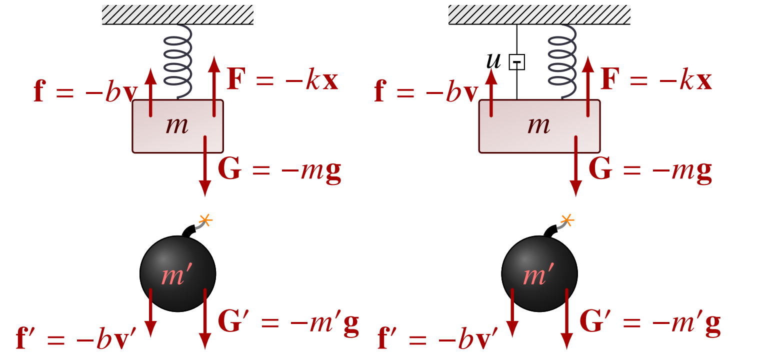
$$m'\ddot{y} = -b\dot{y} - m'g.$$

Suppose $m = m' = 1\text{kg}$, $k = 10\text{N/m}$, $b = 2\text{N}\cdot\text{s/m}$ and $g = 10\text{m/s}^2$. The initial positions are $x(0) = 0\text{m}$, $y(0) = -5\text{m}$, and the initial velocities are $\dot{x}(0) = -12\text{m/s}$ and $\dot{y}(0) = 9\text{m/s}$.

**Question:** Will the mass and the bomb collide at some time $t > 0$?

Plugging the constants into the kinetic equations, we have:

$$\begin{cases} \ddot{x} &= -2\dot{x} - 10x - 10 \\ \dot{x}(0) &= -12 \\ x(0) &= 0 \end{cases}, \quad \begin{cases} \ddot{y} &= -2\dot{y} - 10 \\ \dot{y}(0) &= 9 \\ y(0) &= -5 \end{cases}.$$




(a) The spring-mass system and the bomb

(b) The same system with an extra control

Figure 5.1 Illustration of Examples 14 and 15

Solving the differential equations of $x(t)$ and $y(t)$, we see that

$$x(t) = -\frac{11}{3}\mathrm{e}^{-t}\sin(3t) + \mathrm{e}^{-t}\cos(3t) - 1,\ y(t) = -5t - 7\mathrm{e}^{-2t} + 2.$$

Therefore the collision question is translated to

$$(\exists t)\left((t > 0) \wedge (-\frac{11}{3}\mathrm{e}^{-t}\sin(3t) + \mathrm{e}^{-t}\cos(3t) - 1 = -5t - 7\mathrm{e}^{-2t} + 2)\right).$$

Expanding $\sin(3t), \cos(3t)$ as polynomials in $\sin t$ and $\cos t$, the formula is in the language of $\mathbb{Q}[t, \mathrm{e}^{-t}, \sin t, \cos t]$.

Now set $f_1(t) = t$, $f_2(t) = -5t - 7\mathrm{e}^{-2t} + \frac{11}{3}\mathrm{e}^{-t}\sin(3t) - \mathrm{e}^{-t}\cos(3t) + 3$. Then

$$g_1(t,u) = t, g_2(t,u) = -3\mathrm{e}^{2t}(5t-3)\left(u^2+1\right)^3 - 21\left(u^2+1\right)^3$$
$$+\mathrm{e}^{t}\left(3u^6 + 66u^5 - 45u^4 - 220u^3 + 45u^2 + 66u - 3\right) \in \mathbb{Q}[t, \mathrm{e}^{-t}][u]$$

satisfy that $g_i(t, \tan\frac{t}{2})$ is the numerator of the canonical form of $f_i(t)$ for $i = 1, 2$. Let $g = g_1 \cdot g_2$. The leading coefficient and the discriminant of $g$ are $l = t\left(-3\mathrm{e}^{2t}(5t-3) + 3\mathrm{e}^{t} - 21\right)$ and $\Delta_0 = -826787915366400\mathrm{e}^{4t}t^{10}\left(9\mathrm{e}^{4t}(3-5t)^2 + \mathrm{e}^{2t}(630t - 508) + 441\right)^3 \cdot l$. All real roots of $l$ and $\Delta_0$ are bounded by $[-1, 1]$ (this can be found by [Str12, Algorithm 48]). Therefore $g$ is delineable on $(-\infty, -1)$ and $(+1, +\infty)$. The square-free part $g_0 = g$.

Next, compute $h_\pm = 2\frac{\partial g}{\partial t} \pm \tau\frac{\partial g}{\partial u}(u^2+1)$ (choose $\tau = \frac{15}{16}$). The resultants

$$\delta_\pm = \mathtt{Res}_u(g, h_\pm) = -19680620544000\mathrm{e}^{6t}t^{13}\left(\mathrm{e}^{2t}(5t-3) - \mathrm{e}^{t} + 7\right)\cdot$$
$$\left(\mathrm{e}^{4t}\left(57025t^2 - 55630t + 19249\right) + 2\mathrm{e}^{2t}(61915t - 60734) + 111769\right)^3$$

are not identically zero and all their real roots are also bounded by $[-1, 1]$. So if $\xi(t)$ is a root function of $g$ defined on $(+1, +\infty)$ or $(-\infty, -1)$, then the contraction mapping $T_{\xi,k}$ has a

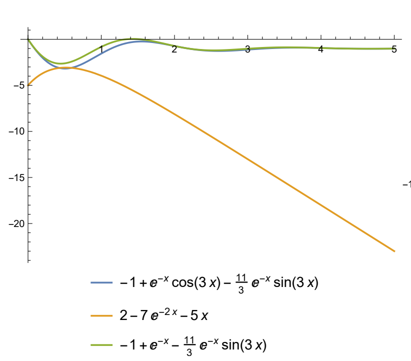



Figure 5.2 Graphs of $x(t)$, $y(t)$ and $z(t)$.

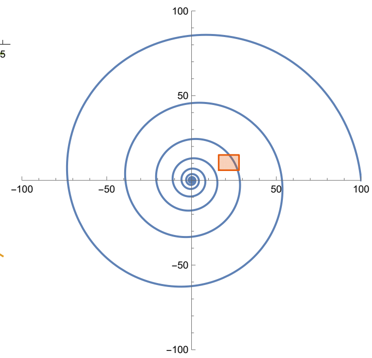


Figure 5.3 The projection of trajectory and the unsafe region to the plane.

unique fixed-point for all non-zero integers $k$.

The singularity at $t = (2k+1)\pi$ ($k \in \mathbb{Z}$) must be treated separately. Notice that $f_2(t) = -7e^{-2t} - 5t - \frac{11}{3}e^{-t}\sin^3(t) - e^{-t}\cos^3(t) + 11e^{-t}\sin(t)\cos^2(t) + 3e^{-t}\sin^2(t)\cos(t) + 3$. Replacing $\sin t, \cos t$ with $0, -1$, the result is $-5t - 7e^{-2t} + e^{-t} + 3$, which is always negative.

Combining all discussions above, we can conclude that it suffices to consider the interval $[-3\pi, 3\pi]$ and the formula

$$(\exists t)((-3\pi \le t \le 3\pi) \wedge (f_1(t) > 0) \wedge (f_2(t) = 0)).$$

In [Str09], the sign testing algorithm and the real root isolation algorithm together give a decision procedure for this formula, which returns `True`. Therefore the mass and the bomb do collide.

**Example 15** (Spring-Mass-Control-Bomb)**.** In this example, the spring-mass system is now equipped with an external control $u$. The force of the control is denoted by $u(t)$. Use $z(t)$ to denote the displacement of the spring-mass-control system. The kinetic equation becomes

$$\ddot{z} = -kz - b\dot{z} - mg + u.$$

Suppose every constant is the same as in Example 14, and the external control is $u(t) = 9e^{-t}$.

**Question:** Does the bomb still hit the mass?

Solving the differential equations, we see that

$$z(t) = e^{-t} - \frac{11}{3}e^{-t}\sin(3t) - 1.$$

And the question becomes

$$(\exists t)\left((t>0)\wedge(\mathrm{e}^{-t}-\frac{11}{3}\mathrm{e}^{-t}\sin(3t)-1=-5t-7\mathrm{e}^{-2t}+2)\right).$$

Run Algorithm 5.1, the result is `False`. In other words, the control force $u(t)$ prevents the mass from hitting the bomb. See also Figure 5.2.

