ABSTRACT

AYYILDIZ AKOĞLU, TÜLAY . Certifying solutions to polynomial systems over $\mathbb{Q}$. (Under the direction of Dr. Agnes Szanto and Dr. Jonathan Hauenstein.)

This dissertation is concerned with certifying that a given point is near an exact root of a polynomial system with rational coefficients.

In Chapter 1, we provide prerequisite background material from algebraic geometry, number theory and matrix theory. Most importantly we introduce the problem of certification and its classical solution with α-theory on well-constrained systems.

In Chapter 2, we establish a method to certify approximate solutions of an overdetermined system with rational coefficients.The difficulty lies in the fact that consistency of overdetermined systems is not a continuous property. Our certification is based on hybrid symbolic-numeric methods to compute an exact rational univariate representation (RUR) of a component of the input system from approximate roots. For overdetermined polynomial systems with simple roots, we compute an initial RUR from approximate roots. The accuracy of the RUR is increased via Newton iterations until the exact RUR is found, which we certify using exact arithmetic. Since the RUR is well-constrained, we can use it to certify the given approximate roots using α-theory. We prove that our algorithms have complexity that are polynomial in the input plus the output size upon successful convergence, and we use worst case upper bounds for termination when our iteration does not converge to an exact RUR.

In Chapter 3, we focus on certifying isolated singular roots. We use a determinantal form of the isosingular deflation, which adds new polynomials to the original system without introducing new variables. The resulting polynomial system is overdetermined, but the roots are now simple, thereby reducing the problem to the overdetermined case.

Finally, in Chapter 4 we propose a method to certify approximate real solutions of polynomial systems using the signature of Hermite matrices. We use approximate roots to construct the Hermite matrices, then rationalize the entries with a preset bound on denominators. Once we ensure that the rationalized Hermite matrices in fact correspond to the given system, one can use the Hermite's theorem to certify a real approximate solutions of the given polynomial system.

Certifying solutions to polynomial systems over $\mathbb{Q}$

by
Tülay Ayyıldız Akoğlu

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

Applied Mathematics

Raleigh, North Carolina

2016

APPROVED BY:

Dr. Agnes Szanto
Co-chair of Advisory Committee

Dr. Jonathan Hauenstein
Co-chair of Advisory Committee

Dr. Seth Sullivant

Dr. Hoon Hong

Dr. Christian Melander

DEDICATION

To my family.

BIOGRAPHY

Tulay Ayyildiz Akoglu was born and raised in Ankara, Turkey. She graduated with a B.S. in Mathematics from Ankara University in 2006. The following year Tulay started her graduate study at the same university, however she earned a scholarship to pursue her graduate study in U.S before completing her study in Ankara. In 2010, she completed a M.S. in Mathematical Sciences at Clemson University. After graduating she enrolled in the Ph.D. program at North Carolina State University. Her post graduation goal is to continue research and learn more about computational algebraic geometry and polynomial optimization.

ACKNOWLEDGEMENTS

First and foremost I want to thank my advisors Agnes Szanto and Jonathan Hauenstein. It has been an honor to be their Ph.D. student. I am forever indebted to Agnes for all her contributions of time, ideas, and funding to make my Ph.D. experience possible. I am also thankful for the excellent example she has provided as a successful woman mathematician and professor. I will forever be thankful to Jonathan for his scientific advice, knowledge and many insightful discussions and suggestions, also solving my computational problems in a minute after I spent days.

I have to thank the members of my Ph.D. committee Seth Sullivant and Hoon Hong, for their excellent input. I am also very grateful to have my friends Seyma Ozcan and Melissa Ngamini at NCSU.

Finally I would like to thank my family, for their never ending support. And special thanks goes to my loving mom Sazimet Ayyildiz, my wonderful husband Kemal Akoglu and the sunshine of my life Baris, and the joy of my life Erdem. Iyi ki varsiniz.

TABLE OF CONTENTS

CHAPTER

1

PRELIMINARIES

1.1 Algebraic geometry background

We start with presenting some basic and fundamental subjects in algebraic geometry. We will mostly follow the notation in [18], [19] and [32].

1.1.1 Ideals and Varieties

Notation 1.1.1. *For the n-tuple $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$ and a collection of n indeterminate $\{x_1, \dots, x_n\}$, the product $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ is called a* **monomial**. *The total degree, often simply referred to as degree, is the sum of the exponents* $\deg(x^\alpha) := \sum_{i=1}^{n} \alpha_i$.
A **polynomial** *is a finite linear combination of monomials $f := \sum_\alpha f_\alpha x^\alpha$. The degree of a polynomial is the maximum degree of its monomials.*
Let $\mathbb{K}$ denote a field and $\mathbb{K}[x_1, \dots, x_n]$ the polynomials in $x_1, \dots, x_n$ with coefficients in $\mathbb{K}$.

Definition 1.1.2. *Let $\mathbf{f} := (f_1, \dots, f_N)$ be a list of polynomials in $\mathbb{K}[x_1, \dots, x_n]$. If $N = n$, the polynomial system $\mathbf{f}$ is called a* **well-constrained system**. *If $N > n$, then it is called an* **overdetermined system**.

Definition 1.1.3. *Let $\bar{\mathbb{K}}$ be an algebraically closed field and $\mathbf{f} := (f_1, \dots, f_N)$ be a list of polynomials in $\mathbb{K}[x_1, \dots, x_n]$, $\boldsymbol{I} := \langle f_1, \dots, f_N \rangle$ the ideal generated by $\mathbf{f}$ in $\mathbb{K}[x_1, \dots, x_n]$. The set of all common solutions* *of $f_1,\dots,f_N$ is called (affine)* **variety** *defined by $f_1,\dots,f_N$, and it is denoted by*

$$\mathbf{V}(f_1,\dots,f_N) := \{\boldsymbol{\xi} = (\xi_1,\dots,\xi_n) \in k^n \ : \ f_i(\boldsymbol{\xi}) = 0 \ i = 1,\dots,N\}$$

Definition 1.1.4. *Let $V \subset \mathbb{K}^n$, the ideal of V is defined*

$$\boldsymbol{I}(V) := \{f \in \mathbb{K}[x_1,\dots,x_n] \ : \ f(a_1,\dots,a_n) = 0 \ \textit{for all} \ (a_1,\dots,a_n) \in V\}.$$

Let I, J be ideals in $\mathbb{K}[x_1,\dots,x_n]$, and V, W be varieties in $\mathbb{K}^n$. Then the inclusion relation is

$$V \subset W \Rightarrow \boldsymbol{I}(W) \subset \boldsymbol{I}(V) \ \text{ and } \ I \subset J \Rightarrow \mathbf{V}(J) \subset \mathbf{V}(I). \qquad (1.1)$$

Definition 1.1.5. *Let V, V_1, V_2 be varieties in $\mathbb{K}^n$. V is* **reducible** *if we can write it as a union of a closed proper subsets*

$$V = V_1 \cup V_2, \quad V_1, V_2 \subsetneq V.$$

V is **irreducible**, *if not reducible, i.e. if whenever $V = V_1 \cup V_2$, then either $V_1 = V$ or $V_2 = V$.*

Now we will show that every variety can be written as a union of irreducible components.

Theorem 1.1.6 ([32, Theorem 6.4])**.** *Let V be a variety in $\mathbb{K}^n$, it can be written as*

$$V = V_1 \cup V_2 \cup \cdots \cup V_k \ , \ k \in \mathbb{N},$$

as a finite union of irreducible variety. This presentation is unique up to permutation provided it is redundant, i.e., $V_i \not\subset V_j$ for any $i \neq j$.

Proof. **(Existence)** If V be irreducible, then $k = 1$. Now assume V is reducible, then there are V_1 and V_1' such that $V = V_1 \cup V_1'$, $V_1, V_1' \subsetneq V$. Then either V_1 and V_1' are irreducible or, after reordering we can write $V_1 = V_2 \cup V_2'$, $V_2, V_2' \subsetneq V_1$. This process can terminate with an expression of V as a union of irreducibles, such that $V \supsetneq V_1 \supsetneq V_2 \dots$, or equivalently there is an infinite ascending sequence of ideals, $I(V) \subsetneq I(V_1) \subsetneq I(V_2) \dots$ by (1.1). However, this contradicts *ascending chain condition* for ideals (see [32], Proposition 2.24).

(Uniqueness) Suppose there are two different representations

$$V = V_1 \cup V_2 \cup \cdots \cup V_{k_1} \text{ and } V = V_1' \cup V_2' \cup \cdots \cup V_{k_2}', \ k_1, k_2 \in \mathbb{N}$$

with any $i \neq j$, $V_i \not\subset V_j$ and $V_i' \not\subset V_j'$. We have $V_j = V_j \cap V = \cup_{i=1}^{k_2} (V_j \cap V_i')$, then $V_j \subset V_i'$ for some i. Similarly, $V_i' \subset V_m$ for some m. The irredundancy assumption implies that $j = m$. Thus we obtain

$V_j \subset V_i' \subset V_j$, so the sets are equal. □

Definition 1.1.7. *The* **radical** *of an ideal I is the set of f, such that there exist $m \in \mathbb{N}_+$ such that f^m is in I. The radical of an ideal I is denoted as $\sqrt{I}$. An ideal I is a* **radical ideal** *if $I = \sqrt{I}$.*

Lemma 1.1.8. *Let V be a variety. Then $\boldsymbol{I}(V)$ is a radical ideal.*

Proof. Assume that $f^m \in \boldsymbol{I}(V)$ for some integer $m \geq 1$. Let $\xi \in V$, then $(f(\xi))^m = 0$, that implies $f(\xi) = 0$. Since ξ is an arbitrary element, we have $f \in \boldsymbol{I}(V)$. □

Definition 1.1.9. *The* **quotient algebra** *in $\mathbb{K}[x_1,\ldots,x_n]/I$ consists of all cosets $[f] := f + I = \{f + q \ : \ q \in I\}$ for $f \in \mathbb{K}[x_1,\ldots,x_n]$.*

Definition 1.1.10. *An ideal $I \subset \mathbb{K}[x_1,\ldots,x_n]$ is a* **zero-dimensional** *ideal if $\mathbb{K}[x_1,\ldots,x_n]/I$ is finite dimensional over $\mathbb{K}$.*

We continue with a very important result relating to the number of points in $\mathbf{V}(I)$ and $\dim \mathbb{K}[x_1,\ldots,x_n]/I$. First, we need the following lemma.

Lemma 1.1.11. *Let $\mathbf{V} = \{\xi_1,\ldots,\xi_d\}$ be a finite set in an algebraically closed field k^n. There exist polynomials $L_i \in \mathbb{K}[x_1,\ldots,x_n]$, $i = 1,\ldots,d$, such that*

$$L_i(\xi_j) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases} \tag{1.2}$$

Proof. Let $\mathbf{V} = \{\xi_1,\ldots,\xi_d\}$, where $\xi_j = (\xi_{j1},\ldots,\xi_{jn}) \in k^n$. we will construct a *Lagrange basis* $\{L_1,\ldots,L_d\}$ for $\mathbf{V}$ satisfying (1.2). We start with L_1, let t_i be a coordinate index where $\xi_{1,t_i} \neq \xi_{i,t_i}$, then we define L_1 as

$$L_1(x_1,\ldots,x_n) := \prod_{i=2}^{n} \frac{x_{t_i} - \xi_{1,t_i}}{\xi_{i,t_i} - \xi_{1,t_i}}. \tag{1.3}$$

Notice that, $L_1(\xi_1) = 1$ and $L_2(\xi_1) = \cdots = L_d(\xi_1) = 0$. Thus (1.2) holds for $L_j(\xi_1)$. The polynomials $L_2,\ldots,L_d$ can be defined similarly, and they will have the desired property. □

Theorem 1.1.12. *Let I be a zero-dimensional ideal. Then $|\mathbf{V}(I)| \leq \dim \mathbb{K}[x_1,\ldots,x_n]/I$. Moreover, if I is radical, equality occurs.*

Proof. Let $\{L_1,\ldots,L_d\}$ for $\mathbf{V}(I)$ be a Lagrange basis as defined in the proof of Lemma (1.1.11). Consider $[L_1],\ldots,[L_d] \in \mathbb{K}[x_1,\ldots,x_n]/I$, and suppose there exists $a_1,\ldots,a_d \in k$ such that

$$a_1[L_1] + \cdots + a_d[L_d] = [0],$$

the definition of quotient algebra implies that $a_1L_1+\cdots+a_dL_d\in\mathscr{I}$. Then we can write

$$a_1L_1(\xi_j)+\cdots+a_dL_d(\xi_j)=0 \text{ for all } \quad j=1,\dots,d.$$

By (1.2), $a_j=0$ for all $j=1,\dots,d$. Therefore $[L_1],\dots,[L_d]$ are linearly independent in $\mathbb{K}[x_1,\dots,x_n]/I$. It means $\dim\mathbb{K}[x_1,\dots,x_n]/I$ is at least d. This completes the first part of the theorem.
Now assume that I is radical and $[f]\in\mathbb{K}[x_1,\dots,x_n]/I$ for a polynomial $f\in\mathbb{K}[x_1,\dots,x_n]$. Let $g:=c_1L_1+\cdots+c_dL_d\in\mathbb{K}[x_1,\dots,x_n]$, where $c_i:=f(\xi_i)\in k$ for all $i=1,\dots,d$. Then $[g]$ is spanned by $[L_1],\dots,[L_d]$ and for all $i=1,\dots,d$, $g(\xi_i)=f(\xi_i)$, which implies $g-f\in I(\mathbf{V})$. Since I is radical, we have $I=I(V)$, thus $[f]=[g]$. This completes the proof. □

1.1.2 Univariate Representations

Consider a polynomial system $\mathbf{f}=(f_1,\dots,f_N)\in\mathbb{Q}[x_1,\dots,x_n]$, and assume that the ideal $\mathscr{I}:=\langle f_1,\dots,f_N\rangle$ is radical and zero dimensional. In this section we will show that, under these assumptions, we can compute an equivalent system, so-called representation, using only univariate polynomials.

Definition 1.1.13. *A polynomial $u\in\mathbb{C}[x_1,\dots,x_n]$* **separates** *$V(\mathscr{I})$, if for all $z,z'\in V(\mathscr{I})$, $z\neq z'\Rightarrow u(z)\neq u(z')$.*

The following [74, Lemma 2.1] guarantees the existence of such elements.

Lemma 1.1.14. *Let V be a finite set in $\mathbb{C}^n$ with $|V|=d$, the finite set of linear forms*

$$\mathscr{T}=\{u_i=x_1+ix_2+\cdots+i^{n-1}x_n,\ \ i=0,1,\dots,(n-1)d(d-1)/2\}$$

contains at least one element that separates V.

Proof. Let z and z' be two distinct points in V, such that $u_i(z)=z_1+iz_2+\cdots+i^{n-1}z_n$ and $u_i(z')=z'_1+iz'_2+\cdots+i^{n-1}z'_n$. The polynomial $\sum_{j=1}^n(z_j-z'_j)T^{j-1}$ is not equal to zero since z and z' are distinct, and it has at most $n-1$ distinct roots. Then $\{u_0,\dots,u_{n-1}\}$ has at least one element u_k such that $u_k(z)\neq u_k(z')$. Since the number of distinct pairs of points in V is $d(d-1)/2$, the set of polynomials $\left\{x_1+ix_2+\cdots+i^{n-1}x_n\ :\ i\in\mathbb{N},0\le i\le(n-1)d(d-1)/2\right\}$ contains at least one element that separates V. □

Definition 1.1.15. *A polynomial $u\in\mathscr{T}$ is called a* **primitive element** *of $V(\mathscr{I})$, if it separates $V(\mathscr{I})$.*

We will define two representations of $\mathbf{f} = (f_1, \ldots, f_N) \in \mathbb{Q}[x_1, \ldots, x_n]$, *Polynomial Univariate Representation (PUR)* and *Rational Univariate Representation (RUR)*, also called *Shape Lemma Representation* and *Kronecker representation*, respectively (see [77]). The history of these notions go back to [53] or possibly even earlier, with modern references, for example, in [40, 3, 74] (see [30] for a detailed historical survey). While RUR can be defined for roots with multiplicities, here we will only consider the case of simple roots.

Definition 1.1.16. *Let $\mathbf{f} = (f_1, \ldots, f_N) \in \mathbb{Q}[x_1, \ldots, x_n]$ for some $N \geq n$, and assume that the ideal $\mathscr{I} := \langle f_1, \ldots, f_N \rangle$ is radical and zero dimensional. Let*

$$\delta := \dim_{\mathbb{C}} \mathbb{Q}[x_1, \ldots, x_n]/\mathscr{I} = |V(\mathscr{I})|.$$

Let $(\lambda_1, \ldots, \lambda_n) \in \mathbb{Q}^n$ and define a primitive element as constructed in Lemma (1.1.14)

$$u(x_1, \ldots, x_n) := \lambda_1 x_1 + \cdots + \lambda_n x_n$$

and let $q, v_1, \ldots, v_n \in \mathbb{C}[T]$ be univariate polynomials in a new variable T over $\mathbb{C}$. We call

$$(u,\ q,\ v_1, \ldots, v_n) \tag{1.4}$$

the **Polynomial Univariate Representation (PUR)** *of a component of $V(\mathscr{I})$ if it satisfies the following properties:*

(1) q is a monic square-free polynomial of degree $d \leq \delta$,

(2) $v_1, \ldots, v_n$ are all degree at most $d - 1$ and satisfy

$$\lambda_1 v_1(T) + \cdots + \lambda_n v_n(T) = T,$$

(3) for all $i = 1, \ldots, N$ we have

$$f_i(v_1(T), \ldots, v_n(T)) \equiv 0 \mod q(T).$$

If, in addition, $u, q, v_1, \ldots, v_n \in \mathbb{Q}(T)$ then we say that they form a PUR for a ratiopnal component of $V(\mathscr{I})$.

For a polynomial $p(T)$, we let $p(T) \bmod q(T)$ be the polynomial of degree less than $d = \deg q$ in the *conjugacy class of* p *with respect to* q*. Using the notation above, consider the polynomials*

$$r_i(T) := v_i(T)q'(T) \mod q(T) \quad i = 1, \dots n,$$

where $q'(t) = dq(T)/dT$*. We call the polynomials*

$$(u,\ q,\ r_1, \dots, r_n) \tag{1.5}$$

the **Rational Univariate Representation (RUR)** *of a component of* $V(\mathscr{I})$*.*

Versions of the following proposition appear in the literature (c.f. [74]), but we include a proof for completeness as appeared in [2].

