ABSTRACT

HARRIS, KATHERINE ELIZABETH. Real Algebraic Geometry with Numerical Homotopy Methods. (Under the direction of Agnes Szanto.)

Many algorithms for determining properties of (real) semi-algebraic sets rely upon the ability to compute smooth points. We present a simple procedure based on computing the critical points of some well-chosen function that guarantees the computation of smooth points in each connected bounded component of an atomic semi-algebraic set. We demonstrate that our technique is intuitive in principal, as it uses well-known tools like the Extreme Value Theorem and optimization via Lagrange multipliers. It also performs well on previously difficult examples, such as the classically difficult singular cusps of a “Thom’s lips" curve. Most importantly, the practical efficiency of this new technique is demonstrated by solving a conjecture on the number of equilibria of the Kuramoto model for the $n = 4$ case.

In the presentation of our approach, we utilize definitions and notation from numerical algebraic geometry. Although our algorithms can also be implemented entirely symbolically, the worst case complexity bounds do not improve on current state of the art methods. Since the usefulness of our technique lies in its computational effectiveness, such as in the examples we described above, it is therefore intuitive for us to present our results under the computational framework which we use in the implementation of the algorithms via existing numerical algebraic geometry software. We also present an approach for finding our “well-chosen" function as described above via isosingular deflation, a known numerical algebraic geometry technique.

In order to apply our approach to non-equidimensional algebraic sets, we perturb using infinitesimals and compute on these perturbed sets. We state and prove results which allow us to compute the limits of these perturbations, including shifting from symbolic infinitesimal perturbations to small constant perturbations which we can track limits of using numerical homotopy techniques. One of the main tools we use throughout our results is polar varieties, based on previous results ([6]) which guarantee the finding of real smooth points (if they exist) under the correct genericity conditions. We also apply our method to design an efficient algorithm to compute the real dimension of algebraic sets, the original motivation for this research.

The use of numerical techniques for algebraic problems raises the question of certification of the computational results. The last chapter is devoted to some results on this problem of certifying solutions, i.e. providing a “certificate" that the approximations resulting from numerical homotopy methods in fact correspond to symbolic solutions. Building on existing work and software `alphaCertified` [26] that work only to certify approximate isolated solutions of well-constrained polynomial systems, we present extensions of these results to underdetermined polynomial systems, with the intention of utilizing them to create a similar software for this case.

Real Algebraic Geometry with Numerical Homotopy Methods

by
Katherine Elizabeth Harris

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Mathematics

Raleigh, North Carolina

2021

APPROVED BY:

Jonathan Hauenstein External Member	Hoon Hong
Ryan Martin	Cynthia Vinzant

Agnes Szanto
Chair of Advisory Committee

DEDICATION

For my parents, Anthony and Denise, who raised me to become the woman I am today, and my sister, Emily, who has been beside me every step of the way.

BIOGRAPHY

Katherine Elizabeth Harris grew up in Derry, New Hampshire, where she attended high school at Pinkerton Academy. There she completed an honors diploma and graduated fifth in her class in May 2011, along with earning honors as a Presidential Scholar and a National Merit Scholar. In addition to her work in the classroom, she played the saxophone in marching band, wind ensemble, and jazz ensemble. Katherine also played four years of varsity basketball and soccer and earned first team all-state honors in both during her senior year.

Katherine attended Bucknell University in Lewisburg, Pennsylvania for her undergraduate studies, graduating Magna Cum Laude while double majoring in Mathematics and German Studies in May 2015. While attending, she completed an honors thesis in the mathematics department under the direction of Nathan Ryan. She was also a member of the Division I women's basketball team and a sister in the Chi Omega women's fraternity Chi Mu chapter.

Katherine began her graduate work at North Carolina State University in August 2015. In addition to conducting research and teaching, she participated in professional development programs to further develop her academic skills, including a 100 hour Teaching and Communication Certificate and a year-long Preparing the Professoriate cohort. She also helped organize education and outreach events as a member and officer of the Association of Women in Mathematics chapter at the school.

Katherine will continue her academic career as an Assistant Professor in the Mathematics and Computer Science Department at Beloit College in Wisconsin in the fall of 2021. She is excited to continue to form personal connections with students and work to make mathematics more accessible.

ACKNOWLEDGEMENTS

The successful completion of my dissertation would not have been possible without the guidance and support of my professional mentors, friends, and family.

I am eternally grateful to my advisor, Agnes Szanto, who I began working with during my third year of graduate school. She started mentoring me when I was lost both professionally and personally, and exhibited great patience and understanding in helping me grow. Agnes modeled the traits of a successful academic and shared her personal experiences to help me believe I could succeed as well. She funded my research and travel to conferences to help me build my professional network and encouraged me to pursue professional development opportunities for my teaching that I was passionate about. Both in my job search and during the writing of my dissertation, Agnes has consistently gone above and beyond in reviewing and revising my written work and presentations to make them better. I can't say thank you enough, Agnes, and I look forward to continuing to have a close relationship in the future.

Along similar lines, Jonathan Hauenstein has acted as an amazing mentor and collaborator in my graduate work. Jon immediately embraced me as a collaborator and constantly reminded me that I was an equal in our work, even when I felt like I was struggling to keep up. He guided and encouraged me to pursue opportunities which would benefit me professionally and helped teach me some intricacies of the academic world. Thank you so much for believing in me even when I didn't, Jon.

Many other faculty and staff members at NC State have been crucial to my success and also put bright spots in my graduate career. David Papp served as an incredible teaching mentor in the Preparing the Professoriate program, and went above and beyond in continuing to mentor me in any way I asked. Vanessa Doriott Anderson provided a seemingly infinite amount of guidance in pedadogy, communication skills, preparing job documents, and general academic success. My committee members, Hoon Hong, Ryan Martin, and Cynthia Vinzant, all provided insight and thoughtful feedback that has improved my work for the years in which they have been there for me. I firmly believe I would not have gotten my future job or finished my dissertation without each and every one of you, thank you all so much.

My time at Bucknell was also filled with mentors who encouraged me to pursue a career in academia when I hadn't even considered the idea. Every professor in the math department provided guidance and knowledge at some point during my four years there. In particular, I am grateful for Nathan Ryan, who introduced me to the process of academic research and graduate school. Thank you for your patience and understanding as I juggled my collegiate athletic career and my future math goals.

One of the biggest joys of graduate school is the friendships made along the way. Jane and Katie took me into their combinatorics homework group in our first year when I was scared to admit I was wrong and ask for help. In the following years, you became not just study buddies, but confidants, shoulders to cry on, cheerleaders, running companions, roommates... the list is countably infinite.

If the only outcome of my graduate experience was meeting you two, it would have been well worth it. Thank you for being there every step of the way.

To my non-academic friends, thank you for putting up with the chaos and uncertainty of my graduate career. Karen, you have never left my side no matter how far I have drifted from myself. Our weekly calls were truly crucial to my sanity. Val, you have been a wealth of knowledge and love since I met you, and living with you has been a joy. And Tara, you know more than anyone the bottom you helped drag me out of. Thank you for always reminding me to be nice to your friend Katherine.

Finally, the unconditional support of my family has been the most important constant in my life. Emily, you are my best friend and biggest cheerleader imaginable. I have always worked to be a good role model for you, and it turns out the more we grow older that you are the best role model for me. Mom, you are the most caring and dedicated educator I know, and our discussions about our work are a joy to me. You have always known what I can do even I when I have doubted myself, and I am so grateful for the solid foundation you provide in my life. And Dad, you have been the biggest fan of all of my pursuits since day one. Your little messages always come at the right time, and I never question your support for my endeavors. I love you all.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER

1

INTRODUCTION

When studying an atomic semi-algebraic set S, i.e.

$$S := \{\mathbf{x} \in \mathbb{R}^n : f_1(\mathbf{x}) = \cdots = f_s(\mathbf{x}) = 0, q_1(\mathbf{x}) > 0, \ldots, q_t(\mathbf{x}) > 0\}$$

for $f_1, \ldots, f_s, q_1, \ldots, q_t \in \mathbb{R}[\mathbf{x}]$, one often first studies the complex variety $V = \{\mathbf{x} \in \mathbb{C}^n : f_1(\mathbf{x}) = \cdots = f_s(\mathbf{x}) = 0\}$, known as its *algebraic relaxation* and deduces properties of S from the properties of V. In particular, if S contains a smooth point and V is irreducible, then S is Zariski dense in V, and so all of the algebraic information of S can be determined from V. Thus, deciding the existence of smooth points in semi-algebraic sets and finding such points is a central problem in real algebraic geometry with many applications. For example, if $\varphi : S \to S'$ is a polynomial map of semi-algebraic sets, then smooth points in $\mathrm{Im}(\varphi)$ are points where the Jacobian of φ has maximal rank within its connected component, called the *typical rank*. Finding real smooth points in each connected component of a semi-algebraic set allows one to compute all typical ranks of real morphisms, an idea which is further explored in [53]. As another example, the existence or non-existence of a real smooth point in an irreducible variety is a well-known criterion for whether the corresponding ideal is real (see the definition of real ideals and the exploitation of this fact in [49]). Similarly, the existence or non-existence of a real smooth point determines where the real dimension of an algebraic set is equal to or less than the complex dimension.

Keeping this general context in mind, this dissertation is organized as follows. In Chapter 2, we present all of the necessary definitions and prior results for the rest of the paper. This includes preliminaries from classical algebraic geometry, real algebraic geometry, and numerical algebraic geometry, as we will utilize the knowledge and context of all three paradigms in our various results.

In Chapter 3, we introduce a new technique to compute smooth points on bounded connected components of atomic semi-algebraic sets. When V is equidimensional, our approach is simple and suggests a natural implementation using numerical homotopy methods. It complements other approaches that compute sample points on real semi-algebraic sets, such as computing the critical points of the distance function, in the sense that our method also guarantees the smoothness of the sample points. We illustrate this advantage on "Thom's lips" in which critical points of the distance function are often at the singularities [57, Ex. 2.3], while our method always computes smooth points. To demonstrate the practical efficiency of our new approach, we present the solution to a conjecture concerning the number of equilibria of the Kuramoto model in the $n = 4$ case given in [58].

Chapter 4 consists of the intermediate results necessary for us to extend our approach from Chapter 3 to the case when V is not equidimensional (i.e. reducible and the components may have different dimensions) by using *infinitesimal* deformations of V and limits. We show that this limiting approach is well-suited for numerical homotopy continuation methods after we translate an infinitesimal real deformation (that may only work for arbitrary small values) into a complex deformation that works along a real arc parameterized by the interval $(0, 1]$. We present a novel technique to construct the "well-chosen" polynomial g, whose critical points are used to compute the smooth points of interest, using deflations, and compare its degree bounds to traditional symbolic approaches (see Proposition 4.4.2).

In Chapter 5, we present our main algorithm for computing a real smooth point on every bounded connected component of an atomic semi-algebraic set that contains one. The result is presented using computational tools from *numerical algebraic geometry* (c.f. [52, 11]), though we note all procedures can be translated to symbolic methods for polynomials with rational coefficients. In particular, Corollary 5.1.4 proves that our REAL SMOOTH POINT ALGORITHM performs well if the depth of the deflations (i.e. the number of iterations) is small.

Chapter 6 applies our method to compute the dimension of real semi-algebraic sets. The difficulty of this problem, compared to its complex counterpart, is that in many cases the real part lies within the singular set of the complex variety containing it, and its real dimension is smaller than the complex one. In terms of worst case complexity bounds of the existing algorithms in the literature, it is an open problem if the real dimension can be computed within the same asymptotic complexity bounds as the complex dimension. The motivation for this research was to try to find an algorithm for the real dimension that has worst case complexity comparable to its complex counterpart. In fact, we did a worst case complexity estimate for a symbolic version, and found that unfortunately it does not improve the existing complexity bounds in the worst case (see [7] and the references therein). This is one of the reasons we wrote the results in a numerical algebraic geometry setting, and give evidence of the efficiency on benchmark problems based on computational experimentation.

Finally, in Chapter 7 we shift our focus to another project involving numerical homotopy methods. Numerical algebraic geometry software that implements algorithms solving systems of polynomial equations often relies heavily on iterations of Newton's method to to compute numerical approximations to symbolic solutions. Shub and Smale's α-theory [50] creates a framework for certifying

the convergence of approximate solutions obtained via Newton's method to the exact solutions of a system. In [26], Hauenstein and Sottile present their software called `alphaCertified` which implements these results for square polynomial systems. In ongoing work, we are extending key results from this context to the case of underdetermined systems, which requires the use of the Moore-Penrose inverse. The implementation of these algorithms will be a complementary software called `alphaCertifiedPlus`.

CHAPTER

2

PRELIMINARIES

2.1 Algebraic Geometry Basics

We begin with establishing some basic algebraic geometry terminology and notation which will help us in our discussion going forward. Our treatment follows that of [23] which can be referenced for more details if necessary.

Definition 2.1.1. A field $\mathbb{K}$ is *algebraically closed* if every non-constant polynomial in $\mathbb{K}[x]$ has a root in $\mathbb{K}$.

We note that in general all of the definitions which follow are often given for any algebraically closed field $\mathbb{K}$; for our purposes, it will suffice to use the algebraically closed field of complex numbers $\mathbb{C}$. We denote by $\mathbb{C}^n$ the n-dimensional affine space over $\mathbb{C}$ and $\mathbb{C}[\mathbf{x}] = \mathbb{C}[x_1, \ldots, x_n]$ the polynomial ring

$$\mathbb{C}[\mathbf{x}] := \left\{ \sum c_\alpha \mathbf{x}^\alpha = \sum c_{(\alpha_1,\ldots,\alpha_n)} x_1^{\alpha_1} \cdots x_n^{\alpha_n} | c_\alpha \in \mathbb{C} \right\}.$$

Definition 2.1.2. Let $f_1, \ldots, f_s \in \mathbb{C}[\mathbf{x}]$. Then

$$\mathcal{V}(f_1, \ldots, f_s) := \{\mathbf{x} \in \mathbb{C}^n : f_1(\mathbf{x}) = \cdots = f_s(\mathbf{x}) = 0\} \tag{2.1}$$

is the set of common zeros of the polynomial system $f_1, \ldots, f_s$ in $\mathbb{C}^n$. We call a set of this form an *algebraic set/affine variety*.

We note some important results about algebraic sets as follows.

Theorem 2.1.3. *The intersection of any collection of algebraic sets in is an algebraic set. The union of a finite collection of algebraic is an algebraic set.*

Theorem 2.1.4. *Every algebraic set can be defined as the common zero set of a finite number of polynomials.*

Furthermore, there exists a correspondence between affine varieties and polynomial ideals which allows us to further explore the properties of both. Given a subset $Z \subset \mathbb{C}^n$, we denote the collection of polynomials that vanish on Z by

$$I(Z) := \{f \in \mathbb{C}[\mathbf{x}] : f(\mathbf{z}) = 0 \text{ for all } \mathbf{z} \in Z\}.$$

We note that by definition the map I reverses inclusions, so $Z \subset Y$ implies that $I(Y) \subset I(Z)$. Furthermore, for any subset $Z \subset \mathbb{C}^n$, $I(Z)$ is an *ideal* of the ring $\mathbb{C}[\mathbf{x}]$.

Lemma 2.1.5. *Given $Z \subset \mathbb{C}^n$, let $X = \mathcal{V}(I(Z))$ be the variety defined by the ideal $I(Z)$. Then $I(Z) = I(X)$ and X is the smallest affine variety in $\mathbb{C}^n$ containing Z.*

Using this result, we introduce the algebro-geometric idea of the closure of a set.

Definition 2.1.6. Given $Z \subset \mathbb{C}^n$, its *closure*, $\overline{Z}$, is the smallest algebraic set in $\mathbb{C}^n$ containing Z. By Lemma 2.1.5,

$$\overline{Z} := \{\mathbf{x} \in \mathbb{C}^n : f(\mathbf{x}) = 0 \text{ for all } f \in I(Z)\} = \mathcal{V}(I(Z)).$$

$Z \subset \mathbb{C}^n$ is *closed* if $Z = \overline{Z}$. $X \subset \mathbb{C}^n$ is *open* if its complement $\mathbb{C}^n \setminus X$ is closed.

By Theorem 2.1.3, the closed subsets of $\mathbb{C}^n$ as given by Definition 2.1.6 satisfy the standard axioms of a topological space. We call this topology the *Zariski topology*, for Oscar Zariski, and note that it also induces the same topology on any subset $Z \subset \mathbb{C}^n$. In particular, given an algebraic set $V \subset \mathbb{C}^n$, closed subsets of V are exactly the closed subsets of $\mathbb{C}^n$ contained in V.

Using this topology, we establish some further results which will help us going forward.

Theorem 2.1.7. *Given an algebraic set $V \subset \mathbb{C}^{n+m}$ and the projection map*

$$\begin{aligned} &\pi : \mathbb{C}^{n+m} \to \mathbb{C}^m \\ &(x_1, \ldots, x_n, y_1, \ldots, y_m) \mapsto (y_1, \ldots, y_m), \end{aligned}$$

then

$$\overline{\pi(V)} = \mathcal{V}\Big(I(V) \cap \mathbb{C}[y_1, \ldots, y_m]\Big).$$

The above theorem is a statement about the images of algebraic sets under projection, which plays a key role in the study of *elimination theory* in algebraic geometry. We explore the basic idea with an example.

Example 2.1.8. Consider the algebraic set $V = \mathcal{V}(f) = \{(x,y) \in \mathbb{C}^2 : f = x^2 + y^2 - 1 = 0\} \subset \mathbb{C}^n$ and the projection map $\pi : \mathbb{C}^2 \to \mathbb{C}$ defined by $\pi(x,y) = x$. Then $\overline{\pi(V)} = \mathcal{V}(g)$ where $g = 0$.

Definition 2.1.9. An algebraic set $V \subset \mathbb{C}^n$ is *reducible* if there exist proper subsets $V_1, V_2 \subsetneq V$ such that

$$V = V_1 \cup V_2.$$

V is *irreducible* if no such decomposition exists.

Theorem 2.1.10. *Given an algebraic set $V \subset \mathbb{C}^n$, there exists a unique finite union of irreducible algebraic subsets of V*

$$V = V_1 \cup V_2 \cup \cdots \cup V_r.$$

We call the V_j in Theorem 2.1.10 the *irreducible components* of V.

The following definition from [53] illustrates how "large" Zariski open sets are under the Zariski topology and how we can utilize them to make statements about properties of an algebraic set that hold "almost everywhere."

Definition 2.1.11. Let V be an irreducible algebraic set as is Definition 2.1.9. Then any nonempty Zariski open $X \subset V$ is *Zariski dense* in V. A property of V is called *generic* if the set of points where that property holds contains a Zariski open subset of V. A point $\mathbf{x} \in V$ where a generic property holds is called a *general* point of V with respect to that property.

Remark 2.1.12. We say that a generic property of an irreducible algebraic set V holds "almost everywhere" because the points where it does not hold must form a proper algebraic subset of V.

There are many equivalent ways of defining the (complex) dimension of an algebraic set. Since we are considering primarily algebraic sets over $\mathbb{C}^n$, here we state the definition which generalizes our intuition of dimension in Euclidean space or a vector space.

Definition 2.1.13. Let $V \subset \mathbb{C}^n$ be an algebraic set. The *dimension* of V, $\dim(V)$, is the maximal length d of the chains $V_0 \subset V_1 \subset \cdots \subset V_d$ of distinct nonempty irreducible algebraic subsets of V. We call V *equidimensional* of dimension d if every irreducible component of V has dimension d.

Definition 2.1.14. The *codimension* of a k-dimensional algebraic set $V \subset \mathbb{C}^n$ is $n-k$.

Definition 2.1.15. Consider a point $\mathbf{z} \in V \subset \mathbb{C}^n$, where V is an algebraic set. The *local dimension* of V at $\mathbf{z}$ is the maximal dimension of every irreducible component V_j of V such that $\mathbf{z} \in V_j$.

Given an algebraic set $V \subset \mathbb{C}^n$ and a point $\mathbf{z} \in V$, the affine tangent space of V at $\mathbf{z}$ is given by

$$T_{\mathbf{z}}V := \left\{ \mathbf{x} \in \mathbb{C}^n : \sum_{j=1}^{n} \left. \frac{\partial f}{\partial x_j} \right|_{\mathbf{z}} (x_j - z_j) = 0 \text{ for all } f \in I(V) \right\}.$$

Definition 2.1.16. Let $V \subset \mathbb{C}^n$ be an algebraic set. A point $\mathbf{z} \in V$ is *smooth/nonsingular* in V if there is a unique irreducible component $V' \subset V$ containing $\mathbf{z}$ such that

$$\dim(T_{\mathbf{z}} V') = \dim(V').$$

We denote by $\mathrm{Sing}(V)$ the set of *singular* (or non-smooth) points in V.

Definition 2.1.17. Given $f \in \mathbb{C}[\mathbf{x}]$, the *gradient* $\nabla f : \mathbb{C}^n \to \mathbb{C}^n$ is defined at the point $\mathbf{z} = (z_1, \ldots, z_n) \in \mathbb{C}^n$ by the vector

$$\nabla f(\mathbf{z}) := \begin{bmatrix} \frac{\partial f}{\partial x_1}(\mathbf{z}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\mathbf{z}) \end{bmatrix}.$$

∇ is also referred to as a *differential operator*.

Definition 2.1.18. Given functions $f_1, \ldots, f_s \in \mathbb{C}[\mathbf{x}]$, the *Jacobian* of $f = \{f_1, \ldots, f_s\}$ is defined by the $s \times n$ matrix

$$Jf(\mathbf{x}) := \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_1(\mathbf{x})}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_s(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_s(\mathbf{x})}{\partial x_n} \end{bmatrix}.$$

Here, we take a turn to briefly discuss homogeneous polynomials. The homogenization of a polynomial in $\mathbb{C}^n$ is important, because it allows us to move from $\mathbb{C}^n$ to $\mathbb{P}^n$, *projective space*.

Definition 2.1.19. *Projective n-space* $\mathbb{P}^n$ (or $\mathbb{P}^n(\mathbb{C})$) is the set of all lines in $\mathbb{C}^{n+1}$ containing the origin.

The construction of projective space is useful, because it allows us to introduce points "at infinity" where parallel lines intersect. This idea allows for some elegant statements of uniform results in algebraic geometry that would otherwise fail, especially for many results in plane geometry. However, the geometric intuition for projective space is often harder to access. Therefore, we introduce the algebra of homogeneous polynomials to obtain a more concrete approach to working with the results utilizing equivalence classes of polynomials in $\mathbb{C}^{n+1}$.

Definition 2.1.20. A polynomial $f \in \mathbb{C}[x_0, x_1, \ldots, x_n]$ is *homogeneous of degree d* in $x_0, x_1, \ldots, x_n$ if it is of the form

$$f = \sum_{|\alpha|=d} c_\alpha \mathbf{x}^\alpha = \sum_{\alpha_0+\alpha_1+\cdots+\alpha_n=d} c_{(\alpha_0,\alpha_1,\ldots,\alpha_n)} x_0^{\alpha_0} x_1^{\alpha_1} \cdots x_n^{\alpha_n}.$$

For our purposes, the defining polynomials of our algebraic set will not always be homogeneous. Therefore, we utilize the standard process of *homogenization*.

Definition 2.1.21. Given a polynomial $f \in \mathbb{C}[x_1, \ldots, x_n]$, the *homogenization of* f *(with respect to* x_0*)* is

$$f^{(h)} := x_0^{\deg(f)} f\left(\frac{x_1}{x_0}, \ldots, \frac{x_n}{x_0}\right).$$

We note that while the homogenization does require us to add another variable to our problem, the benefits this will provide us in our proofs far outweighs the computational cost of this extra variable in practice.

2.2 Real Algebraic Geometry

In this section, we consider how the definitions and results of classical algebraic geometry differ when we examine solution sets to polynomial systems over $\mathbb{R}^n$, rather than $\mathbb{C}^n$. Our treatment follows that of [8], which can be referenced for more details if necessary.

Definition 2.2.1. Let $f_1,\dots,f_s \in \mathbb{R}[\mathbf{x}]$. Then

$$\mathcal{V}_{\mathbb{R}}(f_1,\dots,f_s) := \{\mathbf{x}\in\mathbb{R}^n : f_1(\mathbf{x}) = \cdots = f_s(\mathbf{x}) = 0\}$$

is the common zero set of the polynomial system $f_1,\dots,f_s$ in $\mathbb{R}^n$. We call a set of this form a *real algebraic set*.

We note that this definition conceals some of the difficulties that arise when we study real algebraic sets compared to their complex counterparts. In particular, $\mathbb{R}$ is not an algebraically closed field as in Definition 2.1.1. We illustrate the basic nuance of this difference with a simple example.

Example 2.2.2. Consider the polynomials $f = x^2+y^2+1$. and $g = x^2+y^2+2$. Then $\mathcal{V}_{\mathbb{R}}(f) = \mathcal{V}_{\mathbb{R}}(g) = \emptyset$, but $\mathcal{V}(f) \neq \mathcal{V}(g)$. In other words, these are different algebraic sets, but we cannot distinguish them as *real* algebraic sets.

To explore further, we define topological terms for $\mathbb{R}^n$, which follow the standard Euclidean topology.

Definition 2.2.3. Given $\mathbf{z}\in\mathbb{R}^n$, $r\in\mathbb{R}$, $r>0$, the Euclidean norm is given by $\|\mathbf{z}\| = \sqrt{z_1^2+\cdots z_n^2}$. We define an open ball $B(\mathbf{z},r)\subset\mathbb{R}^n$ by

$$B(\mathbf{z},r) := \{\mathbf{x}\in\mathbb{R}^n : \|\mathbf{z}-\mathbf{x}\|^2 < r^2\}$$

and a closed ball $\overline{B}(\mathbf{z},r)\subset\mathbb{R}^n$ by

$$\overline{B}(\mathbf{z},r) := \{\mathbf{x}\in\mathbb{R}^n : \|\mathbf{z}-\mathbf{x}\|^2 \leq r^2\}.$$

For a set $Z\subset\mathbb{R}^n$, Z is *open* if it is the union of open balls. Z is *closed* if its complement is open. Finally, the *closure* of Z, $\overline{Z}$, is the intersection of all closed subsets of $\mathbb{R}^n$ containing Z.

Definition 2.2.4. Let $f_1,\dots,f_s,q_1,\dots,q_t \in \mathbb{R}[\mathbf{x}]$. Then

$$S := \{\mathbf{x}\in\mathbb{R}^n : f_1(\mathbf{x}) = \cdots = \mathbf{f}_s(\mathbf{x}) = 0, q_1(\mathbf{x}) > 0,\dots,q_t(\mathbf{x}) > 0\} \tag{2.2}$$

is an *atomic semi-algebraic set*. More generally, a set $T\subset\mathbb{R}^n$ is a *semi-algebraic set* if it is a finite union of atomic semi-algebraic sets.

By moving to semi-algebraic sets, we are now considering systems of polynomial equations and inequalities. The following projection theorem for semi-algebraic sets is fundamental to computational algebraic geometry.

Theorem 2.2.5. *[8, Theorem 2.76] The image of a semi-algebraic set under a projection map is semi-algebraic.*

We illustrate with an example why the use of semi-algebraic sets, rather than algebraic sets, is necessary to the statement of Theorem 2.2.5.

Example 2.2.6. Consider $f(x,y) = xy - 1$ and the projection map $\pi : (x,y) \to x$. Then the hyperbola defined by f is a real algebraic set $\mathcal{V}_{\mathbb{R}}(f) \subset \mathbb{R}^2$. However, the image of $\mathcal{V}_{\mathbb{R}}(f)$ under the map π is the set of points satisfying $x \neq 0$, which is a semi-algebraic set.

While for algebraic sets we typically consider the idea of compactness, for semi-algebraic sets we think about what properties hold when they are closed and bounded.

Definition 2.2.7. Let $S_1 \in \mathbb{R}^k$ and $S_2 \in \mathbb{R}^l$ be semi-algebraic sets. A function $f : S_1 \to S_2$ is called *semi-algebraic* if the graph of f in $\mathbb{R}^{k+l}$ is semi-algebraic.

The following result is a useful statement about the preservation of closed and boundedness under mapping.

Theorem 2.2.8. *Let S be a closed and bounded semi-algebraic set and g a continuous semi-algebraic function defined on S. Then the image $g(S)$ is closed and bounded.*

The following result is known as the Curve Selection Lemma.

Lemma 2.2.9. *[8, Theorem 3.19] Let $S \subset \mathbb{R}^n$ be a semi-algebraic set and $\mathbf{x} \in \overline{S}$. Then there exists a continuous semi-algebraic mapping $\gamma : [0,1) \to \mathbb{R}^n$ such that $\gamma(0) = \mathbf{x}$ and $\gamma((0,1)) \subset S$, where a semi-algebraic mapping is defined as in Definition 2.2.7.*

Next we establish some terminology that mirrors the concepts we defined at the end of the previous section.

Definition 2.2.10. A semi-algebraic set $S \subset \mathbb{R}^n$ is *disconnected* if there exist two disjoint closed subsets $C_1, C_2 \subsetneq S$ such that

$$S = C_1 \cup C_2.$$

S is *connected* if no such decomposition exists.

We call C_j as in Definition 2.2.10 the *connected components* of a semi-algebraic set. Intuitively, we should think of connected components of a semi-algebraic set as being the semi-algebraic equivalent to irreducible components of an algebraic set.

Definition 2.2.11. For a semi-algebraic set $S \subset \mathbb{R}^n$, its *real dimension*, $\dim_{\mathbb{R}}(S)$, is the largest k such that there exists an injective semi-algebraic map from $(0,1)^k$ to S.

Definition 2.2.12. Consider a point $\mathbf{z} \in S \subset \mathbb{R}^n$, where S is a semi-algebraic set. The *local real dimension* of S at $\mathbf{z}$ is the maximal real dimension of the closure of every connected component C_j of S such that $\mathbf{z} \in \overline{C_j}$.

We note that by convention, the (real or complex) dimension of the empty set is -1.

Definition 2.2.13. Let $S \subset \mathbb{R}^n$ be an atomic semi-algebraic set as in (2.2). A point $\mathbf{z} \in S$ is *smooth/nonsingular* if it is smooth in the algebraic set $\mathcal{V}(f_1,\ldots,f_s)$ as in Definition 2.1.16.

The following shows that smooth points on each connected component of an atomic semi-algebraic set S can be obtained as projections of smooth points of some real algebraic set.

Proposition 2.2.14. *Let S be an atomic semi-algebraic set as in* (2.2) *and*

$$W := \left\{(\mathbf{x},\mathbf{z}) \in \mathbb{R}^n \times \mathbb{R}^m \,:\, f_1(\mathbf{x}) = \cdots = f_n(\mathbf{x}) = 0,\, z_1^2 q_1(\mathbf{x}) - 1 = \cdots = z_m^2 q_m(\mathbf{x}) - 1 = 0\right\}.$$

If $\mathbf{y} \in W$ is smooth, then $\pi_x(\mathbf{y}) \in S$ is also smooth. Conversely, if $\mathbf{x} \in S$ is smooth, then $(\mathbf{x},\mathbf{z})$ is smooth in W for all $\mathbf{z} = (z_1,\ldots,z_m) \in \mathbb{R}^m$ such that $(\mathbf{x},\mathbf{z}) \in W$.

Proof. Without loss of generality, we can assume that $f := \{f_1,\ldots,f_s\}$ generates a prime ideal. The Jacobian matrix of the polynomial system defining W has the block structure

$$J(\mathbf{x},\mathbf{z}) = \begin{array}{|c|c|}\hline Jf(\mathbf{x}) & 0 \\ \hline * & \operatorname{diag}(2z_i q_i(\mathbf{x})) \\ \hline \end{array}$$

Since for $(\mathbf{x},\mathbf{z}) \in W$ we have $z_i g_i(\mathbf{x}) \neq 0$, the Jacobian matrix $Jf(\mathbf{x})$ has full column rank if and only if $J(\mathbf{x},\mathbf{z})$ has full column rank, which proves the claim. □

The next reduction is to replace an arbitrary real algebraic set with a closed and bounded one.

Proposition 2.2.15. *Let $f_1,\ldots,f_{s-1} \in \mathbb{R}[x_1,\ldots,x_{n-1}]$ and consider $\mathbf{q} = (q_1,\ldots,q_{n-1}) \in \mathbb{R}^{n-1}$. Let $\delta \in \mathbb{R}_+$, introduce a new variable x_n, and consider*

$$f_s := (x_1 - q_1)^2 + \cdots + (x_{n-1} - q_{n-1})^2 + x_n^2 - \delta$$

Then, $\mathcal{V}(f_1,\ldots,f_s) \cap \mathbb{R}^n$ is bounded and

$$\pi_{n-1}\left(\mathcal{V}(f_1,\ldots,f_s) \cap \mathbb{R}^n\right) = \mathcal{V}(f_1,\ldots,f_{s-1}) \cap \left\{\mathbf{z} \in \mathbb{R}^{n-1} \,:\, \|\mathbf{z}-\mathbf{q}\|^2 \leq \delta\right\}$$

where $\pi_{n-1}(x_1,\ldots,x_n) = (x_1,\ldots,x_{n-1})$.