**Example 16.** Show that if $x$ is a real root of the equation $\tanh x=\sin x$, then $\sin^2 x+\cos x>0$. This can be translated as

$$(\forall x)\left((\tanh x=\sin x)\rightarrow(\sin^2 x+\cos x>0)\right).$$

Algorithm 5.1 returns `True`.

**Example 17.** A drone is hovering over a city. The goal of the drone is to land at the origin but it has to avoid some unsafe region $W$. Suppose that the trajectory of the drone is given by

$$\begin{cases} x(t) &= 100\mathrm{e}^{-t/10}\cos t \\ y(t) &= 100\mathrm{e}^{-t/10}\sin t \\ z(t) &= 50-\frac{1}{2}t \end{cases}.$$

And $W$ is a building $\{(x,y,z)|16<x<28, 6<y<15, 0<z<44\}$. Can the drone complete its mission without crashing into $W$?

Plugging $x(t), y(t)$ and $z(t)$ into the defining formula of $W$, we see that entering $W$ is just

$$(\exists t)\left((50-\frac{1}{2}t>0)\wedge(50-\frac{1}{2}t<44)\wedge(100\mathrm{e}^{-\frac{t}{10}}\cos t>16)\wedge\right.$$
$$\left.(100\mathrm{e}^{-\frac{t}{10}}\cos t<28)\wedge(100\mathrm{e}^{-\frac{t}{10}}\sin t>9)\wedge(100\mathrm{e}^{-\frac{t}{10}}\sin t<15)\right).$$

The output of Algorithm 5.1 with this input is `True`. So unfortunately, the drone fails to complete the mission. See also Figure 5.3.

**Example 18.** Show that all the real roots of $\sin x\ln(x^2+x+1)-x=0$ are in $[-2,2]$.

This is just $(\forall x)((\ln\left(x^2+x+1\right)\sin x-x=0)\rightarrow(x^2\leq 4))$, which can be easily verified by Algorithm 5.1.

**Example 19.** Prove that the inequality $\mathrm{e}^{-x^2}\sin x+x^2-x\geq 0$ holds for all $x\in\mathbb{R}$.

Algorithm 5.1 returns `True` for $(\forall x)(\mathrm{e}^{-x^2}\sin x+x^2-x\geq 0)$.

The following examples are taken from [CX23].

**Example 20.** $(\forall x)\left(\left(\left(x^2\cos x-\sin x=0\right)\wedge(x>0)\right)\rightarrow(x-1>0)\right).$

**Example 21.** $(\forall x)(((x-2>0)\wedge(-x^4+(x^4-16)\cos x+2(x^2+4)x\sin x-16>0))\rightarrow(-x+(x+2)\cos x+2>0))$.

**Example 22.** $(\forall x)(\cos x-\sin x+\frac{8}{5}\geq 0)$.

**Example 23.** $(\exists t)((t\cos^2 t)^2+(t\sin t-3)^2-(\frac{6}{5})^2=0)$.

**Example 24.** $(\exists t)((\sin^2 t-\frac{15}{16}=0)\wedge((2\sin t\cos t)^2-\frac{15}{64}=0)\wedge(\sin t>0)\wedge(2\sin t\cos t>0))$.

**Example 25.** $(\forall t)((\sin t\neq 0)\rightarrow((t^2\sin t)^4-(t^2+25)^2(t^2\sin^2 t)^2+(t^2+25)^2(t^2\sin t\cos t)^2\neq 0))$.

We summarize the experiment in Table 5.1. It can be seen that the implementation is quite efficient, all results are returned within a few seconds.

| Examples | 14 | 15 | 16 | 17 | 18 | 19 |
|---|---|---|---|---|---|---|
| Time (s) | 1.531 | 0.969 | 1.797 | 3.562 | 1.891 | 1.438 |
| Examples | 20 | 21 | 22 | 23 | 24 | 25 |
| Time (s) | 0.125 | 0.031 | 0.188 | 0.078 | 8.406 | 2.109 |

Table 5.1 Running Time

# Chapter 6 Summary and Future Work

## 6.1 Summary

This dissertation has systematically revisited cylindrical algebraic decomposition from a geometric perspective, leading to advancements in its theory, algorithms, and applications. The main contributions are summarized as follows:

1. **A Geometric Theory for c.a.d. and a Novel Efficient Algorithm.** We established a fundamental connection between a geometric property (the cardinality of geometric fibers) and a semi-algebraic property (the existence of semi-algebraic continuous sections). This bridged between real and complex algebraic geometry and it provided the theoretical foundation for a new approach. Most importantly, it enabled the systematic exploitation of all available equations during the projection phase for the first time. Based on this theory, we developed a new, simple, and efficient algorithm, with experimental benchmarks confirming its superior performance.
2. **An Effective Criterion for Covering Maps between Real Varieties.** As a deeper investigation inspired by the geometric theory, we proved that for real varieties, a quasi-finite flat morphism with locally constant geometric fibers induces a covering map on the real points. Furthermore, we developed the corresponding algorithms to effectively verify these conditions. This result provides a new and practical tool for studying covering problems in real algebraic geometry, and it is notably applicable even to singular varieties.
3. **Extending the Applicability of c.a.d. to Transcendental Problems.** We identified and formalized a broad class of solvable transcendental problems termed ”Trigonometric Extensions.” By combining the periodicity of trigonometric functions with cylindrical decomposition, we proved that the solution search space for such problems can be reduced from the entire real line to a computable bounded interval. This opens a viable pathway for solving a significant class of problems involving mixed algebraic and transcendental constraints.

## 6.2 Future Work

There are many possible directions for future researches. We list some of them below.

1. **Adapting Improved Projection Operators to the Geometric Context.** There are many improved projection operators besides Hong's [Hon90]. McCallum proposed his simplified projection operator [McC88; McC98] and Brown further improved McCallum's work [Bro01], although their lifting could fail in rare cases (the nullification problem). Lazard came up with another kind of projection operator and he used a slightly different lifting scheme [Laz94]. But he did not provide a proof for an essential claim in his paper. Lazard's c.a.d. construction was only validated recently in [MPP19].
   The key in McCallum's projection is the notion of "order-invariance". He showed that for the co-rank 1 projection from a hypersurface in $R^n$ to $R^{n-1}$, the order of the discriminant (i.e. the Hilbert-Samuel multiplicity of the branch locus) determines the cardinality of geometric fiber. Therefore, McCallum includes only the discriminant in the projection polynomials instead of all subdiscriminants. We conjectured that the stratification in our Geometric c.a.d. algorithm can be coarsened by merging all strata of lower fiber ranks. So only the discriminant locus remains. This is related to an "equi-multiplicity" phenomenon in the singularity theory.

   **Conjecture 2.** Let $\pi : X \to Y$ be a finite flat morphism of reduced $k$-varieties (char $k = 0$). The discriminant $D$ is a locally principal closed subscheme of $Y$. Suppose the Hilbert-Samuel multiplicity $\mathrm{mult}_z(D)$ of $D$ is locally constant along a subscheme $Z$ of $Y$. Further assume that $Z$ and $Y$ are non-singular. Then the geometric fiber cardinality $n(y) = \#X_{\overline{y}}$ is also locally constant along $Z$.

   If this conjecture is true, then one is able to look inside the discriminant to count the geometric fiber. This should significantly accelerate the computation of c.a.d.. Dimca and Rosian showed that for the universal family of monic equations of degree $n$, the geometric fiber cardinality is fully determined by the Hilbert-Samuel multiplicity of the discriminant and the Samuel stratification is Whitney regular [DR84]. Those observations indicate that there might be a connection between the singularity of the discriminant locus and the critical points of the morphism $\pi$. It is possible that an application of projection formula of multiplicities [HIO88, Theorem 6.3] yields the desired result.
2. **Full Quantifier Elimination.** A Quantifier Elimination (QE) algorithm is an algorithm that takes a quantified sentence and returns a quantifier-free sentence that is equivalent to the input. Collins's method is the first practical Quantifier Elimination algorithm for $\mathbb{R}$. It performs an additional step after the construction of c.a.d. to find an equivalent quantifier-free formula.

Strictly speaking, our algorithm here does not directly eliminate quantifiers, because our cells are described by semi-algebraic continuous functions, which are generally not polynomials. However, it does have the ability of describing the regions in an extended language (so-called "Extended Tarski Formula"). In other words, if we allow "root functions" in the language, then our algorithm can give a quantifier elimination procedure.

Collins's original c.a.d. can be refined to give a polynomial-only description of the cells, using more polynomials (`APROJ`, the augmented projection in [Col75]) and Thom's Encoding. A possible line of research is to develop a Geometric c.a.d. theory of quantifier elimination.