Proposition 1.1.17. *Let* $\mathbf{f}$ *and* $\mathscr{I}$ *be as in Definition (1.1.16).*

1. *Let* $V := \{\xi_1, \dots, \xi_d\} \subseteq V(\mathscr{I}) \subset \mathbb{C}^n$*, and denote* $\xi_i = (\xi_{i,1}, \dots, \xi_{i,n})$ *for* $i = 1, \dots, d$*. Fix* $(\lambda_1, \dots \lambda_n) \in \mathbb{Q}^n$ *such that for* $u := \lambda_1 x_1 + \dots + \lambda_n x_n$

$$u(\xi_i) \neq u(\xi_j) \quad i \neq j \in \{1, \dots, d\}. \tag{1.6}$$

Define

$$q(T) := \prod_{i=1}^{d} (T - u(\xi_i)) \in \mathbb{C}[T], \tag{1.7}$$

and for each $j = 1, \dots, n$

$$v_j(T) := \sum_{i=1}^{d} \xi_{i,j} \frac{\prod_{k \neq i}(T - u(\xi_k))}{\prod_{k \neq i}(u(\xi_i) - u(\xi_k))} \in \mathbb{C}[T], \quad j = 1, \dots, n. \tag{1.8}$$

Furthermore, let

$$r_j(T) := \sum_{i=1}^{d} \xi_{i,j} \prod_{k \neq i} (T - u(\xi_k)) \in \mathbb{C}[T], \quad j = 1, \dots, n. \tag{1.9}$$

Then $(u, q, v_1, \dots, v_n)$ *satisfies the 3 properties of a PUR for the component* V *of* $V(\mathscr{I})$*, and* $(u, q, r_1, \dots, r_n)$ *is the corresponding RUR.*

2. *Conversely, assume that* $(u = \lambda_1 x_1 + \dots + \lambda_n x_n, q, v_1, \dots, v_n)$ *satisfy the 3 properties of a PUR in Definition (1.1.16). Then there exists* $V := \{\xi_1, \dots, \xi_d\} \subseteq V(\mathscr{I})$ *of cardinality* $d := \deg(q)$ *such that* u *satisfies (1.6),* q *satisfies (1.7) and* v_j *satisfies (1.8) for* $j = 1, \dots, n$*.*

Proof. To see the first claim, notice that q in (1.7) is monic, has degree $d \leq \delta$, and has no multiple roots by (1.6), thus we get the first property of a PUR.
For the second property, notice that v_j in (1.8) is the unique Lagrange interpolant of degree $\leq d-1$ satisfying $v_j(u(\xi_i)) = \xi_{i,j}$ for $i = 1, \ldots, d$. Thus at $T = u(\xi_i)$ for $i = 1, \ldots, d$ we have that

$$\lambda_1 v_1(u(\xi_i)) + \cdots + \lambda_n v_n(u(\xi_i)) = \lambda_1 \xi_{i,1} + \cdots + \lambda_n \xi_{i,n} = u(\xi_i).$$

Using that $\deg(v_i) \leq d-1$, the uniqueness of the Lagrange interpolation gives the second property of Definition (1.1.16).
For the third property, take again $T = u(\xi_i)$ for $i = 1, \ldots, d$, and notice that, since $\xi_i \in V(\mathscr{I})$,

$$f_t(v_1(u(\xi_i)), \ldots, v_n(u(\xi_i)) = f_t(\xi_{i,1}, \ldots, \xi_{i,n}) = 0.$$

The Chinese remainder theorem gives the third property of Definition 1.1.16.
Finally, at $T = u(\xi_i)$ for $i = 1, \ldots, d$ we have

$$r_j(u(\xi_i)) = \xi_{i,j} \prod_{k \neq i} (u(\xi_i) - u(\xi_k)) = \left[q'(T) v_j(T) \mod q(T) \right]_{T = u(\xi_i)}$$

Since both $r_j(T)$ and $q'(T)v_j(T) \bmod q(T)$ are degree at most $d-1$, the uniqueness of Lagrange interpolation proves that they are equal. Thus $(u, q, r_1, \ldots, r_n)$ is the corresponding RUR of V.

To prove the converse, assume that $(u = \lambda_1 x_1 + \cdots + \lambda_n x_n, q, v_1, \ldots, v_n)$ satisfy the 3 properties of a PUR in Definition (1.1.16). There exists $\alpha_1, \ldots, \alpha_d \in \mathbb{C}$, all distinct, such that

$$q = \prod_{i=1}^{d} (T - \alpha_i).$$

Define for $i = 1, \ldots, d$

$$\xi_i := (v_1(\alpha_i), \ldots, v_n(\alpha_i)) \in \mathbb{C}^n.$$

Then by the third property of the RUR we have $f_t(\xi_i) = 0$ for all $t = 1, \ldots, m$ and $i = 1, \ldots, d$, thus $V := \{\xi_1, \ldots, \xi_d\} \subseteq V(\mathscr{I})$. Finally, from the second property we have

$$u(\xi_i) = \lambda_1 v_1(\alpha_i) + \cdots + \lambda_n v_n(\alpha_i) = \alpha_i,$$

which gives that $u(\xi_i) \neq u(\xi_j)$ and $\xi_i \neq \xi_j$ for $i \neq j$. □

Remark 1.1.18. *Note that, in the literature, a PUR is usually also required to satisfy*

$$\langle q(T),\ x_1 - v_1(T), \ldots, x_n - v_n(T)\rangle = \langle \mathscr{I}, T - (\lambda_1 x_1 + \cdots + \lambda_n x_n)\rangle.$$

while our definition only guarantees $\supseteq$. Thus, the common roots of the ideal on the left correspond to a component of the common roots of $\mathscr{I}$. That is the reason for the name PUR and RUR of a component of $V(\mathscr{I})$.

1.2 Certifying solutions of well-constrained polynomial systems

1.2.1 Homotopy method

Homotopy continuation methods [7, 80] uses symbolic-numeric algorithms to solve polynomial systems. A family of systems, a so-called homotopy, define solution paths, which are followed using path tracking algorithms, e.g., predictor-corrector methods. Even though the homotopy is constructed symbolically, the heart of the algorithm, path tracking algorithms, are numerical algorithms.

The General idea is to start with known solutions of a known start system and then track those solutions as we deform the start system into the system that we wish to solve. This method takes a system $\mathbf{g}(x) = 0$, whose solutions are known, and using a homotopy, e.g.,

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

While t runs from 1 to 0, if paths $\boldsymbol{x}(t)$ are defined “nicely” by the homotopy $H(x, t)$, we end at the solutions of $\mathbf{f}(x)$ at $t = 0$.

1.2.2 The Problem of Certification

Polynomial systems can be solved reliably using numerical homotopy methods, there are various implementations such as PHCPack [85], HOM4PS [17], and Bertini [6]. Numerical methods return numerical approximations to solutions, and all the mentioned implementations validate the solutions heuristically, usually checking Newton iterations at each iteration. Therefore, the output, the approximate solutions of polynomial systems are not certified. Even though the approximate solutions work well in practice, these cannot be used in some applications, specifically in pure mathematics.

The goal is to certify that there is a solution z near an exact root ξ (i.e., if z is in the open ball $\mathscr{B}(\xi, \epsilon) := \{x : \|x - \xi\| < \epsilon\}$ for some small $\epsilon > 0$) of the given system.

1.2.3 Newton Method

Let $\mathbf{f} := (f_1, \ldots, f_n)$ be a system of n polynomials in $\mathbb{C}[x_1, \ldots, x_n]$, with common zeros $\mathbf{V}(\mathbf{f}) := \{\xi \in \mathbb{C}^n \ : \ f_i(\xi) = 0 \ \ i = 1, \ldots, n\}$. Let $J_{\mathbf{f}}(z)$ be the Jacobian matrix of the system $\mathbf{f}$ at $z \in \mathbb{C}^n$. Then the **Newton iteration** is the map $N_{\mathbf{f}} : \mathbb{C}^n \to \mathbb{C}^n$ defined by

$$N_{\mathbf{f}}(z) = z - J_{\mathbf{f}}(z)^{-1}\mathbf{f}(z)$$

as long as $J_{\mathbf{f}}(z)^{-1}$ exists. For $k = 0, 1, \ldots$, the k-th Newton iteration is

$$z^{(k+1)} := N_{\mathbf{f}}(z^{(k)}) = N_{\mathbf{f}}^{k}(z)$$

with initial guess $z^{(0)} := z$.

Definition 1.2.1. *Let* $\mathbf{f} := (f_1, \ldots, f_n)$ *be a system of polynomials in* $\mathbb{C}[x_1, \ldots, x_n]$*. A point* $z \in \mathbb{C}^n$ *is an* **approximate solution** *to* $\mathbf{f}$ *with the* **associated solution** $\xi \in V(\mathbf{f})$ *if, for every* $k \in \mathbb{N}$ *the Newton iteration* $z^{(k)}$ *satisfies*

$$\|z^{(k)} - \xi\| \leq \left(\frac{1}{2}\right)^{2^k - 1} \|z - \xi\| \tag{1.10}$$

where $\|.\|$ *is the usual hermitian norm on* $\mathbb{C}$*, i.e.,* $\|z\| = \|(z_1, \ldots, z_n)\| = (|z_1|^2 + \cdots + |z_n|^2)^{1/2}$*. Then we say the sequence* $\{N_{\mathbf{f}}(z^{(k)}) \ : \ k \in \mathbb{N}\}$ **converges quadratically** *to* ξ*.*

Assumption 1.2.2. *We assume that the Jacobian matrix of the given system* $\mathbf{f} := (f_1, \ldots, f_n) \in \mathbb{C}[x_1, \ldots, x_n]$ *is nonsingular at all* $z^{(i)}$ *for* $i = 0, 1, \ldots$*.*

Definition 1.2.3. *A point* $z \in \mathbb{C}$ *is a* **fixed point** *of* $\mathbf{f}$*, if* $N_{\mathbf{f}}(z) = z$*.*

Remark 1.2.4. *Notice that if* $J_{\mathbf{f}}(z)^{-1}\mathbf{f}(z) = 0$ *then by the definition of Newton iteration,* z *is a fixed point.* $J_{\mathbf{f}}(z)^{-1}\mathbf{f}(z) = 0$ *implies that* $\mathbf{f}(z) = 0$*, by Assumption 1.2.2. When* $\mathbf{f}$ *is a well-constrained system, fixed points are roots. However it may not be true if* $\mathbf{f}$ *is an overdetermined system, we will explain it elaborately in Chapter 2.*

1.2.4 α-theory and certification

We start with some theorems to introduce a criteria for a point $z \in \mathbb{C}$ to be an approximate zero of $\mathbf{f}$. Then introduce a classical solution of certification problem on well-constrained polynomial systems using so-called α-theory as explained in [12] and [37]. First, we need to define some auxiliary quantities.

Definition 1.2.5. *Given a well-constrained polynomial system* $\mathbf{f} = (f_1, \ldots, f_n) \in \mathbb{C}[x_1, \ldots, x_n]$*,* $z \in \mathbb{C}^n$*, and* $k \in \mathbb{N}$ *then*

$$\alpha(\mathbf{f}, z) = \beta(\mathbf{f}, z)\gamma(\mathbf{f}, z) \tag{1.11}$$

where

$$\beta(\mathbf{f}, z) = \|J_{\mathbf{f}}(z)^{-1}\mathbf{f}(z)\| \quad \textit{and} \tag{1.12}$$

$$\gamma(\mathbf{f}, z) = \sup_{k\geq 2} \left\| \frac{J_{\mathbf{f}}(z)^{-1} J_{\mathbf{f}}^{(k)}(z)}{k!} \right\|^{1/k-1}. \tag{1.13}$$

By our Assumption (1.2.2), β is a well defined quantity. As stated in [37], in (1.13), the k-th derivative $J_{\mathbf{f}}^k(\boldsymbol{x})$ to f is the symmetric tensor whose components are the partial derivatives of **f** of order k. It is a linear map from the k-fold symmetric power $S^k\mathbb{C}^n$ of $\mathbb{C}^n$ to $\mathbb{C}^n$. The norm in (1.13) is the operator norm of $J_{\mathbf{f}}(\boldsymbol{x})^{-1} J_{\mathbf{f}}^k(\boldsymbol{x}) : S^k\mathbb{C}^n \mapsto \mathbb{C}^n$, defined with respect to the norm on $S^k\mathbb{C}^n$ that is dual to the standard unitarily invariant norm on homogeneous polynomials,

$$\left\| \sum_{\nu=k} a_\nu x^\nu \right\|^2 := \sum_{|\nu|=k} |a^\nu|^2 / \binom{k}{n}$$

where $\nu = (\nu_1, \ldots, \nu_n)$ is an exponent vector of nonnegative integers with $x^\nu = (x_1^\nu \ldots x_n^\nu)$, $|\nu| = \nu_1 + \cdots + \nu_n$ and $\binom{k}{n} = k!/\nu_1! \ldots \nu_n!$ is the multinomial coefficient.

Remark 1.2.6. *There is a simple polynomial that plays an important role for the rest of this section*

$$\psi(u) = 2u^2 - 4u + 1. \tag{1.14}$$

These facts will help us to obtain some estimates,

(a) $\frac{u}{\psi(u)} < 1$ *for* $0 \leq u \leq \frac{5-\sqrt{17}}{4}$ *using the quadratic formula,*

(b) $\frac{3-\sqrt{7}}{2}$ *is the first positive solution of* $\frac{u}{\psi(u)} = \frac{1}{2}$*,*

(c) $\psi(u) < 1$ *for* $u < 1 - \frac{\sqrt{2}}{2}$.

Now we can begin introducing the main theorems behind the idea of α-theory. The first one gives a relation between an approximate and its associate solutions of a system $\mathbf{f} := (f_1, \ldots, f_n)$ of polynomials in $\mathbb{C}[x_1, \ldots, x_n]$.

Theorem 1.2.7. *Suppose that* $\mathbf{f}(\xi)=0$ *and* $J_{\mathbf{f}}(z)^{-1}$ *exists. If*

$$\|z-\xi\| \le \frac{3-\sqrt{7}}{2\gamma(\mathbf{f},\xi)}$$

then z *is an approximate zero of* $\mathbf{f}$ *with associated zero* ξ.

For the proof of this theorem, we need the following proposition. The detailed proof of this proposition can be found in [12, Chapter 8].

Proposition 1.2.8. *Let* $\mathbf{f}=(f_1,\ldots,f_n)\in\mathbb{C}[x_1,\ldots,x_n]$, $\xi\in V(\mathbf{f})$, $z\in\mathbb{C}^n$ *and* $u=\|z-\xi\|\gamma(\mathbf{f},\xi)$. *Suppose* $u<\frac{5-\sqrt{17}}{4}$. *Then*

(a) $\|N_{\mathbf{f}}(z)-\xi\| < \frac{\gamma(\mathbf{f},\xi)\|z-\xi\|^2}{\psi(u)} = \frac{u\|z-\xi\|}{\psi(u)}$.

(b) $\|N_{\mathbf{f}}^{(k)}(z)-\xi\| \le \left(\frac{u}{\psi(u)}\right)^{2^k-1}\|z-\xi\|$ *for all* $k\ge 0$.

Proof of Theorem 1.2.7. As stated in 1.2.6(b), $\frac{3-\sqrt{7}}{2}$ is the first positive solution of $\frac{u}{\psi(u)}=\frac{1}{2}$. If $u<\frac{3-\sqrt{7}}{2}$, then $\frac{u}{\psi(u)}<\frac{1}{2}$. By (1.10) and 1.2.8(b), if $\|z-\xi\|\gamma(\mathbf{f},\xi)\le\frac{3-\sqrt{7}}{2}$, then z is an approximate zero of $\mathbf{f}$ with associated zero ξ. □

In 1986, Smale [79] introduced the main theorem of α-theory, which provides a certificate that a given point is an approximate solution to $\mathbf{f}$. The proof of this theorem requires some preceding definition and theorems. These theorems will be presented without their proofs to avoid giving additional preliminary propositions and definitions.

Definition 1.2.9. *Suppose that* X *is a complete metric space (i.e., every Cauchy sequence in* X *converges in* X*). A map* $\phi: X\to X$ *satisfying* $d(\phi(x),\phi(y))\le c\,d(x,y)$ *for all* $x,y\in X$, *with* $c<1$ *is called a* **contraction map** *with a* **contraction constant** c.

Theorem 1.2.10. *If*

$$r<\frac{1-\frac{\sqrt{2}}{2}}{\gamma(\mathbf{f},z)},$$

then

(a) for all $z_1\in\mathbb{C}^n$ *with* $\|z_1-z\|<r$,
$\|\frac{\partial N_{\mathbf{f}}(z_1)}{\partial z_1}\| \le \frac{2(\alpha(\mathbf{f},z)+u}{\psi(u)^2}$, $u=r\gamma(\mathbf{f},z)$ *and* ψ *as in (1.14).*

(b) $N_{\mathbf{f}}(\mathscr{B}(r,z))\subset\mathscr{B}(r',N_{\mathbf{f}}(z))$, *where* $r'=\frac{2(\alpha(\mathbf{f},z)+u)}{\psi(u)^2}r$

Corollary 1.2.11. *if $u < 1-(\sqrt{2}/2)$, $c = (2(\alpha(\mathbf{f},z)+u)/\psi(u)^2 < 1$ and $\alpha(\mathbf{f},z)+c\,u \le u$, then N_f is a contraction map of the ball $\mathscr{B}(u/(\gamma(\mathbf{f},z)),z)$ into itself with contraction constant c. Hence there is a unique root ξ of $\mathbf{f}$ in $\mathscr{B}(u/(\gamma(\mathbf{f},z)),z)$ and all $z' \in \mathscr{B}(u/(\gamma(\mathbf{f},z)),z)$ converge to ξ under iteration of $N_{\mathbf{f}}$.*

Corollary (1.2.11) provides a criterion in terms of α and γ for convergence of the Newton iterations by a contraction map in a neighborhood of a point z. Finally, we can state the α-Theory theorem after the following.

Theorem 1.2.12 (Robust α-Theory)**.** *There are positive real numbers α_0 and u_0 such that:*
if $\alpha(\mathbf{f},z) < \alpha_0$, then there is a root ξ of $\mathbf{f}$ such that

$$\mathscr{B}\left(\frac{u_0}{\gamma(\mathbf{f},z)},z\right) \subset \mathscr{B}\left(\frac{3-\sqrt{7}}{2\gamma(\mathbf{f},\xi)},\xi\right)$$

and N_f maps $\mathscr{B}(u_0/\gamma(\mathbf{f},z),z)$ into $\mathscr{B}(u_0/\gamma(\mathbf{f},\xi),\xi)$ with contraction constant less then or equal to $1/2$.

Now we can represent the main theorem behind α-Theory,

Theorem 1.2.13 (**α-Theory**)**.** *Given a well constrained polynomial system $\mathbf{f} = (f_1,\dots,f_n) \in \mathbb{C}[x_1,\dots,x_n]$, and an initial guess $z \in \mathbb{C}^n$. There is a computable constant α_0 such that If $\alpha(\mathbf{f},z) < \alpha_0$, then z is an approximate zero of $\mathbf{f}$. Additionally, $\|z-\xi\| \le 2\beta(f,x)$ where $\xi \in V(\mathbf{f})$ is the associated solution to x.*

The proof of Theorem 1.2.13 follows the Theorems 1.2.7 and 1.2.12.

In [12], one of the best value of the constant α_0 is given as $\frac{13-3\sqrt{17}}{4} \approx 0.157671$. An improvement by Wang and Han to $3-2\sqrt{2} \approx 0.171573$ is reported in [87].

For a well constrained polynomial system $\mathbf{f} = (f_1,\dots,f_n) \in \mathbb{C}[x_1,\dots,x_n]$, and an initial guess $\boldsymbol{x} \in \mathbb{C}^n$, we can compute the constant $\alpha(\mathbf{f},\boldsymbol{x})$, if (1.11) is satisfied then we can state that $\boldsymbol{x}$ **is a certified approximation solution to f**.
In [37], Sotille and Hauenstein showed that one can get an efficient and practical root certification algorithm using α-theory for well-constrained polynomial systems. They implemented the idea on their software package **alphaCertified** in 2011.

1.3 Singularity and multiplicity

The main sources we use here are [7] and [38]. First we need to introduce some basic definitions. Consider a polynomial system $\mathbf{f} = (f_1,\dots,f_N) \in \mathbb{K}[x_1,\dots,x_n]$, and assume that the ideal $\mathscr{I} := \langle f_1,\dots,f_N\rangle$ is zero dimensional.