Remark 2.2.16. The definition of f_s above is based on a standard trick used in real algebraic geometry to make an arbitrary real algebraic set bounded (e.g., see [9]). In general, $V \cap \mathbb{R}^{n-1}$ is embedded into a sphere in $\mathbb{R}^n$ around the origin of radius $1/\zeta$ where ζ is infinitesimal. In the context of homotopy continuation methods described in the following sections, we are only interested in computing points with bounded coordinates, so it is sufficient to embed its intersection with a closed ball around $\mathbf{q}$ of radius $\sqrt{\delta}$ for some fixed $\delta \in \mathbb{R}_+$ (rather than using infinitesimal variables).

Going forward, when we suppose we have a closed and bounded real algebraic set $\mathcal{V}(f_1,\ldots,f_s)\cap\mathbb{R}^n$, we assume that if we were originally given an arbitrary atomic semi-algebraic set, we first applied Propositions 2.2.14 and 2.2.15.

Next, we present some notation and results around perturbations of (real) algebraic sets from [8], which will be crucial to many of our proofs in Chapter 4.

Definition 2.2.17. A *total ordering* on a set A is a binary relation $\leq$ satisfying the properties for all $a,b,c\in A$:

- $\leq$ is reflexive (i.e. $a\leq a$);
- $\leq$ is transitive (i.e. $a\leq b, b\leq c\Rightarrow a\leq c$);
- $\leq$ is anti-symmetric (i.e. $a\leq b, b\leq a\Rightarrow a=b$);
- every two elements $a,b\in A$ are comparable (i.e. $a\leq b$ or $b\leq a$).

An *ordered field* is a field $\mathbb{K}$ along with a total ordering $\leq$ such that for all $x,y,z\in\mathbb{K}$:

- $x\leq y\Rightarrow x+z\leq y+z$;
- $0\leq x, 0\leq y\Rightarrow 0\leq xy$.

For $(\mathbb{K},\leq)$ an ordered field, the subset $C\subset\mathbb{K}$ such that $C=\{x\in\mathbb{K}:0\leq x\}$ is the *positive cone* of $(\mathbb{K},\leq)$.

Definition 2.2.18. An ordered field $(\mathbb{K},\leq)$ is a *real closed field* if its positive cone is the set of squares of elements in $\mathbb{K}$ and every polynomial in $\mathbb{K}[x]$ of odd degree has a root in $\mathbb{K}$.

Definition 2.2.19. Let $\mathbb{K}\subset\mathbb{K}'$ be two ordered fields. An element $\varepsilon\in\mathbb{K}'$ is *infinitesimal over* $\mathbb{K}$ if $|\varepsilon|\leq|x|$ for all $x\in\mathbb{K}$.

In particular, given an ordered field $\mathbb{K}$ and a variable ε, a natural total ordering on $\mathbb{K}\langle\varepsilon\rangle$ is constructed by assuming $\varepsilon>0$ is an infinitesimal over $\mathbb{K}$.

Definition 2.2.20. Let $\mathbb{K}=\mathbb{R}$ or $\mathbb{C}$ and denote by $\mathbb{K}\langle\varepsilon\rangle$ the *field of Puiseux series* over $\mathbb{K}$, i.e.

$$\mathbb{K}\langle\varepsilon\rangle:=\left\{\sum_{i\geq i_0}a_i\varepsilon^{i/q}\ :\ i_0\in\mathbb{Z}, q\in\mathbb{Z}_{>0}, a_i\in\mathbb{K}\right\}.$$

A Puiseux series $z=\sum_{i\geq i_0}a_i\varepsilon^{i/q}\in\mathbb{K}\langle\varepsilon\rangle$ is called *bounded* if $i_0\geq 0$.

We note that the field of real Puiseux series $\mathbb{R}\langle\varepsilon\rangle$ is a real closed field as in Definition 2.2.18.

Theorem 2.2.21. *[8, Corollary 2.98] Since $\mathbb{R}$ is a real closed field, so is $\mathbb{R}\langle\varepsilon\rangle$. Furthermore, $\mathbb{R}\langle\varepsilon\rangle$ is the real closure of $\mathbb{R}(\varepsilon)$.*

In particular, this means we can extend some properties of the real numbers to the field of Puiseux series.

Theorem 2.2.22. *[8, Theorem 2.11] Since $\mathbb{R}\langle\varepsilon\rangle$ is a real closed field, the intermediate value theorem holds on it.*

We also note that for Puiseux series, similar results hold over $\mathbb{C}$ when we replace real closed with algebraically closed.

Theorem 2.2.23. *[9, Theorem 2.92] Since $\mathbb{C}$ is an algebraically closed field, so is $\mathbb{C}\langle\varepsilon\rangle$. Furthermore, $\mathbb{C}\langle\varepsilon\rangle$ is the algebraic closure of $\mathbb{C}(\varepsilon)$.*

Finally, we establish a notation for the concept of an *extension* of a semi-algebraic set $S \subset \mathbb{R}^n$ to $\mathbb{R}\langle\varepsilon\rangle$.

Definition 2.2.24. Given a semi-algebraic set $S \subset \mathbb{R}^n$, the *extension* of S to $\mathbb{R}\langle\varepsilon\rangle^n$, denoted $\text{Ext}(S, \mathbb{R}\langle\varepsilon\rangle^n)$, is the semi-algebraic subset $S' \subset \mathbb{R}\langle\varepsilon\rangle^n$ defined by the same equations and inequalities as S but their solutions considered in $\mathbb{R}\langle\varepsilon\rangle^n$.

2.3 Polar Varieties

In this section, we introduce the idea of *polar varieties*, which will be a main tool in our results in the following chapters. There is extensive literature about different notions of polar varieties. [6] provides a survey, and we explore some of the connections between those results here. We will begin our construction over projective space, $\mathbb{P}^n$, as in Definition 2.1.19.

Definition 2.3.1. For $A, B \subset \mathbb{P}^n$, $\langle A, B\rangle$ is the subvariety of $\mathbb{P}^n$ spanned by A and B. A and B *intersect transversally*, $A \pitchfork B$, if $\langle A, B\rangle = \mathbb{P}^n$. If A and B do not intersect transversally, we write $A \not\pitchfork B$.

For $V \subset \mathbb{P}^n$ of pure codimension p, V_{reg} is the set of all smooth (i.e. regular) points of V, a complex subset of codimension p that is Zariski dense in V. $\text{Sing}(V) = V \setminus V_{\text{reg}}$ is the *singular locus* of V.

Definition 2.3.2. Let $V \subset \mathbb{P}^n$ be a variety. For any linear subvariety $L \subset \mathbb{P}^n$, the *polar variety of V in terms of L* is

$$W_L(V) := \overline{\{\mathbf{p} \in V_{\text{reg}} \setminus L \mid T_{\mathbf{p}}V \not\pitchfork \langle \mathbf{p}, L\rangle \text{ at } \mathbf{p}\}}.$$

In words, $W_L(V)$ is the Zariski closure of the smooth points $\mathbf{p}$ of V which are not also in L, such that the tangent vectors of V at $\mathbf{p}$, along with $\mathbf{p}$ and L, do not all together span $\mathbb{P}^n$. We note that $W_L(V) \subset V$.

For the affine interpretation, let V be the projective closure of a closed subvariety $S \subset \mathbb{C}^n$. Then the *affine polar variety of S in terms of L* is

$$W_L(S) = W_L(V) \cap \mathbb{C}^n.$$

Furthermore, given $V \subset \mathbb{P}^n$ a variety of pure codimension p, we can create a sequence of polar varieties of codimension 1 to $n-p$ in V. To do so, we let $\mathscr{L}$ be a flag of projective linear subvarieties of $\mathbb{P}^n$

$$\mathscr{L} : L^0 \subset \cdots \subset L^{p-1} \subset \cdots \subset L^{n-1} \subset \mathbb{P}^n$$

where k is the dimension of L^k. By our above definition, this gives the sequence of polar varieties

$$\emptyset = W_{L^{n-1}}(V) \subset W_{L^{n-2}}(V) \subset \cdots \subset W_{L^{p-1}}(V) \subset W_{L^{p-2}}(V) = \cdots = W_{L^0}(V) = V.$$

In particular, we note that the non-trivial polar varieties in terms of $\mathscr{L}$ are given by L^{p-1} to L^{n-2}.

Definition 2.3.3. The *i-th polar variety of V with respect to $\mathscr{L}$* is

$$V_i := W_{L^{p+i-2}}(V), 1 \leq i \leq n-p,$$

where $i = \operatorname{codim}(V_i)$ in V.

This concept extends directly to the affine case, giving the definition of the *i-th affine polar variety of V with respect to $\mathscr{L}$* as

$$S_i := W_{L^{p+i-2}}(S), 1 \leq i \leq n-p,$$

where $i = \operatorname{codim}(S_i)$ in S.

Definition 2.3.4. $Q \subset \mathbb{P}^n$ is a variety known as a *hyperquadric* if it is the set of zeros of a nonzero quadratic polynomial $q(x_0, \ldots, x_n)$. More explicitly, Q is the zero set of some

$$q(x_0, \ldots, x_n) := \sum_{i,j} a_{i,j} x_i x_j + \sum_k b_k x_k + c = 0$$

where at least one of the $a_{i,j} \neq 0$.

Definition 2.3.5. Let $L \subset \mathbb{P}^n$ be a linear variety of dimension k and Q a hyperquadric defined by $q(x_1, \ldots, x_n)$. Then $L^\vee$ is the *dual of L with respect to Q*. More explicitly, take $p_0, ..., p_k$ to be a set of points which define L. Then

$$L^\vee := \bigcap_{i=0}^{k} \mathscr{V}\left(\sum_{j=0}^{n} \frac{\partial q}{\partial x_j}(p_i) x_j \right).$$

We note that $\dim(L^\vee) = n-k-1$, since it is the intersection of $k+1$ general hyperplanes in $\mathbb{P}^n$.

Now choose $H \subset \mathbb{P}^n$ a hyperplane not tangent to Q, and restrict our focus to this hyperplane. Since $\dim(H) = n-1$, we can take $y_1, \ldots, y_n$ to be coordinates relative to H. Since H is not tangent to Q, $Q \cap H$ is a hyperquadric, defined by some equation $q^*(y_1, \ldots, y_n)$. Suppose $L \subseteq H$ is a linear variety of dimension k. Then L^* is the *dual of L with respect to $Q \cap H$*. As above, take $p_0, ..., p_k$ to be a

set of points which define L. Then

$$L^* := \bigcap_{i=0}^{k} \mathcal{V}\Big(\sum_{j=1}^{n} \frac{\partial q}{\partial y_j}(p_i) y_j \Big).$$

We note that $\dim(L^*) = n-k-2$, since it is the intersection of $k+2$ general hyperplanes in $\mathbb{P}^n$.

Definition 2.3.6. Let $V \subset \mathbb{P}^n$ be a variety, $Q \subset \mathbb{P}^n$ a hyperquadric, and $H \subset \mathbb{P}^n$ a hyperplane not tangent to Q. Then for any linear subvariety $L \subset \mathbb{P}^n$, the *generalized polar variety of V in terms of L* is

$$\widehat{W_L}(V) := \overline{\{\mathbf{p} \in V_{\text{reg}} \setminus (L \cup H) \mid T_{\mathbf{p}} V \not\cap \langle p, (\langle \mathbf{p}, L \rangle \cap H)^* \rangle \text{ at } \mathbf{p}\}}.$$

In words, $\widehat{W_L}(V)$ is the Zariski closure of the smooth points $\mathbf{p}$ of V which are not also in L or H, such that the tangent vectors of V at $\mathbf{p}$, along with $\mathbf{p}$ and the dual of the linear span of $\mathbf{p}$ and L within H with respect to $Q \cap H$, do not all together span $\mathbb{P}^n$. We note that $\widehat{W_L}(V) \subset V$.

For the affine interpretation, let V be the projective closure of a closed subvariety $S \subseteq \mathbb{C}^n$ and H the hyperplane at infinity of $\mathbb{P}^n$. Then the *affine generalized polar variety of S in terms of L* is

$$\widehat{W_L}(S) := \widehat{W_L}(V) \cap \mathbb{C}^n.$$

Furthermore, given $V \subset \mathbb{P}^n$ a variety of pure codimension p, we can create a sequence of generalized polar varieties of codimension 1 to $n-p$ in V. To do so, we let $\mathscr{L}$ be a flag of projective linear subvarieties of $\mathbb{P}^n$

$$\mathscr{L} : L^0 \subset \cdots \subset L^{n-p-1} \subset \cdots \subset L^{n-1} \subset \mathbb{P}^n$$

where k is the dimension of L^k. By our above definition, this gives the sequence of generalized polar varieties

$$\widehat{W_{L^0}}(V) \subset \widehat{W_{L^1}}(V) \subset \cdots \subset \widehat{W_{L^{n-p-1}}}(V) \subset \widehat{W_{L^{n-p}}}(V) = \cdots = \widehat{W_{L^{n-1}}}(V) = V.$$

In particular, we note that the non-trivial generalized polar varieties in terms of $\mathscr{L}$ are given by L^{n-p-1} to L^0.

Definition 2.3.7. The *i-th generalized polar variety of V with respect to $\mathscr{L}$* is

$$\widehat{V_i} := \widehat{W_{L^{n-p-i}}}(V), 1 \leq i \leq n-p,$$

where $i = \text{codim}(\widehat{V_i})$ in V.

The reduction of the generalized definition to the classic definition for any variety $V \subseteq \mathbb{P}^n$ and hyperquadric $Q \subset \mathbb{P}^n$ relies on an additional restriction for our choice of hyperplane H. Suppose we are concerned with $W_L(V)$, the polar variety of V with respect to some linear variety $L \subseteq \mathbb{P}^n$. Then we must choose our hyperplane H such that $L \subset H$ is a subvariety.

Suppose we have followed the above restriction, and consider $\widehat{W_{L^*}}(V)$. Choose $\mathbf{p} \in V_{\text{reg}} \setminus H$. Since

$L^* \subseteq H$ and $\mathbf{p} \notin H$,

$$\langle \mathbf{p}, L^* \rangle \cap H = L^*.$$

Then

$$\langle \mathbf{p}, (\langle \mathbf{p}, L^* \rangle \cap H)^* \rangle = \langle \mathbf{p}, (L^*)^* \rangle = \langle \mathbf{p}, L \rangle.$$

Thus

$$\widehat{W}_{L^*}(V) = W_L(V).$$

So in the case where the linear variety L associated with our polar variety is such that $L \subset H$, the classic polar variety is equivalent to the generalized polar variety of L^*.

Similarly, for the dual case we choose our hyperplane H such that $L \subset H$ is a subvariety. We consider the generalized polar variety $\widehat{W}_{L^\vee}(V)$ and suppose $\mathbf{p} \in V_{\text{reg}} \setminus (L^\vee \cup H)$. We note that $\langle \mathbf{p}, L^\vee \rangle^\vee \subseteq H$. Then $\langle \mathbf{p}, L^\vee \rangle^\vee \subseteq (\langle \mathbf{p}, L^\vee \rangle \cap H)^\vee \cap H$. Furthermore, due to dimension constraints, we have

$$(\langle \mathbf{p}, L^\vee \rangle \cap H)^\vee \cap H = (\langle \mathbf{p}, L^\vee \rangle \cap H)^\vee \cap H = \langle \mathbf{p}, L^\vee \rangle^\vee.$$

Hence the generalized polar variety definition in this case reduces so that

$$\widehat{W}_{L^\vee}(V) := \overline{\{\mathbf{p} \in V_{\text{reg}} \setminus (L^\vee \cup H) \mid T_{\mathbf{p}} V \not\pitchfork \langle \mathbf{p}, \langle \mathbf{p}, L^\vee \rangle^\vee \rangle \text{ at } \mathbf{p}\}}.$$

We say that $\widehat{W}_{L^\vee}(V)$ the *dual polar variety of V in terms of $L^\vee$*.

We note that in addition to the geometric definitions of polar varieties we have described, it will be more useful for us to develop equivalent explicit definitions, in order to facilitate the computation of polar varieties.

Let $V := \mathcal{V}(f_1^{(h)}, \ldots, f_p^{(h)}) \subset \mathbb{P}^n$ be an algebraic variety of pure codimension p, where $f_1^{(h)}, \ldots, f_p^{(h)} \in \mathbb{C}[x_0, \ldots, x_n]$ are homogeneous as in Definition 2.1.21. Define $S := V \cap \mathbb{C}^n$ and $f = \{f_1, \ldots, f_p\}$ the dehomogenizations of $f_1^{(h)}, \ldots, f_p^{(h)}$ respectively. Then assuming S is nonempty, we have $S = \mathcal{V}(f_1, \ldots, f_p) \subset \mathbb{C}^n$. We note that a point $\mathbf{z} \in S$ is smooth as in Definition 2.1.16 if the $p \times n$ Jacobian matrix

$$Jf(\mathbf{x}) := \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_1(\mathbf{x})}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_p(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_p(\mathbf{x})}{\partial x_n} \end{bmatrix}.$$

has maximal rank p when evaluated at $\mathbf{z}$.

For our context, we suppose H is the hyperplane at infinity of $\mathbb{P}^n$, i.e. when $x_0 = 0$ and the hyperquadric Q is defined by a quadratic form

$$q(x_0, \ldots, x_n) := x_0^2 + \sum_{k=1}^{n} 2 c_k x_0 x_k + \sum_{k=1}^{n} x_k^2$$

where $c_k \in \mathbb{C}$ for $1 \leq k \leq n$. We note that this carefully chosen q gives that $Q \cap H$ is defined by the

quadratic form

$$q^*(\mathbf{x}) := \sum_{k=1}^{n} x_k^2 \in \mathbb{R}[\mathbf{x}].$$

In particular, $Q \cap H \cap \mathbb{R}^n$ is given by q^*, a positive semidefinite form which induces the usual Euclidean distance metric.

Fix $1 \leq i \leq n-p$ and choose a generic $(n-p-i)$-dimensional linear subvariety $L \subset \mathbb{P}^n$. We note that one standard way we could choose L would be to pick spanning points $\mathbf{a}_1, \ldots, \mathbf{a}_{n-p-i+1} \in \mathbb{P}^n$ where $a_{j,0} = 0$ or $a_{j,0} = 1$ and $a_{j,k}$ generic for $0 \leq k \leq n$ and $1 \leq j \leq n-p-i+1$.

Suppose $\mathbf{p} \in V$ is smooth with $p_0 \neq 0$ and $\mathbf{z} \notin L$. Then $\langle \mathbf{p}, L \rangle \cap H$ is a $(n-p-i)$-dimensional linear subvariety given by the $n-p-i+1$ linearly independent points

$$p_0\mathbf{a}_1 - a_{1,0}\mathbf{p}, \ldots, p_0\mathbf{a}_{n-p-i+1} - a_{n-p-i+1,0}\mathbf{p}.$$

Now using $\mathbf{y} := (y_1, \ldots, y_n)$ we rewrite the quadratic form $q_0(\mathbf{x})$ associate with $Q \cap H$ as a bilinear form

$$B(\mathbf{x}, \mathbf{y}) := \sum_{k=1}^{n} x^k y^k \in \mathbb{R}[\mathbf{x}, \mathbf{y}].$$

For $1 \leq j \leq n-p-i+1$, we define

$$l_j := B(p_0 a_{j,1} - a_{j,0} p_1, \ldots, p_0 a_{j,n} - a_{j,0} p_n, x_1, \ldots, x_n) \in \mathbb{C}[x_1, \ldots, x_n]$$

and

$$L_j := p_0 l_j(x_1, \ldots, x_n) - x_0 l_j(p_1, \ldots, p_n) \in \mathbb{C}[x_0, x_1, \ldots, x_n].$$

Then the linear forms $l_1, \ldots, l_{n-p-i+1}$ are linearly independent and give the $(p-i+2)$-dimensional linear variety $(\langle \mathbf{p}, L \rangle \cap H)^*$ in H. Furthermore, $L_1, \ldots, L_{n-p-i+1}$ are linearly independent and vanish at $\mathbf{p}$ and $(\langle \mathbf{p}, L \rangle \cap H)^*$. So $L_1, \ldots, L_{n-p-i+1}$ describe the linear variety $\langle \mathbf{p}, (\langle \mathbf{p}, L \rangle \cap H)^* \rangle$ used to define the generalized polar variety of V in terms of L, $\widehat{W_L}(V)$, as in Definition 2.3.6.

For clarity, we rewrite $L_1, \ldots, L_{n-p-i+1}$ using

$$\begin{aligned} L_j &= p_0 l_j(x_1, \ldots, x_n) - x_0 l_j(p_1, \ldots, p_n) \\ &= -(x_0 - p_0) l_j(x_1, \ldots, x_n) + p_0 l_j(x_1 - p_1, \ldots, x_n - p_n) \\ &= -(x_0 - p_0) l_j(x_1, \ldots, x_n) + p_0 \sum_{k=1}^{n} (p_0 a_{j,k} - a_{j,0} p_k)(x_k - p_k). \end{aligned}$$

Suppose $p_0 = 1$ and let $L'_j := L_j|_{p_0=1}$ for $1 \leq j \leq n-p-i+1$. We note $\mathbf{p} = (p_1, \ldots, p_n)$ is smooth in $S = V \cap \mathbb{C}^n$. So we consider the $(n-i+1) \times n$ Jacobian matrix of $f_1, \ldots, f_p, L'_1, \ldots, L'_{n-p-i+1}$ in the variables $x_1, \ldots, x_n$, i.e.

$$J_i(\mathbf{p}) := \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{p}) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathbf{p}) \\ \vdots & \vdots & \vdots \\ \frac{\partial f_p}{\partial x_1}(\mathbf{p}) & \cdots & \frac{\partial f_p}{\partial x_n}(\mathbf{p}) \\ a_{1,1} - a_{1,0}x_1 & \cdots & a_{1,n} - a_{1,0}x_n \\ \vdots & \vdots & \vdots \\ a_{n-p-i+1,1} - a_{n-p-i+1,0}x_1 & \cdots & a_{n-p-i+1,n} - a_{n-p-i+1,0}x_n \end{bmatrix}.$$

Then the vanishing of all the $(n-i+1)$ minors of $J_i(\mathbf{p})$ is equivalent to the condition $T_{\mathbf{p}}(V) \not\subset \langle \mathbf{p}, (\langle \mathbf{p}, L\rangle \cap H)^* \rangle$ from Definition 2.3.6. So $f_1, \ldots, f_p$ and the $(n-i+1)$ minors of $J_i(\mathbf{p})$ define the generalized affine polar variety $\widehat{W_L}(S)$ outside the singular locus $\mathrm{Sing}(S)$.

For our purposes, we utilize this formulation and reduce to the hypersurface case by taking a sum of squares if necessary. We note that when we reduce to the hypersurface case, $p = 1$. So the ith polar variety is given by f and $L'_1, \ldots, L'_{n-i}$. We observe that for $J_i(\mathbf{x})$, this consists of a generic linear combination of the partial derivatives of f. So without loss of generality (in particular, when we apply a change of variables later in our paper), we can choose the appropriate number of partials to obtain a simplified definition.

In practice, other notions of polar varieties may work better. We chose this presentation for its simplified notation and presentation, following the approach of [46], for conciseness.

Definition 2.3.8. Let $f \in \mathbb{C}[\mathbf{x}]$ be square-free and $V = \mathscr{V}(f) \subset \mathbb{C}^n$. Consider the projections $\pi_i(x_1, \ldots, x_n) = (x_1, \ldots, x_i)$ for $i = 1, \ldots, n$. The *polar variety associated to* π_i of V is defined as

$$\mathrm{crit}(V, \pi_i) := \mathscr{V}\left(f, \frac{\partial f}{\partial x_{i+1}}, \ldots, \frac{\partial f}{\partial x_n}\right) \subset \mathbb{C}^n \quad i = 1, \ldots, n,$$

based on how the polynomials defining this algebraic set correspond to the notion of critical points of a map.

We note that a big difference in how this definition is stated compared to the prior geometric formulation is that it does not exclude the singular locus of an algebraic set V from the polar varieties associated to V. We will address the smoothness of V going forward via a change of variables and perturbations in Chapter 4, so in fact it is natural to make this simplified modification in this context.

2.4 Numerical Homotopy Methods

In this section, we discuss some main ideas from numerical algebraic geometry which guide our problem formulation and computations going forward. This follows the treatment of [11], which can be referenced for more details if necessary.

We consider an algebraic set $V \subset \mathbb{C}^n$ as in Definition 2.1.2 and further assume that it is defined

by a *square* polynomial system, i.e. $s = n$. Let $f := \{f_1, \ldots, f_n\}$. We use the following notation for $\mathscr{V}(f)$

$$f(\mathbf{x}) := \begin{bmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_n(\mathbf{x}) \end{bmatrix} = 0, \tag{2.3}$$

yielding the square Jacobian matrix

$$Jf(\mathbf{x}) := \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_1(\mathbf{x})}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_n(\mathbf{x})}{\partial x_n} \end{bmatrix}.$$

We note that $\mathbf{z} \in \mathscr{V}(f)$ is *smooth* as in Definition 2.1.16 if $\operatorname{rank}(Jf(\mathbf{z})) = n$ and

$$\operatorname{Sing}(\mathscr{V}(f)) \subseteq \{\mathbf{x} \in \mathbb{C}^n : \operatorname{rank}(Jf(\mathbf{x})) < n\}.$$

Definition 2.4.1. Given $\mathbf{z} \in \mathscr{V}(f)$, we call $\mathbf{z}$ an *isolated solution* of f if there is an open ball $B(\mathbf{z}, r)$ as in Definition 2.2.3 such that $\mathbf{z}$ is the only solution to f contained in $B(\mathbf{z}, r)$.

To compute the isolated solutions of the polynomial system (2.3), we follow an approach known as *homotopy continuation*. Here we choose a *start system*, $g(\mathbf{x})$ for some well-chosen $g = \{g_1, \cdots, g_n\}$, whose finite set of solutions are already known. We then create a parameterized family of equations $H(\mathbf{x}, t)$ which continuously deforms the system of $g(\mathbf{x})$ and its solutions to the system of $f(\mathbf{x})$ and its solutions. For our purposes, we define this *homotopy* $H(\mathbf{x}, t)$ using a parameter t which starts at $t = 1$ and moves continuously to end at $t = 0$ such that $H(\mathbf{x}, 1) = g(\mathbf{x})$ and $H(\mathbf{x}, 0) = f(\mathbf{x})$.

Example 2.4.2. The simplest form of homotopy to construct that deforms $g(\mathbf{x})$ into $f(\mathbf{x})$ is a linear homotopy

$$H(\mathbf{x}, t) := t\,g(\mathbf{x}) + (1-t)f(\mathbf{x}).$$

Note that since $t \in \mathbb{C}$, there exist an infinite number of continuous one-real-dimensional mappings going from $t = 1$ to $t = 0$. In order to travel along the paths in a numerically stable way using predictor-corrector methods, we introduce some more structure to the mapping so that our solution paths do not collide in the interval $t \in (0, 1]$. This motivates the following definition of a *good homotopy* for our purposes.

Definition 2.4.3. Suppose $f(\mathbf{x})$ as in (2.3) is the system of polynomial equations we are looking to solve and $g(\mathbf{x})$ is the start system with D distinct solutions. Then a *good homotopy* is a system of infinitely differentiable functions

$$H(\mathbf{x}, t) := \begin{bmatrix} H_1(\mathbf{x}, t) \\ \vdots \\ H_n(\mathbf{x}, t) \end{bmatrix}$$

such that

1. for all $t \in [0,1]$, $H(\mathbf{x}, t)$ is a system of polynomials;
2. for each distinct root of $g(\mathbf{x})$, $\omega_j, 1 \leq j \leq D$ and fixed $t' \in (0,1]$;
   - there exists a smooth map $\phi_j : (0,1] \to \mathbb{C}^n$ such that $\phi_j(1) = \omega_j$;
   - the points $\phi_j(t')$ are smooth isolated solutions of $H(\mathbf{x}, t') = 0$;
   - there exist no two integers $1 \leq j < k < D$ such that $\phi_j(t') = \phi_k(t')$ (i.e. associated solution paths do not cross);
3. every isolated solution of $f(\mathbf{x}) = 0$ is contained in the set of finite limits

$$\left\{ x \in \mathbb{C}^n : \|x\|_2 < \infty \text{ and } x = \lim_{t \to 0} \phi_j(t) \right\}.$$

Example 2.4.4. To give a concrete example of a good homotopy, we consider a *total-degree homotopy*. To solve the system given in (2.3), we construct

$$H(\mathbf{x}, t) := (1-t) \begin{bmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_n(\mathbf{x}) \end{bmatrix} + \gamma t \begin{bmatrix} g_1(\mathbf{x}) \\ \vdots \\ g_n(\mathbf{x}) \end{bmatrix} = 0.$$

In this case, given $d_i = \deg f_i$, $g_1, \ldots, g_n$ are chosen with degrees $d_1, \ldots, d_n$ respectively such that $g = \{g_1, \ldots, g_n\}$ is an "easy" system to solve with $D := d_1 \cdots d_n$ distinct root and isolated roots. One possible example is given by $g_i = z_i^{d_i} - 1$. We also note that $\gamma \neq 0$ is chosen to be a random complex number, which in practice is obtained by picking a number from the uniform distribution in a small band around the unit circle. The introduction of this γ, known as the *gamma trick*, guarantees with probability one that $H(\mathbf{x}, t)$ is a good homotopy [11, Sec. 2.1.4].

The next idea to consider is how to use a homotopy $H(\mathbf{x}, t)$ to track from $t = 1$ to $t = 0$. To formalize this idea, we start with a family of functions

$$H(\mathbf{x}; t) = \begin{bmatrix} H_1(\mathbf{x}; t) \\ \vdots \\ H_n(\mathbf{x}; t) \end{bmatrix} = 0$$

where each H_i is polynomial in the variables $\mathbf{x} \in \mathbb{C}^n$ and analytic in the parameter $t \in \mathbb{C}$. We also have a differentiable map $\mathbf{u} : t \in [0,1] \to \mathbf{x} \in \mathbb{C}^n$ such that, for $t \in (0,1]$:

- $H(\mathbf{u}(t), t) = 0$
- the Jacobian $JH(\mathbf{x}, t)$ with respect to the variables $x_1, \ldots, x_n$ has rank n for all points $(\mathbf{u}(t), t)$.

With this setup given $\mathbf{u}(1) = \mathbf{u}_0$, a prescribed point from the start system, and $\mathbf{u}(0) = \mathbf{u}^*$, the solution of the target system corresponding to this particular path, the problem of path tracking becomes

equivalent to solving the initial value problem with $\mathbf{u}(1) = \mathbf{u}_0$ given by the Davidenko differential equation

$$\frac{\partial H(\mathbf{u}(t),t)}{\partial t} + \sum_{i=1}^{n} \frac{\partial H(\mathbf{u}(t),t)}{\partial x_i} \frac{d u_i(t)}{d t} = 0. \tag{2.4}$$

Using the Jacobian notation, (2.4) is equivalent to

$$\frac{\partial H(\mathbf{u}(t),t)}{\partial t} + JH(\mathbf{u}(t),t)\frac{d\mathbf{u}(t)}{d t} = 0, \tag{2.5}$$

which can be rewritten (since JH($\mathbf{u}$(t),t) has full rank and is invertible) as

$$\frac{d\mathbf{u}(t)}{d t} = -[JH(\mathbf{u}(t),t)]^{-1}\frac{\partial H(\mathbf{u}(t),t)}{\partial t}.$$

There are many tracking algorithms from numerical analysis that solve the problem formulation of (2.5). The simplest version of predictor-corrector tracking one might consider would be a combination of Euler's method and Newton's method. Euler's method computes approximations starting at $\mathbf{u}_0$ at $t_0 = 1$ and steps by

$$\mathbf{u}_{i+1} := \mathbf{u}_i - JH(\mathbf{u}_i, t_i)^{-1}\frac{\partial H(\mathbf{u}_i, t_i)}{\partial t}(t_{i+1} - t_i).$$

We note that while this method takes prediction steps by stepping along the tangent line to the solution path and higher order methods prove more precise in practice. With many of these methods, the next step is a corrector step using Newton's method for $H(\mathbf{u}, t_{i+1})$ starting at $\mathbf{p}_0 = \mathbf{u}_{i+1}$ by

$$\mathbf{p}_{i+1} := \mathbf{p}_i - [JH(\mathbf{p}_i, t_{i+1})]^{-1}H(\mathbf{p}_i, t_{i+1}).$$

Remark 2.4.5. When a solution in our target system is singular, more complicated tracking methods are required in the limit as $t \to 0$. We do not provide all of the details here, but they can be found in [11] and are automatically implemented in the Bertini software utilized later in this paper [10].

2.5 Witness Sets

In this section, we continue our discussion of some main ideas from numerical algebraic geometry following [11]. In particular, we consider positive-dimensional algebraic sets and creating a data structure for them which corresponds to the homotopy methods detailed in the previous section. The key is the notion of *witness sets*, which will rely on the idea of slicing an algebraic set with a general linear space (recalling the idea of general from Definition 2.1.11), seen in the following.

Theorem 2.5.1. *Given an irreducible algebraic set $V \subset \mathbb{C}^n$ of dimension m, a general linear space L of dimension k will intersect V in an algebraic set of dimension $k + m - n$.*

We recall from the previous section that our homotopy path-tracking techniques require a 0-dimensional algebraic set of isolated solutions. Therefore, a natural approach using the result of Theorem 2.5.1 is to intersect an irreducible algebraic set $V \subset \mathbb{C}^n$ of dimension m with an appropriate general linear space L such that the intersection is 0-dimensional, i.e. $k+m-n=0$. Rearranging, we see that choosing a random linear space of codimension m will give the desired result with probability one. This gives the approach for the following construction of a witness set.