3. **Specialized Algorithms.** Although cylindrical algebraic decomposition is a hammer that can crack any problem, there is an expensive cost to wield this hammer. Clearly, a cylindrical algebraic decomposition gives much more than needed if you need just an answer for one problem. Imagine yourself stepping into a neighborhood store intending to buy just a single loaf of bread, only to walk out with a fully loaded shopping cart. This is exactly what cylindrical algebraic decomposition is like. Also, it has been shown that there are doubly-exponential lower bound on worst-case time complexity of c.a.d. [DH88]. Therefore it is crucial to develop special algorithms to solve more specific problems (e.g. Satisfiability, Quantifier Elimination, Real Root Classification, etc).

   A possible way out is to replace the successive cylindrical projections with just one projection. Theorem 4.2.1 provides a solid foundation for this.

# References


[ACM84] Dennis S. Arnon, George E. Collins, and Scott McCallum. "Cylindrical algebraic decomposition. I. The basic algorithm". In: *SIAM J. Comput.* 13.4 (1984), pp. 865–877. ISSN: 0097-5397. DOI: 10.1137/0213054. URL: https://doi.org/10.1137/0213054.

[AM69] M. F. Atiyah and I. G. Macdonald. *Introduction to commutative algebra*. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1969, pp. ix+128.

[AMW08] Melanie Achatz, Scott McCallum, and Volker Weispfenning. "Deciding polynomial-exponential problems". In: *ISSAC 2008—Proceedings of the 2008 International Symposium on Symbolic and Algebraic Computation*. ACM, New York, 2008, pp. 215–222. ISBN: 978-1-59593-904-3. DOI: 10.1145/1390768.1390799. URL: https://doi.org/10.1145/1390768.1390799.

[Arn06] Vladimir I. Arnold. *Ordinary differential equations*. Universitext. Translated from the Russian by Roger Cooke, Second printing of the 1992 edition. Springer-Verlag, Berlin, 2006, pp. ii+334. ISBN: 978-3-540-34563-3; 3-540-34563-9.

[AW00] Hirokazu Anai and Volker Weispfenning. "Deciding linear-trigonometric problems". In: *ISSAC 2000—Proceedings of the 2000 International Symposium on Symbolic and Algebraic Computation (St. Andrews)*. ACM, New York, 2000, pp. 14–22. DOI: 10.1145/345542.345567. URL: https://doi.org/10.1145/345542.345567.

[AW01] Hirokazu Anai and Volker Weispfenning. "Reach set computations using real quantifier elimination". In: 1199. Theory and application in computer algebra (Japanese) (Kyoto, 2000). 2001, pp. 100–113.

[BCR98] Jacek Bochnak, Michel Coste, and Marie-Françoise Roy. *Real algebraic geometry*. Vol. 36. Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]. Translated from the 1987 French original, Revised by the authors. Springer-Verlag, Berlin, 1998, pp. x+430. ISBN: 3-540-64663-9. DOI: 10.1007/978-3-662-03718-8. URL: https://doi.org/10.1007/978-3-662-03718-8.

[BD07] Christopher W. Brown and James H. Davenport. "The complexity of quantifier elimination and cylindrical algebraic decomposition". In: *ISSAC 2007—Proceedings of the 2007 International Symposium on Symbolic and Algebraic Computation*. ACM, New York, 2007, pp. 54–60. ISBN: 978-1-59593-743-8. DOI: 10.1145/1277548.1277557. URL: https://doi.org/10.1145/1277548.1277557.

[Bec91] Eberhard Becker. "Sums of squares and quadratric forms in real algebraic geometry". In: *Cahiers du séminaire d'histoire des mathématiques* 1 (1991), pp. 41–57.

[Bel+10] Paul C. Bell, Jean-Charles Delvenne, Raphaël M. Jungers, and Vincent D. Blondel. "The continuous Skolem-Pisot problem". In: *Theoret. Comput. Sci.* 411.40-42 (2010), pp. 3625–3634. ISSN: 0304-3975,1879-2294. DOI: 10.1016/j.tcs.2010.06.005. URL: https://doi.org/10.1016/j.tcs.2010.06.005.

[BF91] Göran Björck and Ralf Fröberg. "A faster way to count the solutions of inhomogeneous systems of algebraic equations, with applications to cyclic $n$-roots". In: *J. Symbolic Comput.* 12.3 (1991), pp. 329–336. ISSN: 0747-7171,1095-855X. DOI: 10.1016/S0747-7171(08)80153-8. URL: https://doi.org/10.1016/S0747-7171(08)80153-8.

[BFG22] E. Baro, Jose F. Fernando, and J. M. Gamboa. "Spectral maps associated to semialgebraic branched coverings". In: *Rev. Mat. Complut.* 35.1 (2022), pp. 227–264. ISSN: 1139-1138,1988-2807. DOI: 10.1007/s13163-020-00377-5. URL: https://doi.org/10.1007/s13163-020-00377-5.

[BFS15] Magali Bardet, Jean-Charles Faugère, and Bruno Salvy. "On the complexity of the $F_5$ Gröbner basis algorithm". In: *J. Symbolic Comput.* 70 (2015), pp. 49–70. ISSN: 0747-7171,1095-855X. DOI: 10.1016/j.jsc.2014.09.025. URL: https://doi.org/10.1016/j.jsc.2014.09.025.

[BH91] Bruno Buchberger and Hoon Hong. *Speeding-up quantifier elimination by Gröbner bases*. Citeseer, 1991.

[BM05] Christopher W. Brown and Scott McCallum. "On using bi-equational constraints in CAD construction". In: *ISSAC 2005—Proceedings of the 2005 International Symposium on Symbolic and Algebraic Computation*. ACM, New York, 2005, pp. 76–83. ISBN: 1-59593-095-7. DOI: 10.1145/1073884.1073897. URL: https://doi.org/10.1145/1073884.1073897.

[BOKR84] Michael Ben-Or, Dexter Kozen, and John Reif. "The complexity of elementary algebra and geometry". In: *Proceedings of the sixteenth annual ACM symposium on Theory of computing*. 1984, pp. 457–464. DOI: 10.1016/0022-0000(86)90029-2. URL: https://doi.org/10.1016/0022-0000(86)90029-2.

[BPR06] Saugata Basu, Richard Pollack, and Marie-Françoise Roy. *Algorithms in real algebraic geometry*. Second. Vol. 10. Algorithms and Computation in Mathematics. Springer-Verlag, Berlin, 2006, pp. x+662. ISBN: 978-3-540-33098-1; 3-540-33098-4. URL: https://perso.univ-rennes1.fr/marie-francoise.roy/bpr-ed2-posted3.html.

[Bro01] Christopher W. Brown. "Improved projection for cylindrical algebraic decomposition". In: *J. Symbolic Comput.* 32.5 (2001), pp. 447–465. ISSN: 0747-7171,1095-855X. DOI: 10.1006/jsco.2001.0463. URL: https://doi.org/10.1006/jsco.2001.0463.

[Bro03] Christopher W Brown. "QEPCAD B: a program for computing with semi-algebraic sets using CADs". In: *ACM Sigsam Bulletin* 37.4 (2003), pp. 97–108.

[Buc65] Bruno Buchberger. "Ein Algorithmus zum Auffinden der Basiselemente des Restklassenringes nach einem nulldimensionalen Polynomideal". In: *Ph. D. Thesis, Math. Inst., University of Innsbruck* (1965).

[Buc70] B. Buchberger. "Ein algorithmisches Kriterium für die Lösbarkeit eines algebraischen Gleichungssystems". In: *Aequationes Math.* 4 (1970), pp. 374–383. ISSN: 0001-9054,1420-8903. DOI: 10.1007/BF01844169. URL: https://doi.org/10.1007/BF01844169.

[BW94] Eberhard Becker and Thorsten Wörmann. “On the trace formula for quadratic forms”. In: *Contemporary Mathematics* 155 (1994), pp. 271–291. DOI: 10.1090/conm/155/01385. URL: https://doi.org/10.1090/conm/155/01385.

[Cav70] B. F. Caviness. “On canonical forms and simplification”. In: *J. Assoc. Comput. Mach.* 17 (1970), pp. 385–396. ISSN: 0004-5411,1557-735X. DOI: 10.1145/321574.321591. URL: https://doi.org/10.1145/321574.321591.