Definition 1.3.1. *A solution $z \in \mathbb{K}^n$ is said to be* **isolated** *if there is an open ball $\mathscr{B}(z,\epsilon)$ around z for some $\epsilon > 0$ such that the only solution in $\mathscr{B}$ is z. A solution z is called* **nonsingular** *if the Jacobian matrix of* $\mathbf{f}$ *at z is full rank. Otherwise z is called a* **singular** *solution.*

Definition 1.3.2. (Univariate case) *Let $f \in \mathbb{K}[x]$, a univariate polynomial with $f(\xi) = 0$. Then the* **multiplicity** *of ξ is m if*

$$f(\xi) = f'(\xi) = \cdots = f^{(m-1)}(\xi) = 0 \text{ and } f^{(m)}(\xi) \neq 0.$$

The multiplicity of ξ with respect to f is denoted by $\mu(f,\xi)$.
(Multivariate case) *Let* $\mathbf{f} \in \mathbb{K}[x_1,\ldots,x_n]$ *with $N \geq n$. The* **multiplicity** *of ξ with respect to* $\mathbf{f}$ *is defined as*

$$\mu(\mathbf{f},\xi) := \dim \mathscr{O}_\xi / \langle \mathbf{f} \rangle,$$

where $\mathscr{O}_\xi$ is the ring of convergent power series centered at ξ and $\langle \mathbf{f} \rangle$ is the ideal in $\mathscr{O}_\xi$.
(Multiplicity of an irreducible component) *Let* $\mathbf{f} \in \mathbb{K}[x_1,\ldots,x_n]$ *and $V' \subset V(\mathscr{I})$ be an irreducible component of dimension m. Then the* **multiplicity** *of V' with respect to* $\mathbf{f}$ *is*

$$\mu(\mathbf{f}, V') := \mu(\{\mathbf{f}, L_1,\ldots,L_m\},\xi),$$

where $L_1,\ldots,L_m$ are general linear polynomials on k^n and ξ is any point contained in the finite set $V' \cap V(L_1,\ldots,L_m)$.

The univariate and multivariate cases agree on the multivariate definition. We demonstrate that with the following example.

Example 1.3.3. *Consider $f(x) = x^2(x-1)$, where $\mu(f,0) = 2$ and $\mu(f,1) = 1$. For $\xi = 0$, it has convergent power series expansion centered at $\xi = 0$, since $(x-1)$ is nonzero at ξ. This implies $\mathscr{O}_\xi/\langle f \rangle$ is two dimensional with a basis $\{1 + \langle f \rangle, x + \langle f \rangle\}$. Conversely, first two terms in Taylor expansion vanish at zero, i.e., $f(0) = f'(0) = 0$.*

The following definition will provide a base to construct so-called isosingular deflation in Chapter 3.

Definition 1.3.4. *Consider the ideal $\mathscr{I}$ generated by the polynomials* $\mathbf{f}$ *of the original system which has an isolated solution at z^*. We call an ideal $\mathscr{J}$ a* **deflation** *of $\mathscr{I}$ at z^* if $\mathscr{J} \supset \mathscr{I}$, $\mathscr{J} \neq \mathbb{K}[x_1,\ldots,x_n]$ and $1 \leq \mu(\mathscr{J}, z^*) < \mu(\mathscr{I}, z^*)$.*

1.4 Rational Number Reconstruction

Now we will describe how to reconstruct the rational numbers that have bounded denominators and indistinguishable from a given floating point number. In this section we will mainly follow the notation of [73] and [78].

Definition 1.4.1. *Let $a \in \mathbb{R}$, a* **finite continued fraction** *is an expression of the form*

$$a = a_0 + \cfrac{1}{a_1 + \cfrac{1}{\ddots + \cfrac{\vdots}{a_{n-1} + \frac{1}{a_n}}}}$$

where $a_0, a_1, \ldots, a_n$ are all integers with $a_1, \ldots, a_n$ positive. We use the notation $[a_0; a_1, \ldots, a_n]$ to present continued fraction. Here, writing the remainders using the following recurrence relation

$$r_0 := a \ \ and\ r_n := 1/(r_{n-1} - a_{n-1})\ for\ \ n = 1, 2, \ldots$$

gives the formula

$$a_n := \lfloor r_n \rfloor \ for\ n = 0, 1, 2, \ldots$$

where $\lfloor \cdot \rfloor$ is the floor function.

Definition 1.4.2. *Let a be an irrational number, then an* **infinite continued fraction** *can be defined as $[a_0; a_1, \ldots]$ where $a_0, a_1, \ldots$ are all integers with $a_1, \ldots$ positive.*

Definition 1.4.3. *Let $a_0, a_1, \ldots, a_n$ be all integers with $a_1, \ldots, a_n$ positive. Let the sequences $p_0, p_1, \ldots, p_n$ and $q_0, q_1, \ldots, q_n$ be define recursively by*

$$p_0 = a_0 \qquad\qquad q_0 = 1$$
$$p_1 = a_0 a_1 + 1 \qquad\qquad q_1 = a_1$$

and

$$p_k = a_k p_{k-1} + p_{k-2} \qquad\qquad q_k = a_k q_{k-1} + q_{k-2}$$

*for $k = 2, 3, \ldots, n$. Then the k-**th convergent** $C_k = [a_0; a_1, \ldots, a_k]$ is*

$$C_k = p_k/q_k$$

Lemma 1.4.4. *Let $C_k = p_k/q_k$ be the k-th convergent of the continued fraction $[a_0; a_1, \ldots, a_n]$ where $k = 1, \ldots, n$ and p_k and q_k defined as (1.4.3), then*

$$p_k q_{k-1} - p_{k-1} q_k = (-1)^{k-1} \tag{1.15}$$

Proof. We will use induction to proof this lemma.
For $k = 1$, using Definition (1.4.3) we have

$$p_1 q_0 - p_0 q_1 = (a_0 a_1 + 1) \cdot 1 - a_0 a_1 = 1.$$

Now assume that (1.15) is hold for $k \in \mathbb{Z}$ and $k = 1, \ldots, n-1$.

$$p_k q_{k-1} - p_{k-1} q_k = (-1)^{k-1}.$$

Then,

$$\begin{aligned} p_{k+1} q_k - p_k q_{k+1} &= (a_{k+1} p_k + p_{k-1} q_k) - p_k (a_{k+1} q_k + q_{k-1}) \\ &= p_{k-1} q_k - p_k q_{k-1} \\ &= (-1)^k. \end{aligned}$$

shows us, (1.15) is true for $k + 1$. This completes the proof. □

Remark 1.4.5. *Let $C_k = p_k/q_k$ be the k-th convergent of a continued fraction, p_k and q_k are relatively prime, since* $\gcd(p_k, q_k)$ *has to divide $(-1)^{k-1}$ in (1.15).*

The k-th convergent, $C_k = p_k/q_k$ of the irrational number a, is a rational approximation to a. For example, the continued fraction of π is $[3; 7, 15, 1, 292, 1, 1, 1, 2, 1, \ldots]$. Then $C_1 = 3$, $C_2 = 22/7$, $C_3 = 333/106$ and $C_4 = 355/113$ are rational approximations of π. Depending on the desired accuracy and denominator bounds, we can find an appropriate approximation of a given number.

We give the following lemma without proof since its proof is long and not related to any of the concepts we use in this section. However it will be used to prove the following theorem.

Lemma 1.4.6. *Let C_k be the $k-th$ convergent of the finite continued fraction $[a_0; a_1, \dots, a_n]$. Then*

$$C_1 > C_3 > C_5 > \dots \quad \textit{and} \quad C_0 < C_2 < C_4 < \dots,$$

and every odd numbered convergent C_{2i+1} is greater than every even numbered convergent C_{2i}, for $i = 0, 1, 2, \dots$.

The following two theorems show how a convergent as a rational approximation of an irrational number and bounds on the denominator are related.

Theorem 1.4.7. *([73]) Let a be an irrational number and let p_i/q_i, $i = 1, 2, \dots$ be the convergents of the infinite continued fraction of a. If r/s is a rational number, r, s are integers with $s > 0$, and if k is a positive integer such that*

$$|sa - r| < |q_k a - p_k| \quad \textit{then} \quad s \geq q_{k+1}.$$

Proof. Assume that $|sa - r| < |aq_k - p_k|$ and $1 \leq s < q_{k+1}$. Consider the equations,

$$p_k x + p_{k+1} y = r \tag{1.16}$$

$$q_k x + q_{k+1} y = s. \tag{1.17}$$

When we eliminate x, we have $(p_{k+1} q_k - p_k q_{k+1}) y = r q_k - s p_k$. Similarly, when we eliminate y, we have $(p_k q_{k+1} - p_{k+1} q_k) x = r q_{k+1} - s p_{k+1}$. Then by lemma (1.4.4).

$$y = (-1)^k (r q_k - s p_k) \text{ and } x = (-1)^k (s p_{k+1} - r q_{k+1})$$

(i) First, notice that x and y are non zero.
If $x = 0$, then $s p_{k+1} = r q_{k+1}$, since p_{k+1} and q_{k+1} are relatively prime by lemma (1.4.5), q_{k+1} must divide s, that implies $q_{k+1} \leq s$, which contradicts our assumption.
If $y = 0$, then by (1.16) $r = p_k x$ and $s = q_k x$, substitute these values in $|sa - r| = |x||q_k a - p_k| \geq |q_k a - p_k|$ since $|x| \geq 1$, which contradicts our assumption.

(ii) Suppose that $y < 0$, by (1.16), $q_k x = s - q_{k+1} y$ and assumption $s \leq q_{k+1}$, then $q_k x > 0$ and $q_k > 0$ gives us $x > 0$.

(iii) Suppose that $y > 0$, by our assumption, $s < q_{k+1}$, then $s < q_{k+1} \leq q_{k+1} y$. By (1.16), $q_k x = s - q_{k+1} y < 0$, so $x < 0$.

We showed that nonzero x and y has opposite signs. By theorem (1.4.6), either $p_k/q_k < a < p_{k+1}/q_{k+1}$ or $p_{k+1}/q_{k+1} < a < p_k/q_k$. In either case, $q_k a - p_k$ and $q_{k+1}a - p_{k+1}$ have opposite signs. Using equations (1.16)

$$\begin{aligned} |sa - r| &< |(q_k x + q_{k+1} y)a - (p_k x + p_{k+1} y)| \\ &< |(x(q_k - a p_k) + y(q_{k+1} a - p_{k+1})| \end{aligned}$$

we see that $x(q_k a - p_k)$ and $y(q_{k+1}a - p_{k+1})$ have the same sign since $q_k a - p_k$ and $q_{k+1}a - p_{k+1}$ has opposite signs. Thus

$$\begin{aligned} |sa - r| &< |x||q_k - a p_k| + |y||q_{k+1} a - p_{k+1}| \\ &\leq |x||q_k - a p_k| \\ &\leq |q_k - a p_k| \end{aligned}$$

since $|x| \geq 1$. This contradiction completes the proof. □

Theorem 1.4.8. *Let a be an irrational number and if r/s is a rational number, r, s are integers with $s > 0$ such that*

$$|a - r/s| < 1/(2s^2)$$

then r/s is a convergent of the continued fraction expansion of a.

Proof. Assume that r/s is not a convergent of the continued fraction expansion of a. By Theorem (1.4.7), we have

$$|q_k a - p_k| \leq |sa - r| = s|a - r/s| < 1/(2s)$$

since there are convergents p_k/q_k and p_{k+1}/q_{k+1} such that $q_k \leq s < q_{k+1}$.
Then we have

$$|a - p_k/q_k| < 1/(2sq_k)$$

by dividing both sides by q_k. We know that r/s is not equal to p_k/q_k, therefore $sp_k - rq_k$ is a nonzero integer, thus $|sp_k - rq_k| \geq 1$. Using this inequality

$$\begin{aligned}\frac{1}{sq_k} &\leq \frac{|sp_k - rq_k|}{sq_k}\\ &= \left|\frac{p_k}{q_k} - \frac{r}{s}\right|\\ &< \left|a - \frac{p_k}{q_k}\right| + \left|a - \frac{r}{s}\right| \text{(triangle inequality)}\\ &\leq \frac{1}{2sq_k} + \frac{1}{2s^2}\\ &\leq \frac{1}{2s^2}\end{aligned}$$

Hence, we have $q_k > s$, which is a contradiction. □

The following classical result implies that if a number is sufficiently close to a rational number with small denominator, then we can find this latter rational number in polynomial time (c.f. [78, Corrolary 6.3a] or [86, Theorem 5.26]).

Theorem 1.4.9 ([78, 86]). *There exists a polynomial time algorithm which, for a given rational number a and a natural number B tests if there exists a pair of integers (r,s) with $1 \leq s \leq B$ and*

$$|a - r/s| < \frac{1}{2B^2},$$

and if so, finds this unique pair of integers.

Proof. First, assume that r/s and r'/s' are two different rational numbers, such that $1 \leq s \leq B$, $1 \leq s' \leq B$ and $|a - r/s| < 1/(2B^2)$, $|a - r'/s'| < 1/(2B^2)$. Then

$$\begin{aligned}\frac{1}{B^2} &\leq \frac{1}{ss'}\\ &\leq \frac{|rs' - r's|}{ss'} \text{(since } r, r', s, s' \text{ are all integers)}\\ &= \left|\frac{r}{s} - \frac{r'}{s'}\right|\\ &< \left|a - \frac{r}{s}\right| + \left|a - \frac{r'}{s'}\right| \text{(triangle inequality)}\\ &< \frac{1}{2B^2} + \frac{1}{2B^2} = \frac{1}{B^2}\end{aligned}$$

contradicting the assumption.

Suppose r/s exists,

$$\left|a - \frac{r}{s}\right| < \frac{1}{2B^2} \leq \frac{1}{2s^2} \tag{1.18}$$

Now as long as $s \leq B$, it suffices to compute the convergents of a, and to test if any of them satisfies (1.18).

□

Later in this dissertation (see Subsection 2.2.4, Theorem 2.2.7) we give a complexity analysis of rational number reconstruction based on [27].

Remark 1.4.10. *Notice that Theorem (1.4.9) does not guarantee existence of the the pair (r, s) with the given properties, only uniqueness.*

1.5 Some matrix theory

In this part we cover some matrix theory material to provide background for the so-called Hermite method. That method is presented in Chapter 4.
The main sources used here are [63] and [8]. We start with special matrices that are used in the related chapter.

Definition 1.5.1. *A square complex matrix A is* **Hermitian matrix** *(or self-adjoint matrix) if its complex entries that is equal to its own conjugate transpose. It is denoted by*

$$A = \overline{A^T} = A^*.$$

Definition 1.5.2. *A $n \times n$ square matrix H with the following structure is called a* **Hankel** *matrix,*

$$H = \begin{bmatrix} h_0 & h_1 & h_2 & \dots & \dots & h_{n-1} \\ h_1 & h_2 & & & & \vdots \\ h_2 & & & & & \vdots \\ \vdots & & & & & h_{2n-4} \\ \vdots & & & & h_{2n-4} & h_{2n-3} \\ h_{n-1} & \dots & \dots & h_{2n-4} & h_{2n-3} & h_{2n-2} \end{bmatrix}.$$

The $(i, j)-$th entry of H is defined by

$$[H]_{i,j} = [H]_{i+1,j-1} = h_{i+j-2}.$$

Theorem 1.5.3. *Every eigenvalue of a real symmetric matrix is real.*

Proof: Let A be a real symmetric matrix, then A is Hermitian matrix, since $A^* = \overline{A^T} = \overline{A} = A$. Let λ be an eigenvalue and v be the corresponding eigenvector. Then

$$Av = \lambda v$$

multiply both sides with v^*,

$$v^* A v = v^* \lambda v = \lambda v^* v.$$

Then

$$\lambda = \frac{v^* A v}{v^* v}.$$

Using the definition of Hermitian matrix, $(v^* A v)^* = v^* A v$, then $\alpha^* = \alpha$ implies $\alpha \in \mathbb{R}$. Similarly, $(v^* v)^* = v^* (v^*)^* = v^* v$, implies $v^* v \in \mathbb{R}$. Since a ratio of two real numbers is a real number, $\lambda \in \mathbb{R}$.

Diagonalization of matrices will arise in the further parts of this work. We give some definitions and criterion on diagonalization of special kind of matrices.

Definition 1.5.4. *A square matrix A is* **diagonalizable** *if it is similar to a diagonal matrix D, i.e., there exists a nonsingular matrix P such that*

$$P^{-1} A P = D.$$

Moreover, $n \times n$ A is diagonalizable if and only if A has n linearly independent eigenvectors, which form the columns of P, in this case diagonals of D are the eigenvalues of A [63].

Definition 1.5.5. *Square matrices A and B are* **simultaneously diagonalizable** *if there exists a nonsingular matrix P such that*

$$P^{-1} A P = D_A \text{ and } P^{-1} B P = D_B$$

where D_A and D_B are diagonal matrices.

The following theorem will be very handy in Chapter 4.

Theorem 1.5.6 (7.2.16 [63])**.** *Matrices A and B are simultaneously diagonalizable if and only if $AB = BA$.*

We can generalize the previous theorem as follows:

Theorem 1.5.7. *Let $A_1, \ldots, A_n$ be diagonalizable matrices, they are simultaneously diagonalizable if and only if they commute pairwise.*

The Vandermonde matrix is a special matrix which is used broadly in different areas of mathematics. We define it on bases for univariate and multivariate monomials. Note that sometimes its transpose is defined as the Vandermonde matrix.

Definition 1.5.8. *Let $\xi_1, \ldots \xi_m$ be elements in a field $\mathbb{K}$. A $m \times k$* **Vandermonde** *matrix of $\xi_1, \ldots, \xi_m$ with respect to the standard monomial basis $B = \{1, x, \ldots, x^{k-1}\}$ is defined as follows*

$$V = V_B(\xi_1, \ldots, \xi_m) = \begin{bmatrix} 1 & \xi_1 & \ldots & \xi_1^{k-1} \\ 1 & \xi_2 & \ldots & \xi_2^{k-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \xi_m & \ldots & \xi_m^{k-1} \end{bmatrix}.$$

Definition 1.5.9. *Let $\xi_1, \ldots \xi_m$ be elements in a field $\mathbb{K}^n$. Fix a monomial basis $\mathscr{B} = \{x^{\alpha_\mathbf{1}}, \ldots, x^{\alpha_\mathbf{D}}\}$. Then* **multivariate Vandermonde matrix** *of $\xi_1, \ldots \xi_m$ with respect to the basis $\mathscr{B}$ is defined by $[V]_{i,j} := (\xi_\mathbf{i})^{\alpha_\mathbf{j}}$, i.e.,*

$$V = V_{\mathscr{B}}(\xi_1, \ldots, \xi_m) = \begin{bmatrix} \xi_1^{\alpha_1} & \xi_1^{\alpha_2} & \ldots & \xi_1^{\alpha_D} \\ \xi_2^{\alpha_1} & \xi_2^{\alpha_2} & \ldots & \xi_2^{\alpha_D} \\ \vdots & \vdots & \ddots & \vdots \\ \xi_m^{\alpha_1} & \xi_m^{\alpha_2} & \ldots & \xi_m^{\alpha_D} \end{bmatrix}$$

where $(\xi_\mathbf{i})^{\alpha_\mathbf{j}} = (\xi_{i1})^{\alpha_\mathbf{i1}} \ldots (\xi_{in})^{\alpha_\mathbf{in}}$.