Definition 2.5.2. If an algebraic set $V \subset \mathbb{C}^n$ is equidimensional with $\dim(V)=k$, a *witness set* for V is the triple (F, L, W) such that

- $F \subset \mathbb{C}[\mathbf{x}]$ is a *witness system* for V, i.e. each irreducible component of $\mathcal{V}$ is an irreducible component of $\mathcal{V}(F)$,
- $L \subset \mathbb{C}[\mathbf{x}]$ is a *linear system* where $\mathcal{V}(L)$ is a linear space of codimension k that intersects V transversely,
- $W \subset \mathbb{C}^n$ is a *witness point set* which is equal to $V \cap \mathcal{V}(L)$.

We note that the number of points in the witness point set W in the above definition will determine the number of paths we need to track with our homotopy methods from the previous section. Since it has an impact on the complexity of the numerical algebraic geometry computations, we define the following invariant.

Definition 2.5.3. The *degree* of a m-dimensional irreducible algebraic set $V \subset \mathbb{C}^n$ is the number of points in the 0-dimensional intersection of V with a general linear space of codimension m.

Algorithm 2.5.1 MEMBERSHIPTEST

Input: $\mathbf{p} \in \mathbb{C}^n$ and $D = (G, L, W)$ a witness set for some equidimensional algebraic set $V = \mathcal{V}(f_1, \ldots, f_s) \subset \mathbb{C}^n$.

Output: TRUE if $\mathbf{p} \in V$ and FALSE if $\mathbf{p} \notin V$.

1. Choose generic linear polynomials L' with $\mathbf{p} \in \mathcal{V}(L')$ and $\dim(\mathcal{V}(L')) = \dim(\mathcal{V}(L))$.
2. $H(\mathbf{x}, t) := \left[f_1, \ldots, f_s, t\,L(\mathbf{x}) + (1-t)L'(\mathbf{x})\right]$.
3. Compute $W' := \lim_{t \to 0} \mathcal{V}(H(\mathbf{x}, t))$.

   `// Tracking the finitely many homotopy paths` $H(\mathbf{x}, t)$ `starting for` $t = 1$ `at the witness point set` W `computes the set` W' `consisting of the finite witness point set of` $\mathcal{V}(L')$ `at` $t = 0$.
4. If $\mathbf{p} \in W'$, return TRUE. Else, return FALSE.

Although the witness point set provides some information, a witness system is needed to perform additional computations on the limit such as testing membership. MEMBERSHIP TEST ALGORITHM 2.5.1 follows the approach of [11, Section 8.4] in order to do this.

2.6 Isosingular Deflation

Some of the techniques we have described in the previous section encounter difficulties when the multiplicity of an irreducible algebraic set is greater than one. In this section, we follow the approach of [11] to deflate these sets to have multiplicity one. In particular, we focus on the technique of *isosingular deflation*, with further details following from [28].

Definition 2.6.1. Given $f_1,\dots,f_n \in \mathbb{C}[\mathbf{x}]$ and $\mathbf{z} \in \mathbb{C}^n$ an isolated solution, the *multiplicty* of $\mathbf{z}$ with respect to $f_1,\dots,f_n$ is

$$\text{mult}(f_1,\dots,f_s,\mathbf{z}) := \dim \mathscr{O}_{\mathbf{z}}/\langle f_1,\dots f_n\rangle,$$

where $\mathscr{O}_{\mathbf{z}}$ is the ring of convergent power series centered at $\mathbf{z}$ and $\langle f_1,\dots,f_n\rangle$ an ideal in $\mathscr{O}_{\mathbf{z}}$.

We note that if $f_1,\dots,f_n$ generates a square polynomial system, as in Section 2.4, then for each isolated solution $\mathbf{z}$, $\text{mult}(f_1,\dots,f_s,\mathbf{z})$ will be exactly the number of paths ending at $\mathbf{z}$ in a total-degree homotopy.

Definition 2.6.2. Given $V \subset \mathbb{C}^n$ an irreducible component of $\mathscr{V}(f_1,\dots,f_s) \subset \mathbb{C}^n$ of dimension k, the *multiplicity* of V with respect to $f_1,\dots,f_s$ is

$$\text{mult}(f_1,\dots,f_s,V) := \text{mult}(f_1,\dots,f_s,L_1,\dots,L_k,\mathbf{z}),$$

where $L_1,\dots,L_k \in \mathbb{C}[\mathbf{x}]$ are general linear polynomials and $\mathbf{z} \in V \cap \mathscr{V}(L_1,\dots,L_k)$.

Example 2.6.3. Consider $f = x+y-1$ and $g = (x+y-1)^2$. Then the algebraic sets $V := \mathscr{V}(f) = \mathscr{V}(g)$ are both characterized by the line $x+y=1$ in the plane. However, the multiplicity of V with respect to f is 1 and the multiplicity with respect to g is 2.

Definition 2.6.4. Suppose we have the same conditions as in Definition 2.5.2. Then if each irreducible component of V has multiplicity one with respect to F, then F is called a *deflated witness system* and (F,L,W) is a *deflated witness set.*

If we encounter an algebraic set of multiplicity greater than one, as in Example 2.6.3, it is often necessary for us to consider the associated multiplicity one algebraic set. This is especially important for our use of Newton's method to remain numerically stable throughout the tracking of homotopy paths. The process of algorithmically constructing the associated multiplicity one algebraic set is called *deflation.*

Definition 2.6.5. Let $f_1,\dots,f_s \in \mathbb{C}[\mathbf{x}]$, $F_0 = \{f_1,\dots,f_s\}$, and $\mathbf{z} \in \mathscr{V}(F_0) \subset \mathbb{C}^n$. The *isosingular deflation operator* $\mathscr{D}$ is defined via

$$(F_1, \mathbf{z}) := \mathscr{D}(F_0, \mathbf{z})$$

where $F_1 \subset \mathbb{C}[\mathbf{x}]$ consists of F_0 and all $(r+1)\times(r+1)$ minors of the Jacobian matrix JF_0 for F_0 where $r = \operatorname{rank} JF_0(\mathbf{z})$. Thus, $\mathbf{z} \in \mathscr{V}(F_1)$, meaning that we can iterate this operator to construct a sequence of systems $F_j \subset \mathbb{C}[\mathbf{x}]$ with $(F_j, \mathbf{z}) = \mathscr{D}(F_{j-1}, \mathbf{z}) = \mathscr{D}^j(F_0, \mathbf{z})$ for $j \geq 1$.

We say that $F \subset \mathbb{C}[\mathbf{x}]$ is the *isosingular deflation* of F_0 at $\mathbf{z}$ if there exists a minimal $j \geq 0$ such that $(F, \mathbf{z}) = \mathscr{D}^j(F_0, \mathbf{z})$ and $\dim \operatorname{NullSpace}(JF(\mathbf{z})) = \dim_F(\mathbf{z})$, where $\dim_F(\mathbf{z})$ is the maximal dimension of the irreducible components of $\mathscr{V}(F)$ containing $\mathbf{z}$ (called the *local dimension* of $\mathbf{z}$ with respect to F).

Using the deflation operator, we can now formally define the isosingular sets and singular points of our algebraic set in this context.

Definition 2.6.6. Let $f_1,\dots,f_s \in \mathbb{C}[\mathbf{x}]$, $F_0 = \{f_1,\dots,f_s\}$, and $\mathbf{z} \in \mathscr{V}(F_0) \subset \mathbb{C}^n$. Let $\mathscr{D}$ be the isosingular deflation operator defined in Definition 2.6.5. We define

- The *deflation sequence* of F_0 at $\mathbf{z}$ is $\{d_k(F_0, \mathbf{z})\}_{k=0}^{\infty}$ where

  $$d_k(F_0, \mathbf{z}) = \operatorname{dnull}(F_k, \mathbf{z}) := \dim \operatorname{NullSpace} JF_k(\mathbf{z})$$

  with JF_k the Jacobian matrix of F_k with $(F_k, \mathbf{z}) = \mathscr{D}^k(F_0, \mathbf{z})$.

- Let $V \subset \mathscr{V}(F_0)$ be a non-empty irreducible algebraic set. Then V is an *isosingular set* of F_0 if there exists a sequence $\{c_k\}_{k=1}^{\infty}$ such that V is an irreducible component of

  $$\overline{\{\mathbf{z} \in \mathscr{V}(F_0) : d_k(F_0, \mathbf{z}) = c_k, k \in \mathbb{N}\}}.$$

- Let $V \subset \mathscr{V}(F_0)$ be a non-empty irreducible algebraic set. Then $\operatorname{Iso}_{F_0}(V)$ is the unique isosingular set with respect to F_0 containing V such that $\operatorname{Iso}_{F_0}(V)$ and V have the same deflation sequence with respect to F_0.

- Let V be an isosingular set for F_0. The set of *singular points of* V with respect to F_0 is

  $$\operatorname{Sing}_{F_0}(V) = \left\{\mathbf{z} \in V : \{d_k(F_0, \mathbf{z})\}_{k=0}^{\infty} \neq \{d_k(F_0, V)\}_{k=0}^{\infty}\right\}.$$

  Here, $d_k(F_0, V)$ is meant for a generic point in V.

- The *local dimension* of $\mathbf{z}$ with respect to F_0, denoted by $\dim_{F_0}(\mathbf{z})$, is the maximal dimension of the irreducible components of $\mathscr{V}(F_0)$ containing $\mathbf{z}$.

Remark 2.6.7. Since each step of isosingular deflation adds functions to our original f, the dimension of the nullspace being computed cannot decrease. Therefore, the deflation sequence is a nonincreasing sequence of nonnegative integers which must reach its limit after finitely many iterations.

Using the definition, we note that the key to this type of deflation is isosingular sets, i.e. irreducible components of an algebraic set with the same multiplicity/singularity structure. Given a polynomial system $f_1, \ldots, f_s \in \mathbb{C}[\mathbf{x}]$, we observe some properties about its isosingular sets:

1. there are finitely many isosingular sets $V_1, \ldots, V_k \subset \mathcal{V}(f_1, \ldots, f_s)$,
2. given $\mathbf{z} \in \mathcal{V}(f_1, \ldots, f_s)$, there is a unique $i \in \{1, \ldots, k\}$ such that $\mathbf{z} \in V_i$ and the singularity structure of $\mathbf{z}$ and a generic point on V_i are the same,
3. each isosingular set V_i is deflatable, i.e. we can construct a polynomial system g_i such that V_i is an irreducible component of $\mathcal{V}(g_i)$ that is multiplicity one as a component of $g_i^{-1}(0)$.

We note that (2) says that any point in an algebraic set V is a smooth point on a unique isosingular subset of V. Furthermore, (3) suggests an algorithmic approach for computing the isosingular deflation of V: first obtain the isosingular sets of V, and then deflate each individually. We refer to this process as isosingular deflation as well.

We next detail some particular results on isosingular sets which will be important to our methods going forward. To compute deflated witness sets, we will utilize ISOSINGULAR DEFLATION ALGORITHM 2.6.1, which follows the approach of [11, Section 13.2] and [28, Algorithm 6.3].

It is important to consider what properties of the original algebraic set are preserved when we apply isosingular deflation. The following theorem states that the singular points of an algebraic set are preserved under isosingular deflation.

Theorem 2.6.8. *[28, Theorem 5.9] Let V be an isosingular set for F_0 as in Definition 2.6.6. Then if $\mathbf{z} \in V$ and $\mathbf{z} \in \mathrm{Sing}(\mathcal{V}(F_0))$ then $\mathbf{z} \in \mathrm{Sing}_{F_0}(V)$.*

Finally, we have the following theorem which gives an isosingular deflation approach for constructing witness sets of the intersection of a known witness set with another algebraic set.

Theorem 2.6.9. *[29, Theorem 6.2] Given $g_1, \ldots, g_r \in \mathbb{C}[\mathbf{x}]$, let Z be a union of irreducible components of $\mathcal{V}(g_1, \ldots, g_r)$. Suppose $f_1, \ldots, f_s \in \mathbb{C}[\mathbf{y}]$, $F(\mathbf{x}, \mathbf{y}) = \{g_1(\mathbf{x}), \ldots, g_r(\mathbf{x}), f_1(\mathbf{y}), \ldots, f_s(\mathbf{y})\}$, $\Delta = \{(\mathbf{x}, \mathbf{x}) : \mathbf{x} \in \mathbb{C}^n\}$, and $\pi(\mathbf{x}, \mathbf{y}) = \mathbf{x}$. If $A \subset Z \cap \mathcal{V}(f_1, \ldots, f_s)$ is an irreducible component, then there exists a nonempty Zariski open set $U \subset A$ such that for all $\mathbf{p} \in U$, A is an irreducible component of $\pi\Big(\mathrm{Iso}_F((\mathbf{p}, \mathbf{p})) \cap \Delta\Big)$.*

Algorithm 2.6.1 ISOSINGULARDEFLATION

Input: $F_0 = \{f_1, \ldots, f_s\} \subset \mathbb{C}[\mathbf{x}]$, $\mathbf{z} \in \mathscr{V}(F_0) \subset \mathbb{C}^n$.

Output: $F \subset \mathbb{C}[\mathbf{x}]$, an isosingular deflation of F_0 at $\mathbf{z}$.

1. Set $k := 0$.
2. If $d := d_0(F_0, \mathbf{z}) = 0$, exit and return $F := F_0$.
3. Loop
   (a) Set $m := |F_k|$.
   (b) Choose generic $A \in \mathbb{C}^{n-d \times r}$.
   (c) Compute $[R_1, \ldots, R_m] := R := A \cdot F_k$.
   (d) Loop
   Choose generic linear polynomials $L_1, \ldots, L_d \in \mathbb{C}[\mathbf{x}]$ with $\mathbf{z} \in \mathscr{V}(L_1, \ldots, L_d)$ and generic $\lambda := (\lambda_1, \cdots, \lambda_d) \in \mathbb{C}^d$.
   $H(\mathbf{x}, t) := [R_1, \ldots, R_m, L_1 + (1-t)\lambda_1, \ldots, L_d + (1-t)\lambda_d]$.
   If $|\mathscr{V}(H(\mathbf{x}, 1))| = \infty$, go back to start of loop.
   (e) Compute $\mathbf{z}^* := \lim_{t \to 0} \mathbf{z} \in H(\mathbf{x}, t)$ along the homotopy starting at $t = 1$.
   `// Tests whether the deflation sequence has stabilized, sampling` $\mathscr{V}(R)$ `by shifting the general linear space it intersects, tracking` $\mathbf{z} \in \mathscr{V}(R) \cap \mathscr{V}(L_1, \ldots, L_d)$ `when` $t = 1$ `to` $\mathbf{z}^* \in \mathscr{V}(R) \cap \mathscr{V}(L_1 + \lambda_1, \ldots, L_d + \lambda_d)$ `when` $t = 0$.
   (f) If $\mathbf{z}^* \in \mathscr{V}(F_k)$, exit and return $F := F_k$.
   (g) Increment $k := k + 1$.
   (h) Construct F_k by Definition 2.6.5 and return to start of loop.

We provide the following illustrative example for the theorem.

Example 2.6.10. Let $g(x, y, z) = (x + y + z)y$ and $Z = \mathscr{V}(x + y + z)$. Suppose $f(x, y, z) = y$ and note that $Z = Z \cap \mathscr{V}(f) = \mathscr{V}(x + z, y)$ is irreducible. We construct

$$F(x, y, z, x', y', z') = [g(x, y, z) = (x + y + z)y, f(x', y', z') = y'].$$

Choose a generic witness point $\mathbf{p} = (a, 0, -a) \in U \subset A$ for some $a \in \mathbb{C}$. Then we compute the deflation sequence of $(a, 0, -a, a, 0, -a)$ with respect to F as $5, 3, 3, \ldots$ such that

$$\mathrm{Iso}_F((\mathbf{p}, \mathbf{p})) = \{(b, 0, -b, c, 0, d) : b, c, d \in \mathbb{C}\}.$$

The polynomial system defining this 3-dimensional isosingular set is given by adding the 2×2 minors of the Jacobian of F to F, giving

$$G(x, y, z, x', y', z') = \begin{bmatrix} (x+y+z)y \\ y' \\ y \\ x+2y+z \\ y \end{bmatrix}.$$

By Theorem 2.6.9, A is an irreducible component of $\mathcal{V}(G(x, y, z, x, y, z))$ and $G := G(x, y, z, x, y, z)$ suffices as a witness system for A. We note that in this example, removing the redundancies in G would in fact show that $[x+z, y]$ is sufficient as a witness system for A.

CHAPTER

3

COMPUTATION OF REAL SMOOTH POINTS: EQUIDIMENSIONAL CASE

In this chapter, we describe our approach for computing at least one smooth real point on every connected component of an equidimensional atomic semi-algebraic set.

3.1 Related Work

There are many approaches in the literature to compute at least one real point on every connected component of a semi-algebraic set. Methods using projections to obtain a cell decomposition based on sign conditions go back to Collins' Cylindrical Algebraic Decomposition (CAD) algorithm described in [15]. Improved symbolic methods using critical points or generalized critical points of functions along with infinitesimals and randomization can be found in [42, 1, 43, 19]. The current state of the art deterministic symbolic algorithm is given in [8, Alg. 13.3] which computes sample points on each connected component of all realizable sign conditions of a polynomial system and gives a complexity analysis. The most recent application of this technique is in [48, 49] where the authors compute smooth points on real algebraic sets in order to compute the real radical of polynomial systems and analyze complexity. Alternatively, a homotopy-based approach computing the critical points of the distance function from a generic point or a line is presented in [24, 57].

Another line of work has been developed in parallel which specifically focuses on computing critical points while utilizing the tool of *polar varieties*, introduced and developed in [3, 44, 4, 5, 6, 2, 45]. It is important to note, however, that all of these methods are only guaranteed to find real

points on every connected component of a semi-algebraic set, rather than real *smooth* points.

3.2 Main Results

Given an atomic semi-algebraic set, we would like to guarantee the computation of at least one real smooth point on each of its connected components. If necessary, we first apply Proposition 2.2.14 to turn the atomic semi-algebraic set into a real algebraic set. Also, we apply Proposition 2.2.15 to obtain a closed and bounded real algebraic set (if necessary). Therefore, we can now go forward assuming we are considering a closed and bounded real algebraic set.

An ingredient we will use is given by the following theorem that was proven in [39, Theorem 12.6.1].

Theorem 3.2.1. *Let $V \subset \mathbb{C}^n$ be an irreducible algebraic set and let $V_{\mathbb{R}} := V \cap \mathbb{R}^n$. Then*

$$\dim_{\mathbb{R}}(V_{\mathbb{R}}) = \dim_{\mathbb{C}}(V)$$

if and only if there exists some $\mathbf{z} \in V_{\mathbb{R}}$ that is smooth.

To first present the main idea of our approach in its simplest form, we will begin with the case where the algebraic relaxation of our real algebraic set is equidimensional.

Lemma 3.2.2. *Let $f_1, \ldots, f_s \in \mathbb{R}[\mathbf{x}]$ and assume that $V := \mathscr{V}(f_1, \ldots, f_s) \subset \mathbb{C}^n$ is equidimensional of dimension $n-s$. Let $g \in \mathbb{R}[\mathbf{x}]$ such that $\dim\big(V \cap \mathscr{V}(g)\big) < n-s$. Then either $\big(V \setminus \mathscr{V}(g)\big) \cap \mathbb{R}^n = \emptyset$ or g restricted to $V \cap \mathbb{R}^n$ attains a non-zero extreme value on each bounded connected component of $\big(V \setminus \mathscr{V}(g)\big) \cap \mathbb{R}^n$.*

Proof. Suppose $\big(V \setminus \mathscr{V}(g)\big) \cap \mathbb{R}^n \neq \emptyset$ and let C be a bounded connected component of the set $(V \setminus \mathscr{V}(g)) \cap \mathbb{R}^n$. Since $C \not\subset \mathscr{V}(g)$, there exists some $x \in C$ such that $g(x) \neq 0$. Let $\overline{C}$ be the Euclidean closure of C as in Definition 2.2.3. Then $\overline{C} \subset V \cap \mathbb{R}^n$ is closed and bounded, and g vanishes identically on $\overline{C} \setminus C$. By the Extreme Value Theorem, g attains both a minimum and a maximum on $\overline{C}$. Since g is not identically zero on $\overline{C}$, either the minimum or the maximum value of g on $\overline{C}$ must be nonzero, so g attains a non-zero extreme value on C. □

Using this lemma, we will prove the following theorem.

Theorem 3.2.3. *Let V be as in Lemma 3.2.2. Suppose $g \in \mathbb{R}[\mathbf{x}]$ satisfies the following conditions:*

1. $\operatorname{Sing}(V) \cap \mathbb{R}^n \subset \mathscr{V}(g)$;
2. $\dim\big(V \cap \mathscr{V}(g)\big) < n-s$.

Then the set of points where g restricted to $V \cap \mathbb{R}^n$ attains its extreme values intersects each bounded connected component of $\big(V \setminus \operatorname{Sing}(V)\big) \cap \mathbb{R}^n$.

Proof. Suppose $\big(V \setminus \mathrm{Sing}(V)\big) \cap \mathbb{R}^n \neq \emptyset$. By Theorem 3.2.1, $\dim_{\mathbb{R}}(V \cap \mathbb{R}^n) = n-s$. So by (2), $\big(V \setminus \mathscr{V}(g)\big) \cap \mathbb{R}^n \neq \emptyset$. By (1), $\big(V \setminus \mathscr{V}(g)\big) \cap \mathbb{R}^n \subset \big(V \setminus \mathrm{Sing}(V)\big) \cap \mathbb{R}^n$, so the bounded connected components of $\big(V \setminus \mathrm{Sing}(V)\big) \cap \mathbb{R}^n$ are subsets of the bounded connected components of $\big(V \setminus \mathscr{V}(g)\big) \cap \mathbb{R}^n$. By Lemma 3.2.2, g restricted to $V \cap \mathbb{R}^n$ attains a non-zero extreme value on each bounded connected component of $\big(V \setminus \mathscr{V}(g)\big) \cap \mathbb{R}^n$, thus yielding a point in every bounded connected component of $\big(V \setminus \mathrm{Sing}(V)\big) \cap \mathbb{R}^n$. □

The setup and result of this theorem suggest the basic outline of a method for computing a smooth point on every connected component of an equidimensional compact real algebraic set $\mathscr{V}(f_1, \ldots, f_s) \cap \mathbb{R}^n$ (recalling that if we were originally given an atomic semi-algebraic set, we first applied Propositions 2.2.14 and 2.2.15):

1. Construct a g that satisfies (1) and (2) in Theorem 3.2.3;
2. Compute the critical points of g in $\mathscr{V}(f_1, \ldots, f_s) \cap \mathbb{R}^n$ (using Lagrange multipliers);
3. Select the critical points that are not in $\mathscr{V}(g)$.

3.3 Illustrative Examples

In Chapter 5, we give REAL SMOOTH POINT ALGORITHM 5.1.1, which computes real smooth points in this way when $\mathscr{V}(f_1, \ldots, f_s)$ is not necessarily equidimensional by using deformations and limits. However, the same algorithm can be used in the equidimensional case without deformation. For now, we present a few illustrative examples to show the idea, as well as the utility of our results on previously difficult problems, even in the equidimensional case.

Example 3.3.1. An example of a real curve with two singular cusps is often referred to as "Thom's lips," e.g. $f = y^2 - (x(1-x))^3$, as shown in Figure 3.1. A simple choice of g which satisfies the conditions of Theorem 3.2.3 is $g = x(1-x)$. Using Lagrange multipliers to optimize with respect to g results in two points $(0.5, \pm 0.125)$ plotted as red diamonds. Alternatively, the polynomial g can be constructed algorithmically, which we will see in Chapter 4, yielding, e.g. $g = 3(2x-1)(x(1-x))^2 + 2y$ which produces two points plotted as black circles, approximately $(0.5987, 0.1178)$ and $(0.4013, -0.1178)$. Both yield a real smooth point on each of the two connected components of $\big(V \setminus \mathrm{Sing}(V)\big) \cap \mathbb{R}^n$. We note that the first choice of g demonstrates that when $\mathrm{Sing}(V)$ is 0-dimensional, defining g as a product of a coordinate of these points will satisfy the conditions of Theorem 3.2.3. The second choice of g demonstrates the general method described later which works in every dimension.

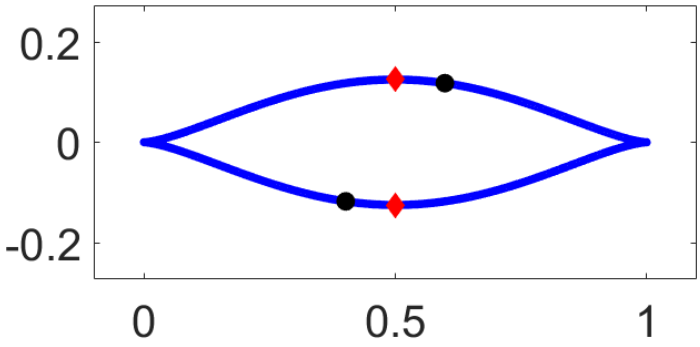
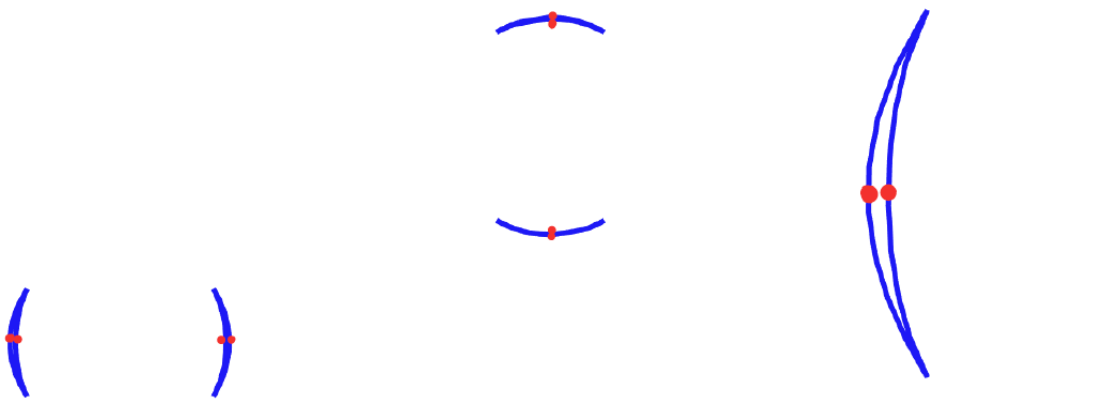


Figure 3.1: "Thom's lips"

The motivation for the above example is not just illustrative of our method, but also how we have improved on previous approaches. We recall that previous methods for computing points on semi-algebraic sets used critical points of the distance function. In [57, Example 2.3], Wu and Reid show how these methods often result in only singular critical points. For "Thom's lips", they prove that if we were to sample points on the curve by finding the critical points of the distance function with respect to some linear function, the probability of the critical points being exactly the singular points of the curve is 0.772. Furthermore, if we increase the steepness of the curve going into the cusps, this probability only increases. Rather than choosing a random linear function to optimize with respect to and only having a certain probability of success, our method constructs a g that will always work.

Example 3.3.2. An example of a real curve with multiple compact components with singular cusps can be constructed by taking $h = f_1^2 + \frac{1}{100} f_2^3$ where $f_1 = (x^2+y^2-1)((x-4)^2+(y-2)^2-1)$, the union of two circles, and $f_2 = (y-1/2)(y+1/2)(x-7/2)(x-9/2)$, the union of four lines, where two of the lines intersect one of the circles twice each and the other two intersect the other circle twice each. Then the curve $V := \mathcal{V}(h)$ has four compact components with two cusps on each component, as shown in Figure 3.2. As in Example 3.3.1, since $\mathrm{Sing}(V)$ is 0-dimensional, defining g as the product of the coordinates of the singular points satisfies Theorem 3.2.3. So we choose $g = (4x^2-3)(4y^2-1)(4x^2-32x+63)(4y^2-16y+13)$ and use Lagrange multipliers to optimize with respect to g, resulting in eight smooth points, seen in red in Figure 3.2 one on each of the connected components of $\big(V \setminus \mathrm{Sing}(V)\big) \cap \mathbb{R}^n$.

This example once again illustrates the simplicity of constructing a g which satisfies the necessary conditions. Furthermore, in [40, Proposition 3.2], Mork and Piene prove that this example is a case where sampling via critical points of the distance function will not work; due to the large number of singular points, sampling in this way will never result in a real smooth point on every connected component of the curve. We observe that our method does result in a real smooth point on each of the 8 connected components.



(a) all 4 components (b) zoomed in on 1 component

Figure 3.2: Real Curve with Singularities on Multiple Components

A final example illustrates how our approach works with a real surface. In particular, we see again that as long as Sing(V) is 0-dimensional, the construction of a g which satisfies Theorem 3.2.3 is straightforward.

Example 3.3.3. An example of a real surface with three singular points coming from semidefinite programming is sometimes referred to as the "Samosa," shown in Figure 3.3. Its defining equation is found by taking the determinant of a 3×3 matrix, resulting in $f = 2xy - x^2 - y^2 - z^2 + 1$. We note that the surface defined by f is not bounded, but by restricting to the bounded component shown in the figure we can apply the theorem. The most straightforward choice for g which satisfies the above conditions is $g = x^2 + y^2 + z^2 - 3$. Using Lagrange multipliers to optimize with respect to this g results in the red points in the figure. Alternatively, we could select a g which satisfies the theorem using the general approach detailed later in Chapter 4. In this case, we use $g = 2xy + 2xz + 2yz - 2x - 2y - 2z$ and optimization with Lagrange multipliers results in the blue points in the figure. Once again, the first choice of g demonstrates that when Sing(V) is 0-dimensional, defining g in a way that intersects each of the points, in this case a sphere, will satisfy conditions of Theorem 3.2.3.

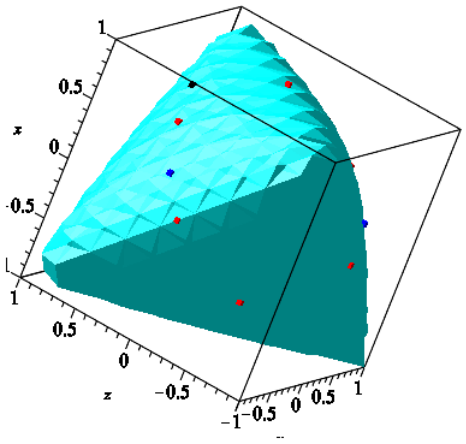


Figure 3.3: “Samosa"

3.4 Application to Kuramoto Model

In this section, we begin with a general description of the Kuramoto model problem, following the treatment of [16], and then describe our success in the proof of a previously conjectured result [58] for the specific case of $n = 4$.

The Kuramoto model, first proposed by Yoshiki Kuramoto in 1975 [32], provides a framework for studying synchronization in a variety of contexts. Specifically, it models the behavior of systems of coupled oscillators, i.e. oscillators that are connected in some way as a part of the same physical system. While Kuramoto originally created the dynamical system to describe systems of chemical and biological oscillators found in nature, the model (or variants) have since been used in a variety of applications across electrical engineering, neuroscience, physics, and even economics. A simple example to keep in mind throughout our study might be describing the oscillating behavior of a series of connected metronomes. The natural phenomenon of these coupled oscillators synchronizing (i.e. syncing their behavior to move at the same oscillation rate) is an intuitively interesting line of study. In fact, being able to accurately predict the synchronization of coupled oscillators becomes crucial in more complicated applications such as electrical engineering, where it is used in conducting stability assessments of the system being modeled.

The standard formulation of the dynamic model for $n \geq 2$ coupled oscillators is the system of first-order ordinary differential equations:

$$\frac{d\theta_i}{dt} = \omega_i - \frac{K}{n}\sum_{j=1}^{n}\sin(\theta_i - \theta_j) \qquad \text{for } i = 1,\ldots,n \tag{3.1}$$

where $K > 0$ is the uniform coupling strength, parameter ω_i is the natural frequency of the ith oscillator, and the unknown $\theta_i(t)$ is the phase angle of the ith oscillator at time t. We note that the

use of the sine function on the phase angle differences between the oscillators naturally models synchronization behavior.

A common generalization of the model allows for non-uniform coupling between oscillators. In particular, for coupling strengths $k=(k_1,\dots,k_n)\in\mathbb{R}^n_{>0}$, the ith and jth oscillators are coupled with strength k_ik_j, resulting in the model

$$\frac{d\theta_i}{dt}=\omega_i-\frac{1}{n}\sum_{j=1}^{n}k_ik_j\sin(\theta_i-\theta_j)\qquad\text{for } i=1,\dots,n. \tag{3.2}$$

We note that the standard Kuramoto model case described in (3.1) is given by the special case where $k=(\sqrt{K},\dots,\sqrt{K})$. Furthermore, the model given by (3.2) naturally lends itself to an equivalent description of the coupling strengths by a symmetric rank one matrix M where the (i,j)th entry is k_ik_j. Therefore, (3.2) is referred to as the rank-one coupled Kuramoto model. We note we could further generalize the model to a coupling matrix of arbitrary rank, a line of study motivated by applications to power flow equations. For our purposes, we study the simplest case of rank one given in (3.1), which corresponds to the equations for a lossless power flow system.