[Che+09] Changbo Chen, Marc Moreno Maza, Bican Xia, and Lu Yang. “Computing cylindrical algebraic decomposition via triangular decomposition”. In: *ISSAC 2009—Proceedings of the 2009 International Symposium on Symbolic and Algebraic Computation*. ACM, New York, 2009, pp. 95–102. ISBN: 978-1-60558-609-0. DOI: 10.1145/1576702.1576718. URL: https://doi.org/10.1145/1576702.1576718.

[Cho99] Timothy Y. Chow. “What is a closed-form number?” In: *Amer. Math. Monthly* 106.5 (1999), pp. 440–448. ISSN: 0002-9890,1930-0972. DOI: 10.2307/2589148. URL: https://doi.org/10.2307/2589148.

[Chu36] Alonzo Church. “An Unsolvable Problem of Elementary Number Theory”. In: *Amer. J. Math.* 58.2 (1936), pp. 345–363. ISSN: 0002-9327,1080-6377. DOI: 10.2307/2371045. URL: https://doi.org/10.2307/2371045.

[CL20] Shiping Chen and Zhong Liu. “Automated proof of mixed trigonometric-polynomial inequalities”. In: *J. Symbolic Comput.* 101 (2020), pp. 318–329. ISSN: 0747-7171,1095-855X. DOI: 10.1016/j.jsc.2019.10.002. URL: https://doi.org/10.1016/j.jsc.2019.10.002.

[CLO05] David A. Cox, John Little, and Donal O’Shea. *Using algebraic geometry*. Second. Vol. 185. Graduate Texts in Mathematics. Springer-Verlag, New York, 2005, pp. xii+572. ISBN: 0-387-20706-6. DOI: 10.1007/b138611. URL: https://link.springer.com/book/10.1007/b138611.

[CLO15] David A. Cox, John Little, and Donal O’Shea. *Ideals, varieties, and algorithms. An introduction to computational algebraic geometry and commutative algebra*. Fourth. Undergraduate Texts in Mathematics. Springer, Cham, 2015, pp. xvi+646. ISBN: 978-3-319-16720-6; 978-3-319-16721-3. DOI: 10.1007/978-3-319-16721-3. URL: https://doi.org/10.1007/978-3-319-16721-3.

[CMM14a] Changbo Chen and Marc Moreno Maza. “An incremental algorithm for computing cylindrical algebraic decompositions”. In: *Computer Mathematics: 9th Asian Symposium (ASCM2009), Fukuoka, December 2009, 10th Asian Symposium (ASCM2012), Beijing, October 2012, Contributed Papers and Invited Talks*. Springer. 2014, pp. 199–221.

[CMM14b] Changbo Chen and Marc Moreno Maza. “Cylindrical algebraic decomposition in the RegularChains library”. In: *Mathematical Software–ICMS 2014: 4th International Congress, Seoul, South Korea, August 5-9, 2014. Proceedings 4*. Vol. 8592. Lecture Notes in Comput. Sci. Springer. Springer, Heidelberg, 2014, pp. 425–433. ISBN: 978-3-662-44199-

2; 978-3-662-44198-5. DOI: 10.1007/978-3-662-44199-2_65. URL: https://doi.org/10.1007/978-3-662-44199-2_65.

[Col75] George E Collins. "Quantifier elimination for real closed fields by cylindrical algebraic decompostion". In: *Automata Theory and Formal Languages: 2nd GI Conference Kaiserslautern, May 20–23, 1975*. Vol. Vol. 33. Lecture Notes in Comput. Sci. Springer. Springer, Berlin-New York, 1975, pp. 134–183.

[Col98] George E Collins. "Quantifier elimination by cylindrical algebraic decomposition—twenty years of progress". In: *Quantifier elimination and cylindrical algebraic decomposition (Linz, 1993)*. Texts Monogr. Symbol. Comput. Springer, Vienna, 1998, pp. 8–23. ISBN: 3-211-82794-3. DOI: 10.1007/978-3-7091-9459-1_2. URL: https://doi.org/10.1007/978-3-7091-9459-1_2.

[Cou01] Louis Couturat. *La logique de Leibniz d'après des documents inédits*. Vol. 1. Ancienne Librairie Germer Baillière, 1901.

[COW16] Ventsislav Chonev, Joël Ouaknine, and James Worrell. "On the Skolem problem for continuous linear dynamical systems". In: *43rd International Colloquium on Automata, Languages, and Programming*. Vol. 55. LIPIcs. Leibniz Int. Proc. Inform. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2016, Art. No. 100, 13. ISBN: 978-3-95977-013-2.

[Cox21] David A Cox. "Stickelberger and the eigenvalue theorem". In: *Commutative Algebra: Expository Papers Dedicated to David Eisenbud on the Occasion of his 75th Birthday*. Springer, Cham, 2021, pp. 283–298. ISBN: 978-3-030-89693-5; 978-3-030-89694-2. DOI: 10.1007/978-3-030-89694-2_8. URL: https://doi.org/10.1007/978-3-030-89694-2_8.

[CX23] Rizeng Chen and Bican Xia. "Deciding first-order formulas involving univariate mixed trigonometric-polynomials". In: *ISSAC 2023—Proceedings of the 2023 International Symposium on Symbolic and Algebraic Computation*. 2023, pp. 145–154.

[DH88] James H. Davenport and Joos Heintz. "Real quantifier elimination is doubly exponential". In: *J. Symbolic Comput.* 5.1-2 (1988), pp. 29–35. ISSN: 0747-7171,1095-855X. DOI: 10.1016/S0747-7171(88)80004-X. URL: https://doi.org/10.1016/S0747-7171(88)80004-X.

[Dic+91] Alicia Dickenstein, Noaï Fitchas, Marc Giusti, and Carmen Sessa. "The membership problem for unmixed polynomial ideals is solvable in single exponential time". In: vol. 33. 1-3. Applied algebra, algebraic algorithms, and error-correcting codes (Toulouse, 1989). 1991, pp. 73–94. DOI: 10.1016/0166-218X(91)90109-A. URL: https://doi.org/10.1016/0166-218X(91)90109-A.

[DK81a] Hans Delfs and Manfred Knebusch. "Semialgebraic topology over a real closed field. I. Paths and components in the set of rational points of an algebraic variety". In: *Math. Z.* 177.1 (1981), pp. 107–129. ISSN: 0025-5874,1432-1823. DOI: 10.1007/BF01214342. URL: https://doi.org/10.1007/BF01214342.

[DK81b] Hans Delfs and Manfred Knebusch. "Semialgebraic topology over a real closed field. II. Basic theory of semialgebraic spaces". In: *Math. Z.* 178.2 (1981), pp. 175–213. ISSN: 0025-5874,1432-1823. DOI: 10.1007/BF01262039. URL: https://doi.org/10.1007/BF01262039.

[DK84] Hans Delfs and Manfred Knebusch. "An introduction to locally semialgebraic spaces". In: *Rocky Mountain J. Math.* 14.4 (1984). Ordered fields and real algebraic geometry (Boulder, Colo., 1983), pp. 945–963. ISSN: 0035-7596,1945-3795. DOI: 10.1216/RMJ-1984-14-4-945. URL: https://doi.org/10.1216/RMJ-1984-14-4-945.

[DK85] Hans Delfs and Manfred Knebusch. *Locally semialgebraic spaces*. Vol. 1173. Lecture Notes in Mathematics. Springer-Verlag, Berlin, 1985, pp. xvi+329. ISBN: 3-540-16060-4. DOI: 10.1007/BFb0074551. URL: https://doi.org/10.1007/BFb0074551.

[DR84] Alexandru Dimca and Rodica Rosian. "The Samuel stratification of the discriminant is Whitney regular". In: *Geom. Dedicata* 17.2 (1984), pp. 181–184. ISSN: 0046-5755. DOI: 10.1007/BF00151505. URL: https://doi.org/10.1007/BF00151505.

[EBD15] Matthew England, Russell Bradford, and James H. Davenport. "Improving the use of equational constraints in cylindrical algebraic decomposition". In: *ISSAC 2015—Proceedings of the 2015 ACM International Symposium on Symbolic and Algebraic Computation*. ACM, New York, 2015, pp. 165–172. ISBN: 978-1-4503-3435-8.

[EBD20] Matthew England, Russell Bradford, and James H. Davenport. "Cylindrical algebraic decomposition with equational constraints". In: *J. Symbolic Comput.* 100 (2020), pp. 38–71. ISSN: 0747-7171,1095-855X. DOI: 10.1016/j.jsc.2019.07.019. URL: https://doi.org/10.1016/j.jsc.2019.07.019.