Definition 1.5.10. *Let $p = x^m + p_{m-1}x^{m-1} + \cdots + p_1 x + p_0$ be a monic univariate polynomial, its* **companion matrix** *is the $m \times m$ matrix defined as*

$$C_p = \begin{bmatrix} 0 & 1 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \\ -p_0 & -p_1 & \ldots & -p_{m-1} \end{bmatrix}.$$

Note that sometimes its transpose is defined as the companion matrix. We list some companion matrix properties collated from [63], [8] and [13]:

- The eigenvalues of C_p are equal to the roots of p.

- Let p has m distinct roots $\xi_1, \ldots, \xi_m$ then the companion matrix of p can be diagonalized as

  $$V^{-1} C_p V = D$$

  with $[D]_{i,i} = \xi_i$ for $i = 1, \ldots, k$ and V is the Vandermonde matrix $V_B(\xi_1, \ldots, \xi_m)$ with respect to the standard monomial basis $B = \{1, x, \ldots, x^{m-1}\}$.

- Let p has m distinct roots $\xi_1, \ldots, \xi_m$ then

  $$Tr(C_p)^s = \sum_{i=1}^{m} \xi_i^s.$$

Definition 1.5.11. *Let* $\mathbf{f} = (f_1, \ldots, f_N) \in \mathbb{K}[x_1, \ldots, x_n]$ *and let* $\mathscr{I} :=< f_1, \ldots, f_N >$*, if* $p \in \mathbb{K}[x_1, \ldots, x_n]/\mathscr{I}$*, define a linear map*

$$M_p : \mathbb{K}[x_1, \ldots, x_n]/\mathscr{I} \mapsto \mathbb{K}[x_1, \ldots, x_n]/\mathscr{I}$$

as the **multiplication map** *of* **f** *such that*

$$M_p(q) = qp \mod \mathscr{I}$$

for $q \in \mathbb{K}[x_1, \ldots, x_n]/\mathscr{I}$.

The multiplication map has the following properties, see [19, Section 2.4]:

- The matrix of the multiplication map $M_p = 0$ when $p \in \mathscr{I}$.
- $M_p = p(M)$
- $M_{p_1+p_2} = M_{p_1} + M_{p_2}$ and $M_{p_1 \cdot p_2} = M_{p_1} \cdot M_{p_2}$
- $M_{p_1} M_{p_2} = M_{p_2} M_{p_1}$

Remark 1.5.12. *Let* M_x *be the multiplication matrix of* x *with respect to the standard basis* $B = \{1, x, \ldots, x^{k-1}\}$*. Explicitly,*

$$M_x : \mathbb{K}[x]/< f > \mapsto \mathbb{K}[x]/< f > \textit{ such that } M_x(p) = px$$

where $f = x^k + f_{k-1}x^{k-1} + \ldots + f_1 x + f_0$*. Then* M_x *has a companion matrix structure. Its columns correspond to the basis* B *and rows correspond to* $xB = \{x, x^2, \ldots, x^k\}$*:*

$$M_x = \begin{array}{c|cccc|} & 1 & x & \dots & x^{k-1} \\ \hline x & 0 & 1 & \dots & 0 \\ x^2 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x^{k-1} & 0 & 0 & \dots & 1 \\ x^k & -f_0 & -f_1 & \dots & -f_{k-1} \end{array}.$$

[19, Section 2.4] includes a theorem that proves the essential role of the matrix of a multiplication map (usually called multiplication matrix) in root finding.

Theorem 1.5.13 ([19]). *Let $\mathscr{I}$ be zero dimensional, $p \in \mathbb{C}[x_1,\dots,x_n]$, and h_p be the minimal polynomial of multiplication matrix M_p. Then for $\lambda \in \mathbb{C}$, the following are equivalent:*

- *λ is a root of the equation $h_p(x) = 0$,*
- *λ is an eigenvalue of the matrix M_p,*
- *λ is a value of the function p on $V(\mathscr{I})$.*

However Theorem 1.5.13 does not reveal how to find those roots. The following theorem describes a procedure to compute the coordinates of the roots. That can be found in [82].

Theorem 1.5.14. *Let k be an algebraically closed field with characteristic zero and let $\mathscr{I}$ be a zero dimensional ideal in $\mathbb{K}[x_1,\dots,x_n]$. Let*

$$V(\mathscr{I}) = \{\xi_i = (\xi_{i1}\dots\xi_{in}) \in \mathbb{K}^n \ : \ i = 1,\dots,D\}$$

and assume that $D = \dim\ \mathbb{K}[x_1,\dots,x_n]/\mathscr{I}$,i.e., $\mathscr{I}$ does not have multiple roots.
Fix a monomial basis

$$N = \{x^{\alpha_{\mathbf{1}}},\dots,x^{\alpha_{\mathbf{D}}}\}.$$

Then the multiplication matrices M are simultaneously diagonalizable such that

$$V^{-1}\ M_f\ V = D_f$$

where $V = V_N(\xi)$ is the Vandermonde matrix and D_f is a diagonal matrix with $[D]_{i,i} = f(\xi_i)$ for $i = 1,\dots,D$.
Moreover, we can find the j-th coordinates of the roots by diagonalizing M_{x_j} to get

$$[D_{x_j}]_{i,i} = \xi_{ji}.$$

The following consequence of the Stickelberger's theorem ([74] and [5]) provides an important property of the trace of a multiplication matrix.

Theorem 1.5.15. *Let $\mathscr{I} \subset \mathbb{K}[x_1,\dots,x_n]$ be a zero dimensional ideal. Then for all $p \in \mathbb{K}[x_1,\dots,x_n]$ trace of the linear (multiplication) map M_p has the property*

$$Tr(M_p) = \sum_{\xi \in V_{\mathbb{C}}(I)} \mu(\xi) p(\xi)$$

where $\mu(\xi)$ is the multiplicity of ξ.

We define one more very well known object. Using Newton identities, one can find powers sums of roots of a polynomial without any knowledge of the roots as shown in the following theorem.

Definition 1.5.16. *The i-**th Newton sum** of the polynomial f is*

$$p_k := \sum_{\xi \in V(f)} \mu(\xi)\xi^k,$$

where $\mu(\xi)$ is the multiplicity of ξ.

Theorem 1.5.17. *Let $z_1,\dots,z_k$ be roots of $f(x) = x^k - f_{k-1}x^{k-1} + \cdots + f_1 x + f_0$, define the power series*

$$p_k = \sum_{i=1}^{k} z_i^k,$$

and the elementary symmetric polynomials are

$$\begin{aligned} &e_0 = 1, \\ &e_1 = z_1 + z_2 + \cdots + z_k, \\ &e_2 = \textstyle\sum_{1\le i<j\le k} z_i z_j, \\ &e_k = z_1 z_2 \cdots z_k, \\ &e_n = 0, \quad \textit{for } n > k. \end{aligned}$$

*Then the **Newton Identities** for all $n \geq 1$ defined by*

$$n e_n = \sum_{i=1}^{n} (-1)^{i-1} e_{n-i} p_i.$$

Then one can find the k-th Newton power sum

$$p_k = \sum_{i=k-n}^{k-1} (-1)^{k-1+i} e_{k-i} p_i,$$

for all $k > n \geq 1$.

The Following definition and theorem are adopted from [5].

Definition 1.5.18. *A* **quadratic form** *with coefficients in $\mathbb{K}$ is defined by*

$$Q(a_1, \dots, a_n) := \sum_{i,j=1}^{n} m_{i,j} a_i a_j$$

with $M = [m_{i,j}]$ a symmetric matrix for $i, j = 1, \dots, n$. If $a = (a_1, \dots, a_n)$, then $Q(a) = a^t M a$.

Theorem 1.5.19 (Sylvester's law of inertia)**.** *Let Q be a quadratic form as defined above,*
(i) *A quadratic form $Q(a)$ of dimension n has always a diagonal expression.*
(ii) *If $\mathbb{K}$ is ordered, the difference between the number of positive coefficients and the number of positive coefficients in a diagonal expression of $Q(a)$ is a well defined quantity.*

1.6 Real Root Counting

There are very well-known methods to count real roots of univariate polynomials. Descartes's rule of signs, and Sturm's Theorem are two classical results we briefly describe in this section.

Descartes's rule of signs is a technique for determining an upper bound on the number of real roots of a polynomial. It is not a complete criterion, because it does not provide the exact number of positive or negative roots. The rule is applied by counting the number of sign changes in the sequence of the polynomial's coefficients. If a coefficient is zero, that term is simply omitted from the sequence.

Definition 1.6.1 (**Descartes's rule of signs**)**.** *Let $f(x) = f_n x^n + \cdots + f_1 x + f_0 \in \mathbb{R}[x]$ with nonzero f_n and f_0. Let v be the number of sign change in the sequence $(f_0, f_1, \dots, f_n)$. Let p be the number of positive real roots of f counted with multiplicity, then*

$$\textit{there exists } m \in \mathbb{Z}_{\geq 0} \textit{ such that } p = v - 2m.$$

Moreover, if we replace $f(x)$ by $f(-x)$ in Descartes's rule then we get a bound on the number of negative real roots of f.

Sturm's Theorem expresses the number of distinct real roots of a polynomial in an interval in terms of the number of changes of signs of the values of the Sturm's sequence at the bounds of the interval. Applied to the interval of all the real numbers, it gives the total number of real roots of the polynomial.

The fundamental theorem of algebra already gives the all number of complex roots, counted with multiplicity. Sturm's theorem counts the number of distinct real roots and locates them in intervals.

Definition 1.6.2. *Let $f(x)$ be a univariate real polynomial. Let $f_0(x) := f(x)$ and $f_1(x) := f'(x)$, then the* **Sturm Sequence** *of f is defined by*

$$\begin{aligned} f_0(x) &= q_1(x)f_1(x) - f_2(x) \\ f_1(x) &= q_2(x)f_2(x) - f_3(x) \\ &\vdots \\ f_{n-2}(x) &= q_{n-1}(x)f_{n-1}(x) - f_n(x). \end{aligned}$$

If f has no multiple roots then f_n is a nonzero constant and the Sturm sequence is

$$(f_0(x), f_1(x), \ldots, f_n).$$

Theorem 1.6.3. [Sturm's Theorem] *Assume $f(x) \in \mathbb{R}[x]$ has no multiple roots, and $f(a) \neq 0$, $f(b) \neq 0$ for $a, b \in \mathbb{R}$. Then the number of real roots of f in $[a, b]$ is equal to*

$$v(a) - v(b)$$

where $v(c)$ is the number of sign variation in the Sturm's sequence at c, i.e., $(f_0(c), \ldots, f_n)$ for any $c \in \mathbb{R}$.

Generalized version of the Sturm's Theorem uses slightly different Sturm sequence with an auxiliary function thus we can omit the no multiple roots assumption.

Theorem 1.6.4. [Generalized Sturm's Theorem] *Assume $f(x), g(x) \in \mathbb{R}[x]$ and $a, b \in \mathbb{R}$ such that $a < b$. Let $(f, f'g, f_2, \ldots, f_n)$ be the Sturm's sequence and $v(c)$ be the number of sign variation in $(f_0(c), \ldots, f_n)$ for any $c \in \mathbb{R}$. Then*

$$v(a) - v(b) = \#\{c \in [a, b] \mid f(c) = 0 \text{ and } g(c) > 0\} - \#\{c \in [a, b] \mid f(c) = 0 \text{ and } g(c) < 0\} \tag{1.19}$$

not counting multiplicity.

CHAPTER

2

CERTIFYING SOLUTIONS TO OVERDETERMINED POLYNOMIAL SYSTEMS OVER $\mathbb{Q}$

2.1 Introduction

2.1.1 Related Work

As we showed in Subsection 1.2.4, α-theory can be used to get efficient root certification algorithm for well-constrained polynomial systems.

For an overdetermined polynomial system $\mathbf{f} := (f_1, \ldots, f_N)$ in $\mathbb{C}[x_1, \ldots, x_n]$ with $N > n$, Dedieu and Shub [24] studied the overdetermined Newton method (also called Gauss-Newton method when N>n) whose iterates are defined by

$$N_{\mathbf{f}}(z) := z - J_{\mathbf{f}}(z)^{\dagger}\mathbf{f}(z)$$

where $J_{\mathbf{f}}(z)^{\dagger}$ is the Moore-Penrose pseudoinverse of $J_{\mathbf{f}}(z)$ ([63], Section 5.12), to determine conditions that guarantee quadratic convergence.

Definition 2.1.1. *Let* $\mathbf{f}$ *be a polynomial system, then* $J_{\mathbf{f}}$ *is a continuous operator from* $\mathbb{C}^n$ *to* $\mathbb{C}^N$. *If* $J_{\mathbf{f}}$ *is one-to-one with closed image (i.e.,* $J_{\mathbf{f}}(z)^{-1} : im(J_{\mathbf{f}}) \to \mathbb{C}^n$ *is continuous), the Moore-Penrose pseudoinverse of* $J_{\mathbf{f}}(z)$ *is defined by*

$$J_{\mathbf{f}}^{\dagger} = (J_{\mathbf{f}}^{*} J_{\mathbf{f}})^{-1} J_{\mathbf{f}}^{*},$$

and if $J_{\mathbf{f}}$ *is onto, the Moore-Penrose pseudoinverse of* $J_{\mathbf{f}}(z)$ *is defined by*

$$J_{\mathbf{f}}^{\dagger} = J_{\mathbf{f}}^{*}(J_{\mathbf{f}} J_{\mathbf{f}}^{*})^{-1}$$

where $J_{\mathbf{f}}^{*}$ *is the conjugate transpose of* $J_{\mathbf{f}}$.

For well-constrained case (with nonsingular Jacobian matrix), as mentioned in Remark 1.2.4, fixed points correspond to roots of $\mathbf{f}$. For overdetermined $\mathbf{f}$, when Newton iterates converges to $z \in \mathbb{C}$, $J_{\mathbf{f}}(z)^{\dagger}\mathbf{f}(z) = 0$. Then $f(z)$ is not necessarily equals zero but z is a critical point of $||\mathbf{f}(z)||^2$.

If our $J_{\mathbf{f}}$ is *onto* then the main and well-known properties of Newton methods are

(i) fixed points correspond to roots of $\mathbf{f}$,

(ii) convergence to fixed points is quadratic.

When derivatives in $J_{\mathbf{f}}$ are *one-to-one*, Newton method may have fixed points which are not roots. Also convergence to these fixed points may fail to be quadratic.

Since the fixed points of $N_{\mathbf{f}}$ may not be solutions to the overdetermined polynomial system $\mathbf{f}$, α-theory approach cannot distinguish whether a point is a root or a local minimum. Thus it cannot certify solutions to overdetermined polynomial systems. The following example illustrates that.

Example 2.1.2. *Consider*

$$\mathbf{f}(x) = \begin{bmatrix} x \\ x^2 + a \end{bmatrix}$$

where $a \in \mathbb{R}$.

$x = 0$ *is a stationary point of*

$$||\mathbf{f}(x)||^2 = x^2 + (x^2 + a)^2 = x^4 + (2a+1)x^2 + a^2,$$

i.e., $||\mathbf{f}(x)||^2$ *attains its minimum at* 0. *If* $a = 0$, $x = 0$ *is a root of* $\mathbf{f}$. *If* $a \neq 0$, *then* $\mathbf{f}(x) \neq 0$. *Newton iterate of* $\mathbf{f}$ *is*

$$N_{\mathbf{f}}(x) = x - \begin{bmatrix} \frac{1}{1+4x^2} & \frac{2x}{1+4x^2} \end{bmatrix} \begin{bmatrix} x \\ x^2 + a \end{bmatrix} = x - \frac{2x^3 + (2a+1)x}{1+4x^2},$$

clearly $x = 0$ is a fixed point since $N_{\mathbf{f}}(0) = 0$. However when $|a| \geq 1/2$, $x = 0$ is not a fixed attractive point.

In [37], Hauenstein and Sottile give a heuristic approach to certify solutions to overdetermined systems. They take two or more randomized square subsystem of the given overdetermined system **f**, then use α-theory as described in Subsection 1.2.4, and compare solutions. They also remark that for their work to give an algorithm, we would need a general bound for the minimum of a positive polynomial on a disk. However they note that all known bounds are too small to be practical.

A closer look at the literature on lower bounds for the minimum of positive polynomials over the roots of zero-dimensional rational polynomial systems reveals that they all reduce the problem to the univariate case and use univariate root separation bounds (see, for example, [15, 43, 14, 44]). This led to the idea of directly using an exact univariate representation for the certification of the input system instead of using universal lower bounds that are often very pessimistic.

In principle, one can compute such a univariate representation using purely algebraic techniques, for example, by solving large linear systems corresponding to resultant or subresultant matrices (see, for example, [81]). However, this purely symbolic method would again lead to worst case complexity bounds. Instead, we propose a hybrid symbolic-numeric approach, using the approximate roots of the system, as well as exact univariate polynomial arithmetic over $\mathbb{Q}$. We expect that our method will make the certification of roots of overdetermined systems practical for cases when the universal lower bounds are too pessimistic, or when the actual size of our univariate representation is significantly smaller than in the worst case.

There is an extended body of literature on using interval arithmetic and optimization methods to certify the existence or non-existence of the solutions of well-constrained systems with guaranteed accuracy, e.g., [51, 65, 76, 89, 49, 67]. Techniques to certify each step of path tracking in homotopy continuation for well-constrained systems using α-theory are presented in [9, 10, 33, 34]. Recently, a certification method of real roots of positive-dimensional systems was studied in [90].

Related to the certification problems under consideration is the problem of finding certified sum of squares (SOS) decompositions of rational polynomials. In [71, 72, 46, 47], they turn SOS decompositions given with approximate (floating point) coefficients into rational ones, assuming that the feasible domain of the corresponding semidefinite feasibility problem has nonempty interior. In [64], they adapt these techniques to the degenerate case, however they also require a feasible solution with rational coordinates to exists. The certification of more general polynomial, semi-algebraic and transcendental optimization problems were considered in [61]. In [25], they compute rational points in semi-algebraic sets and give a method to decide if a polynomial can be expressed as a SOS of rational polynomials. Note that we can straightforwardly translate the certification of approximate roots of overdetermined polynomial systems into polynomial optimization problems over compact convex sets (using a ball around the approximate root), however, we cannot guarantee a rational feasible solution. Instead, we propose to construct the rational univariate representation of several irrational roots that form a rational component of the input system.

The idea of reconstructing exact algebraic or geometric objects from approximations is not new, it includes the computation of the minimal polynomials of approximate algebraic numbers (see for instance [56, 45, 48, 26, 16]), and the computation of a Gröbner basis or geometric representation of algebraic sets (see for example [83, 88, 29, 39, 16, 30, 77, 4, 7]). One of the main difference between our approach and some of the iterative techniques computing a geometric representation that are listed above, is that they use non-Archimedian metric, while here we use the usual Archimedean absolute value of complex numbers. Even though these methods are analogous, they are not equivalent, as shown in [36]. The second difference is the use of lattice basis reduction or interpolation techniques. In [16, 7], they use the algorithm from [48] for the construction of the minimal polynomial of a given approximate algebraic number. However, the algorithm in [48] requires an upper bound for the height of the algebraic number, and this bound is used in the construction of the lattice that they apply LLL lattice basis reduction [56]. To get such bound *a priori*, we would need to use universal bounds for the height. In order to get an incremental algorithm with early termination for the case when the output size is small, one can modify the algorithm in [48] to be incremental, but that would require multiple application of the lattice basis reduction algorithm. Alternatively, one can apply the PSLQ algorithm as in [26], which is incremental and does not require an *a priori* height bound. The main point of the approach in this paper is that we assume to know *all* approximate roots of a rational component. In this case, we can compute the exact RUR much more efficiently, and in parallel, as it is proved in this paper and in [36]. So instead of multiple calls for lattice basis reduction, we propose a cheaper interpolation based lifting and checking technique.