Many questions regarding the behavior of the Kuramoto model involve counting and locating the equilibria of the dynamical system. In the context of coupled oscillators that are constantly moving, equilibria refer to when all of the oscillators are moving at the same frequency. Sometimes this synchronization is referred to as the phase-locked state, since the relative differences between the phases of the oscillators will be constant as they continue to oscillate. In particular, we observe that summing the n differential equations in (3.1) gives the equivalence

$$\sum_{i=1}^{n}\frac{d\theta_i}{dt}=\sum_{i=1}^{n}\omega_i,$$

which implies that necessary conditions for equilibria are

$$\sum_{i=1}^{n}\frac{d\theta_i}{dt}=\sum_{i=1}^{n}\omega_i=0.$$

In a general dynamical model context, these equilibria are referred to as the real solutions to the steady-state equations of the system. Furthermore, since only angle differences between the phases of the oscillators matter, the standard approach to compute equilibria modulo a constant shift is to set $\theta_n=0$.

In order to reformulate the problem as a system of polynomial equations, we apply the trigonometric identity

$$\sin(\theta_i-\theta_j)=\sin(\theta_i)\cos(\theta_j)-\cos(\theta_i)\sin(\theta_j)$$

and set $s_i=\sin(\theta_i)$, $c_i=\cos(\theta_i)$. Combining this with the setup of the uniform rank one coupling

model from (3.1) gives the following polynomial system for the equilibria of the model

$$f(s,c,\omega)=\begin{cases}\omega_i-\frac{K}{n}\sum_{i=1}^{n}s_i c_j-s_j c_i & \text{for } i=1,\dots,n.\\ s_i^2+c_i^2-1\end{cases}$$

The maximum number of equilibria (i.e. real solutions to steady-state equations) for $n\geq 4$ remains an open problem. The following confirms the conjecture in [58] for $n=4$.

Theorem 3.4.1. *The maximum number of equilibria for the Kuramoto model with $n=4$ oscillators is* 10.

The steady-state equations for the $n=4$ Kuramoto model are

$$f_i(\theta;\omega)=\omega_i-\tfrac{1}{4}\textstyle\sum_{j=1}^{4}\sin(\theta_i-\theta_j)=0,\text{ for } i=1,\dots,4$$

parameterized by the natural frequencies $\omega_i\in\mathbb{R}$. Since only the angle differences matter, one can assume $\theta_4=0$ and observe a necessary condition for equilibria is

$$0=f_1+f_2+f_3+f_4=\omega_1+\omega_2+\omega_3+\omega_4,$$

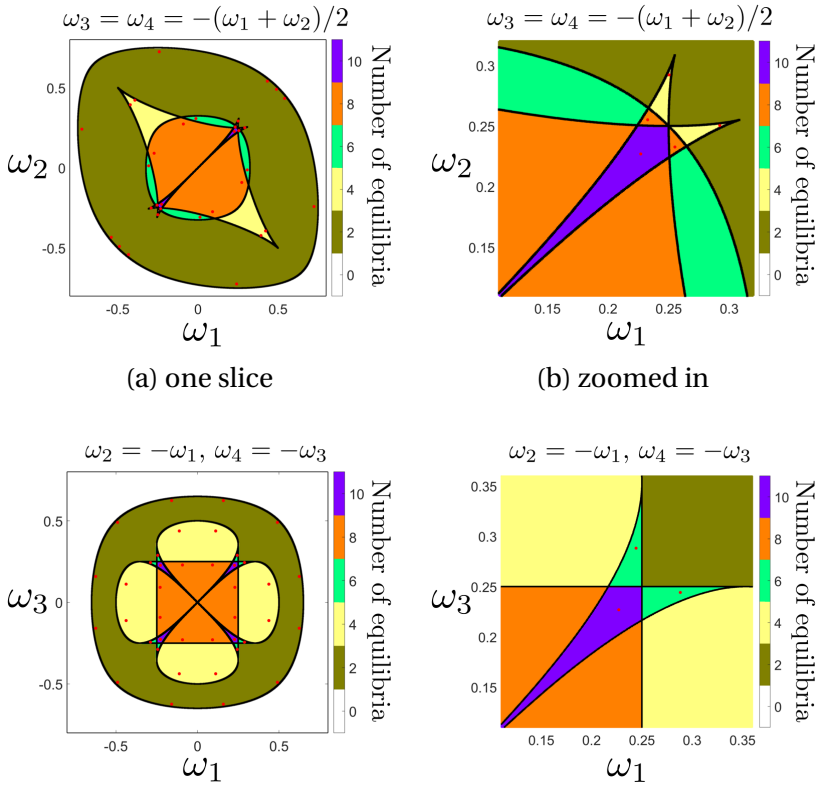
i.e., assume $\omega_4=-(\omega_1+\omega_2+\omega_3)$. Substituting $s_i=\sin(\theta_i)$ and $c_i=\cos(\theta_i)$ yields

$$f(s,c;\omega)=\left\{\omega_i-\tfrac{1}{4}\textstyle\sum_{j=1}^{4}(s_i c_j-s_j c_i),\, s_i^2+c_i^2-1,\text{ for } i=1,2,3\right\}$$

which is a polynomial system with variables $s=(s_1,s_2,s_3)$ and $c=(c_1,c_2,c_3)$, parameters $\omega=(\omega_1,\omega_2,\omega_3)$, and constants $s_4=0$ and $c_4=1$.

The goal is to compute the maximum number of isolated real solutions of $f=0$ as ω varies over $\mathbb{R}^3$. Let $D(\omega)$ be the discriminant polynomial of the system f, a polynomial in ω of degree 48. The number of real solutions of f is constant in each connected component of $\mathbb{R}^3\setminus\mathscr{V}(D)$. Since it is easy to see that there can be no real solutions if $|\omega_i|\geq\frac{n-1}{n}=0.75$, we need to compute at least one interior point in each of the bounded connected components of $\mathbb{R}^3\setminus\mathscr{V}(D)$. Applying Lemma 3.2.2 with $f=0$ and $g=D$, i.e., by computing the real solutions of $\nabla D=0$ and $D\neq 0$, accomplishes this task. Exploiting symmetry and utilizing `Bertini` ([10]), `alphaCertified` ([27]), and `Macaulay2` ([21]) all solutions have been found and certified. In fact, this computation showed that all real critical points of D arose, up to symmetry, along two slices shown in Figure 3.4.

In this figure, the discriminant D is seen as the black lines separating the different colored connected components of the parameter space. We observe that the red points, at least one in each of the bounded connected components, are the real smooth points we have computed via the theory of Lemma 3.2.2. In particularly, we note that as we guaranteed, we result in critical points even in the small components of the parameter space seen in the zoomed in figures. This is notable because previous techniques for searching the parameter space for different components consisted of sampling using a grid. While the grid could be refined, it could not be guaranteed to



(a) one slice (b) zoomed in

(c) other slice (d) zoomed in

Figure 3.4: Compact connected regions and critical points for the Kuramoto model with $n = 4$

pick up smaller components of the space, and one would not know for sure whether they had missed something.

A similar computation then computed and certified the number of real solutions to $f = 0$ at the points ω found above, showing that the maximum number of equilibria is 10. All code used in these computations is available at `dx.doi.org/10.7274/r0-5c1t-jw53`.

CHAPTER

4

PERTURBATIONS OF REAL ALGEBRAIC SETS

In this chapter, we begin to construct the tools for an algorithm using the same ideas as in Section 3.2, but in full generality so the results will hold when $\mathscr{V}(f_1,\ldots,f_s)$ is not equidimensional, i.e. there are some components of dimension greater than $n-s$. To do this, a standard approach (which can be seen in [46] and [7]) is to perturb the defining polynomials of the algebraic set by infinitesimals to obtain an equidimensional variety that has a limit as the infinitesimals approach zero that is equidimensional and contains the $n-s$ dimensional components of $\mathscr{V}(f_1,\ldots,f_s)$.

4.1 Perturbing the Sum of Squares of Polynomials

Assume $f_1,\ldots,f_s \in \mathbb{R}[\mathbf{x}]$ and $\varepsilon > 0$ a real infinitesimal. Let $F = f_1^2+\cdots+f_s^2$ and note that $\mathscr{V}(f_1,\ldots,f_s)\cap \mathbb{R}^n = \mathscr{V}(F)\cap\mathbb{R}^n$. Consider $V_\varepsilon := \mathscr{V}(F-\varepsilon)$ as the perturbed version of the algebraic set $\mathscr{V}(f_1,\ldots,f_s)$. One reason for us to perturb our algebraic set is to obtain smoothness.

Lemma 4.1.1. *[42, Lemma 3.5] V_ε is a smooth hypersurface.*

Recall that $\mathbb{R}\langle\varepsilon\rangle$ denotes the field of Puiseux series over $\mathbb{R}$ as in Definition 2.2.20. The following result [8, Proposition 12.36] states that semi-algebraicity is preserved as ε limits to 0.

Lemma 4.1.2. *Let $S \subset \mathbb{R}\langle\varepsilon\rangle^n$ be a semi-algebraic set. Then $\lim_{\varepsilon\to 0}(S)$ is a closed semi-algebraic set. Furthermore, if S is bounded and connected, then $\lim_{\varepsilon\to 0}(S)$ is connected.*

The statement of the next proposition on the limits of perturbed connected components of a real algebraic set follows the approach of the unpublished work [46], so we restate and prove it here.

Proposition 4.1.3. *Assume $f_1,\dots,f_s \in \mathbb{R}[\mathbf{x}]$, $F = f_1^2+\cdots f_s^2$, and $\varepsilon > 0$ a real infinitesimal. Let $V_\varepsilon := \mathscr{V}(F-\varepsilon) \subset \mathbb{C}\langle\varepsilon\rangle^n$ and $V := \lim_{\varepsilon\to 0} V_\varepsilon \subset \mathbb{C}^n$. Suppose C is a connected component of $V \cap \mathbb{R}^n$. Then:*

1. *there exist connected components $C_{\varepsilon,1},\dots,C_{\varepsilon,l}$ of $V_\varepsilon \cap \mathbb{R}\langle\varepsilon\rangle^n$ such that $C = \cup_{i=1}^{l} \lim_{\varepsilon\to 0} C_{\varepsilon,i}$;*
2. *if C is bounded by some open ball $B \subset \mathbb{R}^n$, then $C_{\varepsilon,i}$ is bounded by and does not intersect $Ext(B \cap \mathbb{R}\langle\varepsilon\rangle^n)$ for $1 \le i \le l$.*

Proof. Let $\mathbf{z} \in C$. Then there exists a connected component $S \subset \mathbb{R}^n \setminus \mathscr{V}(f_1,\dots,f_s)$ such that $\mathbf{z} \in \overline{S}$. Then there exists some $\mathbf{p} \in S$ such that $\mathbf{p} \in B(\mathbf{z}, r)$ for $r > 0$. We note that since F is nonnegative over $\mathbb{R}^n$, $F(\mathbf{p}) > 0$. By Lemma 2.2.9, there exists a continuous semi-algebraic function $\gamma : [0,1] \to \overline{S}$ with $\gamma(0) = \mathbf{z}$, $\gamma(1) = \mathbf{p}$ and $\gamma(t) \in S$ for all $t \in (0,1]$. Note that we have $(F\circ\gamma)(0) = F(\mathbf{z}) = 0$ and $(F\circ\gamma)(1) = F(\mathbf{p}) > 0$.

Let $F' := \mathrm{Ext}(F, \mathbb{R}\langle\varepsilon\rangle^n)$ and $\gamma' := \mathrm{Ext}(\gamma, \mathbb{R}\langle\varepsilon\rangle)$, where Ext is the extension as in Definition 2.2.24. By the infinitesimal property of ε, $\mathrm{Ext}([0,1], \mathbb{R}\langle\varepsilon\rangle)$ includes all Puiseux series with constant term in $[0,1]$. The Intermediate Value Theorem (Theorem 2.2.22) applied to $F'\circ\gamma'$ gives some $t_\varepsilon \in \mathrm{Ext}([0,1], \mathbb{R}\langle\varepsilon\rangle)$ such that $(F'\circ\gamma')(t_\varepsilon) = \varepsilon$. Let $\mathbf{z}_\varepsilon := \gamma'(t_\varepsilon)$. Then $\lim_{\varepsilon\to 0}(\mathbf{z}_\varepsilon) = \mathbf{z}$. Let $C_{\mathbf{z}_\varepsilon}$ be the connected component of $V_\varepsilon \cap \mathbb{R}\langle\varepsilon\rangle^n$ containing $\mathbf{z}_\varepsilon$, and associate $\mathbf{z}$ to that component. Since there are finitely many connected components of $V_\varepsilon \cap \mathbb{R}\langle\varepsilon\rangle^n$, as we run through all $\mathbf{z} \in C$, a subset of these connected components are of the form $C_{\mathbf{z}_\varepsilon}$ with $\lim_{\varepsilon\to 0}(\mathbf{z}_\varepsilon) = \mathbf{z}$ for some $\mathbf{z} \in C$. We denote these components of $V_\varepsilon \cap \mathbb{R}\langle\varepsilon\rangle^n$ by $C_{\varepsilon,1},\dots,C_{\varepsilon,l}$. Clearly $C \subset \cup_{i=1}^{l} \lim_{\varepsilon\to 0} C_{\varepsilon,i}$.

Now suppose $\mathbf{z}' \in \lim_{\varepsilon\to 0} C_{\varepsilon,i}$ for some $1 \le i \le l$. Then there exists some $\mathbf{z}'_\varepsilon \in C_{\varepsilon,i}$ such that $\lim_{\varepsilon\to 0} \mathbf{z}'_\varepsilon = \mathbf{z}'$. We recall that $C_{\varepsilon,i}$ is associated to some $\mathbf{z} \in C$, i.e. there exists some $\mathbf{z}_\varepsilon \in C_{\varepsilon,i}$ such that $\lim_{\varepsilon\to 0} \mathbf{z}_\varepsilon = \mathbf{z}$. Since $C_{\varepsilon,i}$ is connected, there exists some continuous semi-algebraic function $\gamma : [0,1] \to C_{\varepsilon,i}$ such that $\gamma(0) = \mathbf{z}'_\varepsilon$ and $\gamma(1) = \mathbf{z}$ and $\Gamma := \gamma([0,1])$ is a connected semi-algebraic set. By Theorem 2.2.8, Γ is closed and bounded and by Lemma 4.1.2, $\lim_{\varepsilon\to 0}\Gamma$ is connected. Furthermore, we note that $\lim_{\varepsilon\to 0}\Gamma \subset V \cap \mathbb{R}^n$, $\lim_{\varepsilon\to 0}(\gamma(0)) = \mathbf{z}'$, and $\lim_{\varepsilon\to 0}(\gamma(1)) = \mathbf{z}$. Hence $\mathbf{z} \in C$ and $\cup_{i=1}^{l} \lim_{\varepsilon\to 0} C_{\varepsilon,i} \subset C$, thus we have shown (1) as required.

Now suppose C is bounded by some ball $B \subset \mathbb{R}^n$ and does not intersect the boundary of B. Let $\mathbf{z}_\varepsilon \in C_{\varepsilon,i}$ such that $\lim_{\varepsilon\to 0} \mathbf{z}_\varepsilon \in C$. For sake of contradiction, suppose $\mathbf{z}'_\varepsilon \in \overline{C_{\varepsilon,i} \setminus \mathrm{Ext}(B \cap \mathbb{R}\langle\varepsilon\rangle)}$. Since $C_{\varepsilon,i}$ is connected, there exists a continuous semi-algebraic function $\gamma : [0,1] \to C_{\varepsilon,i}$ such that $\gamma(0) = \mathbf{z}_\varepsilon, \gamma(1) = \mathbf{z}'_\varepsilon$ and $\Gamma := \gamma([0,1])$ is a connected semi-algebraic set. The Intermediate Value Theorem (Theorem 2.2.22) applied to the polynomial defining the boundary of B gives some $t_\varepsilon \in [0,1]$ such that $\gamma(t_\varepsilon)$ is in the boundary of $\mathrm{Ext}(B, \mathbb{R}\langle\varepsilon\rangle^n)$. Then $\lim_{\varepsilon\to 0}\gamma(t_\varepsilon)$ is in the boundary of B.

By Theorem 2.2.8, Γ is closed and bounded and by Lemma 4.1.2, $\lim_{\varepsilon\to 0}\Gamma$ is connected. Then $\lim_{\varepsilon\to 0}\Gamma \subset C$ and thus $\lim_{\varepsilon\to 0}\gamma(t_\varepsilon) \in C$. But this contradicts C intersecting the boundary of B. Thus $C_{\varepsilon,i}$ is bounded by (and does not intersect) $\mathrm{Ext}(B, \mathbb{R}\langle\varepsilon\rangle^n)$. □

A natural question might be why it is necessary for us to consider perturbing the sum of squares of the polynomials, rather than each polynomial separately, when we are working over the real numbers. Besides the nonnegativity of F playing a role in the proof of the previous theorem, the following example illustrates what can go wrong over the reals.

Example 4.1.4. This example illustrates that for $s>1$ with $V_\varepsilon := \mathscr{V}(f_1-a_1\varepsilon,\dots,f_s-a_s\varepsilon)\subset\mathbb{C}\langle\varepsilon\rangle^n$ and $V := \lim_{\varepsilon\to 0} V_\varepsilon \subset \mathbb{C}^n$, we may have connected components $C\subset V\cap\mathbb{R}^n$ that can only be extended to the complex part of V_ε, i.e. $C\not\subset\lim_{\varepsilon\to 0}(V_\varepsilon\cap\mathbb{R}\langle\varepsilon\rangle^n)$.

Let $s=n=2$, $f_1=x_1^2+x_2^2-1$, $f_2=-(x_1-2)^2-x_2^2+1$ and $\alpha=(1,1)$. Then with $V_\varepsilon=\mathscr{V}(f_1-\varepsilon,f_2-\varepsilon)\subset\mathbb{C}\langle\varepsilon\rangle^2$ we have

$$\lim_{\varepsilon\to 0} V_\varepsilon\cap\mathbb{R}^2=\{(1,0)\}$$

is a single points, so this point is the only connected component. Note that any points in V_ε has to satisfy $f_1=f_2$, so in particular they will correspond to points on the intersections of the graphs of f_1 and f_2. The graphs of f_1 and f_2 are the surfaces $P_1:=\mathscr{V}(x_3-x_1^2-x_2^2+1)$ and $P_2:=\mathscr{V}(x_3+(x_1-2)^2+x_2^2-1))$, respectively, which are two 3-dimensional parabolas.

Over the complex numbers, P_1 and P_2 intersect in two complex lines

$$\mathscr{V}(x_2\pm i(x_1-1),x_3-2x_1+2)\subset\mathbb{C}^3$$

so whenever $x_3=\varepsilon$, i.e. $x_1=\frac{\varepsilon}{2}+1$, the points $\left(\frac{\varepsilon}{2}+1,\pm i\frac{\varepsilon}{2}\right)$ are in V_ε and

$$\lim_{\varepsilon\to 0}\left(\frac{\varepsilon}{2}+1,\pm i\frac{\varepsilon}{2}\right)=(1,0).$$

This also shows that

$$\lim_{\varepsilon\to 0}\left(V_\varepsilon\cap\mathbb{R}\langle\varepsilon\rangle^n\right)=\emptyset.$$

In particular, over the reals P_1 is a convex parabola with vertex $(0,0,-1)$, P_2 is a concave parabola with vertex $(2,0,1)$, and they tangentially intersect at $(1,0,0)$. So for any real $\varepsilon\neq 0$ we have

$$P_1\cap P_2\cap\mathscr{V}(x_3-\varepsilon)\cap\mathbb{R}^3=\emptyset.$$

On the contrary, for $F:=f_1^2+f_2^2$, if we define $V_\varepsilon:=\mathscr{V}(F-\varepsilon)$, then we proved above that

$$\left(\lim_{\varepsilon\to 0} V_\varepsilon\right)\cap\mathbb{R}^2=\lim_{\varepsilon\to 0}\left(V_\varepsilon\cap\mathbb{R}\langle\varepsilon\rangle^2\right).$$

4.2 A Change of Variables and Perturbed Polar Varieties

We now move to thinking about perturbing polar varieties in particular, recalling from Definition 2.3.8 that for the projections $\pi_i(x_1,\dots,x_n)=(x_1,\dots,x_i)$ for $i=1,\dots,n$, the *polar variety associated to* π_i of $\mathscr{V}(F)$ is defined as

$$\operatorname{crit}(V,\pi_i):=\mathscr{V}\left(F,\frac{\partial F}{\partial x_{i+1}},\dots,\frac{\partial F}{\partial x_n}\right)\subset\mathbb{C}^n\quad i=1,\dots,n.$$

We use the following notation to perform a change of variables.

Definition 4.2.1. Let $F\in\mathbb{Q}[\mathbf{x}]$, $V=\mathscr{V}(F)\subset\mathbb{C}^n$, and $A\in\mathrm{GL}_n(\mathbb{Q})$. Then, we denote $F^A(\mathbf{x}):=F(A\mathbf{x})$, i.e.

$\mathcal{V}(F^A)$ is the image of V via the map $\mathbf{x} \mapsto A^{-1}\mathbf{x}$.

Next, we state some known results on polar varieties which will be used in the proofs of our algorithms. In particular, polar varieties provide a very nice way for us to lower the complex dimension of an algebraic set without losing the real points in the set.

Theorem 4.2.2. *[3, Proposition 3] Let $F \in \mathbb{Q}[\mathbf{x}]$ be non-constant, square-free, and define a smooth algebraic set $V := \mathcal{V}(F) \subset \mathbb{C}^n$. Then there exists a non-empty Zariski open set $\mathscr{A} \subset \mathrm{GL}_n(\mathbb{C})$ such that for all $A \in \mathrm{GL}_n(\mathbb{Q}) \cap \mathscr{A}$ and $1 \leq i \leq n, \mathrm{crit}(V^A, \pi_i)$ is either empty or equidimensional of complex dimension $i-1$.*

We note that in the above reference, the proof of this theorem consists of characterizing the set of matrices for which the result does not hold and showing that those matrices make up a Zariski closed set $\mathrm{GL}_n(\mathbb{C}) \setminus \mathscr{A}$, i.e. the complement of $\mathscr{A}$.

Corollary 4.2.3. *Let F and V be as in Theorem 4.2.2. Suppose ε is an infinitesimal and $V_\varepsilon := \mathcal{V}(F-\varepsilon) \subset \mathbb{C}\langle\varepsilon\rangle^n$ is a smooth algebraic set on the field of Puiseux series as in Definition 2.2.20. Then there exists a non-empty Zariski open set $\mathscr{A} \subset \mathrm{GL}_n(\mathbb{C}\langle\varepsilon\rangle)$ such that for all $A \in \mathrm{GL}_n(\mathbb{Q}) \cap \mathscr{A}$ and $1 \leq i \leq n, \mathrm{crit}(V_\varepsilon^A, \pi_i)$ is either empty or equidimensional of complex dimension $i-1$.*

The proof of 4.2.3 follows from the following lemma.

Lemma 4.2.4. *Let $\mathscr{A} := \mathrm{GL}_n(\mathbb{C}\langle\varepsilon\rangle) \setminus \mathcal{V}(Q)$ be a non-empty Zariski open subset of $\mathrm{GL}_n(\mathbb{C}\langle\varepsilon\rangle)$ defined by some polynomial $Q \in \mathbb{Q}\langle\varepsilon\rangle[a_{i,j}]_{i,j=1}^n$. Then $\lim_{\varepsilon\to 0} \mathscr{A} \cap \mathrm{GL}_n(\mathbb{Q})$ is also a non-empty Zariski open subset of $\mathrm{GL}_n(\mathbb{Q})$.*

Proof. Since $\mathscr{A}$ is non-empty, $Q \neq 0$. Q is a polynomial in the variables $\{a_{i,j}\}_{i,j=1}^n$ with coefficients in $\mathbb{Q}\langle\varepsilon\rangle$, which we can assume without loss of generality are polynomials in $\varepsilon^{\frac{1}{q}}$ for some $q \in \mathbb{N}$ (by multiplying with a possible common denominator of these coefficients). Also, we can assume that Q has minimal degree in $\varepsilon^{\frac{1}{q}}$ among all such polynomials defining $\mathrm{GL}_n(\mathbb{C}\langle\varepsilon\rangle) \setminus \mathscr{A}$. Thus

$$Q = Q_0 + \varepsilon^{\frac{t}{q}} Q_1$$

where $t \in \mathbb{Z}^+, Q_0 \in \mathbb{Q}[a_{i,j}]$, and $Q_1 \in \mathbb{Q}[\varepsilon^{\frac{1}{q}}][a_{i,j}]$. If $Q_0 = 0$ then we have Q_1 has lower degree than Q in $\varepsilon^{\frac{1}{q}}$ and still defines $\mathrm{GL}_n(\mathbb{C}\langle\varepsilon\rangle) \setminus \mathscr{A}$, a contradiction. Therefore, $Q_0 \neq 0$ and

$$\mathcal{V}(Q_0) = \lim_{\varepsilon\to 0} \mathcal{V}(Q) \neq \mathrm{GL}_n(\mathbb{C})$$

so $\lim_{\varepsilon\to 0} \mathscr{A} \cap \mathrm{GL}_n(\mathbb{Q})$ is also a non-empty Zariski open subset of $\mathrm{GL}_n(\mathbb{Q})$. □

The next results relate to the local dimension of algebraic sets after a change of variables. We start with the following result, known commonly as the Noether Normalization Lemma after Emmy Noether from 1926, following the notation of [38].

Lemma 4.2.5. *Let I be an ideal in the polynomial ring $\mathbb{C}[x_1,\ldots,x_n]$ and A defining a suitable change of variables as in Definition 4.2.1 giving $x_1',\ldots,x_n'$. Then $x_1',\ldots,x_d'$ are algebraically independent modulo I and $x_{d+1}',\ldots,x_n'$ are integral over $\mathbb{C}[x_1',\ldots,x_d']$ modulo I.*

We note that in particular, when dealing with affine algebraic sets, the d in the above definition corresponds with the complex dimension of the set, as given in Definition 2.1.13. Furthermore, the following computational result of [36] is proven by characterizing the set of matrices for which the Noether normal position is not achieved.

Theorem 4.2.6. *Let I be as in Lemma 4.2.5. Then there exists a non-empty Zariski open set $\mathscr{A} \subset \mathrm{GL}_n(\mathbb{C})$ such that for all $A \in \mathrm{GL}_n(\mathbb{Q}) \cap \mathscr{A}$, the resulting change of variables $x_1',\ldots,x_n'$ are algebraically independent modulo I and $x_{d+1}',\ldots,x_n'$ are integral over $\mathbb{C}[x_1',\ldots,x_d']$ modulo I.*

The statements of the next two propositions follow the approach of the unpublished work [46], so we restate and prove them here.

Proposition 4.2.7. *Suppose $F \in \mathbb{Q}[\mathbf{x}]$ and $\mathscr{V}(F) \cap \mathbb{R}^n$ is bounded. There exists a non-empty Zariski open set $O \in \mathrm{GL_n}(\mathbb{C})$ such that for $A \in O \cap GL_n(\mathbb{Q})$, if $V^A = \mathscr{V}(F^A)$ and $V_\varepsilon^A := \mathscr{V}(F^A - \varepsilon) \subset \mathbb{C}\langle\varepsilon\rangle^n$ for ε infinitesimal, then*
***(i)** for all $1 \le i \le n$, $\mathrm{crit}(V_\varepsilon^A, \pi_i)$ is either empty or is smooth and equidimensional with complex dimension $i-1$;*
***(ii)** for all $\mathbf{p} \in V^A \cap \mathbb{R}^n$, $\pi_d^{-1}(\pi_d(\mathbf{p})) \cap (V^A \cap \mathbb{R}^n)$ is finite, where d is greater than or equal to the local real dimension of V^A at $\mathbf{p}$.*

Proof. (i) Let $1 \le i \le n$ and suppose $\mathrm{crit}(V_\varepsilon^A, \pi_i)$ is non-empty. We note that V_ε^A is smooth by Lemma 4.1.1. By Corollary 4.2.3, we obtain a non-empty Zariski open set $O_1 \in \mathrm{GL_n}(\mathbb{C}\langle\varepsilon\rangle)$ such that for $A \in O_1 \cap \mathrm{GL_n}(\mathbb{Q})$, $\mathrm{crit}(V_\varepsilon^A, \pi_i)$ is equidimensional with complex dimension $i-1$.

(ii) Since $V^A \cap \mathbb{R}^n$ is semi-algebraic, we can consider it as a union of connected components $C_1,\ldots,C_l$ with corresponding real dimension $d_1,\ldots,d_l$, as in Definition 2.2.11. Then the local real dimension of $V^A \cap \mathbb{R}^n$ at $\mathbf{p}$ is given by $\max_{\mathbf{p}\in\overline{C_i}} d_i$, as in Definition 2.2.12.

Let V_i represent the Zariski closure of each C_i for $1 \le i \le l$. Then the corresponding complex dimensions of $V_1,\ldots,V_l$ are $d_1,\ldots,d_l$. By Theorem 4.2.6, there exists a non-empty Zariski open set $O_{2_i} \in \mathrm{GL_n}(\mathbb{C})$ such that for $A \in O_{2_i} \cap \mathrm{GL_n}(\mathbb{Q})$ and $\mathbf{q} \in \mathbb{C}^{d_i}$, $\pi_{d_i}^{-1}(\mathbf{q}) \cap V_i^A$ is finite. Then for $\mathbf{q} \in \mathbb{R}^{d_i}$, $\pi_{d_i}^{-1}(\mathbf{q}) \cap C_i$ is finite.

Let $\mathbf{p} \in V^A \cap \mathbb{R}^n$ where $A \in O_2 = \cap_{i=1}^l O_{2_i}$. Suppose $d \ge \max_{\mathbf{p}\in\overline{C_i}} d_i$. Then $\pi_d(\mathbf{p}) \in \mathbb{R}^d$. For any $d_i = d$, taking O_2 as defined above with Theorem 4.2.6 guarantees $\pi_d^{-1}(\pi_d(\mathbf{p})) \cap (V^A \cap \mathbb{R}^n)$ is finite. Furthermore, for any d_i strictly less than d, $\pi_d^{-1}(\pi_d(\mathbf{p})) \cap (V^A \cap \mathbb{R}^n)$ is still finite for $A \in O_2$ because $\pi_d^{-1}(\pi_d(\mathbf{p})) \subset \pi_{d_i}^{-1}(\pi_{d_i}(\mathbf{p}))$. Taking $O = O_1 \cap O_2$ completes the proof. □

Now we are ready to state the main result of this section, recalling that it was suggested by the unpublished [46].

Proposition 4.2.8. *Let $F, g_1, \ldots, g_m \in \mathbb{Q}[\mathbf{x}]$ and let ε be infinitesimal. Suppose $F \geq 0$ on $\mathbb{R}^n$, $\mathscr{V}(F) \cap \mathbb{R}^n$ is bounded, and $A \in GL_n(\mathbb{Q})$ with $V^A = \mathscr{V}(F^A)$ and $V_\varepsilon^A := \mathscr{V}(F^A - \varepsilon) \subset \mathbb{C}\langle\varepsilon\rangle^n$ such that **(i)** and **(ii)** from Proposition 4.2.7 hold. Then for $i = 0, \ldots, n$, $\lim_{\varepsilon\to 0} \operatorname{crit}(V_\varepsilon, \pi_i)$ is equidimensional of dimension $i-1$ and for*

$$U := \{\mathbf{x} \in \mathbb{R}^n \,:\, g_1(\mathbf{x}) > 0, \ldots, g_m(\mathbf{x}) > 0\} \subset \mathbb{R}^n$$

and $S := \mathscr{V}(F) \cap U$ we have

$$\left(\lim_{\varepsilon\to 0} \operatorname{crit}(V_\varepsilon, \pi_i)\right) \cap U = S \quad \Leftrightarrow \quad \dim_{\mathbb{R}}(S) \leq i-1.$$

Proof. ($\Rightarrow$) Fix some $0 \leq i \leq n$ and suppose

$$\left(\lim_{\varepsilon\to 0} \operatorname{crit}(V_\varepsilon, \pi_i)\right) \cap U = S.$$

If $S = \emptyset$, then $\dim_{\mathbb{R}}(S) = -1 \leq i-1$ and we are done. So assume $S \neq \emptyset$. Since $S := \mathscr{V}(f) \cap U$, $\mathscr{V}(F) \cap \mathbb{R}^n \neq \emptyset$ and S is bounded by our initial assumption of $\mathscr{V}(F) \cap \mathbb{R}^n$ being bounded.