[EGA $IV_2$] Alexander Grothendieck and Jean Dieudonné. "Éléments de géométrie algébrique. IV. Étude locale des schémas et des morphismes de schémas. II". In: *Inst. Hautes Études Sci. Publ. Math.* 24 (1965), pp. 5–231. ISSN: 0073-8301,1618-1913. URL: http://www.numdam.org/item/PMIHES_1965__24__5_0/.

[EGA $IV_3$] Alexander Grothendieck and Jean Dieudonné. "Éléments de géométrie algébrique. IV. Étude locale des schémas et des morphismes de schémas. III". In: *Inst. Hautes Études Sci. Publ. Math.* 28 (1966), pp. 5–255. ISSN: 0073-8301,1618-1913. URL: http://www.numdam.org/item/PMIHES_1966__28__5_0/.

[EGA $IV_4$] Alexander Grothendieck and Jean Dieudonné. "Éléments de géométrie algébrique: IV. Étude locale des schémas et des morphismes de schémas, Quatrième partie". In: *Inst. Hautes Études Sci. Publ. Math.* 32 (1967), pp. 5–361. ISSN: 0073-8301,1618-1913. URL: http://www.numdam.org/item/PMIHES_1967__32__5_0/.

[EH00] David Eisenbud and Joe Harris. *The geometry of schemes*. Vol. 197. Graduate Texts in Mathematics. Springer-Verlag, New York, 2000, pp. x+294. ISBN: 0-387-98638-3; 0-387-98637-5.

[Ehr50] Charles Ehresmann. "Les connexions infinitésimales dans un espace fibré différentiable". In: *Colloque de topologie, Bruxelles*. Vol. 29. 1950, pp. 55–75.

[EHV92] David Eisenbud, Craig Huneke, and Wolmer Vasconcelos. "Direct methods for primary decomposition". In: *Invent. Math.* 110.2 (1992), pp. 207–235. ISSN: 0020-9910,1432-1297. DOI: 10.1007/BF01231331. URL: https://doi.org/10.1007/BF01231331.

[Eis95] David Eisenbud. *Commutative algebra. With a view toward algebraic geometry*. Vol. 150. Graduate Texts in Mathematics. Springer-Verlag, New York, 1995, pp. xvi+785. ISBN: 0-387-94268-8; 0-387-94269-6. DOI: 10.1007/978-1-4612-5350-1. URL: https://doi.org/10.1007/978-1-4612-5350-1.

[Fau02] Jean-Charles Faugère. "A new efficient algorithm for computing Gröbner bases without reduction to zero ($F_5$)". In: *ISSAC 2002—Proceedings of the 2002 International Symposium on Symbolic and Algebraic Computation*. ACM, New York, 2002, pp. 75–83. DOI: 10.1145/780506.780516. URL: https://doi.org/10.1145/780506.780516.

[Fau10] Jean-Charles Faugère. "FGb: a library for computing Gröbner bases". In: *International Congress on Mathematical Software*. Springer. 2010, pp. 84–87.

[FIS15] Ryoya Fukasaku, Hidenao Iwane, and Yosuke Sato. "Real quantifier elimination by computation of comprehensive Gröbner systems". In: *ISSAC 2015—Proceedings of the 2015 ACM International Symposium on Symbolic and Algebraic Computation*. ACM, New York, 2015, pp. 173–180. ISBN: 978-1-4503-3435-8.

[Gan+15] Ting Gan, Mingshuai Chen, Liyun Dai, Bican Xia, and Naijun Zhan. "Decidability of the reachability for a family of linear vector fields". In: *Automated Technology for Verification and Analysis: 13th International Symposium, ATVA 2015, Shanghai, China, October 12-15, 2015, Proceedings 13*. Springer. 2015, pp. 482–499.

[Gan+16] Ting Gan, Mingshuai Chen, Yangjia Li, Bican Xia, and Naijun Zhan. "Computing reachable sets of linear vector fields revisited". In: *2016 European Control Conference (ECC)*. IEEE. 2016, pp. 419–426.

[GKZ94] Israel M. Gelfand, Mikhail M. Kapranov, and Andrei V. Zelevinsky. *Discriminants, Resultants, and Multidimensional Determinants*. Mathematics: Theory and Application. Springer Science & Business Media, New York, 1994, pp. x+523. ISBN: 0-8176-3660-9. DOI: 10.1007/978-0-8176-4771-1. URL: https://doi.org/10.1007/978-0-8176-4771-1.

[Göd31] Kurt Gödel. "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I". In: *Monatsh. Math. Phys.* 38.1 (1931), pp. 173–198. ISSN: 1812-8076. DOI: 10.1007/BF01700692. URL: https://doi.org/10.1007/BF01700692.

[GP08] Gert-Martin Greuel and Gerhard Pfister. *A* **Singular** *introduction to commutative algebra*. extended. With contributions by Olaf Bachmann, Christoph Lossen and Hans Schönemann, With 1 CD-ROM (Windows, Macintosh and UNIX). Springer, Berlin, 2008, pp. xx+689. ISBN: 978-3-540-73541-0.

[GW20] Ulrich Görtz and Torsten Wedhorn. *Algebraic geometry I. Schemes—with examples and exercises*. Second. Springer Studium Mathematik—Master. Springer Spektrum, Wiesbaden, [2020] ©2020, pp. vii+625. ISBN: 978-3-658-30732-5; 978-3-658-30733-2. DOI: 10.1007/978-3-658-30733-2. URL: https://doi.org/10.1007/978-3-658-30733-2.

[HA28] D. Hilbert and W. Ackermann. *Grundzüge der theoretischen Logik*. Julius Springer, Berlin, 1928.

[Har77] Robin Hartshorne. *Algebraic geometry*. Vol. No. 52. Graduate Texts in Mathematics. Springer-Verlag, New York-Heidelberg, 1977, pp. xvi+496. ISBN: 0-387-90244-9.

[HIO88] M. Herrmann, S. Ikeda, and U. Orbanz. *Equimultiplicity and blowing up*. An algebraic study, With an appendix by B. Moonen. Springer-Verlag, Berlin, 1988, pp. xviii+629. ISBN: 3-540-15289-X. DOI: 10.1007/978-3-642-61349-4. URL: https://doi.org/10.1007/978-3-642-61349-4.

[Hon90] Hoon Hong. "An improvement of the projection operator in cylindrical algebraic decomposition". In: *ISSAC 1990—Proceedings of the 1990 International Symposium on Symbolic and Algebraic Computation*. 1990, pp. 261–264.

[Hon91] Hoon Hong. *Comparison of several decision algorithms for the existential theory of the reals*. Citeseer, 1991.

[HS14] June Huh and Bernd Sturmfels. "Likelihood geometry". In: *Combinatorial algebraic geometry*. Vol. 2108. Lecture Notes in Math. Springer, Cham, 2014, pp. 63–117. ISBN: 978-3-319-04869-7; 978-3-319-04870-3. DOI: 10.1007/978-3-319-04870-3_3. URL: https://doi.org/10.1007/978-3-319-04870-3_3.

[Hua+16] Zongyan Huang, Matthew England, James H Davenport, and Lawrence C Paulson. "Using machine learning to decide when to precondition cylindrical algebraic decomposition with Groebner bases". In: *2016 18th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*. IEEE. 2016, pp. 45–52.

[Jel99] Zbigniew Jelonek. "Testing sets for properness of polynomial mappings". In: *Math. Ann.* 315.1 (1999), pp. 1–35. ISSN: 0025-5831,1432-1807. DOI: 10.1007/s002080050316. URL: https://doi.org/10.1007/s002080050316.

[Kal97] Michael Kalkbrener. "On the stability of Gröbner bases under specializations". In: *J. Symbolic Comput.* 24.1 (1997), pp. 51–58. ISSN: 0747-7171,1095-855X. DOI: 10.1006/jsco.1997.0113. URL: https://doi.org/10.1006/jsco.1997.0113.

[Kem11] Gregor Kemper. *A course in commutative algebra*. Vol. 256. Graduate Texts in Mathematics. Springer, Heidelberg, 2011, pp. xii+246. ISBN: 978-3-642-03544-9. DOI: 10.1007/978-3-642-03545-6. URL: https://doi.org/10.1007/978-3-642-03545-6.