Related literature on certification of singular zeros of polynomial systems include [50, 75, 62, 59, 60]. However, these approaches differ from ours in the sense that they certify singular roots of some small perturbation of the input polynomial system while we certify singular roots of an exact polynomial system with rational coefficients. Recently, the exact certification of the multiplicity structure of isolated singular roots was considered in [35].

Moreover, in a very recent work [84] [Sec. 9.8] authors suggest a similar approach to ours, using numerical approximations to reconstruct an exact polynomial system. The main difference is that we provide complexity bounds on this reconstruction.

2.1.2 Outline of Our Approach

Consider an overdetermined system $\mathbf{f} = (f_1, \ldots, f_N) \in \mathbb{Q}[x_1, \ldots, x_n]$ for some $N > n$, and assume that the ideal $\mathscr{I} := \langle f_1, \ldots, f_N \rangle$ is radical and zero dimensional. Under these assumptions, the Rational Univariate Representation (RUR) for $V(\mathscr{I})$ exists (see Subsection 1.1.2), as well as for any component of $V(\mathscr{I})$ over the rationals. Since the polynomials in the RUR have also rational coefficients, we can hope to compute them exactly, unlike the possibly irrational coordinates of the common roots of $V(\mathscr{I})$. With the exact RUR, which is a well-constrained system of polynomials, we can use α-theory as in Subsection 1.2.4 to certify that a given point is an approximate root for the RUR, and thus for our original system $\mathbf{f}$. Our symbolic-numeric method to compute a RUR for $V(\mathscr{I})$ or for a rational component of $V(\mathscr{I})$ consists of the following steps:

Initialization

(i) Compute approximations of all isolated roots to a given accuracy of a random well-constrained set of linear combinations of the polynomials $f_1, \ldots, f_N$ using homotopy continuation (see Subsection 1.2.1);

(ii) Among the approximate roots computed in Step (i), choose the candidates which could be approximations to roots in $V(\mathscr{I})$ or a rational component of $V(\mathscr{I})$ – for this step, we can only give heuristics on how to proceed;

(iii) Chose a separating linear form (primitive element) for the approximate roots chosen in Step (ii);

Iteration

1. Construct numerically an approximate RUR from the approximate roots chosen in Step (ii) using formulas expressing the RUR in terms of the roots;
2. Using rational number reconstruction (see Section 1.4), find the unique RUR with rational coefficients of bounded denominators that is within a given distance from the approximate RUR computed in Step 1;
3. Check whether the computed exact polynomials form a proper RUR for a component of $V(\mathscr{I})$, using exact arithmetic over $\mathbb{Q}$. If yes, terminate, and return the exact RUR;
4. If the answer is no, then either
   (a) if the size of the coefficients in the RUR computed in Step 3 exceed a preset bound, terminate and return failure. Otherwise:
   (b) apply one step of Newton's iteration to increase the precision of the approximate roots and continue from Step 1 of the Iteration.

Now, we define heights of polynomials over $\mathbb{Q}$ in a way that we can utilize symbolic algorithms over $\mathbb{Z}$ to get bounds (see [42] where it is called projective height).

Definition 2.1.3. *Let $p(T) = T^d + a_{d-1}T^{d-1} + \cdots + a_0 \in \mathbb{Q}[T]$ where each $a_i = n_i/z_i$ for $n_i \in \mathbb{Z}$ and $z_i \in \mathbb{N}$. Let $P(T) \in \mathbb{Z}[T]$ be any integer polynomial that is an integer multiple of $p(T)$, for example we can choose*

$$P(T) = b_d T^d + b_{d-1}T^{d-1} + \cdots + b_0 := \left(\prod_{i=0}^{d-1} z_i\right) p(T) \in \mathbb{Z}[T].$$

Then the height of p is defined as

$$H(p) = H(P) = \frac{\max\{|b_i| \, : \, i = 0,\ldots,d\}}{\gcd(b_i \, : \, i = 0,\ldots,d)}$$

which is clearly independent of the representation of $p(T)$ in $\mathbb{Z}[T]$. Note that if we assume that $\gcd(n_i, z_i) = 1$ for all $i = 0,\ldots,d-1$ then

$$H(p) = \max_i\{|n_i|\,,\, \mathrm{lcm}_j(z_j)\} \geq \max_i\{|n_i|, z_i\}, \tag{2.1}$$

as in [21], so we can use the height to bound the magnitude of the numerators and denominators appearing in the coefficients of our polynomials.
We also define the logarithmic height of p as

$$h(p) := \log H(p).$$

The height and logarithmic height of a rational number $a \in \mathbb{Q}$ is defined to be $H(T-a)$ and $h(T-a)$, respectively.

In Subsection 2.2.3, we prove the following complexity bounds for the iteration part of the above symbolic-numeric algorithm:

- it either terminates successfully returning an exact RUR of a rational component $V \subset V(\mathcal{I})$ after at most
$$\mathcal{O}(\log\log(H^* E_0))$$
itarations, where H^* is the maximal height of the polynomials in the output and E_0 is an upper bound on the Euclidean distance of the initial approximate RUR and the exact RUR;
- or terminates with failure, in which case the iteration converged to some polynomials which either did not have rational coefficients or did not form a proper RUR. In this case, we need at

most

$$\mathcal{O}\left(n\log(D)+\log(nh\delta)+\log\log(E_0nd)\right)$$

iterations to conclude that our iteration would not have converged to an exact RUR of a rational component of $V(\mathscr{I})$. Here, h and D are the maximum of the logarithmic heights and degrees of the polynomials in **f**, respectively, δ is the number of points in $V(\mathscr{I})$, d is the number of points in the rational component $V \subset V(\mathscr{I})$ we compute, and E_0 is as above;

- the k-th iteration has bit complexity polynomial in the input size, but exponential in k (see Theorem 2.2.7 for the exact statement). Using the above bounds for the number of iterations, this implies that once we computed the initial root approximations, the iteration part of our algorithm is polynomial in the input plus the output size upon success, or to detect failure, its complexity is comparable to worst case complexity bounds in the literature.

2.2 The Certification Algorithm

2.2.1 Initialization

The iterative method we use is *locally convergent*, so we need an initial RUR that is sufficiently close to the exact one in order to have convergence. In this subsection, we discuss the algorithm that we propose to find a good initial approximate RUR that we can use as the starting point for our iteration. Using α-theory, we can guarantee that our iteration will be convergent (see below in Subsection 2.2.2.1). However, the RUR that we converge to may not have rational coefficients, or may not be a “proper" RUR satisfying the conditions of Definition 1.1.16, in which case we cannot use it as certificate. In Steps (ii) and (iii) below we give a heuristic method to find an initial RUR that would converge to a proper RUR with rational coefficients.

More precisely, to compute an initial RUR for a rational component of $\mathscr{I}$ we propose the following:

***(i)* Homotopy method on a well-constrained subsystem.** Let $\mathbf{f}=(f_1,\ldots,f_N)\in\mathbb{Q}[x_1,\ldots,x_n]$ for some $N>n$ be the defining equations of $\mathscr{I}$ which is assumed to be zero dimensional and radical. As in [37][Section 3], for any linear map $R:\mathbb{Q}^N\to\mathbb{Q}^n$, which we will also consider as a matrix $R\in\mathbb{Q}^{n\times N}$, we define the well-constrained $n\times n$ system $R(\mathbf{f}):=R\circ\mathbf{f}$. For almost all choices of $R\in\mathbb{Q}^{n\times N}$, the ideal generated by $R(\mathbf{f})$ is zero dimensional and radical. We fix such an $R\in\mathbb{Q}^{n\times N}$ and throughout this paper we use the notation

$$\boldsymbol{F}=(F_1,\ldots,F_n):=R(\mathbf{f}). \tag{2.2}$$

In this step, we assume that by using numerical homotopy algorithms (as mentioned in Subsection 1.2.1), we have computed *approximate roots* for each root in $V(\boldsymbol{F})$, i.e., local Newton's method with respect to $\boldsymbol{F}$ is quadratically convergent starting from these approximate roots (see Subsection 1.2.4 on how to certify approximate roots of well-constrained systems). Using α-theory for $\boldsymbol{F}$, we can estimate the distance from each of these approximate roots to the exact ones. Denote an upper bound for these distances by ε_0.

(ii) **Candidates for roots of a component of $V(\mathscr{I})$.** To find the subset of approximate roots to $V(\boldsymbol{F})$ that approximates the roots in $V(\mathscr{I})$, or a rational component of $V(\mathscr{I})$ containing the root we want to certify, we propose several methods.

The first one is to choose the roots that has residuals for all f_i for $i = 1,\ldots,N$ up to a given tolerance t. The tolerance t can be chosen based on ε_0 defined above in Step (i), and the height and degree of each of the polynomials in **f**.

Another approach could incorporate the ideas in [37, Section 3] to exclude the roots that are *not* approximations of $V(\mathscr{I})$ by comparing the approximate roots of $R(\mathbf{f})$ to the approximate roots of an other random square subsystem $R'(\mathbf{f})$ for some $R' \in \mathbb{Q}^{n\times N}$.

Finally, if we know that $V(\mathscr{I})$ or a rational component of $V(\mathscr{I})$ has small cardinality, then we can check all subsets of the roots computed in Step (i) that have that cardinality as candidates for the common roots of a component of $\mathscr{I}$.

Denote the cardinality of the roots selected in this step by d.

(iii) **Primitive element** In this step, we choose a random rational linear form $u = \lambda_1 x_1 + \cdots + \lambda_n x_n$ from the finite set

$$\mathscr{T} = \left\{x_1 + i x_2 + \cdots + i^{n-1} x_n \;:\; i \in \mathbb{N}, 0 \le i \le (n-1)d(d-1)/2\right\}$$

(see Lemma 1.1.14), and check if it separates the approximate roots chosen in Step **(ii)**. After scaling, we will assume that

$$||u||_\infty = \max(|\lambda_1|,\ldots,|\lambda_n|) \le 1. \tag{2.3}$$

2.2.2 Iteration

2.2.2.1 Increasing the Precision of the RUR using Local Newton Iteration

In Step 1 of the Iteration part of our algorithm, we compute a RUR for the d approximate roots using (1.7) and (1.9). In this section, we discuss the sensitivity of this step to the perturbation of the roots.

More precisely, we give an upper bound for the distance of the approximate RUR after k iterations from the exact one.

Let $\boldsymbol{F} = (F_1,\dots,F_n)$ be as in (2.2). Let $\{\xi_1^{(0)},\dots,\xi_d^{(0)}\} \subset \mathbb{C}^n$ be the the d approximate roots we identified in Step (ii) of Subsection 2.2.1, and let $\{\xi_1^*,\dots,\xi_d^*\} \subset \mathbb{C}^n$ be the corresponding exact roots in $V(\boldsymbol{F})$. For each $i = 1,\dots,d$ and $k \geq 0$ we define the $(k+1)$-th Newton iterate by

$$\xi_i^{(k+1)} := \xi_i^{(k)} - J_{\boldsymbol{F}}(\xi_i^{(k)})^{-1} F(\xi_i^{(k)}),$$

where $J_{\boldsymbol{F}}(\xi_i^{(k)})$ is the $n \times n$ Jacobian matrix of $\boldsymbol{F}$ evaluated at $\boldsymbol{z}_i^{(k)}$, which we assume to be invertible. Then, using our assumption in Step (i) of Subsection 2.2.1, namely that for all $i = 1,\dots,d$,

$$\|\xi_i^{(0)} - \xi_i^*\|_2 \leq \varepsilon_0,$$

and that the Newton iteration is quadratically convergent from each $\xi_i^{(0)}$ to ξ_i^*, by Definition 1.10 we get that

$$\|\xi_i^{(k)} - \xi_i^*\|_2 \leq \varepsilon_0 \left(\frac{1}{2}\right)^{2^k - 1}.$$

Next, we analyze the possible loss of precision when applying (1.7) and (1.9) in the computation of the approximate RUR. Note that to get the error bounds below, we assume that there is no roundoff error in our computations, only the approximation error from the roots.

Proposition 2.2.1. *Using the above notation, denote by $(u, q^{(k)}, r_1^{(k)},\dots,r_n^{(k)})$ the approximate RUR corresponding to $\{\xi_1^{(k)},\dots,\xi_d^{(k)}\}$, and suppose that $(u, q^*, r_1^*,\dots,r_n^*)$ is the exact RUR corresponding to $\{\xi_1^*,\dots,\xi_d^*\}$. Assume that $\|u\|_\infty \leq 1$. Then for all $i = 0,\dots,d-1,\ j = 1,\dots,n$ we have*

$$|\mathrm{coeff}_{T^i}(q^{(k)} - q^*)|, |\mathrm{coeff}_{T^i}(r_j^{(k)} - r_j^*)| < d(2e)^{d/2} C^{d-1} \varepsilon_0 \left(\frac{1}{2}\right)^{2^k - 1}, \tag{2.4}$$

where $C = \max\{C^{(k)}, C^*\}$ such that

$$|\xi_{i,j}^{(k)}| \leq C^{(k)},\ \ |\xi_{i,j}^*| \leq C^*, \textit{ for all } i = 0,\dots,d,\ j = 1,\dots,n. \tag{2.5}$$

Proof. Using the formula in (1.7), coefficients of $q^{(k)}(T)$ and $q^*(T)$ are polynomials in the coordinates of $\{\xi_1^{(k)},\dots,\xi_d^{(k)}\}$ and $\{\xi_1^*,\dots,\xi_d^*\}$ respectively. $\mathrm{coeff}_{T^i}(q^{(k)})$ and $\mathrm{coeff}_{T^i}(q^{(k)})$ have at most $\binom{d}{\lfloor \frac{d}{2} \rfloor}$ terms, each has at most $(d-i)$ product and each with coefficients ≤ 1 in absolute value.

Let $C^{(k)} := \max\{|\xi_{i,j}^{(k)}|\}$, and $C^* := \max\{|\xi_{i,j}^*|\}$, and using (2.3),

$$|\mathrm{coeff}_{T^i}(q^{(k)} - q^*)| < \binom{d}{\lfloor \frac{d}{2} \rfloor}((C^{(k)})^{d-i} - (C^*)^{d-i}).$$

Using that for real numbers x and y, $x^n - y^n < (x-y)ny^{n-1}$ when $0 < x < y$ for any natural number n, and $\binom{d}{t} \leq \left(\frac{de}{t}\right)^t$, with e the natural base, we obtain

$$|\mathrm{coeff}_{T^i}(q^{(k)} - q^*)| < \left(\frac{de}{\lfloor d/2 \rfloor}\right)^{\lfloor d/2 \rfloor} (C^{(k)} - C^*)(d-i)C^{d-i-1}.$$

Now by Definition 1.10, and $\left(\frac{de}{\lfloor d/2 \rfloor}\right)^{\lfloor d/2 \rfloor} < (2e)^{d/2}$, we get the error bound for the coefficients in the RUR is as claimed

$$|\mathrm{coeff}_{T^i}(q^{(k)} - q^*)| < (2e)^{d/2} \varepsilon_0 \left(\frac{1}{2}\right)^{2^k - 1} dC^{d-1}, \tag{2.6}$$

for all $i = 0, \ldots, d$.

Similarly, using the formula (1.9) for $j = 1, \ldots, n$, $r_j(T) = \sum_{i=1}^d \xi_{i,j} \prod_{m \neq i}(T - u(\xi_m))$ the coefficients of $r_j^{(k)}(T)$ and $r_j^*(T)$ are polynomials in the coordinates of $\{\xi_1^{(k)}, \ldots, \xi_d^{(k)}\}$ and $\{\xi_1^*, \ldots, \xi_d^*\}$ respectively. First consider $\prod_{m \neq i}(T - u(\xi_m))$ part which is very similar to $q(T)$, with degree $d-1$ instead of d with coefficients ≤ 1 in absolute value. Then using the equations and definitions of $C^{(k)}$, C^* and C above we get

$$\begin{aligned} |\mathrm{coeff}_{T^i}(r_j^{(k)} - r_j^*)| &< d(C^{(k)} - C^*)\binom{d-1}{\lfloor \frac{d-1}{2} \rfloor}((C^{(k)})^{d-1-i} - (C^*)^{d-1-i}) \\ &< \left(\frac{de}{\lfloor d/2 \rfloor}\right)^{\lfloor d/2 \rfloor} (C^{(k)} - C^*)dC^{d-i-1}. \end{aligned}$$

Therefore, the right hand side of (2.6) is an upper bound of $|\mathrm{coeff}_{T^i}(r_j^{(k)} - r_j^*)|$ for all $j = 1, \ldots, n$ and $i = 0, \ldots, d$. □

Remark 2.2.2. *Since C, d and ε_0 are constant throughout the iteration, we see that the error in the coefficients in the RUR converges to zero as $k \to \infty$.*

2.2.2.2 Rational Number Reconstruction

In this subsection, we show how to find the exact RUR of a rational component of $V(\mathscr{I})$ once we computed a sufficiently close approximation of it. The main idea is that we can reconstruct the unique rational numbers that have bounded denominators and indistinguishable from the coefficients of the polynomials in our approximate RUR within their accuracy estimates. Then, we can use purely symbolic methods to check whether the RUR with the bounded rational coefficients is an exact RUR for a component of $V(\mathscr{I})$.

Since the coordinates of the approximate roots are given as floating point complex numbers, we can consider them as Gaussian rational in $\mathbb{Q}(i)$, and the same is true for the coefficients of the approximate RUR computed from these roots in Subsection 2.2.2.1. However, since the exact RUR has rational coefficients, we will neglect the imaginary part of the coefficients of the approximate RUR. Therefore, we will assume that the coefficients of the approximate RUR are in $\mathbb{Q}$, given as floating point numbers.

In Theorem 1.4.9, to compute the pair $(r,s)\in\mathbb{Z}^2$ for each coefficient appearing in the approximate RUR computed in the previous subsections, we use the bound $B\in\mathbb{N}$ such that $\frac{1}{2B^2}\cong E$, where E denotes our estimate of the accuracy of our approximate RUR from (2.4). Thus, we can define

$$B := \left\lceil (2E)^{-1/2} \right\rceil. \tag{2.7}$$

For efficient computation of the rational number reconstruction, we can use continued fractions as described in Section 1.4.

Remark 2.2.3. *Theorem 1.4.9 does not guarantee the existence of the pair (r,s) with the given properties, only uniqueness. In case the rational number reconstruction algorithm for some coefficient c returns that there is no rational number within distance E with denominators at most B, we will need to improve the precision E (which will increase the bound B on the denominator). This is done by applying further local Newton steps on our approximates. As described in Theorem 2.2.5 below, if the bound B we obtained this way is larger than an a priori bound, we terminate our iteration and conclude that it did not converge to a RUR of a rational component.*

2.2.3 Termination

One key task is to decide whether to terminate our iterations or increase the accuracy of the approximate RUR as described in Sections 2.2.2.1.

Let $(u,q,r_1,\ldots,r_n)$ be the rational polynomials with bounded denominators as computed in Subsection 2.2.2.2. In this step, we will symbolically check the 3 properties of Definition 1.1.16 that defines a PUR of a component of $V(\mathscr{I})$. Note that since we are given a RUR not a PUR, we will use that

$$v_j = \frac{r_j}{q'} \mod q \quad j=1,\ldots,n.$$

We can either compute $1/q' \mod q$ if it exists, or equivalently we can check the polynomial equalities

$$\begin{gathered}\gcd(q,q')=1,\\ \lambda_1 r_1+\cdots\lambda_n r_n \equiv Tq' \mod q,\\ \mathrm{numer}\Big(f_t\left(\tfrac{r_1}{q'},\ldots,\tfrac{r_n}{q'}\right)\Big)\equiv 0 \mod q \quad t=1,\ldots,N.\end{gathered} \tag{2.8}$$

The latter is more efficient if $\deg f_t < d = \deg q$. Since the coefficients of the polynomials during this computation grow in size, we will use a small primes modular approach as in [27][Section 6.7], which can be efficiently parallelized.