By Lemma 4.2.10, $\operatorname{crit}(V_\varepsilon, \pi_i) \neq \emptyset$. Then by Proposition 4.2.7, $\operatorname{crit}(V_\varepsilon, \pi_i)$ is equidimensional of complex dimension $i-1$. Then $\lim_{\varepsilon\to 0}(\operatorname{crit}(V_\varepsilon, \pi_i))$ has complex dimension $i-1$. So the real dimension of $\lim_{\varepsilon\to 0}(\operatorname{crit}(V_\varepsilon, \pi_i))$ is $\leq i-1$ and thus

$$\left(\lim_{\varepsilon\to 0} \operatorname{crit}(V_\varepsilon, \pi_i)\right) \cap U = S$$

has real dimension $\leq i-1$.
($\Leftarrow$) Now suppose $\dim_{\mathbb{R}}(S) \leq i-1$. Since $\operatorname{crit}(V_\varepsilon, \pi_i) \subset V_\varepsilon$, $\lim_{\varepsilon\to 0} V_\varepsilon = \mathscr{V}(F)$, and $\mathscr{V}(F) \cap U = S$,

$$\left(\lim_{\varepsilon\to 0} \operatorname{crit}(V_\varepsilon, \pi_i)\right) \cap U \subset S.$$

If $S = \emptyset$,

$$S \subset \left(\lim_{\varepsilon\to 0} \operatorname{crit}(V_\varepsilon, \pi_i)\right) \cap U$$

and we are done. So suppose $S \neq \emptyset$ and take $\mathbf{z} = (z_1, \ldots, z_n) \in S$. Since $S = \mathscr{V}(F) \cap U$, $\mathbf{z} \in U$ and $\mathbf{z} \in \mathscr{V}(F) \cap \mathbb{R}^n$. Furthermore, the local real dimension of S at $\mathbf{z}$ is $\leq i-1$, so the local real dimension of $\mathscr{V}(F) \cap \mathbb{R}^n$ at $\mathbf{z}$ is also $\leq i-1$.

Define F' as the function F where the first $i-1$ coordinates have been evaluated at the first $i-1$ coordinate values of $\mathbf{z}$,i.e.

$$F' := F(x_i, \ldots, x_n) = F(z_1, \ldots, z_{i-1}, x_i, \ldots, x_n).$$

We note that since F is nonnegative over $\mathbb{R}^n$, F' is nonnegative over $\mathbb{R}^{n-i+1}$. Also define $\mathbf{z}' := (z_i, \ldots, z_n)$, $V' := \mathscr{V}(F')$ and $V'_\varepsilon := \mathscr{V}(F' - \varepsilon) \subset \mathbb{C}\langle\varepsilon\rangle^{n-i+1}$, and the canonical projection $\varphi_i(\mathbf{x}) = x_i$ and the respective $\varphi'_i(x_i, \ldots, x_n) = x_i$. Note that $\mathbf{z}'$ is isolated in $V' \cap \mathbb{R}^{n-i+1}$ since $\pi_{i-1}^{-1}(\pi_{i-1}(\mathbf{z})) \cap \mathscr{V}(F) \cap \mathbb{R}^n$ is finite by **(ii)** of Proposition 4.2.7.

Applying Lemma 4.2.9 to $\mathbf{z}'$ and V' we get some $\mathbf{z}'_\varepsilon \in \operatorname{crit}(V'_\varepsilon, \varphi'_i)$ such that $\lim_{\varepsilon\to 0} \mathbf{z}'_\varepsilon = (z_i, \ldots, z_n)$. Define $\mathbf{z}_\varepsilon := (z_1, \ldots, z_{i-1}, \mathbf{z}'_\varepsilon)$ and $V^*_\varepsilon = V_\varepsilon \cap \pi_{i-1}^{-1}(z, \ldots, z_{i-1})$. Then $\mathbf{z}_\varepsilon \in \operatorname{crit}(V^*_\varepsilon, \varphi_i)$. By Lemma 4.2.11, $\mathbf{z}_\varepsilon \in \operatorname{crit}(V_\varepsilon, \pi_i)$. Since $\lim_{\varepsilon\to 0} \mathbf{z}_\varepsilon = \mathbf{z}$,

$$S \subset \left(\lim_{\varepsilon\to 0} \operatorname{crit}(V_\varepsilon, \pi_i)\right) \cap U.$$

□

The following lemmas were used in the proof of the above proposition.

Lemma 4.2.9. *Assume $f_1, \ldots, f_s \in \mathbb{R}[\mathbf{x}], F = f_1^2 + \cdots f_s^2$, and $\varepsilon > 0$ a real infinitesimal. Let $V_\varepsilon := \mathscr{V}(F - \varepsilon) \subset \mathbb{C}\langle\varepsilon\rangle^n$ and $V := \lim_{\varepsilon\to 0} V_\varepsilon \subset \mathbb{C}^n$. Suppose $\mathbf{z} \in V \cap \mathbb{R}^n$ and there exists a neighborhood $B(\mathbf{z}, r) \subset \mathbb{R}^n$ for some $r > 0$ such that $B(\mathbf{z}, r) \cap V \cap \mathbb{R}^n$ is a finite set. Then there exists $\mathbf{z}_\varepsilon \in \operatorname{crit}(V_\varepsilon, \pi_1)$ such that $\lim_{\varepsilon\to 0} \mathbf{z}_\varepsilon = \mathbf{z}$.*

Proof. Since $B(\mathbf{z}, r) \cap V \cap \mathbb{R}^n$ is a finite set, there exists some $r' > 0$ such that $\mathbf{z}$ is the only point in $B(\mathbf{z}, r') \cap V \cap \mathbb{R}^n$. So $\{\mathbf{z}\}$ is a bounded connected component of $V \cap \mathbb{R}^n$. Then by Proposition 4.1.3, there exist connected components $C_{\varepsilon,1}, \ldots, C_{\varepsilon,l}$ of $V_\varepsilon \cap \mathbb{R}\langle\varepsilon\rangle^n$ such that $\{\mathbf{z}\} = \cup_{i=1}^l \lim_{\varepsilon\to 0} C_{\varepsilon,i}$ and $C_{\varepsilon,i}$ is bounded by and does not intersect the boundary of $\operatorname{Ext}(B(\mathbf{z}, r') \cap \mathbb{R}\langle\varepsilon\rangle^n)$ for $1 \le i \le l$.

Then for $1 \le i \le l$, we have $C_{\varepsilon,i} \subset \operatorname{Ext}(B(\mathbf{z}, r') \cap \mathbb{R}\langle\varepsilon\rangle^n)$ and is closed and bounded. By the Extreme Value Theorem, $C_{\varepsilon,i} \cap \operatorname{crit}(V_\varepsilon, \pi_1) \neq \emptyset$. Since $\lim_{\varepsilon\to 0} C_{\varepsilon,i)} = \{\mathbf{z}\}$, all $\mathbf{z}_\varepsilon \in C_{\varepsilon,i} \cap \operatorname{crit}(V_\varepsilon, \pi_1) \neq \emptyset$ are such that $\lim_{\varepsilon\to 0} \mathbf{z}_\varepsilon = \mathbf{z}$, and we are done. □

Lemma 4.2.10. *Assume $f_1, \ldots, f_s \in \mathbb{R}[\mathbf{x}], F = f_1^2 + \cdots f_s^2$, and $\varepsilon > 0$ a real infinitesimal. Let $V_\varepsilon := \mathscr{V}(F - \varepsilon) \subset \mathbb{C}\langle\varepsilon\rangle^n$ and $V := \lim_{\varepsilon\to 0} V_\varepsilon \subset \mathbb{C}^n$ with $V \cap \mathbb{R}^n$ nonempty and bounded. Then $\operatorname{crit}(V_\varepsilon, \pi_i)$ is nonempty and intersects each bounded connected components of $V_\varepsilon \cap \mathbb{R}\langle\varepsilon\rangle^n$ for all $1 \le i \le n$.*

Proof. Since $V \cap \mathbb{R}^n$ is nonempty, there exists some nonempty connected component $C \subset V \cap \mathbb{R}^n$. Let $\mathbf{z} \in C$. Since $V \cap \mathbb{R}^n$ is bounded, there exists some $r > 0$ such that the $C \subset B(\mathbf{z}, r) \subset \mathbb{R}^n$ and C does not intersect the boundary of $B(\mathbf{z}, r)$. So by Proposition 4.1.3, there exist connected components $C_{\varepsilon,1}, \ldots, C_{\varepsilon,l}$ of $V_\varepsilon \cap \mathbb{R}\langle\varepsilon\rangle^n$ such that $C = \cup_{j=1}^l \lim_{\varepsilon\to 0} C_{\varepsilon,j}$ and $C_{\varepsilon,j}$ is bounded by and does not intersect the boundary of $\operatorname{Ext}(B(\mathbf{z}, r) \cap \mathbb{R}\langle\varepsilon\rangle^n)$ for $1 \le j \le l$.

Then for $1 \le j \le l$, we have $C_{\varepsilon,j} \subset \operatorname{Ext}(B(\mathbf{z}, r) \cap \mathbb{R}\langle\varepsilon\rangle^n)$ and is closed and bounded. Hence $C_{\varepsilon,j} \cap \operatorname{crit}(V_\varepsilon, \pi_1) \neq \emptyset$. Since by definition $\operatorname{crit}(V_\varepsilon, \pi_1) \subset \operatorname{crit}(V_\varepsilon, \pi_i)$, we are done. □

Lemma 4.2.11. *Let $F \in \mathbb{C}[\mathbf{x}]$ and $\alpha = (\alpha_1, \ldots, \alpha_{i-1}) \in \mathbb{C}^i$. Suppose $V_{i,\alpha}$ is the algebraic set $\mathscr{V}(F) \cap \pi_{i-1}^{-1}(\alpha)$ and φ_i is the projection defined by $\varphi(\mathbf{x}) = x_i$. Then*

$$\operatorname{crit}(V_{i,\alpha}, \varphi_i) \subset \operatorname{crit}(\mathscr{V}(F), \pi_i).$$

Proof. We recall that by definition

$$\operatorname{crit}(\mathscr{V}(F), \pi_i) = \mathscr{V}\left(F, \frac{\partial F}{\partial x_{i+1}}, \ldots, \frac{\partial F}{\partial x_n}\right).$$

By how we have defined $V_{i,\alpha}$, $\text{crit}(V_{i,\alpha}, \varphi_i)$ is the algebraic set defined by the polynomials $f, x_1 - \alpha_1, \ldots, x_{i-1} - \alpha_{i-1}$ and the maximal minors of the Jacobian matrix of the polynomials. Then in fact,

$$\text{crit}(V_{i,\alpha}, \varphi_i) = \mathscr{V}\left(F, x_1 - \alpha_1, \ldots, x_{i-1} - \alpha_{i-1}, \frac{\partial F}{\partial x_{i+1}}, \ldots, \frac{\partial F}{\partial x_n}\right)$$

and we are done. □

In Chapter 5, we will use the results of Proposition 4.2.7 and Proposition 4.2.8 to develop a genericity assumption (corresponding to the definition of "generic" given in Definition 2.1.11) and establish a framework with which to construct a probabilistic algorithm.

4.3 Shifting from Infinitesimals to Complex Perturbations

In this section, we establish results in order to formulate our algorithms so they can be implemented not only purely symbolically, but also in a numerical algebraic geometry context. Here we track our perturbed set to its limit variety by employing homotopy continuation while our perturbation constant follows a complex arc towards zero.

To this end, we first shift from the paradigm of real infinitesimals to arbitrarily small real numbers, as established by the following result from real algebraic geometry.

Theorem 4.3.1. *[8, Proposition 3.17] A result holds over $\mathbb{R}\langle\varepsilon\rangle$ if and only if there exists some $e_0 \in \mathbb{R}$ such that it also holds for all $e \in (0, e_0) \cap \mathbb{R}$.*

For our purposes, we also want to establish that we are able to make this switch in terms of witness set computations as well. We note that witness set computations are done over the complex numbers. Therefore, the results in this section can be formulated in terms of a more general perturbation setup, based on the following result from Faugére et al. on perturbing the defining polynomials of an algebraic set.

Lemma 4.3.2. *[19, Lemma 1] Let $f_1, \ldots, f_s \in \mathbb{R}[\mathbf{x}]$ and fix $l \leq s$ and $\{i_1, \ldots, i_l\} \subset \{1, \ldots, s\}$. Then there exists a Zariski closed subset $\mathscr{A} \times \mathscr{E} \subset \mathbb{C}^s \times \mathbb{C}$ such that for all $\mathbf{a} := (a_1, \ldots, a_s) \in \mathbb{R}^s \setminus \mathscr{A}$ and $e \in \mathbb{R} \setminus \mathscr{E}$, the ideal generated by the polynomials $f_{i_1} - e\,a_{i_1}, \ldots, f_{i_l} - e\,a_{i_l}$ is a radical equidimensional ideal and $\mathscr{V}(f_{i_1} - e\,a_{i_1}, \ldots, f_{i_l} - e\,a_{i_l})$ is either empty or smooth of dimension $n - l$.*

Using the above result, we define a genericity assumption which holds over the complex numbers.

Definition 4.3.3. Consider polynomials $f_1, \ldots, f_s \in \mathbb{R}[\mathbf{x}]$ and point $\mathbf{a} = (a_1, \ldots, a_s) \in \mathbb{Q}^s$. We say that $f_1, \ldots f_s$ and $\mathbf{a}$ satisfy Assumption `(A)` if

`(A)`: There exists $e_0 > 0$ such that for all $0 < e \leq e_0$, the polynomials $f_1 - e\,a_1, \ldots, f_s - e\,a_s$ generate a radical equidimensional ideal and $V_e^{\mathbf{a}} := \mathscr{V}(f_1 - e\,a_1, \ldots, f_s - e\,a_s)$ is smooth and has dimension $n - s$.

Assumption (A) in Definition 4.3.3 guarantees the existence of $e_0 > 0$; however, in practice this number can be arbitrarily small. Instead of trying to compute an e_0 that works for a given system $f_1, \ldots, f_s$, the next result shows that we can choose a generic $\xi \in \mathbb{C}$ with $|\xi| = 1$ to replace e_0 with ξ and e with $t\xi$, where $t \in (0,1]$.

Proposition 4.3.4. *Let $f_1, f_2, \ldots, f_s \in \mathbb{R}[\mathbf{x}]$, $\mathbf{a} = (a_1, \ldots, a_s) \in \mathbb{Q}^s$ and let ε be infinitesimal. Assume that $V_\varepsilon^{\mathbf{a}} := \mathcal{V}(f_1 - \varepsilon a_1, \ldots, f_s - \varepsilon a_s) \subset \mathbb{C}\langle\varepsilon\rangle^n$ is smooth and equidimensional of dimension $n-s$. Then for all but finitely many $\xi \in \mathbb{C}$ with $|\xi| = 1$ and for all $t \in (0,1]$, $V_{t\xi}^{\mathbf{a}} := \mathcal{V}(f_1 - t\xi a_1, \ldots, f_s - t\xi a_s) \subset \mathbb{C}^n$ is smooth and equidimensional of dimension $n-s$ and in that case we have*

$$\lim_{\varepsilon \to 0} V_\varepsilon^{\mathbf{a}} = \lim_{t \to 0} V_{t\xi}^{\mathbf{a}}.$$

Proof. First, we show that for all but a finite number of choices of $\xi \in \mathbb{C}$, $V_\xi^{\mathbf{a}} = \mathcal{V}(f_1 - \xi a_1, \ldots, f_s - \xi a_s)$ is smooth. Note that from our assumptions on $V_\varepsilon^{\mathbf{a}}$ we get that $f_1, \ldots, f_s$ and $\mathbf{a}$ satisfy Assumption (A) for some $\mathbf{e}_0 > 0$. Consider the ideal using new variables x_0, z and $\lambda_1, \ldots, \lambda_s$:

$$\begin{aligned} I := \langle f_1^{(h)} - a_1 z x_0^{\deg(f_1)}, \ldots, f_s^{(h)} - a_s z x_0^{\deg(f_s)} \rangle \\ + \langle (\lambda_1 \nabla(f_1) + \ldots + \lambda_s \nabla(f_s))^{(h)} \rangle. \end{aligned}$$

Here $g^{(h)}$ denotes the homogenization of $g \in \mathbb{R}[x_1, \ldots, x_n]$ by the variable x_0 as in Definition 2.1.21 and ∇ is the differential operator in the variables $x_1, \ldots, x_n$ as in Definition 2.1.17. Thus I is bihomogeneous in the variables $(\lambda_1, \ldots \lambda_s)$ and $(x_0, \ldots, x_n)$. Then the projection of $X(I) \subset \mathbb{P}^n \times \mathbb{P}^s \times \mathbb{C}$ onto $\mathbb{C}$ is a Zariski closed subset of $\mathbb{C}$, and since e_0 is not in the projection, the projection is not $\mathbb{C}$, thus a finite set Z. Clearly, for $\xi \in \mathbb{C} \setminus Z$ and for all $\mathbf{p} \in V_\xi^{\mathbf{a}}$, the Jacobian of $f_1 - \xi a_1, \ldots, f_s - \xi a_s$ at $\mathbf{p}$ has rank s, thus $V_\xi^{\mathbf{a}}$ is smooth and equidimensional of dimension $n-s$. This also implies that for all but finitely many $\xi \in \mathbb{C}$ with $|\xi| = 1$ and for all $t \in (0,1]$ we have that $V_{t\xi}^{\mathbf{a}} = \mathcal{V}(f_1 - t\xi a_1, \ldots, f_s - t\xi a_s)$ is smooth and equidimensional.

Fix $\xi \in \mathbb{C} \setminus Z$ with $|\xi| = 1$ so $V_{t\xi}^{\mathbf{a}}$ is smooth and equidimensional. To prove the second claim, let $L_1, \ldots, L_{n-s} \in \mathbb{C}[\mathbf{x}]$ be linear polynomials such that $\mathscr{L} = \mathcal{V}(L_1, \ldots, L_{n-s})$ is a generic linear space of codimension $n-s$ which intersects both $\lim_{\varepsilon \to 0} V_\varepsilon^{\mathbf{a}}$ and $\lim_{t \to 0} V_{t\xi}^{\mathbf{a}}$ transversely. By our assumptions, both $V_\varepsilon^{\mathbf{a}} \cap \mathscr{L}$ and $V_{t\xi}^{\mathbf{a}} \cap \mathscr{L}$ are finite.

Then since $\mathscr{L}$ does not depend on either ε or t,

$$\begin{aligned} \lim_{\varepsilon \to 0} \left(V_\varepsilon^{\mathbf{a}} \cap \mathscr{L} \right) = \lim_{\varepsilon \to 0} V_\varepsilon^{\mathbf{a}} \cap \mathscr{L} \quad \text{and} \\ \lim_{t \to 0} \left(V_{t\xi}^{\mathbf{a}} \cap \mathscr{L} \right) = \lim_{t \to 0} V_{t\xi}^{\mathbf{a}} \cap \mathscr{L} \end{aligned}$$

Since $\mathscr{L}$ is a generic linear space which intersects both $\lim_{\varepsilon \to 0} V_\varepsilon^{\mathbf{a}}$ and $\lim_{t \to 0} V_{t\xi}$ transversely, $\lim_{\varepsilon \to 0} V_\varepsilon^{\mathbf{a}} \cap \mathscr{L} = \lim_{t \to 0} V_{t\xi}^{\mathbf{a}} \cap \mathscr{L}$ implies $\lim_{\varepsilon \to 0} V_\varepsilon^{\mathbf{a}} = \lim_{t \to 0} V_{t\xi}^{\mathbf{a}}$.

So it is sufficient to prove that

$$\lim_{\varepsilon\to 0}\left(V_\varepsilon^{\mathbf{a}}\cap\mathscr{L}\right)=\lim_{t\to 0}\left(V_{t\xi}^{\mathbf{a}}\cap\mathscr{L}\right)$$

to achieve the desired result.

Let $H\subset\mathbb{R}[\mathbf{x},\varepsilon]$ be the system

$$H:=H(\mathbf{x},\varepsilon)=\left[f_1-\varepsilon a_1,\ldots,f_s-\varepsilon a_s,L_1,\ldots,L_{n-s}\right].$$

Let $S\subset\mathbb{C}\langle\varepsilon\rangle^n$ be the finite set of bounded solutions of $H=0$, where bounded is as defined for Puiseux series in Definition 2.2.20. Then for all $\mathbf{x}(\varepsilon)\in S$, let $\lim_{\varepsilon\to 0}\mathbf{x}(\varepsilon)=\mathbf{x}_0\in\mathbb{C}^n$. Furthermore, by the definition of H, $\lim_{\varepsilon\to 0}S=\lim_{\varepsilon\to 0}\left(V_\varepsilon^{\mathbf{a}}\cap\mathscr{L}\right)$.

Since $\varepsilon>0$ is a real infinitesimal, each $\mathbf{x}(\varepsilon)$ has an interval of convergence $(0,\varepsilon_x)\subset\mathbb{R}$ for some $\varepsilon_x>0$. Choose $\varepsilon_0>0$ such that $\varepsilon_0<\min_{x\in S}\varepsilon_x$. Now we make a switch, and instead of considering $\mathbf{x}(\varepsilon)\in S$ an element $\mathbb{C}\langle\varepsilon\rangle^n$, we consider $\mathbf{x}$ as a function $\mathbb{C}\to\mathbb{C}^n$ which is well-defined for $z\in\mathbb{C}$ with $|z|\le\varepsilon_0$. Abusing the notation, we denote by $\mathbf{x}$ both the Puiseux series and the corresponding complex function.

Recall that if a pair $(z^*,\mathbf{x}^*)\in\mathbb{C}\times\mathbb{C}^n$ has the property that $H(\mathbf{x}^*,z^*)=0$ and $\det JH(\mathbf{x}^*,z^*)=0$, where JH is the Jacobian matrix of H with respect to the $\mathbf{x}$ variables, then z^* is a critical point and $\mathbf{x}^*$ is a branch point for for $H(\mathbf{x},z)=0$. Let $\mathscr{C}$ denote the set of all critical points of $H(\mathbf{x},z)=0$. Then, since $|S|<\infty$, we know $|\mathscr{C}|<\infty$.

Now let $z\in\mathscr{C}$. Then there exists some $\xi_z\in\mathbb{C}$ with $|\xi_z|=1$ such that for $t\in\mathbb{R}$, the path $\xi_z t$ passes through z, so that $\mathbf{x}(t\xi_z)\in\mathbb{C}^n$ has some branching point. Let $Z=\{\xi_z : z\in\mathscr{C}\}\subset\mathscr{S}_1$, since $|\mathscr{C}|<\infty$, $|Z|<\infty$. Then, for any $\xi\in\mathbb{C}\setminus Z$ with $|\xi|=1$, we have that $\mathbf{x}(t\xi)\in\mathbb{C}^n$ for $t\in(0,1]$ does not pass through branching points. Since $\mathbb{C}\setminus Z$ is Zariski dense in $\mathbb{C}$, the same holds for generic $\xi\in\mathbb{C}$ with $|\xi|=1$.

So let $\xi\in\mathbb{C}\setminus Z$ with $|\xi|=1$ and $H_\xi\subset\mathbb{C}^{n+1}$ be the homotopy defined by the system

$$H_\xi:=H_\xi(\mathbf{x},t)=\left[f_1-t\xi a_1,\ldots,f_s-t\xi a_s,L_1\ldots,L_{n-s}\right].$$

The limit points of the solutions of H_ξ are $\lim_{t\to 0}\left(V_{t\xi}^{\mathbf{a}}\cap\mathscr{L}\right)$. Let $T\subset\mathbb{C}^n$ be the roots of $H_\xi(\mathbf{x},1)$. Then $|T|=|V_\varepsilon^{\mathbf{a}}\cap\mathscr{L}|<\infty$. Furthermore, by the above argument the homotopy paths for H_ξ are exactly described by the points in $V_\varepsilon^{\mathbf{a}}\cap\mathscr{L}\subset\mathbb{C}\langle\varepsilon\rangle^n$ by replacing ε with $t\xi$. Hence,

$$\lim_{\varepsilon\to 0}\left(V_\varepsilon^{\mathbf{a}}\cap\mathscr{L}\right)=\lim_{t\to 0}\left(V_{t\xi}^{\mathbf{a}}\cap\mathscr{L}\right).$$

□

The above result gives a proof of correctness for WITNESS POINTS IN LIMITS ALGORITHM 4.3.1 which computes a witness point set (as in Definition 2.5.2) of a limit with algebraic probability one.

Algorithm 4.3.1 WITNESSPOINTSINLIMITS

Input: $f_1,\ldots,f_s \in \mathbb{R}[\mathbf{x}]$, $\mathbf{a}=(a_1,\ldots,a_s)\in\mathbb{Q}^s$

Output: flag =TRUE if $\mathcal{V}(f_1-a_1e,\ldots,f_s-a_se)$ is 0-dimensional for sufficiently small $e>0$ and the finite set of points in $W:=\lim_{e\to 0+}\mathcal{V}(f_1-a_1e,\ldots,f_s-a_se)$, flag =FALSE otherwise and $W=\emptyset$.

1. Loop
   (a) Choose generic $\xi\in\mathbb{C}$ with $|\xi|=1$.
   (b) Define $H_\xi(\mathbf{x},t):=\left[f_1-t\,\xi\,a_1,\ldots,f_s-t\,\xi\,a_s\right]$.
   (c) If $|\mathcal{V}(H_\xi(\mathbf{x},1))|=\infty$, exit loop and return flag = FALSE, $W=\emptyset$.
   (d) Compute $\lim_{t\to 0}\mathcal{V}(H_\xi(\mathbf{x},t))$ via a homotopy starting at $t=1$.
   (e) If no branch points were hit during homotopy tracking, exit loop and return flag = TRUE, $W=\lim_{t\to 0}\mathcal{V}(H_\xi(\mathbf{x},t))$.

A difficulty that arises is that the limit points may be singular, arising from multiple paths converging to the same limit point. This is demonstrated in the following example.

Example 4.3.5. For $f_1=x_1x_2$, $f_2=x_1x_2-x_1^2$, and $\mathbf{a}=(1,1/2)$, there are two paths that both limit to $(0,0)\subset\mathcal{V}(f_1,f_2)=\mathcal{V}(x_1)$.

Another difficulty that may arise is more delicate, in that the witness system $f=(f_1,\ldots,f_s)$ for the original algebraic set $\mathcal{V}(f)$ is not a witness system for $\lim_{t\to 0}V^{\mathbf{a}}_{t\xi}$. This is demonstrated in the following example.

Example 4.3.6. Consider $f(x,y)=(xy,xy-x)\subset\mathbb{C}[x,y]$ and $\mathbf{a}=(1,0)$. Then $f(x,y,t)=(xy-t,xy-x)\subset\mathbb{C}[x,y,t]$ gives $V^{\mathbf{a}}_{t\xi}$. We note that $\lim_{t\to 0}f(x,y,t)=\{(0,1,0)\}$. But $f(x,y,t)|_{t=0}=f(x,y)=\mathcal{V}(x)$ is not a witness system for $\{(0,1)\}$. We compute

$$Jf(x,y)=\begin{bmatrix} y & x\\ y-1 & x\end{bmatrix}.$$

For $\mathbf{p}=(0,1)$, $\operatorname{rank}(Jf(0,1))=1$. So by Definition 2.6.6, $F=[xy,xy-x,x]$. But $\mathcal{V}(F)=\mathcal{V}(x)$ is still not a witness system for $\{(0,1)\}$.

Instead, we have to do isosingular deflation of $f(x,y,t)$, as in Steps (1) and (2) of Algorithm 4.3.2. Here, we compute

$$Jf(x,y,t)=\begin{bmatrix} y & x & -1\\ y-1 & x & 0\end{bmatrix}.$$

For $\mathbf{p}=(0,1,0)$, $\operatorname{rank}(Jf(0,1,0))=1$. Applying Theorem 2.6.9 gives a deflated witness system $F=[xy,xy-x,x,y-1]$ for $\lim_{t\to 0}V^{\mathbf{a}}_{t\xi}$.

Applications of isosingular deflation as discussed in Section 2.6 can help us resolve both of these

types of difficulties. DEFLATED WITNESS SYSTEM ALGORITHM 4.3.2 computes a deflated witness system for irreducible components of a variety defined as a limit.

Algorithm 4.3.2 DEFLATEDWITNESSSYSTEM

Input: $f_1,\ldots,f_s \in \mathbb{R}[\mathbf{x}]$, $\mathbf{a}=(a_1,\ldots,a_s)\in\mathbb{Q}^s$, and $\mathbf{p}\in V := \lim_{e\to 0+} \mathscr{V}(f_1-a_1e,\ldots,f_s-a_se)$, a generic point on a unique irreducible component $V_{\mathbf{p}}$ of V.

Output: A deflated witness system $G\subset\mathbb{R}[\mathbf{x}]$ for $V_{\mathbf{p}}$.

1. Define $F_0(\mathbf{x},t) := (f_1-a_1t,\ldots,f_s-a_st)\in\mathbb{R}[\mathbf{x},t]^s$ and $\mathbf{q} := (\mathbf{p},0)\in\mathbb{R}^{n+1}$.
2. $F :=$ IsosingularDeflation$(F_0,\mathbf{q})$. `// See Algorithm 2.6.1`
3. Define $G_0(\mathbf{x}) := F(\mathbf{x},0)$.
4. $G :=$ IsosingularDeflation$(G_0,\mathbf{p})$. `// See Algorithm 2.6.1`
5. Return G.

Theorem 4.3.7. *Let $f_1,\ldots,f_s$, $\mathbf{a}$, and $\mathbf{p}$ as in the input of* ALGORITHM 4.3.2. *Then G, computed by* ALGORITHM 4.3.2, *satisfies the output specifications.*

Proof. Since $V_{\mathbf{p}}$ is an irreducible component of V, there exists an irreducible component $Z\subset \mathscr{V}(F_0(\mathbf{x},t))\subset\mathbb{C}^{n+1}$ such that $V_{\mathbf{p}}\times\{0\}$ is an irreducible component of $Z\cap\mathscr{V}(t)$ which is an intersection. Hence, one can apply the isosingular deflation approach applied to intersections in Theorem 2.6.9. Although Theorem 2.6.9 would deflate $H_0(\mathbf{x},t,t') := (F_0(\mathbf{x},t),t')$ at $\mathbf{q}' := (\mathbf{p},0,0)$, the simplicity of the intersection together with t' contained in H_0 easily shows that one obtains an equivalent deflation as deflating $F_0(\mathbf{x},t)$ at $\mathbf{q}=(\mathbf{p},0)$, say $F(\mathbf{x},t)$. Therefore, $V_{\mathbf{p}}$ must be an irreducible component of $\mathscr{V}(F^*(\mathbf{x},0))$ so $G_0(\mathbf{x}) := F(\mathbf{x},0)$ is a witness system for $V_{\mathbf{p}}$. Since G_0 need not be a deflated witness system for $V_{\mathbf{p}}$, one deflates G_0 at $\mathbf{p}$ to yield a deflated witness system G for $V_{\mathbf{p}}$. □

We note that the isosingular deflation operator from Section 2.6 applied above above uses all appropriate minors for constructing the sequences of polynomial systems. One could utilize alternative deflation approaches such as those found in [18, 20, 25, 34, 35] to possibly simplify the construction of the witness system.

4.4 Computation of g

The final key tool required to compute a real smooth point on every connected component of an algebraic set V is a "well-chosen" polynomial g that satisfies the conditions of Theorem 3.2.3, i.e., $\mathrm{Sing}(V)\cap\mathbb{R}^n\subset\mathscr{V}(g)$ and $\dim(V\cap\mathscr{V}(g))<\dim(V)$. There exist symbolic methods to compute

such a g for an irreducible variety V. For example, [48, Lemma 4.3] computes the defining equation w of a generic projection $\overline{\pi(V)}$ that is a hypersurface. Then, g can be taken to be one of the partial derivatives of w. This idea could be extended to the case when V is not equidimensional using infinitesimal deformations and limits (c.f., [47]). In COMPUTATION OF G ALGORITHM 4.4.1, we provide a new approach based on isosingular deflation, as discussed in Section 2.6, which computes several g's depending on the isosingular deflation sequence of the irreducible components.

Theorem 4.4.1. *Let $f_1,\ldots,f_s$, $\mathbf{a}$, $V_e^{\mathbf{a}}$, and V be as in the input and output specifications of* ALGORITHM 4.4.1. *Then,* ALGORITHM 4.4.1 *is correct.*

Proof. By our assumption on the genericity of L, each point $\mathbf{p}\in W$ is a generic point of a unique irreducible components $V_{\mathbf{p}}$ of V containing $\mathbf{p}$. Based on the output of Algorithm 4.3.2, assume that for any $\mathbf{p}\in W$, in Step (3d) we compute $G_j\subset\mathbb{R}[\mathbf{x}]$ such that the irreducible component $V_{\mathbf{p}}\subset V$ containing $\mathbf{p}$ is an irreducible component of $\mathcal{V}(G_j)$, $f_1,\ldots,f_s\in G_j$, $G_j(\mathbf{p})=0$ and $\operatorname{rank} JG_j(\mathbf{p})=s$. Then, $G_j\subset\mathbb{R}[\mathbf{x}]$ computed in Step (3d) deflates all generic points of $V_{\mathbf{p}}$. Step (4) adds all other points from W which are deflated by G_j. In particular, every other point on $V_{\mathbf{p}}$ contained in W will be added to W_j. Hence, (G_j,L,W_j) is a deflated witness set for a union of irreducible components of V, denoted by V_j, proving (ii). Since $\bigcup_j W_j=W$, we also get $\bigcup_j V_j=V$, which proves (iii). If $\mathbf{y}\in\operatorname{Sing}(V_j)$, then $\operatorname{rank}(JG_j(\mathbf{y}))<s$ so all $s\times s$ minors of $JG_j(\mathbf{y})$ vanish. Hence, $g_j(\mathbf{y})=\det(M(\mathbf{y}))=0$ proving (iv). Conversely, for any $\mathbf{p}'\in W_j$, some $s\times s$ minor of $JG_j(\mathbf{p}')$ does not vanish at $\mathbf{p}'$. Since g_j is a generic choice of combinations of all such minors, $g_j(\mathbf{p}')\neq 0$ for all $\mathbf{p}'\in W_j$. By Assumption (A), $V=\lim_{e\to 0}V_e^{\mathbf{a}}$ is equidimensional of dimension $n-s$, so for all $\mathbf{p}'\in W$, $\dim V_{\mathbf{p}'}=n-s$. Since g_j does not vanish identically on $V_{\mathbf{p}'}$ for any $\mathbf{p}'\in W_j$, we get $\dim(V_j\cap\mathcal{V}(g_j))<n-s$, proving (v).