[KL91] Teresa Krick and Alessandro Logar. "An algorithm for the computation of the radical of an ideal in the ring of polynomials". In: *Applied algebra, algebraic algorithms and error-correcting codes (New Orleans, LA, 1991)*. Vol. 539. Lecture Notes in Comput. Sci. Springer, Berlin, 1991, pp. 195–205. ISBN: 3-540-54522-0. DOI: 10.1007/3-540-54522-0_108. URL: https://doi.org/10.1007/3-540-54522-0_108.

[Kne89] Manfred Knebusch. *Weakly semialgebraic spaces*. Vol. 1367. Lecture Notes in Mathematics. Springer-Verlag, Berlin, 1989, pp. xx+376. ISBN: 3-540-50815-5. DOI: 10.1007/BFb0084987. URL: https://doi.org/10.1007/BFb0084987.

[Lac03] M. Laczkovich. “The removal of $\pi$ from some undecidable problems involving elementary functions”. In: *Proc. Amer. Math. Soc.* 131.7 (2003), pp. 2235–2240. ISSN: 0002-9939,1088-6826. DOI: 10.1090/S0002-9939-02-06753-9. URL: https://doi.org/10.1090/S0002-9939-02-06753-9.

[Lak91] Y. N. Lakshman. “A single exponential bound on the complexity of computing Gröbner bases of zero-dimensional ideals”. In: *Effective methods in algebraic geometry (Castiglioncello, 1990)*. Vol. 94. Progr. Math. Birkhäuser Boston, Boston, MA, 1991, pp. 227–234. ISBN: 0-8176-3546-7.

[Lan66] Serge Lang. *Introduction to transcendental numbers*. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1966, pp. vi+105.

[Lap06] Santiago Laplagne. “An algorithm for the computation of the radical of an ideal”. In: *ISSAC 2006—Proceedings of the 2006 International Symposium on Symbolic and Algebraic Computation*. ACM, New York, 2006, pp. 191–195. ISBN: 1-59593-276-3. DOI: 10.1145/1145768.1145802. URL: https://doi.org/10.1145/1145768.1145802.

[Laz94] Daniel Lazard. “An improved projection for cylindrical algebraic decomposition”. In: *Algebraic geometry and its applications: collections of papers from Shreeram S. Abhyankar’s 60th birthday conference*. Springer, New York, 1994, pp. 467–476. ISBN: 0-387-94176-2.

[Liu02] Qing Liu. *Algebraic geometry and arithmetic curves*. Vol. 6. Oxford Graduate Texts in Mathematics. Translated from the French by Reinie Erné, Oxford Science Publications. Oxford University Press, Oxford, 2002, pp. xvi+576. ISBN: 0-19-850284-2.

[LL91] Y. N. Lakshman and Daniel Lazard. “On the complexity of zero-dimensional algebraic systems”. In: *Effective methods in algebraic geometry (Castiglioncello, 1990)*. Vol. 94. Progr. Math. Birkhäuser Boston, Boston, MA, 1991, pp. 217–225. ISBN: 0-8176-3546-7.

[LPY01] Gerardo Lafferriere, George J. Pappas, and Sergio Yovine. “Symbolic reachability computation for families of linear vector fields”. In: *J. Symbolic Comput.* 32.3 (2001), pp. 231–253. ISSN: 0747-7171,1095-855X. DOI: 10.1006/jsco.2001.0472. URL: https://doi.org/10.1006/jsco.2001.0472.

[LR07] Daniel Lazard and Fabrice Rouillier. “Solving parametric polynomial systems”. In: *J. Symbolic Comput.* 42.6 (2007), pp. 636–667. ISSN: 0747-7171,1095-855X. DOI: 10.1016/j.jsc.2007.01.007. URL: https://doi.org/10.1016/j.jsc.2007.01.007.

[LSED22] Huu Phuoc Le and Mohab Safey El Din. “Solving parametric systems of polynomial equations over the reals through Hermite matrices”. In: *J. Symbolic Comput.* 112 (2022), pp. 25–61. ISSN: 0747-7171,1095-855X. DOI: 10.1016/j.jsc.2021.12.002. URL: https://doi.org/10.1016/j.jsc.2021.12.002.

[LSW19] Dong Lu, Yao Sun, and Dingkang Wang. “A survey on algorithms for computing comprehensive Gröbner systems and comprehensive Gröbner bases”. In: *J. Syst. Sci. Complex.* 32.1 (2019), pp. 234–255. ISSN: 1009-6124,1559-7067. DOI: 10.1007/s11424-019-8357-z. URL: https://doi.org/10.1007/s11424-019-8357-z.

[Mac16] Angus Macintyre. “Turing meets Schanuel”. In: *Ann. Pure Appl. Logic* 167.10 (2016), pp. 901–938. ISSN: 0168-0072,1873-2461. DOI: 10.1016/j.apal.2015.10.003. URL: https://doi.org/10.1016/j.apal.2015.10.003.

[Man20] Frédéric Mangolte. *Real algebraic varieties*. Springer Monographs in Mathematics. Translated from the 2017 French original [3727103] by Catriona Maclean. Springer, Cham, 2020, pp. xviii+444. ISBN: 978-3-030-43104-4; 978-3-030-43103-7. DOI: 10.1007/978-3-030-43104-4. URL: https://doi.org/10.1007/978-3-030-43104-4.

[Mat73] Yu.V. Matijasevich. “On recursive unsolvability of Hilbert’s tenth problem”. In: *Logic, methodology and philosophy of science, IV (Proc. Fourth Internat. Congress, Bucharest, 1971)*. Vol. Vol. 74. Stud. Logic Found. Math. North-Holland, Amsterdam-London, 1973, pp. 89–110.

[Mat89] Hideyuki Matsumura. *Commutative ring theory*. Second. Vol. 8. Cambridge Studies in Advanced Mathematics. Translated from the Japanese by M. Reid. Cambridge University Press, Cambridge, 1989, pp. xiv+320. ISBN: 0-521-36764-6.

[May89] Ernst Mayr. “Membership in polynomial ideals over Q is exponential space complete”. In: *STACS 89 (Paderborn, 1989)*. Vol. 349. Lecture Notes in Comput. Sci. Springer, Berlin, 1989, pp. 400–406. ISBN: 3-540-50840-6. DOI: 10.1007/BFb0029002. URL: https://doi.org/10.1007/BFb0029002.

[MB09] Scott McCallum and Christopher W. Brown. “On delineability of varieties in CAD-based quantifier elimination with two equational constraints”. In: *ISSAC 2009—Proceedings of the 2009 International Symposium on Symbolic and Algebraic Computation*. ACM, New York, 2009, pp. 71–78. ISBN: 978-1-60558-609-0. DOI: 10.1145/1576702.1576715. URL: https://doi.org/10.1145/1576702.1576715.

[McC01] Scott McCallum. “On propagation of equational constraints in CAD-based quantifier elimination”. In: *ISSAC 2001—Proceedings of the 2001 International Symposium on Symbolic and Algebraic Computation*. ACM, New York, 2001, pp. 223–230. DOI: 10.1145/384101.384132. URL: https://doi.org/10.1145/384101.384132.

[McC88] Scott McCallum. “An improved projection operation for cylindrical algebraic decomposition of three-dimensional space”. In: *J. Symbolic Comput.* 5.1-2 (1988), pp. 141–161. ISSN: 0747-7171,1095-855X. DOI: 10.1016/S0747-7171(88)80010-5. URL: https://doi.org/10.1016/S0747-7171(88)80010-5.

[McC98] Scott McCallum. “An improved projection operation for cylindrical algebraic decomposition”. In: *Quantifier Elimination and Cylindrical Algebraic Decomposition*. Texts Monogr. Symbol. Comput. Springer, Vienna, 1998, pp. 242–268. ISBN: 3-211-82794-3.

[McC99] Scott McCallum. “On projection in CAD-based quantifier elimination with equational constraint”. In: *ISSAC 1999—Proceedings of the 1999 International Symposium on Symbolic and Algebraic Computation (Vancouver, BC)*. ACM, New York, 1999, pp. 145–149. DOI: 10.1145/309831.309892. URL: https://doi.org/10.1145/309831.309892.

[Mil80] James S. Milne. *Étale cohomology*. Vol. No. 33. Princeton Mathematical Series. Princeton University Press, Princeton, NJ, 1980, pp. xiii+323. ISBN: 0-691-08238-3.

[MM82] Ernst W. Mayr and Albert R. Meyer. "The complexity of the word problems for commutative semigroups and polynomial ideals". In: *Adv. in Math.* 46.3 (1982), pp. 305–329. ISSN: 0001-8708. DOI: 10.1016/0001-8708(82)90048-2. URL: https://doi.org/10.1016/0001-8708(82)90048-2.