If $(u,q,r_1,\ldots,r_n)$ does not satisfy some of the properties in (2.8), then either the accuracy of our approximate RUR was insufficient to define the exact RUR in Subsection 2.2.2.2, or the iteration does not converge to a RUR of a rational component of $V(\mathscr{I})$. To decide which case we are in, we will use *a priori* upper bounds on the heights of the coefficients of a RUR of a rational component of $V(\mathscr{I})$.

Below, we review some of the known upper bounds that we can use in our estimates. The best known upper bounds for the logarithmic heights of the polynomials in the RUR of $V(\mathscr{I})$ are as follows.

First, we give a bound for the logarithmic bound of the primitive element $u=\lambda_1 x_1+\cdots+\lambda_n x_n$ using Definition 2.1.3 and Lemma 1.1.14 :

$$\begin{aligned} h(\lambda_i) = h(T-\lambda_i) &\text{ for all } i=1,\ldots,n.\\ &=\log(i^{n-1})\\ &<(n-1)\log((n-1)d(d-1)/2)\\ &<(n-1)\log(nd^2)\\ &<2(n-1)\log(nd)\end{aligned}$$

Thus we can state that

$$\max(h(\lambda_1),\ldots,h(\lambda_n))\le 2(n-1)\log(nd).$$

Assume that the input polynomials $f_1,\ldots,f_N$ have degree at most D and logarithmic height at most h. To use the bound in [21], which assumes that x_1 is a primitive element, we take

$$(f_1,\ldots,f_N,\lambda_1 x_1+\cdots+\lambda_n x_n - T)$$

as our input, with a logarithmic height upper bound

$$h' := 2(n-1)\log(nd) + h.$$

Then, just as in [21], using an arithmetic Bézout theorem in [52][Cor. 2.10] and [21][Thm. 1], the logarithmic heights of the polynomials $q, r_1, \dots, r_n$ in a RUR of $V(\mathscr{I})$ are bounded by

$$\begin{aligned} h(q), h(r_i) &\le (nh' + (2n+3)\log(n+1))D^n + 5\log(n+3)nD \\ &= [n(2(n-1)\log(nd) + h) + (2n+3)\log(n+1)]D^n + 5\log(n+3)nD \\ &= [2n(n-1)(\log(n) + \log(d)) + nh + (2n+3)\log(n+1)]D^n + 5\log(n+3)nD \\ &< [nh + 2n(n-1)\log(d) + (2n^2+5n+3)\log(n+3)]D^n \\ &< [nh + 2dn^2 + (2n^2+5n+3)\log(n+3)]D^n \\ &< [nh + 4n^2 log(nd)]D^n \text{ for } n > 3. \end{aligned}$$

Finally, for $i = 1, \dots n$ and $n > 3$ we obtained

$$h(q), h(r_i) \le (nh + 4n^2 log(nd))D^n \tag{2.9}$$

See also similar bounds for the logarithmic height of the polynomials in the RUR in [28, 77].

To get a bound for the height of the polynomials in a RUR for a rational component of $V(\mathscr{I})$, first notice that if $(u, q, r_1, \dots, r_n)$ is a RUR of $V(\mathscr{I})$ and $(u, \tilde{q}, \tilde{r}_1, \dots, \tilde{r}_n)$ is a RUR of a rational component of $V(\mathscr{I})$ then

$$\tilde{q} \mid q \text{ and } \tilde{r}_i = (r_i \mod \tilde{q}) \quad i = 1, \dots, n.$$

Here the division is in $\mathbb{Q}[T]$, but after multiplying with least common denominators, it can be considered in $\mathbb{Z}[T]$ without changing the heights of the polynomials. We use Gelfand's inequality for the height of a polynomial divisor of an integer polynomial [42][Prop. B.7.3] (other bounds can also be used, a survey of the known bounds of factors in $\mathbb{Z}[x]$ be found in [1]). For $P, Q \in \mathbb{Z}[T]$ such that P is a divisor of Q, we have

$$H(P) \le e^{\deg(Q)} H(Q) \text{ or equivalently } h(P) \le \deg(Q) + h(Q),$$

thus $h(\tilde{q}) \le \delta + h(q)$. Furthermore, at the k-th step of division with remainder of r_i by $\tilde{q}$ gives that coefficients of $h(\tilde{r}_i)$ has degree $k-1$ in coefficients of $\tilde{q}$ and degree 1 in coefficients of r_i, and the division algorithm has total of $(\delta-1)-(d-1)=\delta-d$ steps. Thus we get the following bound

$$h(\tilde{r}_i)\leq h(r_i)+(\delta-d)h(\tilde{q})+(\delta-d).$$

Substituting $h(\tilde{q})\leq\delta+h(q)$ gives

$$\begin{aligned} h(\tilde{r}_i) &\leq h(r_i)+(\delta-d)(\delta+h(q))+(\delta-d)\\ &\leq \max(h(q),h(r_i))+\max(h(q),h(r_i))(\delta-d)+\delta(\delta-d)+(\delta-d)\\ &\leq \max(h(q),h(r_i))(\delta-d+1)+(\delta-d)(\delta+1)\\ &\leq \max(h(q),h(r_i))(\delta-d+1)+\delta^2+\delta(1-d)-d\\ &\leq \delta^2\max(h(q),h(r_i)) \text{ since } d\geq 1. \end{aligned}$$

This, combined with (2.9), gives the following upper bound for the logarithmic heights of the polynomials in an exact RUR $\tilde{q}(T),\tilde{r}_1(T),\ldots,\tilde{r}_n(T)$ of a rational component of $V(\mathscr{I})$:

$$h(\tilde{q}),h(\tilde{r}_i)\leq(nh+4n^2log(nd))\delta^2D^n \quad i=1,\ldots n. \tag{2.10}$$

Once we have an *a priori* bound for the heights of the polynomials in an exact RUR of a rational component of $V(\mathscr{I})$, we can use (2.1) and check if the bound B for the denominators used in the rational number reconstruction in Subsection 2.2.2.2 exceeds the *a priori* bound from (2.10). If that is the case, we conclude that the iteration did not converge to an exact RUR of a rational component of $V(\mathscr{I})$ and terminate our algorithm. Otherwise, continue to increase the accuracy of our approximation.

We summarize this subsection in the following theorems:

Theorem 2.2.4. *Let $\mathscr{I}$ be as above. Assume that $u,q^*,r_1^*,\ldots,r_n^*$ is an exact RUR of a rational component of $V(\mathscr{I})$. Define the maximum height*

$$H^*:=max\{H(q^*),H(r_1^*),\ldots,H(r_n^*)\}.$$

Assume that an approximate RUR, $u,q,r_1,\ldots,r_n$, satisfies

$$\|q(T)-q^*(T)\|_2,\|r_i(T)-r_i^*(T)\|_2\leq E<\frac{1}{2(H*)^2} \quad i=1,\ldots,n, \tag{2.11}$$

for some $E>0$, and let $\hat{q},\hat{r}_1,\ldots,\hat{r}_n$ obtained via rational number reconstruction on the coefficients of

$q, r_1, \dots, r_n$ using bound $B := \lceil (2E)^{-1/2} \rceil > H^*$. Then

$$\hat{q} = q^*, \hat{r}_1 = r_1^*, \dots, \hat{r}_n = r_n^*.$$

Proof. Note that the coefficients of $q^*, r_1^*, \dots, r_n^*$ have denominator at most $H^* < B$ by (2.1). Since the 2-norm gives an upper bound for the infinity norm, by our assumptions, all coefficients of $q^*, r_1^*, \dots, r_n^*$ are at most distance E from the corresponding coefficient of $q, r_1, \dots, r_n$. By Theorem 1.4.9, for each coefficient of $q, r_1, \dots, r_n$, there is at most one rational number with denominator bounded by $B = \lceil (2E)^{-1/2} \rceil > H^*$ within the distance of

$$\frac{1}{2B^2} = \frac{1}{2\lceil (2E)^{-1/2} \rceil^2} \leq \frac{1}{2\left((2E)^{-1/2}\right)^2} = E.$$

Therefore the rational reconstruction must be equal to the exact RUR. □

The next theorem considers the converse statement.

Theorem 2.2.5. *Let $\mathbf{f} \in \mathbb{Q}[x_1, \dots, x_n]^N$ be as above and assume that h and D are the maximum logarithmic height and degree of the polynomials in $\mathbf{f}$, respectively. Also, let δ be the number of the common roots of $\mathbf{f}$ in $\mathbb{C}^n$. Assume that we have an upper bound E for the accuracy of our approximate RUR $q, r_1, \dots, r_n$ from (2.4), and denote $d := \deg(q)$. Let $B := \lceil (2E)^{-1/2} \rceil$ and assume that*

$$\log(B) \geq (nh + 4n^2 log(nd))\delta^2 D^n.$$

Let $\hat{q}, \hat{r}_1, \dots, \hat{r}_n$ be obtained via rational number reconstruction from the coefficients of $q, r_1, \dots, r_n$ using the bound B for the denominators. If any of the 3 properties in (2.8) are not satisfied, then there is no exact RUR of a rational component of $V(\mathcal{I})$ with primitive element u within the distance of E from $u, q, r_1, \dots, r_n$.

Proof. If there was an exact RUR of a rational component of $V(\mathcal{I})$ within E from $q, r_1, \dots, r_n$, the logarithmic heights of its coefficients would be bounded by $(nh+4n^2 log(nd))\delta^2 D^n$ as in (2.10). The rational number reconstruction algorithm would have found this exact RUR, which is a contradiction. □

2.2.4 Complexity

Next, we give asymptotic bounds for the number of iterations needed in the "best case" and in the "worst case".

Theorem 2.2.6. *Let* $\mathbf{f} \in \mathbb{Q}[x_1, \ldots, x_n]^N$ *and* $\mathscr{I} = \langle \mathbf{f} \rangle$ *be zero dimensional and radical.*

1. *Assume that* $u, q^*, r_1^*, \ldots, r_n^*$ *is an exact RUR for a rational component of* $V(\mathscr{I})$ *and assume that* $u, q^{(0)}, r_1^{(0)}, \ldots, r_n^{(0)}$ *is an initial approximate RUR which quadratically converges to* $q^*, r_1^*, \ldots, r_n^*$ *using local Newton iteration as in Subsection 2.2.2.1. Then the number of iterations needed to find* $q^* r_1^*, \ldots, r_n^*$ *is asymptotically bounded by*
$$\mathcal{O}(\log\log(H^* E_0))$$
where $d = \deg(q)$, $H^* = max\{H(q^*), H(r_1^*), \ldots, H(r_n^*)\}$ *is the height of the output, and* $E_0 := d(2e)^{d/2} C^{d-1} \varepsilon_0$ *is the upper bound for the distance of the initial RUR from the exact RUR given in (2.4) for* $k = 0$.

2. *Assume that* $q^{(0)}, r_1^{(0)}, \ldots, r_n^{(0)}$ *is an initial approximate RUR which quadratically converges to the polynomials* $q^*, r_1^*, \ldots, r_n^*$*, but these polynomials do not form a RUR for a rational component of* $V(\mathscr{I})$*, i.e., either they have irrational coefficients, or they do not satisfy the properties in (2.8). In this case, we need*
$$\mathcal{O}\left(n\log(D) + \log(nh\delta) + \log\log(E_0 nd)\right)$$
iterations to conclude that our iteration did not converge to an exact RUR of a rational component of $V(\mathscr{I})$*. Here,* h *and* D *are the maximum of the logarithmic heights and degrees of the polynomials in* $\mathbf{f}$*, respectively,* δ *is the number of roots in* $V(\mathscr{I})$*, and* d, E_0 *are as above.*

Proof. 1. By Theorem 2.2.4, to successfully terminate the algorithm with the exact RUR we need to achieve accuracy of $\leq \frac{1}{2(H*)^2}$. Thus, using (2.4) with $E_0 := d(2e)^{d/2} C^{d-1} \varepsilon_0$, we need that
$$E_0 \left(\frac{1}{2}\right)^{2^k - 1} \leq \frac{1}{2(H*)^2},$$
or equivalently
$$k \geq \log_2 \log_2(4E_0(H^*)^2)$$
which is satisfied if
$$k \geq c_1 \log_2 \log_2(E_0 H^*)$$
for some constant $c_1 \leq 2$.

2. By Theorem 2.2.5, to terminate the algorithm we need to have
$$\log(B) \geq (4n^2 \log(nd) + nh)\delta^2 D^n,$$

where $B = \left\lceil E_0\left(\frac{1}{2}\right)^{2^k-1}\right\rceil^{-1/2}$ with $E_0 = d(2e)^{d/2}C^{d-1}\varepsilon_0$ as in (2.4). Thus we need that

$$-\frac{1}{2}\log(E_0) + \frac{2^k-1}{2}\log(2) \geq (4n^2\log(nd) + nh)\delta^2 D^n$$
$$(2^k-1)\frac{\log(2)}{2} \geq (4n^2\log(nd)+nh)\delta^2 D^n + \frac{1}{2}\log(E_0)$$

Here we use the inequality of arithmetic and geometric means, which is

$$\frac{x+y}{2} \geq \sqrt{xy} \Rightarrow \log(x+y) \geq \log(2) + \frac{\log(x)+\log(y)}{2}.$$

Now we go back to our inequality,

$$\begin{aligned}
2^k - 1 &\geq \log_2(e)[(8n^2\log(nd)+2nh)\delta^2 D^n + \log(E_0)] \\
k &\geq \log(\log_2(e)[(8n^2\log(nd)+2nh)\delta^2 D^n + \log(E_0)] + 1) \\
k &\geq \log_2(e)[\log(2) + 1 + \frac{1}{4}\log\log(E_0\sqrt{nd}) + \frac{n}{4}\log D + \frac{1}{8}\log(n^3 h\delta^4)] \\
k &\geq 1 + \frac{\log_2(e)}{4}(\log\log(E_0\sqrt{nd}) + n\log D + \log(\delta^2\sqrt{n^3h}))
\end{aligned}$$

This is satisfied if

$$k \geq 2\log_2(e)\big(n\log(D) + \log(nh\delta) + \log\log(E_0 nd)\big) + 1.$$

□

Finally we give an asymptotic bound for the computational time of the k-th iteration in the binary model.

Theorem 2.2.7. *Let $\mathbf{f} = (f_1,\dots,f_N) \in \mathbb{Q}[x_1,\dots,x_n]$ and $\mathscr{I}$ as above, and assume that polynomials in $\mathbf{f}$ has degrees at most D and logarithmic height at most h. Assume further that the maximal lengths of the arithmetic straightline programs over $\mathbb{Q}$ evaluating any of the polynomials in $\mathbf{f}$ is bounded by A. Let ε_0 be the maximal distance of the initial approximations of the roots computed by the homotopy method in Step (i). Let d be the number of roots of the component of $V(\mathscr{I})$ selected in Step (ii), and C be an upper bound for the absolute values of the coordinates of the roots in the component of $V(\mathscr{I})$ as in (2.5). Then, for each $k \geq 0$, the number of bit operations that the k-th iteration takes is asymptotically bounded by*

$$\tilde{\mathscr{O}}\big(2^{2k}nd\log^2(1/\varepsilon_0) + 2^k ND^2d^3hA\log(dC/\varepsilon_0) + 2^k dn^2Ah\log(1/\varepsilon_0)\big),$$

where $\tilde{\mathcal{O}}$ is the usual soft-O notation that ignores logarithmic factors.

Proof. We consider each step which is summarized in the introduction:

1. Construct numerically an approximate RUR for the approximate roots using formula (1.7) and (1.9). Let
$$L_k := (2^k - 1)\log_2\left(\frac{1}{\varepsilon_0}\right)$$
be the length of the floating point numbers in iteration k, using all correct digits of the coordinates of roots. Since the coefficients of u are rational and thus exact, we can assume that we turn them into floating point numbers of length L_k. Here, we assume that addition and multiplication of floating point numbers of length L takes $\mathcal{O}(L)$ binary operations. Complexity: Evaluation the formulas in (1.7) and (1.9) requires
$$\mathcal{O}\left(M(d)\log(d)L_k\right) \tag{2.12}$$
binary operations, using for example the [27][Algorithm 10.3], where $M(d)$ denotes the arithmetic complexity of the multiplication of two degree d polynomials. Note that $M(d) \in \tilde{\mathcal{O}}(d)$ (c.f. [27]).

2. Rational number reconstruction for each coefficients of the RUR. Complexity: The cost estimate that we use is from [27] for the Euclidean Algorithm in $\mathbb{Z}$, which gives the asymptotic bound for each rational number reconstruction to be $\mathcal{O}\left(\log^2(B)\right)$, and we have $d(n+1)$ coefficients to reconstruct.

   Here by Proposition 2.2.1, an upper bound for the absolute values of the coefficients is $E_0\left(\frac{1}{2}\right)^{2^k-1}$. Then by (2.7), we obtain $B = \frac{1}{\sqrt{2E_0(\frac{1}{2})^{2^k-1}}}$. Thus $B = \frac{2^{2^k-2}}{\sqrt{d(2e)^{d/2}C^{d-1}\varepsilon_0}}$ is an upper bound for the absolute values of the denominators of the coefficients. $\log B$ is approximately $2^k\log(1/\sqrt{\varepsilon_0})$, which implies $(\log B)^2 \leq 2^{2k}\log^2(1/\varepsilon_0)$. Since we have total of $d(n+1)$ coefficients, this gives a term
$$\mathcal{O}(2^{2k}nd\log^2(1/\varepsilon_0)) \tag{2.13}$$
in the overall complexity.
   We also need an upper bound for the logarithmic height $\hat{h}$ of the polynomials in the RUR that we compute in this step. We have
$$\hat{h} \in \mathcal{O}(2^k d\log(\frac{dC}{\varepsilon_0})), \tag{2.14}$$

since $d(2e)^{d/2}C^d$ is a bound for the absolute value of the coefficients of the polynomials in (1.7) and (1.9), which gives a bound for the numerators, and $B^d \leq 2^{d(2^k-2)}(1/\varepsilon_0)^{d/2}$ is a bound for the common denominator of the coefficients. Then the logarithmic height is less than $\log(d(2e)^{d/2}C^d 2^{d(2^k-2)}(1/\varepsilon)^{-d/2})$, which implies the asymptotic bound (2.14).

3. Check whether the computed exact polynomials form a proper RUR for a component of $V(\mathscr{I})$, using exact arithmetic over $\mathbb{Q}$. This involves the following steps:

   - Check if $\gcd(q,q')=1$: We can use the fast gcd algorithm in [27][Chap. 11] in complexity $\tilde{\mathscr{O}}(d^2+d(h+\hat{h}))$ (c.f. [27][Cor. 11.14]).
   - Check if $\lambda_1 r_1+\cdots+\lambda_n r_n \equiv Tq' \mod q$; the complexity dominated by the other checks.
   - Check if $f_i(r_1/q',\ldots,r_n/q') \equiv 0 \mod q$ for $i=1,\ldots,N$. Here we give a bound for evaluating the numerators of $f_i(r_1/q',\ldots,r_n/q')$, which have degrees at most $D(d-1)$ and logarithmic heights at most $Dh\hat{h}A$, and then reduce them modulo q. This can be done in complexity $\tilde{\mathscr{O}}(Nd^2D^2h\hat{h}A)$, using again the bounds in [27][Cor. 11.14].