To prove the first claim in (vi), note that each V_i is a union of $(n-s)$-dimensional irreducible components of V and sample points from the irreducible components of V are uniquely assigned to one W_j. Then for $i\neq j$, V_i and V_j cannot share an irreducible component, so their intersection is lower dimensional.

To prove the second claim in (vi) we use Theorem 2.6.8 as follows. Let $\mathbf{y}\in V_i\cap V_j$. Suppose that X is an irreducible component of V_i and Y is an irreducible component of V_j such that $\mathbf{y}\in X\cap Y$. Let $\xi\in\mathbb{C}$ be generic with $|\xi|=1$, t a complex variable, and denote $f_\xi^{\mathbf{a}}=f_\xi^{\mathbf{a}}(x,t):=(f_1-a_1t\xi,\ldots,f_s-a_st\xi)$. Then, $X\times\{0\}$ and $Y\times\{0\}$ are irreducible varieties of $\mathbb{C}^{n+1}$ and both are subsets of $V(f_\xi^{\mathbf{a}})\subset\mathbb{C}^{n+1}$. Therefore, each is contained in a unique isosingular set of $f_\xi^{\mathbf{a}}$ denoted by $\operatorname{Iso}_{f_\xi^{\mathbf{a}}}(X\times\{0\})$ and $\operatorname{Iso}_{f_\xi^{\mathbf{a}}}(Y\times\{0\})$, respectively. Let $F_i(x,t)$ and $F_j(x,t)$ be their corresponding deflated witness systems, respectively. If $F_i=F_j$, i.e. the two isosingular sets of $f_\xi^{\mathbf{a}}$ are the same, then $\operatorname{Iso}_{F_j(x,0)}(X)\neq\operatorname{Iso}_{F_j(x,0)}(Y)$ (otherwise $X=Y$) so $\mathbf{y}\in\operatorname{Sing}_{F_j(x,0)}(Y)$. Note that by the DEFLATED WITNESS SYSTEM ALGORITHM 4.3.2, $G_j(x)$ is the deflation of $F_j(x,0)$ at a generic point of V_j. This implies by Theorem 2.6.8 that $\mathbf{y}\in\operatorname{Sing}_{G_j}(Y)$ and $g_j(\mathbf{y})=0$.

If $F_i\neq F_j$, then $(\mathbf{y},0)$ is in the intersection of two different isosingular sets so $(\mathbf{y},0)$ has a different deflation sequence than $Y\times\{0\}$, i.e., $(\mathbf{y},0)\in\operatorname{Sing}_{f_\xi^{\mathbf{a}}}(Y\times\{0\})$. By Theorem 2.6.8, we have that $(\mathbf{y},0)\in\operatorname{Sing}_{F_j}(Y\times\{0\})$. Denoting the Jacobian by $J:=JF_j(x,t)$, we have that $\operatorname{rank}J(\mathbf{y})<s$ with $\operatorname{rank}J(\mathbf{y}')=s$ for all generic $\mathbf{y}'\in Y$. Consider $J':=JF_j(x,0)$. (i.e. column of J corresponding to ∂t removed). Note

that $Jf(x)$ is a submatrix of J', since $f \subset F_j(x,0)$. If $\operatorname{rank} J'(\mathbf{y}') = s$ for generic $\mathbf{y}' \in Y$, then $G_j = F_j(x,0)$, $\mathbf{y} \in \operatorname{Sing}_{G_j}(Y)$, and $g_j(\mathbf{y}) = 0$. If $\operatorname{rank} J'(\mathbf{y}') < s$ for generic $\mathbf{y}' \in Y$, we claim that $\operatorname{rank} J'(\mathbf{y}) < \operatorname{rank} J'(\mathbf{y}')$ for generic $\mathbf{y}' \in Y$. First note that both $\operatorname{rank} Jf(\mathbf{y}) \leq s-1$ and $\operatorname{rank} Jf(\mathbf{y}') \leq s-1$ for $f = (f_1, \ldots, f_s)$, so without loss of generality (after maybe some Gaussian elimination on these Jacobian matrices), we assume that $\nabla f_1(y) = \nabla f_1(y') = 0$. Note that the ∂t column of $J = JF_j(x,t)$ has the only possibly non-zero constant entries in the rows corresponding to $f_1 - a_1 t\xi, \ldots, f_s - a_s t\xi$. Then for a generic $\mathbf{y}' \in Y$ we have $\operatorname{rank} J'(\mathbf{y}') = s-1$, since among all $s \times s$ minors of $J(\mathbf{y}')$ some has to be non-zero, and the only possible non-zeros are the ones that are a_1 times the $(s-1)\times(s-1)$ minors of $J'(\mathbf{y}')$, thus we must have $a_1 \neq 0$ and $\operatorname{rank} J'(\mathbf{y}') = s-1$. On the other hand, the $s \times s$ minors of $J(\mathbf{y})$ contain all $(s-1)\times(s-1)$ minors of $J'(\mathbf{y})$ times a_1, so all these minors of $J'(\mathbf{y})$ must be zero. This implies that $\operatorname{rank} J'(\mathbf{y}) < s-1$. Thus, $\operatorname{rank} J'(\mathbf{y}) < \operatorname{rank} J'(\mathbf{y}')$. In particular, $\mathbf{y} \in \operatorname{Sing}_{F_j(x,0)}(Y)$ and by Theorem 2.6.8, $\mathbf{y} \in \operatorname{Sing}_{G_j}(Y)$ which implies that $g_j(\mathbf{y}) = 0$. This proves (vi), and the theorem. □

One advantage of the approach using isosingular deflation is that, in many problems, the number of iterations in the deflation process is a small constant (zero or one). In this case, the degrees of the polynomials in the output of both DEFLATED WITNESS SET ALGORITHM 4.3.2 and COMPUTATION OF G ALGORITHM 4.4.1 are comparable to the maximal degree of the input polynomials $f_1, \ldots, f_s$. On the other hand, the degree of the polynomial w computed in the symbolic approach in [48, Lemma 4.3] mentioned above is the degree of V bounded by the product of the degrees of the input polynomials. Nonetheless, the disadvantage of our approach is that in the worst case, we need as many iterations in the deflation as the multiplicity of the points and this may result polynomials that are higher degree than w. We have the following bound on the degree of g as a function on the number of iterations in the deflation:

Proposition 4.4.2. *Let $f = (f_1, \ldots, f_s)$ and $\mathbf{a} = (a_1, \ldots, a_s) \in \mathbb{Q}^s$ such that $V_e^{\mathbf{a}} := \mathcal{V}(f_1 - a_1 e, \ldots, f_s - a_s e)$ satisfies Assumption (A). Let $D := \max_{i=1}^{s}\{\deg(f_i)\}$ and fix $\mathbf{p} \in V := \lim_{e\to 0} V_e^{\mathbf{a}}$. If* ALGORITHM 4.3.2 *takes k iterations of the isosingular deflation to output $G \subset \mathbb{R}[\mathbf{x}]$, the degrees of the polynomials in G are bounded by $s^k D$. Furthermore, if $g(\mathbf{x}) := \det(M(\mathbf{x})) \in \mathbb{R}[\mathbf{x}]$ where $M(\mathbf{x})$ is a $s \times s$ submatrix of $JG(\mathbf{x})$, then $\deg(g) \leq s^{k+1} D$.*

Proof. The first claim follows from the fact that each iteration of the deflation algorithm adds the minors of the Jacobian of the polynomials in the previous iteration, and these minors have size less than s. Thus, the degrees of polynomials added to the system in each iteration are at most s times the degrees of the polynomials in the previous iteration. The second claim follows from the first. □

Algorithm 4.4.1 COMPUTATIONOFG

Input: $f_1,\ldots,f_s \in \mathbb{R}[\mathbf{x}]$, $\mathbf{a}=(a_1,\ldots,a_s)\in\mathbb{Q}^s$.

Output: $\{(g_j,D_j) : j=1,\ldots,r\}$ such that for all $i\neq j\in\{1,\ldots,r\}$, $V_e^{\mathbf{a}} := \mathcal{V}(f_1-a_1e,\ldots,f_s-a_se)$ and $V := \lim_{e\to 0+} V_e^{\mathbf{a}}$:

(i) $g_j \in \mathbb{R}[\mathbf{x}]$, $G_j, L \subset \mathbb{R}[\mathbf{x}]$, and $W_j \subset V$.

(ii) $D_j := (G_j, L, W_j)$ is a deflated witness set of some $V_j \subset V$, where V_j is a union of irreducible components of V;

(iii) $V = \bigcup_{j=1}^{r} V_j$

(iv) $\mathrm{Sing}(V_j) \subseteq \mathcal{V}(g_j)$

(v) $\dim(V_j \cap \mathcal{V}(g_j)) < n-s$

(vi) $\dim(V_i \cap V_j) < n-s$ and $V_i \cap V_j \subseteq \mathcal{V}(g_j)$.

1. Loop

   (a) Choose a generic system $L \subset \mathbb{R}[\mathbf{x}]$ of $n-s$ linear polynomials.

   (b) $\mathbf{a}' := (a_1,\ldots,a_s,0,\ldots,0)$.

   (c) (flag, W) := WitnessPointsInLimits($\{f_1,\ldots,f_s,L\},\mathbf{a}'$). `// See Algorithm 4.3.1`

   (d) If flag =TRUE, exit loop.

2. Set $j := 1$.

3. Loop

   (a) Pick some $\mathbf{p}\in W$.

   (b) $W_j := \{\mathbf{p}\}$.

   (c) Update $W := W \setminus \{\mathbf{p}\}$.

   (d) G_j := DeflatedWitnessSystem($\{f_1,\ldots,f_s\},\mathbf{a},\mathbf{p}$). `// See Algorithm 4.3.2`
   `//` $G_j \subset \mathbb{R}[\mathbf{x}]$ `is a witness system for the irreducible component` $V_{\mathbf{p}} \subset V$ `containing` $\mathbf{p}$ `such that` $f_1,\ldots,f_s \in G_j$, $G_j(\mathbf{p})=0$ `and` $\mathrm{rank}\, JG_j(\mathbf{p}) = s$.

   (e) For all $\mathbf{p}'\in W$

   If $G_j(\mathbf{p}')=0$ and $\mathrm{rank}\, JG_j(\mathbf{p}')=s$, then

   Update $W_j := W_j \cup \{\mathbf{p}'\}$ and $W := W\setminus\{\mathbf{p}'\}$.

   (f) Compute $g_j(\mathbf{x}) := \det(M(\mathbf{x}))$, where M is a generic rational linear combination of all $s\times s$ submatrices of $JG_j(\mathbf{x})$.

   (g) If $W\neq\emptyset$, increment $j := j+1$.

4.5 Finite Critical Points of g

In this section, we establish a result characterizing when a function g will have a finite number of critical points over an algebraic set. This is an adaptation of Theorem 36 and Lemma 37 from [30].

Definition 4.5.1. Given $f_1,\dots,f_s,g\in\mathbb{R}[\mathbf{x}]$. We say that $\mathbf{x}\in\mathbb{C}^n$ is a *critical point* of g for $\mathscr{V}(f_1,\dots,f_s)$ if $\mathbf{x}\in V(f_1,\dots,f_s)$ and

$$\nabla g(\mathbf{x})\in \operatorname{span}_{\mathbb{C}}\big(\nabla f_1(\mathbf{x}),\dots,\nabla f_s(\mathbf{x})\big),$$

where ∇ denotes the gradient operation as in Definition 2.1.17.

We need the following corollary of Sard's theorem from [52, Theorem A.6.1]. It uses the notion of *quasi-projective* sets, which are the intersection inside some projective space of a Zariski-open and a Zariski-closed subset. Let X_{reg} denote the set of smooth points in X.

Theorem 4.5.2. *Let $f(\mathbf{x})=0$ denote a system of n algebraic functions on an irreducible quasiprojective set X. Then there is a Zariski openset $U\subset\overline{f(X)}\subset\mathbb{C}^n$ such that for $\mathbf{y}\in U$, $\mathscr{V}(f(\mathbf{x})-\mathbf{y})\cap X_{\text{reg}}$ is smooth of dimension equal to the corank of f, i.e. $\dim X-\dim\overline{f(X)}$. Moreover, he Jacobian matrix of f is of rank equal to $\dim X-\dim\overline{f(X)}$ at all points of $\mathscr{V}(f(\mathbf{x})-\mathbf{y})\cap X_{\text{reg}}$.*

Theorem 4.5.3. *Let $f_1,\dots,f_s\in\mathbb{R}[\mathbf{x}]$ and assume that $\mathscr{V}(f_1,\dots,f_s)\subset\mathbb{C}^n$ is a smooth equidimensional algebraic set of dimension $n-s$. Let $g_0\in\mathbb{R}[\mathbf{x}]$. Then there exists a Zariski closed proper subset $\mathscr{S}$ of $\mathbb{C}^n$ with $\dim(\mathscr{S})<n$ such that for all $\mathbf{c}=(c_1,\dots,c_n)\in\mathbb{R}^n\setminus\mathscr{S}$ the polynomial*

$$g:=g_0\cdot\big((x_1-c_1)^2+\cdots+(x_n-c_n)^2+1\big)\in\mathbb{R}[\mathbf{x}]$$

has finitely many critical points for $\mathscr{V}(f_1,\dots,f_s)$ where g does not vanish.

Proof. Let V be an irreducible component of $\mathscr{V}(f_1,\dots,f_s)$. By our assumptions, $\dim(V)=n-s$ and V is smooth. We will prove that g has finitely many critical points for $\mathscr{V}(f_1,\dots,f_s)$ that lie in $V\setminus\mathscr{V}(g)$, and since this will be true for all irreducible components of $\mathscr{V}(f_1,\dots,f_s)$, we get the claim of the theorem.

We can assume that

$$\dim(V\cap\mathscr{V}(g))<n-s$$

otherwise, since V is irreducible, $V\subset\mathscr{V}(g)$ and there is nothing to prove.

To simplify the notation, define for $\mathbf{c}=(c_1,\dots,c_n)\in\mathbb{R}^n$

$$U_{\mathbf{c}}(\mathbf{x}):=(x_1-c_1)^2+\cdots+(x_n-c_n)^2+1.$$

Then

$$\nabla g(\mathbf{x})=U_{\mathbf{c}}(\mathbf{x})\nabla g_0(\mathbf{x})+g_0(\mathbf{x})\nabla U_{\mathbf{c}}(\mathbf{x})$$

Thus, a point $\mathbf{x}\in V$ is a critical point of g for $\mathscr{V}(f_1,\dots,f_s)$ if and only if

$$U_{\mathbf{c}}(\mathbf{x})\nabla g_0(\mathbf{x})+g_0(\mathbf{x})\nabla U_{\mathbf{c}}(\mathbf{x})\in\operatorname{span}_{\mathbb{C}}\big(\nabla f_1(\mathbf{x}),\dots,\nabla f_s(\mathbf{x})\big),$$

This implies that $\mathbf{x} \in V$ is a critical point of g for $\mathscr{V}(f_1, \ldots, f_s)$ such that $g(\mathbf{x}) \neq 0$ if and only if there exists $\lambda = (\lambda_1, \ldots, \lambda_s) \in \mathbb{C}^s$ such that

$$\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = \frac{U_{\mathbf{c}}(\mathbf{x})}{g_0(x)} \begin{bmatrix} \partial_{x_1} g_0(\mathbf{x}) \\ \vdots \\ \partial_{x_n} g_0(\mathbf{x}) \end{bmatrix} + 2 \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} - \lambda_1 \begin{bmatrix} \partial_{x_1} f_1(\mathbf{x}) \\ \vdots \\ \partial_{x_n} f_1(\mathbf{x}) \end{bmatrix} \cdots - \lambda_s \begin{bmatrix} \partial_{x_1} f_s(\mathbf{x}) \\ \vdots \\ \partial_{x_n} f_s(\mathbf{x}) \end{bmatrix}.$$

Define $p_i : W \to \mathbb{C}$ for $i = 1, \ldots, n$,

$$p_i(\mathbf{x}, t, \lambda) := t\,\partial_{x_i} g_0(\mathbf{x}) + 2x_i - \lambda_1 \partial_{x_i} f_1(\mathbf{x}) - \cdots - \lambda_s \partial_{x_i} f_s(\mathbf{x})$$

where

$$W := \{(\mathbf{x}, t, \lambda) \in V \times \mathbb{C}^{s+1} \mid g(\mathbf{x}) \neq 0, t \neq 0\}.$$

Thus, $\mathbf{x} \in V \setminus \mathscr{V}(g)$ is a critical point of g for $\mathscr{V}(f_1, \ldots, f_s)$ if and only if there exists $(t, \lambda) \in \mathbb{C}^{s+1}$ such that $(\mathbf{x}, t, \lambda)$ satisfies

$$t = \frac{U_{\mathbf{c}}(\mathbf{x})}{g_0(\mathbf{x})} \text{ and } p_i(\mathbf{x}, t, \lambda) = c_i \quad i = 1, \ldots, n.$$

First we prove that for $\mathbf{p} = (p_1, \ldots, p_n) : W \to \mathbb{C}^n$, $\mathbf{p}$ is dominant. For all $\mathbf{x}^* \in V$ and for $t = \lambda_1 = \cdots = \lambda_s = 0$ we have

$$J_W p(\mathbf{x}^*, 0, 0) = [2 \cdot I_{n-s} | \nabla g_0(\mathbf{x}^*)| - Jf(\mathbf{x}^*)],$$

where $J_W p$ is the Jacobian of $\mathbf{p}$ in a local parametrization of W at $(\mathbf{x}^*, 0, 0)$. By our assumtion on V, $\operatorname{rank} Jf(\mathbf{x}^*) = s$, thus $\operatorname{rank} J_W p(\mathbf{x}^*, 0, 0) \geq n$. This implies that the image of $\mathbf{p}$ is n-dimensional, thus $\mathbf{p}$ is dominant. Since W inherits the irreducibility and and smoothness of V, we get that $\overline{\mathbf{p}(W)} = \mathbb{C}^n$.

We can apply Theorem 4.5.2 for $\mathbf{p}$, so there exists a Zariski closed subset $\mathscr{S}$ of $\mathbb{C}^n$ and such that for all $\mathbf{c} \in \mathbb{C}^n \setminus \mathscr{S}$ for $W_1 := \{(\mathbf{x}, t, \lambda) \in W \mid \mathbf{p}(\mathbf{x}, t, \lambda) = c\}$ we have

$$\dim(W_1) = \dim(W) - n = 1$$

using that $\dim(W) = n - s + s + 1 = n + 1$ by our assumption that $\dim(V \cap \mathscr{V}(g)) < n - s$.

Fix $\mathbf{c} \in \mathbb{R}^n \setminus \mathscr{S}$. Next we show that that

$$\dim\left\{(\mathbf{x}, t, \lambda) \in W_1 \,:\, q(\mathbf{x}, t) := U_{\mathbf{c}}(\mathbf{x}) - t\,g_0(x) = 0\right\} = 0.$$

By the previous lemma applied to $q(\mathbf{x}, t) : W_1 \to \mathbb{C}$, if the above dimension is not 0 then 0 is a critical value of the function $q(\mathbf{x}, t)$. If we have such a critical value, then there exists $(\mathbf{x}^*, t^*, \lambda^*) \in W_1$ such

that $\nabla q(\mathbf{x}^*, t^*) = 0$, i.e.

$$\begin{bmatrix} \partial_{x_1} q(\mathbf{x}^*, t^*) \\ \vdots \\ \partial_{x_n} q(\mathbf{x}^*, t^*) \\ g_0(\mathbf{x}^*) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}.$$

Thus we must have $g_0(\mathbf{x}^*) = 0$. However, $W_1 \subset W$, so for $(\mathbf{x}^*, t^*, \lambda^*) \in W_1$ we have $g_0(\mathbf{x}^*) \neq 0$, a contradiction.

This implies that for any $\mathbf{c} \in \mathbb{R}^n \setminus \mathscr{S}$, the solution set of $p_i(\mathbf{x}, t, \lambda) = c_i$ for $i = 1, \ldots, n$ and the equation $t = \frac{U_{\mathbf{c}}(\mathbf{x})}{g_0(\mathbf{x})}$ is a zero dimensional subset $Z \subset W$. The set $\{\mathbf{x} : (\mathbf{x}, t, \lambda) \in Z\}$ is the finite set of critical points of g for $\mathscr{V}(f_1, \ldots, f_s)$ in $V \setminus \mathscr{V}(g)$.

□

CHAPTER

5

COMPUTATION OF REAL SMOOTH POINTS: GENERAL CASE

In this chapter, we put together all of the results from Chapter 4 to present an algorithm that computes real smooth points for the limits of these perturbed varieties, giving us real smooth points on each connected component of our real algebraic set, regardless of whether it is equidimensional to start.

5.1 Computation of Real Smooth Points

We first define two genericity assumptions, informed by our results from Chapter 4, in particular Proposition 4.2.7 and Theorem 4.5.3. Recall that $\text{crit}(V, \pi_i)$ is the polar variety of the algebraic set V with respect to the projection π_i as in Definition 2.3.8.

Definition 5.1.1. Consider polynomial F with $\mathscr{V}(F) \cap \mathbb{R}^n$ bounded, matrix $A \in \text{GL}_n(\mathbb{Q})$. Define $V^A = \mathscr{V}(F^A)$ and $V_e^A := \mathscr{V}(F^A - e) \subset \mathbb{C}^n$ for some constant $e > 0$. We say that F and A satisfy Assumption (B) if:

(1): there exists $e_0 > 0$ such that for all $0 < e \leq e_0$ and all $1 \leq i \leq n$, $\text{crit}(V_e^A, \pi_i)$ is either empty or is smooth and equidimensional with complex dimension $i-1$;

(2): for all $\mathbf{p} \in V^A \cap \mathbb{R}^n$, $\pi_d^{-1}(\pi_d(\mathbf{p})) \cap (V^A \cap \mathbb{R}^n)$ is finite, where d is greater than or equal to the local real dimension of V^A at $\mathbf{p}$;

Algorithm 5.1.1 REALSMOOTHPOINT

Input: $f=\left(f_1,\ldots,f_s\right)\subset\mathbb{Q}[x_1,\ldots,x_n]$, $i\in\{1,\ldots,n\}$, $n\geq 2$.

Output: $S\subset\mathbb{R}^n$, a finite set containing smooth points in each $(i-1)$-dim bounded connected component of $\mathscr{V}_{\mathbb{R}}\left(f\right)$

1. Define $F:=f_1^2+\cdots+f_s^2$ and $e_1:=(1,0,\ldots,0)$.
2. Choose generic $A\in GL_n(\mathbb{Q})$.
3. $\left\{\left(g_1,D_1\right),\ldots,\left(g_r,D_r\right)\right\}:=\mathsf{ComputationOfg}\left(F^A,\frac{\partial F^A}{\partial x_{i+1}},\ldots,\frac{\partial F^A}{\partial x_n},\mathbf{e}_1\right)$. `// See Algorithm 4.4.1`
   `//` D_j `is a deflated witness set for some` V_j `a union of irreducible components of` $\mathrm{crit}(V^A,\pi_i):=\lim_{e\to 0^+}\mathrm{crit}(V_e^A,\pi_i)$ `where` $\mathrm{crit}(V_e^A,\pi_i):=\mathscr{V}(F^A-e,\frac{\partial F}{\partial x_{i+1}},\ldots,\frac{\partial F}{\partial x_n})$ `for` e `a parameter.`
4. For $j=1,\ldots,r$
   (a) Loop
   $$L^{(j)}:=\left\{F^A,\frac{\partial F}{\partial x_{i12}},\ldots,\frac{\partial F}{\partial x_n},g_jz-1\right\}\cup\left\{\frac{\partial g_j}{\partial x_k}+\frac{\partial F^A}{\partial x_k}+\sum_{t=i+1}^{n}\lambda_t\frac{\partial^2F}{\partial x_t\partial x_k}:k=1,\ldots,n\right\}.$$
   `//` $L^{(j)}$ `is the Lagrange multiplier system in variables` $x_1,\ldots,x_n,z,\lambda_1,\ldots,\lambda_s$.
   $(\mathrm{flag},U_j):=\mathsf{WitnessPointsInLimits}\left(L^{(j)},\mathbf{e}_1\right)$. `// See Algorithm 4.3.1`
   If flag =TRUE, exit loop.
   Choose generic $c\in\mathbb{Q}^n$.
   $\overline{g_j}:=g_j\cdot\left((x_1-c_1)^2+\cdots(x_n-c_n)^2+1\right)$.
   Restart loop with $g_j:=\overline{g_j}$.
   (b) Compute $T_j:=U_j\backslash\mathscr{V}\left(g_j\right)\cap\mathbb{R}^n$.
   (c) Set $S_j:=\emptyset$.
   (d) For each $\mathbf{p}\in T_j$
   If $\mathsf{MembershipTest}\left(\mathbf{p},D_j\right)=$ TRUE, then `// See Algorithm 2.5.1`
   $S_j:=S_j\cup\left\{\mathbf{p}\right\}$.
5. Return $S:=\bigcup_{j=1}^{r}S_j$.

Definition 5.1.2. Consider polynomials $F,g\in\mathbb{Q}[\mathbf{x}]$ and constant $\mathbf{c}=(c_1,\ldots,c_n)\in\mathbb{Q}^n$. Define $V_e:=\mathscr{V}(F-e)\subset\mathbb{C}^n$ for some constant $e>0$. We say that F,g and $\mathbf{c}$ satisfy Assumption `(C)` if:

(C): There exists $e_0 > 0$ such that for all $0 < e \leq e_0$, all $1 \leq i \leq n$, the polynomial

$$\overline{g} := g \cdot \left((x_1 - c_1)^2 + \cdots + (x_n - c_n)^2 + 1\right) \in \mathbb{R}[\mathbf{x}]$$

has finitely many critical points for the polar variety $\operatorname{crit}(V_e, \pi_i)$ where g does not vanish.

The following theorem and corresponding proof establish the correctness of the main result of this chapter, REAL SMOOTH POINT ALGORITHM 5.1.1.

Theorem 5.1.3. *Fix $n, i, f_1, \ldots, f_s$ as in the input of* ALGORITHM 5.1.1. *Assume $A \in GL_n(\mathbb{Q})$ such that A and $F = f_1^2 + \cdots + f_s^2$ satisfy Assumption* (B) *as in Definition 5.1.1. Also, for each $j = 1, \ldots, r$, in Step (4) of* ALGORITHM 5.1.1 *we assume that F^A, g_j and $\mathbf{c}$ satisfy Assumption* (C) *as in Definition 5.1.2. Then* ALGORITHM 5.1.1 *is correct. Furthermore, if $S = \emptyset$, then $\mathcal{V}(f_1, \ldots, f_s) \cap \mathbb{R}^n$ has no bounded connected components of dimension $i-1$. If $S \neq \emptyset$, then $\mathcal{V}(f_1, \ldots, f_s) \cap \mathbb{R}^n$ has some connected components (possibly unbounded) of dimension $i-1$.*

Proof. By Assumption (B), $\operatorname{crit}(V_e^A, \pi_i)$ is smooth and equidimensional of dimension $i-1$ for all sufficiently small $e > 0$. We apply Proposition 4.1.3 (over $\mathbb{C}$ instead of $\mathbb{R}$) to show that the set $\operatorname{crit}(V^A, \pi_i) := \lim_{e \to 0} \operatorname{crit}(V_e^A, \pi_i) \subset \mathbb{C}^n$ is a Zariski closed set that is either equidimensional of dimension $i-1$ or empty. Assume that $\{(g_j, (G_j, L, W_j)) : j = 1, ..., r\}$ satisfies output specifications (i)-(vi) of Algorithm 4.4.1. Fix $j \in \{1, \ldots, r\}$ and let $V_j \subset \operatorname{crit}(V^A, \pi_i)$ be the union of irreducible components of $\operatorname{crit}(V^A, \pi_i)$ with witness set (G_j, L, W_j). First we establish that U_j defined in Step (4a) is finite. We note that since $\operatorname{crit}(V_e^A, \pi_i)$ is smooth and equidimensional for all sufficiently small $e > 0$. If $|U_j| = \infty$ then we redefine g_j with a generic $\mathbf{c} \in \mathbb{Q}^n$. Using Assumption (C), we get that the redefined U_j is finite and the loop will terminate.

Next, since $\dim(V_j \cap \mathcal{V}(g_j)) < i-1$ by (v) in Algorithm 4.4.1, either $(V_j \setminus \mathcal{V}(g_j)) \cap \mathbb{R}^n = \emptyset$ or for each bounded connected component C of $V_j \cap \mathbb{R}^n$ where g_j is not identically zero, there exists $\mathbf{z} \in U_j \cap C$ such that $g_j(\mathbf{z}) \neq 0$. Suppose $(V_j \setminus \mathcal{V}(g_j)) \cap \mathbb{R}^n \neq \emptyset$. Let $C_1, \ldots, C_t \subset V_j \cap \mathbb{R}^n$ be the bounded connected components of $V_j \cap \mathbb{R}^n$ where g_j is not identically zero. Fix $m \in \{1, \ldots, t\}$. Since each C_m is compact, the distance from C_m to C_k is positive for each $m \neq k$. Also, for all sufficiently small e, $V_e^A \cap \mathbb{R}^n$ is also compact. Since $C_m \subset V \cap \mathbb{R}^n$ is compact, Proposition 4.1.3 shows that there exist connected components $C_{m,1}^{(e)}, \ldots, C_{m,s_m}^{(e)}$ of $V_e^A \cap \mathbb{R}^n$ for all sufficiently small $e > 0$ such that $C_m = \bigcup_{l=1}^{s_m} \lim_{e \to 0^+} C_{m,l}^{(e)}$, each $C_{m,l}^{(e)}$ is bounded, and since C_m and C_j has positive distance for $m \neq j$, also by Proposition 4.1.3 we have that

$$\cup_{l=1}^{s_m} C_{m,l}^{(e)} \cap \cup_{l=1}^{s_j} C_{j,l}^{(e)} = \emptyset$$

for all $j \neq m$. For each $l = 1, \ldots, s_m$, let $\mathscr{S}_{m,l}^{(e)} := \pi_x(\mathcal{V}(L^{(j)})) \cap C_{m,l}^{(e)}$. By Lemma 3.2.2, $\mathscr{S}_{m,l}^{(e)} \neq \emptyset$ and it contains all points in $C_{m,l}^{(e)}$ where g_j takes its extreme values. Let $\mathscr{S}_m := \bigcup_{l=1}^{s_m} \lim_{e \to 0} \mathscr{S}_{m,l}^{(e)}$. Since $\mathscr{S}_{m,l}^{(e)}$ is bounded for all sufficiently small e, none of the limit points escape to infinity. Suppose that for all $\mathbf{z} \in \mathscr{S}_i$ we have $g_j(\mathbf{z}) = 0$. Since C_i is compact, by the Extreme Value Theorem, g_j attains both a minimum and a maximum on C_i. Since g_j is not identically zero on C_i, either the minimum or the

maximum value of g_j on C_i must be nonzero. Let $\mathbf{z}^* \in C_i$ such that $|g_j(\mathbf{z}^*)| > 0$. Let $\mathbf{z}_e^* \in C_{i,l}^{(e)}$ for some $l = 1, \ldots, s_i$ such that $\lim_{e\to 0} \mathbf{z}_e^* = \mathbf{z}^*$. Then for any $\mathbf{z} \in \mathscr{S}_i$, if $\mathbf{z}_e \in \mathscr{S}_i^{(e)}$ such that $\lim_{e\to 0} \mathbf{z}_e = \mathbf{z}$, then for sufficiently small e we have that $|g_j(\mathbf{z}_e^*)| > |g_j(\mathbf{z}_e)|$ by $\lim_{e\to 0} g_j(\mathbf{z}_e) = g_j(\mathbf{z}) = 0$. Since $\mathscr{S}_i$ is finite, we can choose a common e_0 value for all $\mathbf{z} \in \mathscr{S}_i$ so that if $0 < e < e_0$ then $|g_j(\mathbf{z}_e^*)| > |g_j(\mathbf{z}_e)|$ for all $\mathbf{z}_e \in \mathscr{S}_i^{(e)}$. Thus, $\mathscr{S}_i^{(e)}$ could not contain all points of $C_{i,l}^{(e)}$ for $l = 1, \ldots, s_i$ where g_j takes its extreme values, a contradiction. So this proves $\lim_{e\to 0} \pi_x(\mathscr{V}(L_e^{(j)})) \cap C_i = U_j \cap C_i$ contains a point $\mathbf{z} \in C_i$ such that $g_j(\mathbf{z}) \neq 0$.