[MMA] Wolfram Research, Inc. *Mathematica, Version 13.3*. Champaign, IL. 2023.

[Möl88] H. Michael Möller. "On the construction of Gröbner bases using syzygies". In: *J. Symbolic Comput.* 6.2-3 (1988). Computational aspects of commutative algebra, pp. 345–359. ISSN: 0747-7171,1095-855X. DOI: 10.1016/S0747-7171(88)80052-X. URL: https://doi.org/10.1016/S0747-7171(88)80052-X.

[Mor11] Guillaume Moroz. "Properness defects of projection and minimal discriminant variety". In: *J. Symbolic Comput.* 46.10 (2011), pp. 1139–1157. ISSN: 0747-7171,1095-855X. DOI: 10.1016/j.jsc.2011.05.013. URL: https://doi.org/10.1016/j.jsc.2011.05.013.

[MPP19] Scott McCallum, Adam Parusiński, and Laurentiu Paunescu. "Validity proof of Lazard's method for CAD construction". In: *J. Symbolic Comput.* 92 (2019), pp. 52–69. ISSN: 0747-7171,1095-855X. DOI: 10.1016/j.jsc.2017.12.002. URL: https://doi.org/10.1016/j.jsc.2017.12.002.

[MR10] Ernst W. Mayr and Stephan Ritscher. "Degree bounds for Gröbner bases of low-dimensional polynomial ideals". In: *ISSAC 2010—Proceedings of the 2010 International Symposium on Symbolic and Algebraic Computation*. New York, NY, USA: Association for Computing Machinery, 2010, pp. 21–27. ISBN: 9781450301503. DOI: 10.1145/1837934.1837945. URL: https://doi.org/10.1145/1837934.1837945.

[MR13] Ernst W. Mayr and Stephan Ritscher. "Dimension-dependent bounds for Gröbner bases of polynomial ideals". In: *J. Symbolic Comput.* 49 (2013), pp. 78–94. ISSN: 0747-7171,1095-855X. DOI: 10.1016/j.jsc.2011.12.018. URL: https://doi.org/10.1016/j.jsc.2011.12.018.

[MW12] Scott McCallum and Volker Weispfenning. "Deciding polynomial-transcendental problems". In: *J. Symbolic Comput.* 47.1 (2012), pp. 16–31. ISSN: 0747-7171,1095-855X. DOI: 10.1016/j.jsc.2011.08.004. URL: https://doi.org/10.1016/j.jsc.2011.08.004.

[MW96] Angus Macintyre and A. J. Wilkie. "On the decidability of the real exponential field". In: *Kreiseliana*. A K Peters, Wellesley, MA, 1996, pp. 441–467. ISBN: 1-56881-061-X.

[Neu05] John von Neumann. *John von Neumann: selected letters*. Vol. 27. History of Mathematics. With a foreword by P. Lax and an introduction by Marina von Neumann Whitman, Edited and with a preface and introductory comments by Miklós Rédei. American Mathematical Society, Providence, RI; London Mathematical Society, London, 2005, pp. xxvi+301. ISBN: 0-8218-3776-1. DOI: 10.1090/hmath/027. URL: https://doi.org/10.1090/hmath/027.

[Ped91] Paul Christian Pedersen. *Counting real zeros*. Ph.D. Thesis. New York University, 1991, p. 184. URL: http://gateway.proquest.com/openurl?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&res_dat=xri:pqdiss&rft_dat=xri:pqdiss:9134683.

[PR17] Daniel Perrucci and Marie-Françoise Roy. “Elementary recursive quantifier elimination based on Thom encoding and sign determination”. In: *Ann. Pure Appl. Logic* 168.8 (2017), pp. 1588–1604. ISSN: 0168-0072,1873-2461. DOI: 10.1016/j.apal.2017.03.001. URL: https://doi.org/10.1016/j.apal.2017.03.001.

[PRS93] Paul Pedersen, Marie-Françoise Roy, and Aviva Szpirglas. “Counting real zeros in the multivariate case”. In: *Computational algebraic geometry*. Vol. 109. Progr. Math. Springer. Birkhäuser Boston, MA, 1993, pp. 203–224. ISBN: 0-8176-3678-1. DOI: 10.1007/978-1-4612-2752-6_15. URL: https://doi.org/10.1007/978-1-4612-2752-6_15.

[Ric68] Daniel Richardson. “Some undecidable problems involving elementary functions of a real variable”. In: *J. Symbolic Logic* 33 (1968), pp. 514–520. ISSN: 0022-4812,1943-5886. DOI: 10.2307/2271358. URL: https://doi.org/10.2307/2271358.

[Ric97] Daniel Richardson. “How to recognize zero”. In: *J. Symbolic Comput.* 24.6 (1997), pp. 627–645. ISSN: 0747-7171,1095-855X. DOI: 10.1006/jsco.1997.0157. URL: https://doi.org/10.1006/jsco.1997.0157.

[RT15] Jose Israel Rodriguez and Xiaoxian Tang. “Data-discriminants of likelihood equations”. In: *ISSAC 2015—Proceedings of the 2015 ACM International Symposium on Symbolic and Algebraic Computation*. ACM, New York, 2015, pp. 307–314. ISBN: 978-1-4503-3435-8.

[RT17] Jose Israel Rodriguez and Xiaoxian Tang. “A probabilistic algorithm for computing data-discriminants of likelihood equations”. In: *J. Symbolic Comput.* 83 (2017), pp. 342–364. ISSN: 0747-7171,1095-855X. DOI: 10.1016/j.jsc.2016.11.017. URL: https://doi.org/10.1016/j.jsc.2016.11.017.

[Sch24] Claus Scheiderer. *A course in real algebraic geometry—positivity and sums of squares*. Vol. 303. Graduate Texts in Mathematics. Springer, Cham, 2024, pp. xviii+404. ISBN: 978-3-031-69212-3; 978-3-031-69213-0. DOI: 10.1007/978-3-031-69213-0. URL: https://doi.org/10.1007/978-3-031-69213-0.

[Sch88] Niels Schwartz. “Open locally semi-algebraic maps”. In: *J. Pure Appl. Algebra* 53.1-2 (1988), pp. 139–169. ISSN: 0022-4049,1873-1376. DOI: 10.1016/0022-4049(88)90017-5. URL: https://doi.org/10.1016/0022-4049(88)90017-5.

[Sch89] Niels Schwartz. *The basic theory of real closed spaces*. Vol. 77. 397. 1989, pp. viii+122. DOI: 10.1090/memo/0397. URL: https://doi.org/10.1090/memo/0397.

[Sch94] Claus Scheiderer. *Real and étale cohomology*. Vol. 1588. Lecture Notes in Mathematics. Springer-Verlag, Berlin, 1994, pp. xxiv+273. ISBN: 3-540-58436-6. DOI: 10.1007/BFb0074269. URL: https://doi.org/10.1007/BFb0074269.

[SEDS04] Mohab Safey El Din and Éric Schost. "Properness defects of projections and computation of at least one point in each connected component of a real algebraic set". In: *Discrete Comput. Geom.* 32.3 (2004), pp. 417–430. ISSN: 0179-5376,1432-0444. DOI: 10.1007/s00454-004-1107-5. URL: https://doi.org/10.1007/s00454-004-1107-5.

[SS06] Akira Suzuki and Josuke Sato. "A simple algorithm to compute comprehensive Gröbner bases using Gröbner bases". In: *ISSAC 2006—Proceedings of the 2006 ACM International Symposium on Symbolic and Algebraic Computation*. ACM, New York, 2006, pp. 326–331. ISBN: 1-59593-276-3. DOI: 10.1145/1145768.1145821. URL: https://doi.org/10.1145/1145768.1145821.

[SS88] Günter Scheja and Uwe Storch. *Lehrbuch der Algebra. Teil 2*. Mathematische Leitfäden. [Mathematical Textbooks]. Unter Einschluss der linearen Algebra. [Including linear algebra]. B. G. Teubner, Stuttgart, 1988, p. 816. ISBN: 3-519-02212-5. DOI: 10.1007/978-3-322-80092-3. URL: https://doi.org/10.1007/978-3-322-80092-3.

[Sta02] Anna Stasica. "An effective description of the Jelonek set". In: *J. Pure Appl. Algebra* 169.2-3 (2002), pp. 321–326. ISSN: 0022-4049,1873-1376. DOI: 10.1016/S0022-4049(01)00063-9. URL: https://doi.org/10.1016/S0022-4049(01)00063-9.