   Substituting (2.14) gives bounds $d^2+dh+2^kd^2\log(\frac{Cd}{\varepsilon_0})$ and $ND^2dh2^kd^2\log(\frac{Cd}{\varepsilon_0})A$. Total checking process is dominated by

$$\tilde{\mathscr{O}}(2^kND^2d^3hA\log(\frac{Cd}{\varepsilon_0})). \tag{2.15}$$

4. One Newton iteration on each of the d approximate roots. Complexity: $\mathscr{O}(dn^2AhL_k)$ bit operation. As we substitute L_k, we have

$$\mathscr{O}(2^kdn^2A\log(1/\varepsilon_0)). \tag{2.16}$$

Finally, we note that the asymptotic bound given in the claim of the theorem dominates the bounds given in each steps. That total asymptotic bound is

$$\tilde{\mathscr{O}}\left(2^{2k}nd\log^2(1/\varepsilon_0)+2^kND^2d^3hA\log(Cd/\varepsilon_0)+2^kdn^2A\log(1/\varepsilon_0)\right).$$

□

2.3 An illustrative example

We illustrate our approach described in this chapter on a simple example. Maple code used for the example can be obtained from `www.math.ncsu.edu/~aszanto/code.html`.

To demonstrate the approach, consider the polynomial system

$$\mathbf{f}(x_1,x_2,x_3)=\begin{bmatrix} x_1^2+x_2^2-1 \\ 8x_1-16x_2^2+17 \\ x_1-x_2^2-x_3-1 \\ 64x_1x_2+16x_2 \end{bmatrix}.$$

It is easy to verify that **f** has two regular roots. A randomization of **f** consists of 3 quadratics which has 4 regular solutions, two of which can be shown to not correspond to roots of **f** using Subsection 2.2.1, part **(ii)**. We start with the following numerical approximations for the $d=2$ points of interest:

$$z_1=(-0.250,0.968,-2.188) \text{ and } z_2=(-0.250,-0.968,-2.188)$$

with error bound $\varepsilon=0.002$. From these numerical approximations, we see that we can take the primitive element to be $u=x_2$.

Using exact arithmetic, the initial RUR corresponding to this setup is

$$\begin{aligned} q(T)&=T^2-14641/15625,\\ r_1(T)&=121/250,\\ r_2(T)&=15/8,\\ r_3(T)&=-35T/8. \end{aligned}$$

At $k=1$, with denominator bound $B=16$ and error tolerance ε, we obtain

$$\begin{aligned} q(T)&=T^2-15/16,\\ r_1(T)&=-T/2,\\ r_2(T)&=15/8,\\ r_3(T)&=-35T/8. \end{aligned}$$

Since

$$\gcd(q,q')=1,$$

$$r_2\equiv Tq' \mod q(T)$$

$$\operatorname{numer}\left(\mathbf{f}\left(\frac{r_1}{q'},\frac{r_2}{q'},\frac{r_3}{q'}\right)\right)=\begin{bmatrix} -15(16T^2-15) \\ 15(16T^2-15) \\ -15(16T^2-15) \\ 0 \end{bmatrix}\equiv 0 \mod q(T)$$

we have proven that **f** has (at least) 2 roots which form a rational component. The corresponding well-constrained system from this RUR is

$$\begin{bmatrix} x_1+1/4 \\ x_2^2-15/16 \\ x_3+35/16 \end{bmatrix}.$$

CHAPTER

3

CERTIFYING SOLUTIONS TO SINGULAR POLYNOMIAL SYSTEMS OVER $\mathbb{Q}$

3.1 Introduction

3.1.1 Related Work

Consider a polynomial system $\mathbf{f} = (f_1, \dots, f_N) \in k[x_1, \dots, x_n]$, and assume that the ideal $\mathscr{I} := \langle f_1, \dots, f_N \rangle$ is zero dimensional. Let $n = N$, $(k+1)$-th Newton iterate of $\mathbf{f}$ be

$$z_{k+1} := z_k - J_{\mathbf{f}}(z_k)^{-1}\mathbf{f}(z_k)$$

with an initial guess z_0. Recall from Subsection 1.3 that a solution z^* of $\mathbf{f}$ is singular if the Jacobian matrix is singular at z^*, i.e., det $J_f(z^*) = 0$ or rank $\mathrm{J_f(z^*)} < \mathrm{N}$. The quadratic convergence property of Newton's Method does not hold for singular solutions. The rate of convergence is linear since the ratios

$$\|z_{k+1} - z^*\| / \|z_k - z^*\|$$

tend to the limit $\mu(\mathbf{f}, z^*)/(\mu(\mathbf{f}, z^*)+1)$ for $k = 0, 1, \dots$ where $\mu(\mathbf{f}, z^*)$ is the multiplicity of z^* as a root of $\mathbf{f}$. Moreover, it is usually difficult to obtain the singular solution as the same accuracy as a usual nonsingular solution (e.g., [23, 31]).
For example, let $f(x) = x^2$, then Newton's method converges to the origin linearly. The following classical example demonstrates a worse local behavior of Newton Method for singular systems:

Example 3.1.1. *[31] Let*

$$\mathbf{f}(x, y) = \begin{bmatrix} \frac{29}{16}x^3 - 2xy \\ y - x^2 \end{bmatrix},$$

only solution of $\mathbf{f}$ *is the origin. However, repeated application of Newton's method to the system* $\mathbf{f}$ *diverges starting at any point off of the line* $x = 0$ *other than the origin.*

If the given polynomial system $\mathbf{f}$ has singular roots, we cannot guarantee neither the quadratic convergence of Newton's Method nor obtaining a desired accuracy within a reasonable number of Newton iterations. Therefore α-theory (see Subsection 1.2.4) cannot be used to certify its solutions.

Deflation is a method to deal with non-reduced solution sets, means regularizing irreducible components of multiplicity greater than 1. Since Newton's method is only reliable on reduced solution sets, deflation is an important tool for certification purposes. The main idea of deflation for an isolated singular solution was introduced in [69, 68], which is basically differentiating the multiplicity away. In [57] and [55] symbolic and numerical algorithms presented for an isolated point that they showed terminated.

Before defining deflation processes, we need to introduce some basic notations and definitions. Consider a given system $\mathbf{g} = (g_1, \dots, g_N) \in \mathbb{Q}[x_1, \dots, x_n]$ and a root z in $V(g)$.

Let d be the dimension of the null space of the Jacobian matrix of $\mathbf{g}$ evaluated at z and it is denoted by

$$d = \mathrm{dnull}(\mathbf{g}, \mathrm{z}) := \dim \mathrm{null}\, \mathrm{J}_{\mathbf{g}}(\mathrm{z}).$$

Let

$$\ell = \binom{n}{n-d+1} \cdot \binom{N}{n-d+1}$$

and $\{\sigma_1, \dots, \sigma_\ell\}$ be the index set of all $(n-d+1) \times (n-d+1)$ submatrices of an $N \times n$ matrix. If $d = \max\{0, n-N\}$, then $\ell = 0$ in which case we know that z is a smooth point on an irreducible solution component of dimension d.

Definition 3.1.2. *Let* $\mathscr{D}$ *be the* **deflation operator** *such that*

$$\mathscr{D}(\mathbf{g}, z) := (\mathbf{g}_{\det}, z_{\det})$$

where $z_{\text{det}} = z$ *and*

$$\mathbf{g}_{\text{det}} = \begin{bmatrix} \mathbf{g} \\ \det J_{\sigma_1}\mathbf{g} \\ \vdots \\ \det J_{\sigma_\ell}\mathbf{g} \end{bmatrix},$$

the matrix $J_\sigma \mathbf{g}$ *is the submatrix of the Jacobian* $J\mathbf{g}$ *indexed by* σ.

Notice that $\mathbf{g}_{\text{det}}$ consists of polynomials in $\mathbb{Q}[x_1,\ldots,x_n]$. The construction guarantees that $\mathbf{V}(\langle \mathbf{g}_{\text{det}}\rangle) \subset \mathbf{V}(\langle \mathbf{g}\rangle)$, since there is no new variable in $\mathbf{g}$ but only new polynomials. Since the deflation operator $\mathscr{D}$ will be repeatedly applied, we write

$$\mathscr{D}^k(\mathbf{g}, z) := \mathscr{D}(\mathscr{D}^{k-1}(\mathbf{g}, z))$$

to mean k successive iterations with $\mathscr{D}^0(\mathbf{g}, z) := (\mathbf{g}, z)$.

Definition 3.1.3. *The* **deflation sequence** *of* $\mathbf{g}$ *at* z *is the sequence* $\{d_k(\mathbf{g}, z)\}_{k=0}^{\infty}$ *where*

$$d_k(\mathbf{g}, z) := \text{dnull}(\mathscr{D}^{\text{k}}(\mathbf{g}, \text{z})) \;\; \textit{for} \;\; \text{k} \geq 0.$$

The dimension of the null space cannot increase since we add more polynomials to $\mathbf{g}$. Hence, the deflation sequence is a nonincreasing sequence of integers greater than or equal 0. Then the deflation sequence must reach its limit after finitely many iterations. The following theorem, guarantees the termination of deflation.

That is, there are integers $d_\infty(\mathbf{g}, z) \geq 0$ and $s \geq 0$ so that $d_t(\mathbf{g}, z) = d_\infty(\mathbf{g}, z)$ for all $t \geq s$. When z is isolated, s is bounded above by the depth as well as multiplicity [22, 38, 57]. The limit $d_\infty(\mathbf{g}, z)$ is called the *isosingular local dimension* of z with respect to $\mathbf{g}$. The *isosingular points* are those for which their isosingular local dimension is zero so that, after finitely many iterations, isosingular deflation has regularized the root, i.e., constructed a polynomial system for which the point is a regular root. Clearly, such a system must consist of at least n polynomials, but will typically be overdetermined.

Now we will illustrate the idea of isosingular deflation with a very basic example from [7], Section 13.2.

Example 3.1.4. *Let* $f(x, y, z) = x^2 + y^2 + z^2$ *, then* $\xi = (0,0,0)$ *is a root with multiplicity* $\mu(f, \xi) = 2$. *We will deflate* f *at the origin* ξ *using isosingular deflation.* $J_f(x, y, z) = [2x \;\; 2y \;\; 2z]$ *and* $J_f(\xi)$ *is the* *zero matrix, thus we add three* 1×1 *minors of* $J_f(x, y, z)$ *to our given system* f.

$$\mathbf{g}(x, y, z) = \begin{bmatrix} x^2 + y^2 + z^2 \\ 2x \\ 2y \\ 2z \end{bmatrix}$$

We have $\xi = (0, 0, 0)$, *and* $\mu(\mathbf{g}, \xi) = 1$, *so it is nonsingular with respect to* $\mathbf{g}$.
Now let us find its deflation sequence, $d_0(\mathbf{f}, \xi) = \dim \operatorname{null}(\mathscr{D}^0(\mathrm{f}, \xi)) = \dim \operatorname{null}(\mathbf{g}, \xi) = 3$ *and* $d_1(\mathbf{f}, \xi) = d_2(\mathbf{f}, \xi) = \cdots = 0$. *Thus the deflation sequence of the origin is* $3, 0, 0, \ldots$. *That implies we deflate the given polynomial only after one isosingular deflation iteration, since it reaches its limit at the second deflation.*

3.1.2 Our Approach

We consider the problem of certifying isolated singular roots of a rational polynomial system. Due the behavior of Newton's method near singular roots, standard techniques in α-theory can not be applied to certify such roots even if the polynomial system is well-constrained. The key tool to handle such multiple roots is called *deflation* (see Subsection 1.3). Deflation techniques "regularize" the system thereby creating a new polynomial system which has a simple root corresponding to the multiple root of the original system [69, 68, 22, 57, 58, 38]. In this work, we will focus on using a determinantal form of the *isosingular deflation* [38], in which one simply adds new polynomials to the original system without introducing new variables. The new polynomials are constructed based on exact information that one can obtain from a numerical approximation of the multiple root. In particular, if the original system had rational coefficients, the new polynomials which remove the multiplicity also have rational coefficients. Thus, using this technique, we have reduced the given system to the case of an overdetermined system over $\mathbb{Q}$ in the original set of variables that has a simple root.

As summarized in Section 1.3 we will use isosingular deflation to certify isolated singular roots of a rational polynomial system. The problem will be reduced to the case of certifying simple roots to overdetermined systems, which was discussed in Chapter 2. Due to this reduction, we can extend this approach to all points which can be regularized by isosingular deflation, called *isosingular points*. Since every isolated multiple root is an isosingular point, this method applies to multiple roots. However, isosingular points need not be isolated as it will be demonstrated in the following example by the origin with respect to the Whitney umbrella $x^2 - y^2 z = 0$.

Example 3.1.5. *As a demonstration, consider the Whitney umbrella defined by* $\mathbf{g}(x,y,z)=x^2-y^2z$. *Following ([38, Example 5.12]), the deflation sequence for the origin is* $\{3,2,0,0,\dots\}$ *showing that the origin is not isolated but is an isosingular point. In particular, it takes two iterations to construct a polynomial system for which the origin is a regular root. Since* $J_{\mathbf{g}}(0)$ *is identically zero, the first iteration appends all partial derivatives, say*

$$\mathbf{g}'(x,y,z)=\begin{bmatrix}\mathbf{g}(x,y,z)\\ x\\ yz\\ y^2\end{bmatrix}.$$

Since $J'_{\mathbf{g}}(0)$ *has rank 1, the original formulation of* $\mathscr{D}_{\text{det}}$ *will append* 18 2×2 *minors of* $J'_{\mathbf{g}}$. *However, with our modification, we only need to add the* 6 *minors which arise by submatrices that, in this case, include the unique nonzero element of* $J'_{\mathbf{g}}(0)$. *For this example, it is easy to verify that the ideal of the resulting regularizing polynomial system is equal to* $\langle x,y,z\rangle$.

Since the resulting polynomial systems have rational coefficients, it immediately follows that every point in a zero-dimensional rational component must have the same deflation sequence. This can be used to partition the set of points under consideration into subsets and run the certification procedure described in Chapter 2 independently on each subset.

The construction of the deflation sequence and the resulting regularized system is an exact process that depends upon z. In our situation where z is only known approximately, we use the numerical approximations to compute exact numbers, namely the nonnegative integers arising as the dimensions of various linear subspaces which form the deflation sequence.

Remark 3.1.6. *One drawback with the deflation* $\mathscr{D}$ *is the number of minors used in each iteration, namely* $\ell=\binom{n}{n-d+1}\cdot\binom{m}{n-d+1}$. *Since the codimension of the set of* $m\times n$ *matrices of rank* $n-d$ *is* $c=d(m+d-n)$, *we will adjust* $\mathscr{D}_{\text{det}}$ *to use exactly* c *minors as follows. Since* $d=\text{dnull}(\mathbf{g},\text{z})$, *we can select an invertible* $(n-d)\times(n-d)$ *submatrix of* $J_{\mathbf{g}}(z)$. *Rather than using all of* $\{\sigma_1,\dots,\sigma_\ell\}$, *we only use the* c *many which contain our selected invertible submatrix. In particular, with this setup, the tangent space of these* c *many minors is equal to the tangent space of all* ℓ *minors at* z.

With this specialized construction, one now needs to be cautious that two points with the same deflation sequence can fail to be regularized by the system constructed by the other. However, all points in the same zero-dimensional rational component will still be regularized simultaneously. In particular, by comparing ranks of various submatrices, one may be able to produce a finer partition of the points under consideration before independently certifying each collection.

3.2 Certification

Given $\mathbf{f} = (f_1, \ldots, f_N) \in \mathbb{Q}[x_1, \ldots, x_n]$ and a subset $V \subset V(\mathbf{f})$ consisting of isosingular points, the process for certification proceeds as follows.

1. Deflation sequences. Compute the deflation sequences for each of the points in V. If each point is an isosingular point, then isosingular deflation will terminate and produce a regularized system for each point. If one is not an isosingular point, one can apply the tests developed in [38, Section 6] to determine that the sequence has stabilized with the point having a positive isosingular local dimension. Remove all such points from V and partition the remaining points based on their deflation sequences and common regularizing polynomial systems, say $V_1, \ldots, V_k$.

2. Certify each V_i. Associated with each V_i is a polynomial system $\mathbf{f}^{(i)}$ having rational coefficients that must be either well-constrained or overdetermined. If it is well-constrained, simply apply standard α-theoretic techniques for certification. If overdetermined, use the approach presented in Chapter 2 for certification.

Successfully completing the certification proves that the points under consideration are indeed isosingular points of **f**, i.e., the isosingular local dimension is zero. However, as currently formulated, this does not yield any information about the embedded dimension of the points in the original system, e.g., deciding if the point is isolated or not. Furthermore, even if one knows that a given point is isolated, this approach currently does not yield information about its multiplicity. The latter problem was addressed in [35].

3.3 Examples

We illustrate our approach described in Chapter 2 and Chapter 3 on several examples. Maple code used for these examples can be obtained from `www.math.ncsu.edu/~aszanto/code.html`.

3.3.1 An illustrative example

To demonstrate the approach, consider the polynomial system

$$\mathbf{g}(x_1, x_2, x_3) = \begin{bmatrix} x_1^2 + x_2^2 - 1 \\ 8x_1 - 16x_2^2 + 17 \\ x_1 - x_2^2 - x_3 - 1 \end{bmatrix}.$$

It is easy to verify that **g** has two roots of multiplicity 2. Thus, after appending $\det J\mathbf{g}$, we are interested in the overdetermined polynomial system

$$\mathbf{f}(x_1, x_2, x_3) = \begin{bmatrix} \mathbf{g}(x_1, x_2, x_3) \\ 64x_1x_2 + 16x_2 \end{bmatrix}$$

which has two regular roots.
The rest of the example is the same as Example 2.3.

3.3.2 Caprasse System

A common benchmark system is the Caprasse system which is a well-constrained system with regular and multiple roots. The system, presented below, has 24 regular roots and 8 roots of multiplicity four [66].

$$\mathbf{g} = \begin{bmatrix} x_1^3x_3 - 4x_1^2x_2x_4 - 4x_1x_2^2x_3 - 2x_2^3x_4 - 4x_1^2 - 4x_1x_3 + 10x_2^2 + 10x_2x_4 - 2 \\ x_1x_3^3 - 4x_1x_3x_4^2 - 4x_2x_3^2x_4 - 2x_2x_4^3 - 4x_1x_3 + 10x_2x_4 - 4x_3^2 + 10x_4^2 - 2 \\ 2x_1x_2x_4 + x_2^2x_3 - 2x_1 - x_3 \\ x_1x_4^2 + 2x_2x_3x_4 - x_1 - 2x_3 \end{bmatrix}.$$

Since the system is well-constrained, numerical approximations for the 24 regular roots can be certified using standard α-theory. Here, we consider certifying the multiple roots. At each of these multiple roots, $J_\mathbf{g}$ has rank 2 with the lower right 2×2 block having full rank. Thus, we consider the system **f** constructed by appending the four 3×3 minors of $J_\mathbf{g}$ containing the lower right block to **g**.