Next, let $S_j = U_j \setminus \mathscr{V}(g_j) \cap \mathbb{R}^n \cap V_j$ and $S = \bigcup_{j=1}^r S_j$ as in Steps (4) and (5). Since $\text{crit}(V^A, \pi_i) = \bigcup_{j=1}^r V_j$ and for each $j = 1, \ldots, r$, $\text{Sing}(V_j) \subset \mathscr{V}(g_j)$, $V_k \cap V_j \subset V(g_j)$ for all $k \neq j$ by (iii)-(vi) in ALGORITHM 4.4.1, these points are smooth in $V_j \cap \mathbb{R}^n$, and also smooth in $\text{crit}(V^A, \pi_i) \cap \mathbb{R}^n$. Thus if $S \neq \emptyset$, by Theorem 3.2.1 and Proposition 4.2.8, $\mathscr{V}(f_1, \ldots, f_s) \cap \mathbb{R}^n$ must have dimension $i-1$ connected components. Conversely, if $\mathscr{V}(f_1, \ldots, f_s) \cap \mathbb{R}^n$ has a bounded connected component of dimension $i-1$, then there exists $j \in \{1, \ldots, r\}$ such that $V_j \cap \mathbb{R}^n$ has a bounded connected component of dimension $i-1$. By Theorem 3.2.1, this component has real smooth points. In fact, these real smooth points form a semi-algebraic set that has also dimension $i-1$. However, since $\dim\big(V_j \cap \mathscr{V}(g_j)\big) < i-1$, g_j does not vanish on all real smooth points of this component, but it vanishes on the singular points. By the above argument $U_j \cap \mathbb{R}^n \cap V_j$ must contain points where g_j is not zero, thus S_j and S are not empty. □

Using Proposition 4.4.2, we can bound the number of homotopy paths followed in Step (3) in the REAL SMOOTH POINT ALGORITHM 5.1.1, which is the bottleneck of our method. Note that the number of iterations r is at most $\deg(V) \leq D^n$. Thus the MEMBERSHIP TEST ALGORITHM 2.5.1 utilized in Step (4) of the REAL SMOOTH POINT ALGORITHM 5.1.1 follows at most $|W_j| = \deg(V_j) \leq \deg(V) \leq D^n$ homotopy paths.

Corollary 5.1.4. *Let $f_1, \ldots, f_s \in \mathbb{Q}[\mathbf{x}]$, $A \in GL_n(\mathbb{Q})$ such that A and $F = f_1^2 + \cdots + f_s^2$ satisfy Assumption* (B). *Consider the zero-dimensional polynomial system $L^{(j)}$ for some fixed $j \in \{1, \ldots, r\}$ as in* ALGORITHM 5.1.1. *Then, the number of complex roots of $L^{(j)}$ is bounded by $\deg(g_j)^n D^s \leq s^{(k_j+1)n} D^{n+s}$, where D is as above when we assume that $\deg(g_j) \geq D$ and k_j is the number of iterations of the isosingular deflation needed to compute G_j using* ALGORITHM 4.3.2.

Remark 5.1.5. We assumed in Theorems 3.2.3 and Algorithm 5.1.1 that g is a polynomial, but we can straightforwardly extend the results to $g : \mathbb{R}^n \to \mathbb{R}$ differentiable functions as long as $\frac{\partial g}{\partial x_i}$ for $i = 1, \ldots, n$ are rational functions.

CHAPTER

6

COMPUTING REAL DIMENSION

In this chapter, we apply the REAL SMOOTH POINT ALGORITHM 5.1.1 to compute the real dimension of real algebraic sets with the main idea as follows. Using Theorem 3.2.1, if we find a real smooth point, we find the real dimension to be the same as the complex one. If there are no real smooth points, we conclude that the real dimension is smaller than the complex dimension. In that case, we need to lower the complex dimension in a way that we do not lose any real points inside the variety. One approach is to replace the variety by its singular set which, for hypersurfaces, one simply adds all partial derivatives. However, recursively adding minors of the Jacobian matrix for higher codimension varieties can cause a drastic increase in the degree of the polynomials utilized. Here we apply an alternative technique using a sequence of polar varieties.

6.1 Related Work

The real dimension problem has been widely studied with both for the purpose of determining that property for a given semi-algebraic set, and also for use in algorithms which take the quantity as an input parameter (e.g. [8, Alg. 13.3] mentioned previously). The current state of the art deterministic algorithm is given by [8, Alg. 14.10] computing all realizable sign conditions of a polynomial system. This approach improves on previous work in [55] to obtain a complexity result with a better dependence on the number of polynomials in the input by utilizing a block elimination technique first proposed in [22]. Recent work has been presented giving probabilistic algorithms utilizing polar varieties which improve on complexity bounds even further in [46, 7]. We use a benchmark family from [7] to demonstrate the efficiency of our method.

One can also compute the real dimension by computing the real radical of a semi-algebraic set,

first studied in [12] with improvements and implementations in [41, 59, 54, 14]. The most recent implementation can be found in [48, 49] as mentioned previously. Their approach is shown to be efficient in the case when the polynomial system is smooth, but the iterative computation of singularities of singularities can increase the complexity significantly in the worst case. An alternative method using semidefinite programming techniques was proposed by [56, 37]. Finally, methods for numerically computing homologies are given in [17, 13] but only apply to the smooth case.

6.2 Numerical Real Dimension Algorithm

Our real dimension algorithm is as follows.

Algorithm 6.2.1 NUMERICAL REAL DIMENSION

Input: $f_1,\ldots,f_s \in \mathbb{Q}[x_1,\ldots,x_n]$ such that $\mathscr{V}(f_1,\ldots,f_s)\cap\mathbb{R}^n$ is compact and $n \geq 2$.

Output: The real dimension of $\mathscr{V}(f_1,\ldots,f_s)\cap\mathbb{R}^n$.

1. Let $i := n$.
2. Loop
   (a) $S := \mathsf{RealSmoothPoint}(f_1,\ldots,f_s,i)$. `// See Algorithm 5.1.1`
   `//` $S \subset \mathbb{R}^n$ `contains smooth points in` $\mathscr{V}(f_1,\ldots,f_s)\cap\mathbb{R}^n$.
   (b) If $S \neq \emptyset$, exit loop and return $i-1$.
   (c) Increment $i := i-1$.
   (d) If $i = 0$, exit loop and return -1.

Theorem 6.2.1. *Let $n \geq 2$, $f_1,\ldots,f_s \in \mathbb{Q}[\mathbf{x}]$ such that $\mathscr{V}(f_1,\ldots,f_s)\cap\mathbb{R}^n$ is compact. Assume that the conditions of Theorem 5.1.3 are satisfied for $1 \leq i \leq n$. Then* ALGORITHM 6.2.1 *is correct.*

Proof. By assumption, Theorem 5.1.3 gives the correctness of REAL SMOOTH POINT ALGORITHM 5.1.1 in Step (1). We prove by induction on $n-i<n$ that we have the following loop invariant in Step (2): $\dim(\mathscr{V}(F^A)\cap\mathbb{R}^n) \leq i-1$. This is true when $n-i=0$. Assume it is true for $n-i<n$, and we are in Step (2a) with $i>0$. By Proposition 4.2.8, $V\cap\mathbb{R}^n = \mathscr{V}(F^A)\cap\mathbb{R}^n$ for $V := \lim_{e\to 0}\operatorname{crit}\big(\mathscr{V}(F^A-e),\pi_i\big)$ since $\dim(\mathscr{V}(F^A)\cap\mathbb{R}^n)\leq i-1$ by the inductive hypothesis. In Step (2b) if $S\neq\emptyset$, V has a real smooth point by Theorem 5.1.3, so by Theorem 3.2.1 we have $\dim(V\cap\mathbb{R}^n) = \dim V = i-1$ and we return this value. If $S=\emptyset$, the compactness of $V\cap\mathbb{R}^n$ and Theorem 5.1.3 implies that there are no real smooth points on V, so $\dim(V\cap\mathbb{R}^n) < \dim V = i-1$. In this case we proceed: if $i-1=0$ then we return -1 concluding that $\mathscr{V}(F^A)\cap\mathbb{R}^n=\emptyset$, or we return to Step (2a) with $i-1>0$ maintaining the loop invariant. □

6.3 Implementation on a Benchmark Family of Problems

A benchmark family from [7] are hypersurfaces $\mathscr{V}(f_n) \subset \mathbb{C}^n$ for $n \geq 3$ such that

$$f_n(\mathbf{x}) = \left(\sum_{j=1}^n x_j^2\right)^2 - 4\sum_{j=1}^n \left(x_j x_{j+1}\right)^2 \tag{6.1}$$

where $x_{n+1} = x_1$. Since f_n is homogeneous, one knows $\dim \mathscr{V}(f_n) \cap \mathbb{R}^n = \dim(\mathscr{V}(f_n, s_n) \cap \mathbb{R}^n) + 1$ where $s_n = \sum_{j=1}^n x_j^2 - 1$ in which $\mathscr{V}(f_n, s_n) \cap \mathbb{R}^n$ is compact. The cases $3 \leq n \leq 6$ were solved in [7] with the following considering $3 \leq n \leq 8$. All code used in these computations is available at `dx.doi.org/10.7274/r0-5c1t-jw53` with the timings reported using `Bertini` ([10]) on an AMD Opteron 6378 2.4 GHz processor using one (serial) or 64 (parallel) cores.

For $n = 3$ with $g = \partial f_3/\partial x_1$, one obtains smooth points on $\mathscr{V}(f_3) \cap \mathbb{R}^3$ thereby showing $\dim \mathscr{V}(f_3) \cap \mathbb{R}^3 = 2$ in about a second in serial.

For $n = 4$, $\mathscr{V}(f_4)$ has multiplicity 2 with respect to f_4 since

$$f_4(x_1, x_2, x_3, x_4) = \left(x_1^2 - x_2^2 + x_3^2 - x_4^2\right)^2.$$

Trivially, a deflated witness system for $\mathscr{V}(f_4)$ is $G = x_1^2 - x_2^2 + x_3^2 - x_4^2$. For $g = x_1 x_2$, one obtains smooth points on $\mathscr{V}(f_4) \cap \mathbb{R}^4$ showing $\dim \mathscr{V}(f_4) \cap \mathbb{R}^4 = 3$ in about a second in serial.

For $n = 5, \ldots, 8$, with $g = \partial f_n/\partial x_1$, one does not obtain smooth points on $\mathscr{V}(f_n) \cap \mathbb{R}^n$ showing $\dim \mathscr{V}(f_n) \cap \mathbb{R}^n < n-1$. Therefore, one can move down the dimensions searching for real smooth points using perturbed polar varieties, similarly to Step (2) of Algorithm 6.2.1. Nonsingular real points are first found at dimension 2, i.e., $\dim \mathscr{V}(f_n) \cap \mathbb{R}^n = 2$. In fact, at dimension 2, the polar variety contains various irreducible components of degree 2 and testing one is enough to confirm the existence of a smooth real point. Table 6.1 lists the total computation time using parallel processing.

Table 6.1: Summary of benchmark problem (6.1) for $5 \leq n \leq 8$

n	$\dim \mathscr{V}(f_n) \cap \mathbb{R}^n$	Time (min)
5	2	3.63
6	2	5.73
7	2	34.81
8	2	159.81

CHAPTER

7

CERTIFYING SOLUTIONS OF UNDERDETERMINED SYSTEMS

In this chapter, we detail and extend the previous results of Shub and Smale's α-theory for certifying solutions to polynomial systems to the context of underdetermined systems, i.e. polynomial systems that consist of more variables than polynomials.

7.1 α-Theory Preliminaries for Underdetermined Systems

Since this chapter follows a significantly different vein than the others, we begin with a section providing some background and definitions relevant to this chapter only.

We start by describing spaces of polynomial systems together with a unitarily invariant metric, as in the approach of [50]. Let $\mathscr{P}_{\mathbf{d}}$ be the linear space of all polynomial systems $f : \mathbb{C}^n \to \mathbb{C}^m$, $f = (f_1, \ldots, f_m)$ where each f_i a polynomial of n-variables $\mathbf{x} = (x_1, \ldots, x_n)$ of degree at most d_i with $\mathbf{d} = (d_1, \ldots, d_m)$ and $d_i \geq 0$. Then we can think of each $f_i : \mathbb{C}^n \to \mathbb{C}$ as

$$f_i(\mathbf{x}) = \sum_{|\alpha| \leq d_i} a_\alpha \mathbf{x}^\alpha = \sum_{|\alpha| \leq d_i} a_\alpha x_1^{\alpha_1} \cdots x_n^{\alpha_n}$$

where $\alpha = (\alpha_1, \ldots, \alpha_n)$ and $|\alpha| = \sum \alpha_i$. For $n > m$, we call f an *underdetermined* system because it has less polynomials than variables.

Let $\mathscr{H}_{\mathbf{d}}$ be the homogenization of $\mathscr{P}_{\mathbf{d}}$, so that $f^{(h)} : \mathbb{C}^{n+1} \to \mathbb{C}^m$ is such that each $f_i^{(h)}$ is homoegeneous of degree d_i, obtained by an additional variable x_0 (as in Definition 2.1.21). Suppose $\mathbf{0} \in \mathscr{H}_{\mathbf{d}}$

so that it is a linear space. Then the natural linear isomorphism of the homogenization is given by $\Phi : \mathscr{P}_{\mathbf{d}} \to \mathscr{H}_{\mathbf{d}}$ such that $\Phi(f) = (\Phi_1(f_1), \dots \Phi_n(f_n))$ with

$$\Phi_i(f_i)(\mathbf{x}) = \sum_{\alpha} a_\alpha \mathbf{x}^\alpha x_0^{d_i - |\alpha|}$$

and the inverse of the map is obtained by taking $x_0 = 1$.

Let $U(n+1)$ be the unitary group of degree $n+1$. One common representation of this group we could think of is $(n+1) \times (n+1)$ unitary matrices with the group operation of matrix multiplication.

Definition 7.1.1. A square matrix $U \in \mathbb{C}^{(n+1)\times(n+1)}$ is *unitary* if $UU^* = U^*U = I$ where U^* is the adjoint (i.e. conjugate transpose) of U.

For homogeneous polynomials $f_1^{(h)}, f_2^{(h)} : \mathbb{C}^{n+1} \to \mathbb{C}$ of degree d given by $f_1^{(h)} = \sum_\alpha a_\alpha \mathbf{x}^\alpha$ and $f_2^{(h)} = \sum_\alpha b_\alpha \mathbf{x}^\alpha$, let

$$\langle f_1^{(h)}, f_2^{(h)} \rangle := \sum_{|\alpha|=d} a_\alpha b_\alpha \left(\frac{\alpha_1! \cdots \alpha_n!}{d!} \right).$$

This inner product natually induces the norm on homogenous polynomials of degree d denoted by

$$\left\| \sum_{|\alpha|=d} a_\alpha \mathbf{x}^\alpha \right\|^2 := \sum_{|\alpha|=d} \frac{|a_\alpha|^2}{\binom{d}{\alpha}}$$

where $\binom{d}{\alpha}$ is the multinomial coefficient $\binom{d}{\alpha_1, \dots, \alpha_n}$.

Now suppose $f := \{f_1, \dots, f_n\}$ and $g := \{g_1, \dots, g_n\}$ such that $f, g \in \mathscr{H}_{\mathbf{d}}$. Then

$$\langle f, g \rangle := \sum_i \langle f_i, g_i \rangle$$

is a Hermitian inner product.

Theorem 7.1.2. *[31, Theorem 4.1] This inner product is invariant on $\mathscr{H}_{\boldsymbol{d}}$ under the group action of $U(n+1)$, i.e. $\langle F \circ U^{-1}, G \circ U^{-1} \rangle = \langle F, G \rangle$ for all $F, G \in \mathscr{H}_{\boldsymbol{d}}$ and $U : \mathbb{C}^{n+1} \to \mathbb{C}^{n+1}$ unitary.*

By the isomorphism Φ, this also induces a Hermitian structure and its corresponding norm on $\mathscr{P}_{\mathbf{d}}$, so that for any polynomial $f_i : \mathbb{C}^n \to \mathbb{C}$, the norm is defined with respect to its homogenization by

$$\|f_i\|^2 := \sum_{|\alpha| \le d} |a_\alpha|^2 \frac{\alpha!(d - |\alpha|)!}{d!}.$$

Similarly, for a polynomial system $f : \mathbb{C}^n \to \mathbb{C}^m$, we define

$$\|f\|^2 := \sum_{i=1}^m \|f_i\|^2$$

and for a point $\mathbf{x} \in \mathbb{C}^n$ we define

$$\|\mathbf{x}\|_1^2 := 1 + \|\mathbf{x}\|^2 = 1 + \sum_{i=1}^{n} |x_i|^2.$$

Returning to our polynomial system $f : \mathbb{C}^n \to \mathbb{C}^m$, we denote the *Jacobian* matrix of the system f at any point $\mathbf{x} \in \mathbb{C}^n$ by the notation $Df(\mathbf{x})$, where the (i, j)-th entry of the matrix is $\frac{\partial f_i}{\partial x_j}(\mathbf{x})$. We note that this is the same Jacobian matrix $Jf(\mathbf{x})$ we have used in the previous chapters, but we follow convention and use an alternative notation here. Importantly, for $n > m$, this matrix is not square, and therefore does not have a well-defined inverse matrix $Df(\mathbf{x})^{-1}$.

We follow the approach of [51, Section 1C] and introduce a pseudoinverse for the non-square $Df(\mathbf{x})$ with full rank m.

Definition 7.1.3. Suppose A is an $m \times n$ matrix with full rank m. Then the *Moore-Penrose inverse* of A, denoted by $A^\dagger$, is defined to be the $n \times m$ matrix

$$A^\dagger = A^*(AA^*)^{-1}.$$

We note that $AA^\dagger = I$ and $A^\dagger A$ is the orthogonal projection onto $(\ker A)^\perp$ (i.e. the orthogonal complement of the kernel of A).

Using this pseudoinverse, Newton's method constructs a sequence of points $\mathbf{x}_1, \mathbf{x}_2, \ldots$ by the formula

$$\mathbf{x}_k = \mathbf{x}_{k-1} - Df(\mathbf{x}_{k-1})^\dagger f(\mathbf{x}_{k-1})$$

for $k \in \mathbb{N}$ in an effort to approximate the common zeros of f, $\mathcal{V}(f) := \{\xi \in \mathbb{C}^n | f(\xi) = 0\}$.

Formally, we define a *Newton iteration* for our underdetermined system f at $\mathbf{x}$. Let $N_f : \mathbb{C}^n \to \mathbb{C}^n$ be the map such that

$$N_f(\mathbf{x}) = \mathbf{x} - Df(\mathbf{x})^\dagger f(\mathbf{x})$$

for $\mathbf{x} \in \mathbb{C}^n$. We note that for $n = m$, N_f is a usual iteration of Newton's method. Furthermore, for $n > m$, N_f is well-defined with the above pseudoinverse formula whenever $Df(x)$ is surjective (i.e. has full rank m).

Lemma 7.1.4. *[51, Section 1C] For $\mathbf{x} \in \mathbb{C}^n$ such that $Df(\mathbf{x})$ surjective, $\mathbf{x}$ is a fixed point of the map N_f if and only if $f(\mathbf{x}) = 0$.*

Repeated applications of the map $N_f(\mathbf{x})$ allow us to defined approximate solutions of f. For $k \in \mathbb{N}$, let

$$N_f^k(\mathbf{x}) := \underbrace{N_f \circ \cdots \circ N_f(\mathbf{x})}_{k \text{ times}}$$

be the kth Newton iteration of f starting at $\mathbf{x}$.

Definition 7.1.5. For a fixed $\mathbf{x} \in \mathbb{C}^n$ we say that the sequence of points $N_f^k(\mathbf{x}) \in \mathbb{C}^n$ for $k \in \mathbb{N}$ obtained via Newton's method *converges quadratically* to $\xi \in \mathbb{C}^n$ an exact solution of f if

$$\left\| N_f^k(\mathbf{x}) - \xi \right\| \leq \left(\frac{1}{2}\right)^{2^k - 1} \|\mathbf{x} - \xi\|.$$

Then we say that $\mathbf{x} \in \mathbb{C}^n$ is an *approximate solution* of f.

Smale's α-theory consists of results for proving when a point $\mathbf{x} \in \mathbb{C}^n$ is an approximate solution of f. These results are given in terms of constants $\beta(f,\mathbf{x}), \gamma(f,\mathbf{x})$, and $\alpha(f,\mathbf{x})$. For $Df(\mathbf{x})$ surjective, these are defined by

$$\begin{aligned}
\beta(f,\mathbf{x}) &:= \left\| Df(\mathbf{x})^\dagger f(\mathbf{x}) \right\| \\
\gamma(f,\mathbf{x}) &:= \max_{k>1} \left\| \frac{Df(\mathbf{x})^\dagger D^k f(\mathbf{x})}{k!} \right\|^{\frac{1}{k-1}} \\
\alpha(f,\mathbf{x}) &:= \beta(f,\mathbf{x})\gamma(f,\mathbf{x}).
\end{aligned}$$

If $Df(\mathbf{x})$ is not surjective and $\mathbf{x} \notin \mathscr{V}(f)$, we define $\alpha(f,\mathbf{x}) := \beta(f,\mathbf{x}) := \gamma(f,\mathbf{x}) := \infty$. For consistency, if $Df(\mathbf{x})$ is not surjective, but $\mathbf{x} \in \mathscr{V}(f)$, we defined $\alpha(f,\mathbf{x}) := \beta(f,\mathbf{x}) := 0$ and $\gamma(f,\mathbf{x}) := \infty$. We explain the notation in the definition of $\gamma(f,\mathbf{x})$ further in the next section.

7.2 Establishing Shub-Smale Constant Bounds

The main theorem of α-theory gives an upper bound on $\alpha(f,\mathbf{x})$ for quadratic Newton convergence, which Shub and Smale also state for the underdetermined case.

Theorem 7.2.1. *[51, Theorem C1] Let $f : \mathbb{C}^n \to \mathbb{C}^m$ and $\mathbf{x} \in \mathbb{C}^n$. If*

$$\alpha(f,\mathbf{y}) < \frac{13 - 3\sqrt{17}}{4} \approx 0.157671,$$

then $\mathbf{x} \in \mathbb{C}^n$ is an approximate solution of f.

We note that our definition of $\alpha(f,\mathbf{x})$ depends directly on $\gamma(f,\mathbf{x})$, so an upper bound on $\gamma(f,\mathbf{x})$ should be our goal. However, the definition of $\gamma(f,\mathbf{x})$ is complicated, so we should first understand that.

$D^k f(\mathbf{x})$ is the kth-derivative of the system $f(\mathbf{x})$. Following the definitions of Lang in [33, Chapter 5], this is a tensor whose components are the partial derivatives of f of order k. For each $f_i(\mathbf{x})$ in the system, this looks like the k-fold symmetric power $S^k\mathbb{C}^n$. Since there are m polynomials in our system, the tensor consists of m of these k-fold components. When we consider $Df(\mathbf{x})^\dagger D^k f(\mathbf{x})$, we note in particular that $Df(\mathbf{x})^\dagger$ has m column vectors in $\mathbb{C}^n$. So to compute this product, we apply each column vector k times to the respective k-fold symmetric power, and result in a scalar in $\mathbb{C}$. Repeating this for each column vector gives a resulting vector in $\mathbb{C}^m$.

Therefore, in the definition of $\gamma(f,\mathbf{x})$ we are considering the map $Df(\mathbf{x})^\dagger D^k f(\mathbf{x}) : S^k\mathbb{C}^n \to \mathbb{C}^m$. The operator norm of this k-fold linear map is taken with respect to the norm on $S^k\mathbb{C}^n$ that is dual to the standard unitarily invariant norm on homogenous polynomials that we previously discussed [26]. We utilize the definition of this operator norm explicitly in the proof of Proposition 7.3.4.

We note that $\gamma(f,\mathbf{x})$ defined in this matter can be expensive to compute, so instead we work to find a practical upper bound via condition numbers. Following the approach of [50], we modify the definition of the condition number slightly in order to obtain a scaled version of the standard definition which is more useful in our results.

Definition 7.2.2. Let $\Delta(\omega)$ be the $m\times m$ diagonal matrix whose ith entry is ω_i. Then for $f:\mathbb{C}^n\to\mathbb{C}^m$ a polynomial system where f_i has degree d_i, $\mathbf{x}\in\mathbb{C}^n$, and $Df(\mathbf{x})$ surjective, we define the *condition number*

$$\mu(f,\mathbf{x}) := \max\left\{1, \left\|f\right\| \cdot \left\| Df(\mathbf{x})^\dagger \cdot \Delta\left(d_i^{\frac{1}{2}} \left\|\mathbf{x}\right\|_1^{d_i-1}\right)\right\|\right\}.$$

We note that this definition is similar to one given in [50, I-3], but modified to include the pseudoinverse. We can now state our main result, a γ bound for the underdetermined case analogous to [50, I-3. Proposition 3] for the square case.

Theorem 7.2.3. *Let $f:\mathbb{C}^n\to\mathbb{C}^m$ be a polynomial system with $d_i=\deg f_i$ and $D=\max d_i$. Then for $\mathbf{x}\in\mathbb{C}^n$ such that $Df(\mathbf{x})$ is surjective,*

$$\gamma(f,\mathbf{x}) \le \frac{\mu(f,\mathbf{x})D^{\frac{3}{2}}}{2\left\|\mathbf{x}\right\|_1}.$$

7.3 Proving the γ Bound

This section follows the approach of [50, III-1], including some direct results reproven with more detail for clarity. We first prove a couple of inequalities for how the unitarily invariant norm interacts with homogeneous polynomials.

Proposition 7.3.1. *[50, III-1. Proposition 1] Let $g:\mathbb{C}^{n+1}\to\mathbb{C}$ be a homogeneous polynomial of degree d. Then for $\mathbf{x}\in\mathbb{C}^{n+1}$,*

$$\frac{|g(\mathbf{x})|}{\left\|\mathbf{x}\right\|^d} \le \left\|g\right\|.$$

Proof. Let $\mathbf{y}=(\left\|\mathbf{x}\right\|,0,\ldots,0)\in\mathbb{C}^{n+1}$. Define $U:\mathbb{C}^{n+1}\to\mathbb{C}^{n+1}$ represented by a matrix with first row vector $\frac{\overline{\mathbf{x}}}{\left\|\mathbf{x}\right\|}$ and the other rows given by any set of vectors extending to an orthonormal basis. For example, if $n=1$ one possible U would be

$$\begin{pmatrix} \frac{\overline{x_0}}{\left\|\mathbf{x}\right\|} & \frac{\overline{x_1}}{\left\|\mathbf{x}\right\|} \\ \frac{-x_1}{\left\|\mathbf{x}\right\|} & \frac{x_0}{\left\|\mathbf{x}\right\|} \end{pmatrix}.$$

By definition, U is a unitary automorphism such that $U\mathbf{x} = \mathbf{y}$. Now define the homogeneous polynomial $h = g \circ U^{-1}$ such that

$$h(\mathbf{x}) = g(U^{-1}\mathbf{x}) = \sum_{\alpha} a_\alpha \mathbf{x}^\alpha.$$

Then we compute

$$\frac{|g(\mathbf{x})|}{\|\mathbf{x}\|^d} = \frac{|(g \circ U^{-1})(U\mathbf{x})|}{\|\mathbf{x}\|^d} = \frac{|h(\mathbf{y})|}{\|\mathbf{y}\|^d} = \frac{|a_{(d,0,\dots,0)}|\,\|\mathbf{x}\|^d}{\|\mathbf{y}\|^d} = |a_{(d,0,\dots,0)}| \leq \|h\|.$$

Since the norm as defined in the preliminaries is unitarily invariant, $\|h\| = \|g\|$. So we have

$$\frac{|g(\mathbf{x})|}{\|\mathbf{x}\|^d} \leq \|g\|$$

as required.

□

We now extend this result to a polynomial system.

Proposition 7.3.2. *[50, III-1. Proposition 2] If $f \in \mathscr{H}_{\mathbf{d}}$, then*

$$\left\|\Delta\left(\|\mathbf{x}\|^{-d_i}\right) \cdot f(\mathbf{x})\right\| \leq \|f\|.$$

Proof. We recall $f : \mathbb{C}^{n+1} \to \mathbb{C}^m$ is a polynomial system with $d_i = \deg f_i$. By Proposition 7.3.1, we have that for $i = 1, \dots, m$,

$$\frac{|f_i(\mathbf{x})|}{\|\mathbf{x}\|^{d_i}} \leq \|f_i\|.$$

Squaring both sides gives

$$\left(\|\mathbf{x}\|^{-d_i} f_i(\mathbf{x})\right)^2 \leq \|f_i\|^2.$$

Then we sum over i and simplify using definitions

$$\sum_{i=1}^{m} \left(\|\mathbf{x}\|^{-d_i} f_i(\mathbf{x})\right)^2 \leq \sum_{i=1}^{m} \|f_i\|^2$$
$$\left\|\Delta\left(\|\mathbf{x}\|^{-d_i}\right) \cdot f(\mathbf{x})\right\|^2 \leq \|f\|^2$$

Taking the square root of both sides of the inequality gives our result.

□

Now we consider how derivatives of homogeneous polynomials interact with the unitarily invariant norm.