[Stacks] The Stacks project authors. *The Stacks project*. https://stacks.math.columbia.edu. 2026.

[Str08] Adam Strzeboński. "Real root isolation for exp-log functions". In: *ISSAC 2008—Proceedings of the 2008 International Symposium on Symbolic and Algebraic Computation*. ACM, New York, 2008, pp. 303–313. ISBN: 978-1-59593-904-3. DOI: 10.1145/1390768.1390810. URL: https://doi.org/10.1145/1390768.1390810.

[Str09] Adam Strzeboński. "Real root isolation for tame elementary functions". In: *ISSAC 2009—Proceedings of the 2009 International Symposium on Symbolic and Algebraic Computation*. ACM, New York, 2009, pp. 341–350. ISBN: 978-1-60558-609-0. DOI: 10.1145/1576702.1576749. URL: https://doi.org/10.1145/1576702.1576749.

[Str11] Adam Strzeboński. "Cylindrical decomposition for systems transcendental in the first variable". In: *J. Symbolic Comput.* 46.11 (2011), pp. 1284–1290. ISSN: 0747-7171,1095-855X. DOI: 10.1016/j.jsc.2011.08.009. URL: https://doi.org/10.1016/j.jsc.2011.08.009.

[Str12] Adam Strzeboński. "Real root isolation for exp-log-arctan functions". In: *J. Symbolic Comput.* 47.3 (2012), pp. 282–314. ISSN: 0747-7171,1095-855X. DOI: 10.1016/j.jsc.2011.11.004. URL: https://doi.org/10.1016/j.jsc.2011.11.004.

[Tar51] Alfred Tarski. *A decision method for elementary algebra and geometry*. 2nd ed. University of California Press, Berkeley-Los Angeles, Calif., 1951, pp. iii+63. ISBN: 9780598987778.

[Tur36] A. M. Turing. "On Computable Numbers, with an Application to the Entscheidungsproblem". In: *Proc. London Math. Soc. (2)* 42.3 (1936), pp. 230–265. ISSN: 0024-6115. DOI: 10.1112/plms/s2-42.1.230. URL: https://doi.org/10.1112/plms/s2-42.1.230.

[Vas97] Wolmer V. Vasconcelos. "Flatness testing and torsionfree morphisms". In: *J. Pure Appl. Algebra* 122.3 (1997), pp. 313–321. ISSN: 0022-4049,1873-1376. DOI: 10.1016/S0022-4049(97)00062-5. URL: https://doi.org/10.1016/S0022-4049(97)00062-5.

[Wan74] Paul S. Wang. "The undecidability of the existence of zeros of real elementary functions". In: *J. Assoc. Comput. Mach.* 21 (1974), pp. 586–589. ISSN: 0004-5411,1557-735X. DOI: 10.1145/321850.321856. URL: https://doi.org/10.1145/321850.321856.

[WBD12] David Wilson, Russell Bradford, and James H. Davenport. "Speeding up cylindrical algebraic decomposition by Gröbner bases". In: *International Conference on Intelligent Computer Mathematics*. Springer. 2012, pp. 280–294.

[WDM20] Dongming Wang, Rina Dong, and Chenqi Mou. "Decomposition of polynomial sets into characteristic pairs". In: *Math. Comp.* 89.324 (2020), pp. 1993–2015. ISSN: 0025-5718,1088-6842. DOI: 10.1090/mcom/3504. URL: https://doi.org/10.1090/mcom/3504.

[Wei88] Volker Weispfenning. "The complexity of linear problems in fields". In: *J. Symbolic Comput.* 5.1-2 (1988), pp. 3–27. ISSN: 0747-7171,1095-855X. DOI: 10.1016/S0747-7171(88)80003-8. URL: https://doi.org/10.1016/S0747-7171(88)80003-8.

[Wei92] Volker Weispfenning. "Comprehensive Gröbner bases". In: *J. Symbolic Comput.* 14.1 (1992), pp. 1–29. ISSN: 0747-7171,1095-855X. DOI: 10.1016/0747-7171(92)90023-W. URL: https://doi.org/10.1016/0747-7171(92)90023-W.

[Wei98] Volker Weispfenning. "A new approach to quantifier elimination for real algebra". In: *Quantifier Elimination and Cylindrical Algebraic Decomposition*. Texts Monogr. Symbol. Comput. Springer, Vienna, 1998, pp. 376–392. ISBN: 3-211-82794-3.

[XLY15] Ming Xu, Zhi-Bin Li, and Lu Yang. "Quantifier elimination for a class of exponential polynomial formulas". In: *J. Symbolic Comput.* 68 (2015), pp. 146–168. ISSN: 0747-7171,1095-855X. DOI: 10.1016/j.jsc.2014.09.015. URL: https://doi.org/10.1016/j.jsc.2014.09.015.

# Publications and Other Achievements

## Published Papers

Chen, R. (2026). A geometric approach to cylindrical algebraic decomposition. *Mathematics of Computation*, 95(360), 2025-2059.
DOI: 10.1090/mcom/4099

Chen, R. (2025). The geometry of cylindrical algebraic decomposition, *ACM Communications in Computer Algebra* 58(3), 67-71.
DOI: 10.1145/3717582.3717586

Chen, R., and Xia, B. (2024, July). Reduction of transcendental decision problems over the reals. In *Proceedings of the 2024 International Symposium on Symbolic and Algebraic Computation* (pp. 56-64).
DOI: 10.1145/3666000.3669675

Chen, R., Li, H., Xia, B., Zhao, T., and Zheng, T. (2024). Isolating all the real roots of a mixed trigonometric-polynomial. *Journal of Symbolic Computation*, 121, 102250.
DOI: 10.1016/j.jsc.2023.102250

Chen, R., and Xia, B. (2023, July). Deciding first-order formulas involving univariate mixed trigonometric-polynomials. In *Proceedings of the 2023 International Symposium on Symbolic and Algebraic Computation* (pp. 145-154).
DOI: 10.1145/3597066.3597104

## Preprints

Chen, R. (2026). An Effective Criterion for Covering Maps Between Real Varieties. *Manuscript*.
arXiv:2602.21708

Chen, R., Hong, H., and Yang, J. (2025). Stratifying Discriminant Hypersurface. *Manuscript*.
arXiv:2509.25820

Chen, R. (2025). What Kind of Morphisms Induces Covering Maps over a Real Closed Field?. *Manuscript*.
arXiv:2502.05834

## Projects

- National Key R&D Program of China (No. 2024YFA1014003) (2024/12-2027/11)
- National Key R&D Program of China (No. 2022YFA1005102) (2022/12-2027/11)
- National Natural Science Foundation of China (No. 61732001) (2018/01-2022/12)

## Talks

- *"A geometric approach to cylindrical algebraic decomposition"*, KLMM seminar, Chinese Academy of Sciences, Beijing, China. April 2026
- *"An effective criterion for covering maps between real algebraic varieties"*, CM2026, Jiangsu University, Zhenjiang, China. April 2026
- *"A geometric approach to cylindrical algebraic decomposition"*, PKU Applied Math Lunch Seminar, Peking University, Beijing, China. May 2025
- *"A geometric approach to cylindrical algebraic decomposition"*, NCSU Symbolic Computation Webinar, North Carolina State University (Virtual). April 2025
- *"Reduction of Transcendental Decision Problems over the Reals"*, ISSAC '24, North Carolina State University, Raleigh, NC, USA. July 2024
- *"The geometry of cylindrical algebraic decomposition"*, CM2024, Fujian Normal University, Fuzhou, China. June 2024
- *"Deciding first-order formulas involving univariate mixed trigonometric-polynomials"*, ISSAC '23, The Arctic University of Norway, Tromsø, Norway. July 2023

## Awards

- 2025-2026, Peking University School of Mathematical Sciences Departmental PhD Scholarship
- 2024-2025, The Third Prize of Peking University Scholarship
- 2024-2025, Award for Scientific Research, Peking University
- 2024-2025, Peking University Presidential Scholarship
- 2024, The Second Prize of Excellent Poster Presentation, PKU-SMS
- 2023-2024, The Third Prize of Peking University Scholarship
- 2023-2024, Award for Scientific Research, Peking University
- 2023-2024, Peking Univeristy PhD Presidential Scholarship
- 2023, The Second Prize of Excellent Poster Presentation, PKU-SMS
- 2022-2023, Peking University School of Mathematical Sciences Departmental PhD Scholarship
- 2021-2022, Peking University PhD Scholarship