Lemma 7.3.3. *[50, III-1. Lemma 1] Let $g:\mathbb{C}^{n+1}\to\mathbb{C}$ be a homogeneous polynomial, $\mathbf{x}\in\mathbb{C}^{n+1}$, and $U:\mathbb{C}^{n+1}\to\mathbb{C}^{n+1}$ a unitary automorphism. Then*

$$D\left[g\circ U^{-1}\right](U\mathbf{x}) = Dg(\mathbf{x})\cdot U^{-1}.$$

Proof. Consider the $(n+1)\times(n+1)$ inverse matrix $U^{-1}=(v_{i,j})_{i,j=0}^{n}$. Let $\mathbf{r}_0,\mathbf{r}_1,\ldots,\mathbf{r}_n$ be the row vectors of U^{-1} and $\mathbf{c}_0,\mathbf{c}_1,\ldots,\mathbf{c}_n$ be the column vectors of U^{-1}. Then

$$U^{-1}\mathbf{x} = (\mathbf{r}_0\cdot\mathbf{x},\mathbf{r}_1\cdot\mathbf{x},\ldots,\mathbf{r}_n\cdot\mathbf{x}).$$

So

$$\left[g\circ U^{-1}\right](\mathbf{x}) = g(U^{-1}\mathbf{x}) = g(\mathbf{r}_0\cdot\mathbf{x},\mathbf{r}_1\cdot\mathbf{x},\ldots,\mathbf{r}_n\cdot\mathbf{x}).$$

Using the chain rule for multivariate functions and simplifying using the entries of U^{-1}, we obtain the ith partial derivative

$$\begin{aligned}
\frac{\partial}{\partial x_i}\left[g\circ U^{-1}\right](\mathbf{x}) &= \frac{\partial}{\partial x_i}\left(g(\mathbf{r}_0\cdot\mathbf{x},\mathbf{r}_1\cdot\mathbf{x},\ldots,\mathbf{r}_n\cdot\mathbf{x})\right)\\
&= \sum_{j=0}^{n}\frac{\partial}{\partial x_j}\left(g(\mathbf{r}_0\cdot\mathbf{x},\mathbf{r}_1\cdot\mathbf{x},\ldots,\mathbf{r}_n\cdot\mathbf{x})\right)\cdot\frac{\partial}{\partial x_i}\left(\mathbf{r}_j\cdot\mathbf{x}\right)\\
&= \sum_{j=0}^{n}\frac{\partial}{\partial x_j}g(U^{-1}\mathbf{x})\cdot\frac{\partial}{\partial x_i}\left(\mathbf{r}_j\cdot\mathbf{x}\right)\\
&= \frac{\partial}{\partial x_0}g(U^{-1}\mathbf{x})v_{0,i}+\cdots+\frac{\partial}{\partial x_n}g(U^{-1}\mathbf{x})v_{n,i}\\
&= Dg(U^{-1}\mathbf{x})\cdot\mathbf{c}_i.
\end{aligned}$$

Then

$$\begin{aligned}
D\left[g\circ U^{-1}\right](\mathbf{x}) &= \left(Dg(U^{-1}\mathbf{x})\cdot\mathbf{c}_0, Dg(U^{-1}\mathbf{x})\cdot\mathbf{c}_1,\ldots,Dg(U^{-1}\mathbf{x})\cdot\mathbf{c}_n\right)\\
&= Dg(U^{-1}\mathbf{x})\cdot U^{-1}.
\end{aligned}$$

Substituting $U\mathbf{x}$ for $\mathbf{x}$ gives

$$D\left[g\circ U^{-1}\right](U\mathbf{x}) = Dg(U^{-1}(U\mathbf{x}))\cdot U^{-1} = Dg(\mathbf{x})\cdot U^{-1}.$$

□

Proposition 7.3.4. *[50, III-1. Proposition 3] Let $g:\mathbb{C}^{n+1}\to\mathbb{C}$ be a homogeneous polynomial of degree d. Given $\mathbf{x},\mathbf{y}_1,\ldots,\mathbf{y}_k\in\mathbb{C}^{n+1}$,*

$$\left\|D^k g(\mathbf{x})(\mathbf{y}_1,\ldots,\mathbf{y}_k)\right\| \le d(d-1)\cdots(d-k+1)\left\|g\right\|\|\mathbf{x}\|^{d-k}\left\|\mathbf{y}_1\right\|\cdots\left\|\mathbf{y}_k\right\|.$$

Proof. Without loss of generality, suppose $\left\|\mathbf{y}_1\right\| = 1$. Define $U : \mathbb{C}^{n+1} \to \mathbb{C}^{n+1}$ represented by a matrix with first column vector $\overline{\mathbf{y}_1}$ and the other columns given by any set of vectors extending to an orthonormal basis. Then U is a unitary automorphism such that $U\mathbf{y}_1 = \mathbf{e}_0$, the standard basis vector $(1,0,\ldots,0)$. Now consider $h = g \circ U^{-1}$ such that

$$h(\mathbf{x}) = g(U^{-1}\mathbf{x}) = \sum_{|\alpha|=d} a_\alpha x_0^{\alpha_0} x_1^{\alpha_1} \cdots x_n^{\alpha_n}.$$

Then

$$Dh(\mathbf{x}) \cdot \mathbf{e}_0 = \sum_{|\alpha|=d, \alpha_0 \neq 0} \alpha_0 a_\alpha x_0^{\alpha_0 - 1} x_1^{\alpha_1} \cdots x_n^{\alpha_n}.$$

We note that $Dh(\mathbf{x})$ is a homogeneous polynomial of degree $d-1$. So using the unitarily invariant norm, we compute

$$\begin{aligned} \|Dh(\mathbf{x}) \cdot \mathbf{e}_0\|_{(d-1)} &= \|D\left[h \circ U\right](\mathbf{x}) \cdot \mathbf{e}_0\|_{(d-1)} \\ &= \left\|Dh(U\mathbf{x}) \cdot U\mathbf{y}_1\right\|_{(d-1)} \\ &= \left\|D\left[g \circ U^{-1}\right](U\mathbf{x}) \cdot U\mathbf{y}_1\right\|_{(d-1)}. \end{aligned}$$

Applying Lemma 7.3.3 gives

$$D\left[g \circ U^{-1}\right](U\mathbf{x}) \cdot U\mathbf{y}_1 = Dg(\mathbf{x}) \cdot U^{-1} \cdot U\mathbf{y}_1 = Dg(\mathbf{x}) \cdot \mathbf{y}_1.$$

So

$$\left\|Dg(\mathbf{x}) \cdot \mathbf{y}_1\right\|_{(d-1)} = \|Dh(\mathbf{x}) \cdot \mathbf{e}_0\|_{(d-1)}.$$

Using properties of the norm and squaring for simplicity gives

$$\begin{aligned} \left\|Dg(\mathbf{x}) \cdot \mathbf{y}_1\right\|_{(d-1)}^2 &= \sum_{|\alpha|=d, \alpha_0 \neq 0} |\alpha_0|^2 |a_\alpha|^2 \frac{(\alpha_0 - 1)!\alpha_1! \cdots \alpha_n!}{(d-1)!} \\ &= d \sum_{|\alpha|=d, \alpha_0 \neq 0} |\alpha_0| |a_\alpha|^2 \frac{\alpha_0! \cdots \alpha_n!}{d!} \\ &\leq d^2 \sum_{|\alpha|=d} |a_\alpha|^2 \frac{\alpha_0! \cdots \alpha_n!}{d!} \\ &= d^2 \left\|g\right\|^2 \\ &= d^2 \left\|g\right\|^2 \left\|\mathbf{y}_1\right\|^2 \end{aligned}$$

where the last step holds since $\left\|\mathbf{y}_1\right\| = 1$. Taking the square root of both sides of the inequality gives

$$\left\|Dg(\mathbf{x})(\mathbf{y}_1)\right\|_{(d-1)} \leq d \left\|g\right\| \left\|\mathbf{y}_1\right\|.$$

To show this in general, we proceed by induction as we did in our base case, assuming that

$$\left\|D^{k-1}g(\mathbf{x})(\mathbf{y}_1,\ldots,\mathbf{y}_{k-1})\right\|_{(d-k+1)} \leq d(d-1)\cdots(d-k+2)\left\|g\right\|\left\|\mathbf{y}_1\right\|\cdots\left\|\mathbf{y}_{k-1}\right\|.$$

Without loss of generality, suppose $\left\|\mathbf{y}_k\right\| = 1$. Define $U : \mathbb{C}^{n+1} \to \mathbb{C}^{n+1}$ represented by a matrix with first column vector $\overline{\mathbf{y}_k}$ and the other columns given by any set of vectors extending to an orthonormal basis. Then U is a unitary automorphism such that $U\mathbf{y}_k = \mathbf{e}_0$, the standard basis vector $(1,0,\ldots,0)$. Now consider $h = D^{k-1}g \circ U^{-1}$ such that

$$h(\mathbf{x}) = D^{k-1}g(U^{-1}\mathbf{x}) = \sum_{|\alpha|=d-k+1} a_\alpha x_0^{\alpha_0} x_1^{\alpha_1}\cdots x_n^{\alpha_n}.$$

Then

$$Dh(\mathbf{x})\cdot\mathbf{e}_0 = \sum_{|\alpha|=d-k+1,\alpha_0\neq 0} \alpha_0 a_\alpha x_0^{\alpha_0-1} x_1^{\alpha_1}\cdots x_n^{\alpha_n}.$$

We note that $Dh(\mathbf{x})$ is a homogeneous polynomial of degree $d-k$. So using the unitarily invariant norm, we compute

$$\begin{aligned}\|Dh(\mathbf{x})\cdot\mathbf{e}_0\|_{(d-k)} &= \|D[h\circ U](\mathbf{x})\cdot\mathbf{e}_0\|_{(d-k)}\\ &= \left\|Dh(U\mathbf{x})\cdot U\mathbf{y}_k\right\|_{(d-k)}\\ &= \left\|D\left[D^{k-1}g\circ U^{-1}\right](U\mathbf{x})\cdot U\mathbf{y}_k\right\|_{(d-k)}.\end{aligned}$$

Applying Lemma 7.3.3 gives

$$D\left[D^{k-1}g\circ U^{-1}\right](U\mathbf{x})\cdot U\mathbf{y}_k = D^k g(\mathbf{x})\cdot U^{-1}\cdot U\mathbf{y}_k = D^k g(\mathbf{x})\cdot\mathbf{y}_k.$$

So

$$\left\|D^k g(\mathbf{x})\cdot\mathbf{y}_k\right\|_{(d-k)} = \|Dh(\mathbf{x})\cdot\mathbf{e}_0\|_{(d-k)}.$$

Using properties of the norm and squaring for simplicity gives

$$\begin{aligned}
\left\|D^{k-1}(\mathbf{x})\cdot\mathbf{y}_k\right\|^2_{(d-k)} &= \sum_{|\alpha|=d+k-1,\alpha_0\neq 0} |\alpha_0|^2|a_\alpha|^2\frac{(\alpha_0-1)!\alpha_1!\cdots\alpha_n!}{(d-k+2)!}\\
&= (d-k+1)\sum_{|\alpha|=d+k-1,\alpha_0\neq 0} |\alpha_0||a_\alpha|^2\frac{\alpha_0!\cdots\alpha_n!}{(d-k+1)!}\\
&\leq (d-k+1)^2\sum_{|\alpha|=d+k-1} |a_\alpha|^2\frac{\alpha_0!\cdots\alpha_n!}{(d-k+1)!}\\
&= (d-k+1)^2\left\|D^{k-1}g(\mathbf{x})(\mathbf{y}_1,\ldots,\mathbf{y}_{k-1})\right\|^2\\
&\leq (d-k+1)^2\left[d(d-1)\cdots(d-k+2)\left\|g\right\|\left\|\mathbf{y}_1\right\|\cdots\left\|\mathbf{y}_{k-1}\right\|\right]^2\\
&= (d-k+1)^2\left[d(d-1)\cdots(d-k+2)\left\|g\right\|\left\|\mathbf{y}_1\right\|\cdots\left\|\mathbf{y}_{k-1}\right\|\right]^2\left\|\mathbf{y}_k\right\|^2
\end{aligned}$$

where the last step holds since $\left\|\mathbf{y}_k\right\| = 1$. Taking the square root of both sides of the inequality gives

$$\left\|D^k g(\mathbf{x})(\mathbf{y}_1,\ldots,\mathbf{y}_k)\right\|_{(d-k)} \leq d(d-1)\cdots(d-k+1)\left\|g\right\|\left\|\mathbf{y}_1\right\|\cdots\left\|\mathbf{y}_k\right\|.$$

Finally, we note $\left\|D^k g(\mathbf{x})(\mathbf{y}_1,\ldots,\mathbf{y}_k)\right\|_{(d-k)}$ is the norm in terms of a homogeneous polynomial of degree $d-k$. However, we want a statement in terms of the dual operator norm, which is defined by

$$\left\|D^k g(\mathbf{x})(\mathbf{y}_1,\ldots,\mathbf{y}_k)\right\| := \sup_{\mathbf{x}\in\mathbb{C}^{n+1}}\frac{|D^k g(\mathbf{x})(\mathbf{y}_1,\ldots,\mathbf{y}_k)|}{\|\mathbf{x}\|^k}.$$

For any $\mathbf{x}\in\mathbb{C}^{n+1}$, we have

$$\left\|D^k g(\mathbf{x})(\mathbf{y}_1,\ldots,\mathbf{y}_k)\right\| \leq \frac{|D^k g(\mathbf{x})(\mathbf{y}_1,\ldots,\mathbf{y}_k)|}{\|\mathbf{x}\|^k}.$$

Applying Proposition 7.3.1 gives

$$\left\|D^k g(\mathbf{x})(\mathbf{y}_1,\ldots,\mathbf{y}_k)\right\| \leq \frac{\left\|D^k g(\mathbf{x})(\mathbf{y}_1,\ldots,\mathbf{y}_k)\right\|_{(d-k)}\|\mathbf{x}\|^d}{\|\mathbf{x}\|^k} = \left\|D^k g(\mathbf{x})(\mathbf{y}_1,\ldots,\mathbf{y}_k)\right\|_{(d-k)}\|\mathbf{x}\|^{d-k}.$$

Substituting in our result for $\left\|D^k g(\mathbf{x})(\mathbf{y}_1,\ldots,\mathbf{y}_k)\right\|_{(d-k)}$ gives

$$\left\|D^k g(\mathbf{x})(\mathbf{y}_1,\ldots,\mathbf{y}_k)\right\| \leq d(d-1)\cdots(d-k+1)\left\|g\right\|\|\mathbf{x}\|^{d-k}\left\|\mathbf{y}_1\right\|\cdots\left\|\mathbf{y}_k\right\|.$$

□

We now put together these results to obtain a degree bound on these norms of derivatives of homogeneous polynomial systems. Once again, we begin with a bound on a single polynomial.

Lemma 7.3.5. *[50, III-1. Lemma 4] Let $g:\mathbb{C}^{n+1}\to\mathbb{C}$ be a homogeneous polynomial of degree d. Then*

for every $k > 1$,

$$\left(\frac{\left\|D^k g(\mathbf{x})(\mathbf{y}_1,\ldots,\mathbf{y}_k\right\|}{d^{\frac{1}{2}}\|\mathbf{x}\|^{d-k}\,k!\left\|g\right\|\left\|\mathbf{y}_1\right\|\cdots\left\|\mathbf{y}_k\right\|}\right)^{\frac{1}{k-1}} \leq \frac{d^{\frac{1}{2}}(d-1)}{2}.$$

Proof. We recall that Proposition 7.3.4 gave us

$$\left\|D^k g(\mathbf{x})(\mathbf{y}_1,\ldots,\mathbf{y}_k)\right\| \leq d(d-1)\cdots(d-k+1)\left\|g\right\|\|\mathbf{x}\|^{d-k}\left\|\mathbf{y}_1\right\|\cdots\left\|\mathbf{y}_k\right\|.$$

Dividing through by the right hand side gives

$$\frac{\left\|D^k g(\mathbf{x})(\mathbf{y}_1,\ldots,\mathbf{y}_k)\right\|}{d(d-1)\cdots(d-k+1)\left\|g\right\|\|\mathbf{x}\|^{d-k}\left\|\mathbf{y}_1\right\|\cdots\left\|\mathbf{y}_k\right\|} \leq 1.$$

Multiplying both sides of the inequality by $\frac{d(d-1)\cdots(d-k+1)}{d^{1/2}k!}$ and taking the $k-1$-th root gives

$$\left(\frac{\left\|D^k g(\mathbf{x})(\mathbf{y}_1,\ldots,\mathbf{y}_k)\right\|}{d^{\frac{1}{2}}k!\left\|g\right\|\|\mathbf{x}\|^{d-k}\left\|\mathbf{y}_1\right\|\cdots\left\|\mathbf{y}_k\right\|}\right)^{\frac{1}{k-1}} \leq \left(\frac{d(d-1)\cdots(d-k+1)}{d^{\frac{1}{2}}k!}\right)^{\frac{1}{k-1}}.$$

We consider the term on the right hand side of the inequality. We note that for $d \geq k > 2$ integers,

$$\left(\prod_{i=1}^{k-1}\frac{d-i}{i+1}\right)^{\frac{1}{k-1}} > \frac{d-k}{k+1},$$

since each of the $k-1$ terms in the product is larger than the term on the right hand side of the inequality. Then for $2 \leq k < d$ we have

$$\left(\frac{\frac{d(d-1)\cdots(d-k+1)}{d^{1/2}k!}}{\frac{d(d-1)\cdots(d-k)}{d^{1/2}(k+1)!}}\right)^k = \frac{d^{\frac{1}{2(k-1)}}\left(\prod_{i=1}^{k-1}\frac{d-i}{i+1}\right)^{\frac{1}{k-1}}}{\frac{d-k}{k+1}} > 1.$$

So

$$\left(\frac{d(d-1)\cdots(d-k+1)}{d^{\frac{1}{2}}k!}\right)^{\frac{1}{k-1}}$$

in terms of k is a decreasing function for $2 \leq k < d$. Hence it's maximum occurs at $k = 2$,

$$\frac{d^{\frac{1}{2}}(d-1)}{2},$$

and thus we result in

$$\left(\frac{\left\|D^k g(\mathbf{x})(\mathbf{y}_1,\ldots,\mathbf{y}_k)\right\|}{d^{\frac{1}{2}}k!\left\|g\right\|\|\mathbf{x}\|^{d-k}\left\|\mathbf{y}_1\right\|\cdots\left\|\mathbf{y}_k\right\|}\right)^{\frac{1}{k-1}} \leq \frac{d^{\frac{1}{2}}(d-1)}{2}.$$

□

Now we extend to a bound for a polynomial system, similarly to how we did with Proposition 7.3.2.

Proposition 7.3.6. *[50, III-1. Theorem 1] Let $f \in \mathscr{H}_{\boldsymbol{d}}$ such that $f : \mathbb{C}^{n+1} \to \mathbb{C}^m$ is a polynomial system with $d_i = \deg f_i$ and $D = \max(d_i)$. For $\mathbf{x} \in \mathbb{C}^{n+1}$,*

$$\left(\frac{\left\| \Delta\left(\|\mathbf{x}\|^{d_i-k} d_i^{\frac{1}{2}} \right)^{-1} D^k f(\mathbf{x}) \right\|}{k! \left\| f \right\|} \right)^{\frac{1}{k-1}} \leq \frac{D^{\frac{3}{2}}}{2}.$$

Proof. Using the definitions of $\Delta(\omega)$ and the unitarily invariant norm on a homogeneous system in the numerator, we simplify

$$\begin{aligned} \left(\frac{\left\| \Delta\left(\|\mathbf{x}\|^{d_i-k} d_i^{\frac{1}{2}} \right)^{-1} D^k f(\mathbf{x}) \right\|}{k! \left\| f \right\|} \right)^{\frac{1}{k-1}} &= \left(\sum_{i=1}^{m} \left(\frac{\left\| \left(\|\mathbf{x}\|^{d_i-k} d^{\frac{1}{2}} \right)^{-1} D^k f_i(\mathbf{x}) \right\|}{k! \left\| f \right\|} \right)^2 \right)^{\frac{1}{2(k-1)}} \\ &\leq \left(\sum_{i=1}^{m} \left(\frac{\left\| D^k f_i(\mathbf{x}) \right\|}{\|\mathbf{x}\|^{d_i-k} k! \left\| f \right\| d^{\frac{1}{2}}} \right)^2 \right)^{\frac{1}{2(k-1)}}. \end{aligned}$$

For each term in the summation, we multiply by the relevant $\left\| f_i \right\| / \left\| f_i \right\|$ and apply Lemma 7.3.5 to compute

$$\left(\sum_{i=1}^{m} \left(\frac{\left\| D^k f_i(\mathbf{x}) \right\|}{\|\mathbf{x}\|^{d_i-k} k! \left\| f \right\| d^{\frac{1}{2}}} \right)^2 \right)^{\frac{1}{2(k-1)}} \leq \left(\sum_{i=1}^{m} \left(\left(\frac{d_i^{\frac{1}{2}}(d_i-1)}{2} \right)^{k-1} \frac{\left\| f_i \right\|}{\left\| f \right\|} \right)^2 \right)^{\frac{1}{2(k-1)}}.$$

Again applying the definition of the unitarily invariant norm on a polynomial system and using that $D = \max(d_i)$, we simplify as follows

$$\begin{aligned} \left(\sum_{i=1}^{m} \left(\left(\frac{d_i^{\frac{1}{2}}(d_i-1)}{2} \right)^{k-1} \frac{\left\| f_i \right\|}{\left\| f \right\|} \right)^2 \right)^{\frac{1}{2(k-1)}} &\leq \left(\left(\frac{D^{\frac{1}{2}}(D-1)}{2} \right)^{k-1} \frac{\left\| f \right\|}{\left\| f \right\|} \right)^{\frac{1}{k-1}} \\ &= \frac{D^{\frac{1}{2}}(D-1)}{2} \\ &\leq \frac{D^{\frac{3}{2}}}{2}. \end{aligned}$$

□

Finally, this brings us to the proof of our main result, Theorem 7.2.3, which states

$$\gamma(f,\mathbf{x}) \leq \frac{\mu(f,\mathbf{x})D^{\frac{3}{2}}}{2\|\mathbf{x}\|_1}.$$

We note that up to this point in this section, our results have been stated in terms of homogeneous polynomials. However, in the first section of the chapter we defined the norm on any polynomial to be the one defined on its canonical homogenization. Therefore, all of the results can be extended to the non-homogeneous case, provided that we utilize this definition of the norm and substitute $\|\mathbf{x}\|_1$ for $\|\mathbf{x}\|$ to account for the switch from $\mathbf{x} \in \mathbb{C}^{n+1}$ to $\mathbf{x} \in \mathbb{C}^n$. We prove our main result below in this context to allow for the widest possible application.

Proof of Theorem 7.2.3. Let $f : \mathbb{C}^n \to \mathbb{C}^m$ be a polynomial system with $d_i = \deg f_i$ and $D = \max d_i$. Suppose $\mathbf{x} \in \mathbb{C}^n$ such that $Df(\mathbf{x})$ is surjective. We recall by definition that

$$\gamma(f,\mathbf{x}) = \max_{k>1} \left\| \frac{Df(\mathbf{x})^\dagger D^k f(\mathbf{x})}{k!} \right\|^{\frac{1}{k-1}}.$$

We rewrite the right hand side of the equation

$$\begin{aligned}
&\max_{k>1} \left\| \frac{Df(\mathbf{x})^\dagger D^k f(\mathbf{x})}{k!} \right\|^{\frac{1}{k-1}} \\
&= \max_{k>1} \left\| \frac{f \cdot Df(\mathbf{x})^\dagger \cdot \Delta\left(d_i^{\frac{1}{2}}\right) \cdot \Delta\left(\|\mathbf{x}\|_1^{d_i-1}\right) \cdot \Delta\left(d_i^{-\frac{1}{2}}\right) \cdot \Delta\left(\|\mathbf{x}\|_1^{-(d_i-k)}\right) D^k f(\mathbf{x})}{f \cdot k!} \right\|^{\frac{1}{k-1}} \cdot \frac{1}{\|\mathbf{x}\|_1} \\
&\leq \max_{k>1} \left\| f \cdot Df(\mathbf{x})^\dagger \cdot \Delta\left(d_i^{\frac{1}{2}}\right) \cdot \Delta\left(\|\mathbf{x}\|_1^{d_i-1}\right) \right\|^{\frac{1}{k-1}} \cdot \left\| \frac{\Delta\left(d_i^{-\frac{1}{2}}\right) \cdot \Delta\left(\|\mathbf{x}\|_1^{-(d_i-k)}\right) D^k f(\mathbf{x})}{f \cdot k!} \right\|^{\frac{1}{k-1}} \cdot \frac{1}{\|\mathbf{x}\|_1} \\
&\leq \max_{k>1} \left(\|f\| \left\| Df(\mathbf{x})^\dagger \cdot \Delta\left(d_i^{\frac{1}{2}} \|\mathbf{x}\|_1^{d_i-1}\right) \right\| \right)^{\frac{1}{k-1}} \cdot \left(\frac{\left\| \Delta\left(\|\mathbf{x}\|^{d_i-k} d_i^{\frac{1}{2}}\right)^{-1} D^k f(\mathbf{x}) \right\|}{k! \|f\|} \right)^{\frac{1}{k-1}} \cdot \frac{1}{\|\mathbf{x}\|_1}
\end{aligned}$$

Applying the definition of $\mu(f,\mathbf{x})$ and Proposition 7.3.6 gives

$$\begin{aligned}
&\max_{k>1}\left(\|f\|\left\|Df(\mathbf{x})^{\dagger}\cdot\Delta\left(d_i^{\frac{1}{2}}\|\mathbf{x}\|_1^{d_i-1}\right)\right\|\right)^{\frac{1}{k-1}}\cdot\left(\frac{\left\|\Delta\left(\|\mathbf{x}\|^{d_i-k}d_i^{\frac{1}{2}}\right)^{-1}D^k f(\mathbf{x})\right\|}{k!\|f\|}\right)^{\frac{1}{k-1}}\cdot\frac{1}{\|\mathbf{x}\|_1}\\
&\leq\max_{k>1}\mu(f,\mathbf{x})^{1/(k-1)}\frac{D^{\frac{3}{2}}}{2\|\mathbf{x}\|_1}\\
&\leq\frac{D^{\frac{3}{2}}}{2\|\mathbf{x}\|_1},
\end{aligned}$$

where the last step follows since $\mu(f,\mathbf{x})\geq 1$. □

BIBLIOGRAPHY

[1] P. Aubry, F. Rouillier, and M. Safey El Din. Real solving for positive dimensional systems. *J. Symbolic Comput.*, 34(6):543–560, 2002.

[2] B. Bank, M. Giusti, J. Heintz, G. Lecerf, G. Matera, and P. Solernó. Degeneracy loci and polynomial equation solving. *Foundations of Computational Mathematics*, 15(1):159–184, 2015.

[3] B. Bank, M. Giusti, J. Heintz, R. Mandel, and G. M. Mbakop. Polar varieties and efficient real equation solving: the hypersurface case. In *Proceedings of the 3rd International Conference on Approximation and Optimization in the Caribbean (Puebla, 1995)*, page 13. Benemérita Univ. Autón. Puebla, Puebla, 1997.

[4] B. Bank, M. Giusti, J. Heintz, and L. M. Pardo. Generalized polar varieties and an efficient real elimination procedure. *Kybernetika (Prague)*, 40(5):519–550, 2004.

[5] B. Bank, M. Giusti, J. Heintz, and L. M. Pardo. On the intrinsic complexity of point finding in real singular hypersurfaces. *Inform. Process. Lett.*, 109(19):1141–1144, 2009.

[6] B. Bank, M. Giusti, J. Heintz, M. Safey El Din, and E. Schost. On the geometry of polar varieties. *Applicable Algebra in Engineering, Communication and Computing*, 21(1):33–83, Jan 2010.

[7] I. Bannwarth and M. Safey El Din. Probabilistic algorithm for computing the dimension of real algebraic sets. In *ISSAC'15—Proceedings of the 2015 ACM International Symposium on Symbolic and Algebraic Computation*, pages 37–44. ACM, New York, 2015.

[8] S. Basu, R. Pollack, and M.-F. Roy. *Algorithms in real algebraic geometry*, volume 10 of *Algorithms and Computation in Mathematics*. Springer-Verlag, Berlin, second edition, 2006.

[9] S. Basu, R. Pollack, and M.-F. Roy. Computing the dimension of a semi-algebraic set. *Journal of Mathematical Sciences*, 134(5):2346–2353, May 2006.

[10] D. J. Bates, J. D. Hauenstein, A. J. Sommese, and C. W. Wampler. Bertini: Software for numerical algebraic geometry. Available at `bertini.nd.edu`.

[11] D. J. Bates, J. D. Hauenstein, A. J. Sommese, and C. W. Wampler. *Numerically solving polynomial systems with Bertini*, volume 25 of *Software, Environments, and Tools*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2013.

[12] E. Becker and R. Neuhaus. Computation of real radicals of polynomial ideals. In *Computational algebraic geometry (Nice, 1992)*, volume 109 of *Progr. Math.*, pages 1–20. Birkhäuser Boston, Boston, MA, 1993.

[13] P. Bürgisser, F. Cucker, and P. Lairez. Computing the homology of basic semialgebraic sets in weak exponential time. *J. ACM*, 66(1):Art. 5, 30, 2019. [Publication date initially given as 2018].

[14] C. Chen, J. H. Davenport, J. P. May, M. Moreno Maza, B. Xia, and R. Xiao. Triangular decomposition of semi-algebraic systems. *J. Symbolic Comput.*, 49:3–26, 2013.

[15] G. E. Collins. Quantifier elimination for real closed fields by cylindrical algebraic decompostion. In H. Brakhage, editor, *Automata Theory and Formal Languages*, pages 134–183, Berlin, Heidelberg, 1975. Springer Berlin Heidelberg.

[16] O. Coss, J. D. Hauenstein, H. Hong, and D. K. Molzahn. Locating and counting equilibria of the Kuramoto model with rank-one coupling. *SIAM J. Appl. Algebra Geom.*, 2(1):45–71, 2018.

[17] F. Cucker, T. Krick, and M. Shub. Computing the homology of real projective sets. *Found. Comput. Math.*, 18(4):929–970, 2018.

[18] B. H. Dayton and Z. Zeng. Computing the multiplicity structure in solving polynomial systems. In *ISSAC'05*, pages 116–123. ACM, New York, 2005.

[19] J.-C. Faugère, G. Moroz, F. Rouillier, and M. Safey El Din. Classification of the perspective-three-point problem, discriminant variety and real solving polynomial systems of inequalities. In *ISSAC'08 – Proceedings of the 2008 ACM International Symposium on Symbolic and Algebraic Computation*, pages 79–86, New York, NY, USA, 2008. ACM.

[20] M. Giusti and J.-C. Yakoubsohn. Multiplicity hunting and approximating multiple roots of polynomial systems. In *Recent advances in real complexity and computation*, volume 604 of *Contemp. Math.*, pages 105–128. Amer. Math. Soc., Providence, RI, 2013.

[21] D. R. Grayson and M. E. Stillman. Macaulay2, a software system for research in algebraic geometry. Available at `http://www.math.uiuc.edu/Macaulay2/`.

[22] D. Y. Grigor′ev and N. N. Vorobjov, Jr. Solving systems of polynomial inequalities in subexponential time. *J. Symbolic Comput.*, 5(1-2):37–64, 1988.

[23] B. Hassett. *Introduction to algebraic geometry*. Cambridge University Press, 2007.

[24] J. D. Hauenstein. Numerically computing real points on algebraic sets. *Acta Applicandae Mathematicae*, 125(1):105–119, Jun 2013.

[25] J. D. Hauenstein, B. Mourrain, and A. Szanto. On deflation and multiplicity structure. *J. Symbolic Comput.*, 83:228–253, 2017.

[26] J. D. Hauenstein and F. Sottile. Algorithm 921: Alphacertified: Certifying solutions to polynomial systems. *ACM Trans. Math. Softw.*, 38(4), Aug. 2012.

[27] J. D. Hauenstein and F. Sottile. Algorithm 921: alphaCertified: certifying solutions to polynomial systems. *ACM Trans. Math. Software*, 38(4):Art. 28, 20, 2012.

[28] J. D. Hauenstein and C. W. Wampler. Isosingular sets and deflation. *Found. Comput. Math.*, 13(3):371–403, 2013.

[29] J. D. Hauenstein and C. W. Wampler. Unification and extension of intersection algorithms in numerical algebraic geometry. *Appl. Math. Comput.*, 293:226–243, 2017.

[30] H. Hong, J. Rohal, M. S. E. Din, and E. Schost. Connectivity in semi-algebraic sets i, 2020.

[31] E. Kostlan. *On the Distribution of Roots of Random Polynomials*, pages 419–431. Springer US, New York, NY, 1993.

[32] Y. Kuramoto. Self-entrainment of a population of coupled non-linear oscillators. *Lect. Notes Phys.*, 39:420–422, 1975.

[33] S. Lang. *Real and Functional Analysis*. Graduate Texts in Mathematics. Springer New York, 2012.

[34] A. Leykin, J. Verschelde, and A. Zhao. Newton's method with deflation for isolated singularities of polynomial systems. *Theoret. Comput. Sci.*, 359(1-3):111–122, 2006.

[35] A. Leykin, J. Verschelde, and A. Zhao. Higher-order deflation for polynomial systems with isolated singular solutions. In *Algorithms in algebraic geometry*, volume 146 of *IMA Vol. Math. Appl.*, pages 79–97. Springer, New York, 2008.

[36] A. Logar. A computational proof of the noether normalization lemma. *Applied Algebra, Algebraic Algorithms and Error-Correcting Codes*, page 259–273, 1989.

[37] Y. Ma, C. Wang, and L. Zhi. A certificate for semidefinite relaxations in computing positive-dimensional real radical ideals. *J. Symbolic Comput.*, 72:1–20, 2016.

[38] I. G. Macdonald and I. G. Macdonald. *Introduction to commutative algebra*. Addison-Wesley, 1969.

[39] M. Marshall. *Positive polynomials and sums of squares*, volume 146 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2008.

[40] H. C. Mork and R. Piene. Polars of real singular plane curves. *Algorithms in Algebraic Geometry The IMA Volumes in Mathematics and its Applications*, page 99–115, 2008.

[41] R. Neuhaus. Computation of real radicals of polynomial ideals. II. *J. Pure Appl. Algebra*, 124(1-3):261–280, 1998.

[42] F. Rouillier, M.-F. Roy, and M. Safey El Din. Finding at least one point in each connected component of a real algebraic set defined by a single equation. *Journal of Complexity*, 16(4):716 – 750, 2000.

[43] M. Safey El Din. Testing sign conditions on a multivariate polynomial and applications. *Math. Comput. Sci.*, 1(1):177–207, 2007.

[44] M. Safey El Din and E. Schost. Polar varieties and computation of one point in each connected component of a smooth algebraic set. In *Proceedings of the 2003 International Symposium on Symbolic and Algebraic Computation*, pages 224–231. ACM, New York, 2003.

[45] M. Safey El Din and P.-J. Spaenlehauer. Critical point computations on smooth varieties: degree and complexity bounds. In *Proceedings of the 2016 ACM International Symposium on Symbolic and Algebraic Computation*, pages 183–190. ACM, New York, 2016.

[46] M. Safey El Din and E. Tsigaridas. A probabilistic algorithm to compute the real dimension of a semi-algebraic set. *CoRR*, abs/1304.1928, 2013.

[47] M. Safey El Din and E. Tsigaridas. Personal communication. ICERM - Nonlinear Algebra Program, 2018.

[48] M. Safey El Din, Z.-H. Yang, and L. Zhi. On the complexity of computing real radicals of polynomial systems. In *ISSAC'18—Proceedings of the 2018 ACM International Symposium on Symbolic and Algebraic Computation*, pages 351–358. ACM, New York, 2018.

[49] M. Safey El Din, Z.-H. Yang, and L. Zhi. Computing real radicals and S-radicals of polynomial systems. *Journal of Symbolic Computation*, 2019.

[50] M. Shub and S. Smale. Complexity of bezout's theorem i: Geometric aspects. *Journal of the American Mathematical Society*, 6(2):459–501, 1993.

[51] M. Shub and S. Smale. Complexity of bezout's theorem iv: Probability of success; extensions. *SIAM Journal on Numerical Analysis*, 33(1):128–148, 1996.

[52] A. J. Sommese and C. W. Wampler, II. *The numerical solution of systems of polynomials*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2005. Arising in engineering and science.

[53] F. Sottile. Real algebraic geometry for geometric constraints. In *Handbook of geometric constraint systems principles*, Discrete Math. Appl. (Boca Raton), pages 273–285. CRC Press, Boca Raton, FL, 2019.

[54] S. J. Spang. A zero-dimensional approach to compute real radicals. *Comput. Sci. J. Moldova*, 16(1):64–92, 2008.

[55] N. Vorobjov. Complexity of computing the local dimension of a semialgebraic set. *J. Symbolic Comput.*, 27(6):565–579, 1999.

[56] F. Wang. *Computation of Real Radical Ideals by Semidefinite Programming and Iterative Methods*. PhD thesis, University of Western Ontario, 2016.

[57] W. Wu and G. Reid. Finding points on real solution components and applications to differential polynomial systems. In *ISSAC 2013—Proceedings of the 38th International Symposium on Symbolic and Algebraic Computation*, pages 339–346. ACM, New York, 2013.

[58] X. Xin, T. Kikkawa, and Y. Liu. Analytical solutions of equilibrium points of the standard Kuramoto model: 3 and 4 oscillators. In *2016 American Control Conference (ACC)*, pages 2447–2452, 2016.

[59] G. Zeng. Computation of generalized real radicals of polynomial ideals. *Sci. China Ser. A*, 42(3):272–280, 1999.