From the numerical approximations of the 8 points z_i that we computed using `Bertini` [6], we see that $u = x_1 - x_2 + 2x_3 - 2x_4$ is a primitive element. Starting the numerical approximations correct to 10 digits, we obtain the following RUR after one Newton iteration:

$$\begin{aligned} q(T) &= 1/3(T^2+3)(3T^2+1)(T^2-12T+39)(T^2+12T+39) \\ r_1(T) &= 6240 + 1568T^2 - (6176/3)T^4 + (160/3)T^6 \\ r_2(T) &= 1560 - 3688T^2 - (1256/3)T^4 - (40/3)T^6 \\ r_3(T) &= -9984 - 13952T^2 - (4096/3)T^4 + (128/3)T^6 \\ r_4(T) &= -1560 + 3688T^2 + (1256/3)T^4 + (40/3)T^6 \end{aligned}$$

This RUR shows that the 8 roots arise from 4 rational components, each of degree 2, with splitting field $\mathbb{Q}[\sqrt{3}]$.

We also experimented with starting the computation using approximate roots that are less accurate. Our experiment indicated that we needed to start with points that were accurate to at least 6 digits to find the exact RUR in one iteration. Using the same u, if we started with approximate roots that had 4, 3, or 2 correct digits, then we needed 2, 2, and 3 Newton iterations, respectively.

We note that as long as we are guaranteed that our initial numerical approximations are in the basins of quadratic convergence, regardless of the number of correct digits in the initial data, we will eventually reach an exact RUR. The initial accuracy only influences the number of additional Newton iterations that is needed to reach the required accuracy that provides an exact RUR.

We also compared the sizes of a PUR and the above RUR for this example. With the same setup, obtain the following exact PUR:

$$
\begin{aligned}
q(T) &= 1/3(T^2+3)(3T^2+1)(T^2-12T+39)(T^2+12T+39)\\
v_1(T) &= -(1709/4874688)T^7+(44779/2089152)T^5-(295969/696384)T^3-(3483881/1624896)T\\
v_2(T) &= -(1529/4874688)T^7+(42151/2089152)T^5-(299149/696384)T^3-(1861229/1624896)T\\
v_3(T) &= (1619/4874688)T^7-(43465/2089152)T^5+(297559/696384)T^3+(3485003/1624896)T\\
v_4(T) &= (1529/4874688)T^7-(42151/2089152)T^5+(299149/696384)T^3+(1861229/1624896)T.
\end{aligned}
$$

Clearly, the modular division by q' significantly increases the size of the coefficients in the PUR, compared to the RUR above.

3.3.3 Two cyclic systems

A common family of benchmark examples are the cyclic-n systems [11]. For $n \geq 2$, the cyclic-n system is

$$
\mathbf{f}_n = \left[\begin{array}{ll} \sum_{j=1}^{n}\prod_{k=1}^{D} x_{j+k} & \text{for } D = 1,\ldots,n-1 \\ \prod_{k=1}^{n} x_k - 1 & \end{array}\right]
$$

where $x_{n+\ell} = x_\ell$ for all $\ell = 1,\ldots,n$.

Below, we demonstrate our approach on the instances $n = 4$ and $n = 9$.

For $n = 4$, the solutions of $\mathbf{f}_4 = 0$ lie on two irreducible curves, with 8 embedded points which are isosingular points. We deflate these points simultaneously by appending the four 3×3 minors of $J_{\mathbf{f}_4}$ containing the first and last rows, and second and third columns. By using numerical approximations computed by `Bertini`, we see that we can use the primitive element $u = x_1 + 2x_2 - x_3 + 3x_4$. Starting the numerical approximations correct to 5 digits, this yields the following certified exact RUR after one Newton iteration:

$$\begin{aligned} q(T) &= (T-1)(T+1)(T-3)(T+3)(T^2+1)(T^2+9) \\ r_1(T) &= 16(7T^4-27) \\ r_2(T) &= 8(-13T^4-27) \\ r_3(T) &= 16(-7T^4+27) \\ r_4(T) &= 8(13T^4+27) \end{aligned}$$

Now, for $n=9$, we consider the overdetermined system

$$\mathbf{f}(x) = \begin{bmatrix} \mathbf{f}_9(x) \\ x_1 - x_4 \\ x_1 - x_7 \end{bmatrix}$$

motivated by Example 9 of [22, Section 7]. The solutions of this system provide witness points on some of the dimension 2 components of the cyclic-9 system $\mathbf{f}_9(x)$.

The degree of the ideal generated by **f** is 162 with `Bertini` computing 54 regular points and 54 double points. The following uses the primitive element $u = x_1 + 2x_2 - x_3 + 2x_5 + x_6 - x_8$.

For the 54 regular points, we compute

$$\begin{aligned} q(T) = \; & (T^2+T-101)(T^4-T^3+102T^2+101T+10201) \\ & (T^2+19T+79)(T^4-19T^3+282T^2-1501T+6241) \\ & (T^2-17T+61)(T^4+17T^3+228T^2+1037T+3721) \\ & (T^{12}-2356T^9+5057697T^6-1161599884T^3+243087455521) \\ & (T^{12}-304T^9+1122717T^6+313211504T^3+1061520150601) \\ & (T^{12}+1802T^9+3020223T^6+409019762T^3+51520374361), \end{aligned}$$

and we do not display here $r_i(T)$ for $i = 1, \ldots, 9$ due to space. We obtained this result using 70 digit floating-point arithmetic to help control the round-off errors of the computation with the rational number reconstruction succeeding when the approximate roots were accurate to 14 digits. This RUR computation proves that the overdetermined system **f**, has at least 54 regular roots which decompose into 9 rational components: three each of degree 2, 4, and 12. Hence, this prove that there are at least 54 cyclic-9 roots with $x_1 = x_4 = x_7$.

For the 54 double points, by comparing ranks of 8×8 submatrices of $J_{\mathbf{f}}$, we are able to partition into 3 subsets of size 18. For each of these subsets, we added one polynomial arising from the isosingular deflation to **f**. The results of the RUR computation produced the following polynomials $q(T)$:

$$\begin{aligned} q(T) &= (T^2-11T+19)(T^4+11T^3+102T^2+209T+361) \\ &\quad (T^{12}+704T^9+488757T^6+4828736T^3+47045881) \end{aligned}$$

$$\begin{aligned} q(T) &= (T^2+13T+31)(T^4-13T^3+138T^2-403T+961) \\ &\quad (T^{12}-988T^9+946353T^6-29433508T^3+887503681) \end{aligned}$$

$$\begin{aligned} q(T) &= (T^2+T-11)(T^4-T^3+12T^2+11T+121) \\ &\quad (T^{12}-34T^9+2487T^6+45254T^3+1771561). \end{aligned}$$

Just as above, we do not list the corresponding polynomials $r_i(T)$ due to space and the computations used 70 digit floating-point arithmetic with the the rational number reconstruction succeeding when the approximate roots were accurate to 14 digits. This RUR computation proves that **f** has at least 54 singular roots which are isosingular points and decompose into 9 rational components: three each of degree 2, 4, and 12. Hence, this prove that there are at least 54 singular cyclic-9 roots with $x_1 = x_4 = x_7$.

3.3.4 A family with clustered roots

We close with a family of overdetermined polynomial systems that have a cluster of roots near the origin. As noted above, as long as we are given a numerical approximation in each quadratic convergence basin of each root, the algorithm is insensitive to the distance between the roots, but sensitive to the input and output sizes.

For example, for any nonzero M, we consider the system

$$\mathbf{f}_M = \begin{bmatrix} Mx-1 \\ 3xy^2+8x^2y+6xz^2-3yz^2+3y^2z+4x^3+4z^3-6xyz \\ 10xz^2-6x^2y-xy^2+6x^2z+yz^2+3y^2z-8x^3-10z^3+3xyz \\ 4xy^2+4x^2y-8xz^2+18x^2z+2yz^2-8x^3+4z^3-5xyz \\ 4xy^2-9x^2y+4x^2z-4yz^2+10x^3-8z^3+4xyz \end{bmatrix}.$$

For generic $M \in \mathbb{C}$, a randomization of $\mathbf{f}_M$ has 14 roots, but $\mathbf{f}_M$ has only 5 roots. The 5 roots of $\mathbf{f}_M$ converge to the origin as $|M| \to \infty$. Thus, for large M, the 5 roots of $\mathbf{f}_M$ are clustered around the origin.

We consider 3 cases of M, namely $M = 10, 10^3, 10^5$, for which the five roots of $\mathbf{f}_M$ are all contained in a ball centered at the origin of radius $3.7/M$. Using the primitive element $u = y$, we obtain the following polynomials $q(T)$ in an exact RUR for these choices of M:

$$\begin{aligned}
&M = 10:\\
q(T) &= (9900000T^5 - 2062500T^4 + 279875T^3 - 180025T^2 + 12105T - 2351)/9900000\\
&M = 10^3:\\
q(T) &= (99000000000000000T^5 - 206250000000000T^4 + 279875000000T^3 - 1800250000T^2\\
&\quad + 1210500T - 2351)/99000000000000000\\
&M = 10^5:\\
q(T) &= (990000000000000000000000000T^5 - 20625000000000000000000T^4\\
&\quad + 279875000000000000T^3 - 18002500000000T^2 + 121050000T\\
&\quad - 2351)/990000000000000000000000000
\end{aligned}$$

In our computations, we needed the roots to be approximated to at least 15, 35, and 55 digits, respectively, in order to obtain the RUR to certify the corresponding 5 roots of $\mathbf{f}_M$.

CHAPTER

4

CERTIFYING REAL SOLUTIONS OF POLYNOMIAL SYSTEMS

4.1 Introduction

Consider polynomial system $\mathbf{f} = (f_1, \dots, f_N) \in \mathbb{K}[x_1, \dots, x_n]$. If $N = n$ approximate roots of the system can be certified with α-theory as explained in Section 1.2. Let $\mathbb{K} = \mathbb{Q}$, and $\mathbf{f}$ be overdetermined, i.e., $N > n$, assuming the ideal $\mathscr{I} := \langle f_1, \dots, f_N \rangle$ is radical and zero dimensional, the system has a well-constrained Rational Univariate Representation (RUR). In this case, the method introduced in Chapter 2 can be used to certify solutions of $\mathbf{f}$ over rationals. If the given rational polynomial system is singular, then one can use the isosingular deflation method given in Chapter 3 to regularize isolated singular solutions. The deflated system is most likely an overdetermined system with additional new polynomials. Therefore the method which is detailed in Chapter 2 can certify the approximate solutions.

Alternatively using Hermite matrices, one can certify real roots of $\mathbf{f}$ within a neighborhood. The goal of this chapter is to present the details of this method, both in the univariate and multivariate cases. The definition of the fundamental terms used here can be found in Section 1.5. We mainly follow the notation and approach from [19], [5] and [70].

4.2 Hermite matrices

We begin with Hermite's bilinear map and Hermite's quadratic form, then we point how it is related to real root locating.

Definition 4.2.1. *The Hermite's bilinear map :*

$$\begin{aligned} H_g(\mathscr{I})\colon \mathbb{R}[x_1,\dots,x_n]/\mathscr{I} &\to \mathbb{R} \\ (p,q) \quad &\mapsto Tr(M_{pgq}) \end{aligned}$$

where $\mathscr{I} =< f_1,\dots,f_N >$ is a zero dimensional ideal with $\mathbf{f} = (f_1,\dots,f_N)$ and g are in $\mathbb{R}[x_1,\dots,x_n]$. M_{pgq} defines the multiplication map by pgq as in Definition 1.5.11.

If we fix a basis for the finite dimensional vector space $\mathbb{R}[x_1,\dots,x_n]/\mathscr{I}$, the Hermite matrix is the symmetric matrix corresponding to Hermite's bilinear form with respect to this basis. Thus, We can define a Hermite matrix as the following way. Note that by abuse of notation both the bilinear map and its matrix denoted by $H_g(\mathscr{I})$.

Definition 4.2.2. *Let $\mathbf{f} = (f_1,\dots,f_N)$, f_i and $g \in \mathbb{R}[x_1,\dots,x_n]$ for all $i = 1,\dots N$, and $\mathscr{I} =< f_1,\dots,f_N >$ be a zero dimensional ideal, and $\mathscr{B} = \{x^{\alpha_1},\dots,x^{\alpha_D}\}$ be a monomial basis (normal set) of $\mathbb{R}[x_1,\dots,x_n/\mathscr{I}$ then*

$$[H_g(\mathscr{I})]_{i,j} := Tr(M_{x^{\alpha_i} g x^{\alpha_j}})$$

where M_p is the multiplication matrix of p with respect to the basis $\mathscr{B}$.

Observation 4.2.3. *Let $\mathbf{f} = (f_1,\dots,f_N)$ and g be in $\mathbb{R}[x_1,\dots,x_n]$ and $\mathscr{I} =< f_1,\dots,f_N >$ be a zero dimensional ideal, then the Hermite matrix $H_g(\mathscr{I})$ is a real and symmetric matrix. Moreover, if $\mathbf{f}$ and g are in $\mathbb{Q}[x_1,\dots,x_n]$, then the Hermite matrix $H_g(\mathscr{I})$ is a symmetric matrix with rational entries. In these cases, the eigenvalues of the Hermite matrix $H_g(\mathscr{I})$ are real by Theorem 1.5.3.*

Let $\mathbf{f} = (f_1,\dots,f_N) \in \mathbb{R}[x_1,\dots,x_n]$ and $\mathscr{B} = \{x^{\alpha_1},\dots,x^{\alpha_D}\}$ be a monomial basis for $\mathbb{R}[x_1,\dots,x_n]/\mathscr{I}$ where $\mathscr{I} =< f_1,\dots,f_N >$. Then Hermite Matrices can be expressed in terms of the Vandermonde matrix with respect to basis $\mathscr{B}$ as follows:

Theorem 4.2.4. *Let $\mathbf{f} = (f_1,\dots,f_N), g \in \mathbb{R}[x_1,\dots,x_n]$ and $\mathscr{I} =< f_1,\dots,f_N >$ be a zero dimensional ideal. Let $\xi_1,\xi_2,\dots,\xi_m$ be the common roots of $\mathbf{f}$ such that $\xi_i \in \mathbb{C}^n$, for $i = 1\dots m$ (here each root listed as many times as their multiplicity). Then the Hermite matrix of $\mathscr{I}$ with respect to g is*

$$H_g(\mathscr{I}) := V^T G V \tag{4.1}$$

where $V = V_{\mathscr{B}}(\xi_1, \xi_2, \ldots, \xi_m)$ is the Vandermonde matrix of $\xi_1, \xi_2, \ldots, \xi_m \in \mathbb{C}^n$ with respect to a monomial basis $\mathscr{B}$ and G is an $m \times m$ diagonal matrix with $[G]_{ii} = g(\xi_i)$ for $i = 1, \ldots, m$.

Proof. Let $\mathbf{f} = (f_1, \ldots, f_N), g \in \mathbb{R}[x_1, \ldots, x_n]$ and $\mathscr{I} =< f_1, \ldots, f_N >$ be a zero dimensional ideal. 4.2.2 defines the entries of the Hermite matrix $H_g(\mathscr{I})$ by

$$[H_g(\mathscr{I})]_{i,j} = Tr(M_{x^{\alpha_i} g x^{\alpha_j}}), \tag{4.2}$$

where M_p is the multiplication matrix of p with respect to the monomial basis $\mathscr{B} = \{x^{\alpha_1}, \ldots, x^{\alpha_D}\}$. Then we yield the following by Theorem 1.5.15,

$$[H_g(\mathscr{I})]_{i,j} = \sum_{\xi \in V_{\mathbb{C}}(\mathscr{I})} \mu(\xi) \xi^{\alpha_i} g(\xi) \xi^{\alpha_j}, \tag{4.3}$$

where $\mu(\xi)$ is the multiplicity of ξ. Now we can write the left hand side in a matrix form with the Vandermonde matrix $V = V_{\mathscr{B}}(\xi_1, \xi_2, \ldots, \xi_m)$ of $\xi_1, \xi_2, \ldots, \xi_m \in V_{\mathbb{C}}(\mathscr{I})$ with respect to the monomial basis $\mathscr{B}$ by

$$H_g(\mathscr{I}) := V^T G V,$$

where G is an $m \times m$ diagonal matrix with $[G]_{ii} = g(\xi_i)$ for $i = 1, \ldots, m$.

□

Definition 4.2.5. *Let A be a real and symmetric matrix, then the* **signature** *of A is*

$$\sigma(A) := (\#of\ \ positive\ \ eigenvalues) - (\#of\ \ negative\ \ eigenvalues)$$

In 1856, Hermite [41] stated the following theorem in univariate case.

Theorem 4.2.6 (Hermite Theorem)**.** *Let $f, g \in \mathbb{R}[x]$,*

$$\sigma(H_g(f)) = N_+ - N_-.$$

where $N_+ := \#\{f \in \mathbb{R} \mid f(x) = 0$ and $g(x) > 0\}$ and $N_- := \#\{f \in \mathbb{R} \mid f(x) = 0$ and $g(x) < 0\}$

Proof. Define a quadratic form(see Definition1.5.18) associated to the symmetric matrix $H_g(f)$,

$$a^T H_g(f) a := \sum_{\xi \in V(f)} \mu(\xi) g(\xi) (a_{k-1}\xi^{k-1} + \ldots + a_1 \xi + a_0)^2$$

where $a(\xi) := (a_{k-1}\xi^{k-1} + \ldots + a_1 \xi + a_0)$ is a linear form, for all $\xi \in V(f)$. We can separate the sum of real and non-real roots of f as

$$a^T H_g(f)a = \sum_{\xi\in\mathbb{R}} \mu(\xi)g(\xi)a(\xi)^2 + \sum_{\xi,\xi^*\in\mathbb{C}-\mathbb{R}} \mu(\xi)(g(\xi)a(\xi)^2 + g(\xi)a(\xi^*)^2). \tag{4.4}$$

We can write $\mu(\xi)g(\xi) = (\alpha(\xi) + i\ \beta(\xi))^2$ where $\alpha(\xi), \beta(\xi) \in \mathbb{R}$. Then we have

$$a'(\xi) \quad = \quad \sum_{i=1}^{k}(\alpha(\xi)Re(\xi^i) - \beta(\xi)Im(\xi^i))a_i, \tag{4.5}$$

$$a''(\xi) \quad = \quad \sum_{i=1}^{k}(\alpha(\xi)Im(\xi^i) + \beta(\xi)Re(\xi^i))a_i, \tag{4.6}$$

such that

$$\mu(\xi)(g(\xi)a(\xi)^2 + g(\xi^*)a(\xi^*)^2) = 2a'(\xi)^2 - 2a''(\xi)^2.$$

Since linear forms in 4.4 associated to the real roots, a' and a'' are linearly independent, thus by Sylvester's law of inertia (Theorem 1.5.19),

$$\sigma(H_g(f)) = \sigma\left(\sum_{\xi\in\mathbb{R}} \mu(\xi)g(\xi)a(\xi)^2\right)$$

with linearly independent linear forms $a(\xi)$. We have

$$\sigma(H_g(f)) = \sum_{\xi\in\mathbb{R}, f(\xi)=0} \text{sign}(g(\xi)).$$

□

More detailed proof and further readings can be found on [5].

The classical univariate Hermite theorem is generalized to the multivariate case by Pedersen, Roy and Szpirglas [70], this theorem also proved in [5], [19]:

Theorem 4.2.7 (Multivariate Hermite Theorem)**.** *Let* $\mathbf{f} = (f_1, \ldots, f_N)$*,* $f_i \in \mathbb{R}[x_1, \ldots, x_n]$ *for all* $i = 1, \ldots, N$ *and* $g \in \mathbb{R}[x_1, \ldots, x_n]$*, with* $\mathscr{I} =< f_1, \ldots, f_N >$ *is a zero dimensional ideal. Let* $H_g(\mathscr{I})$ *be the Hermite matrix of* $\mathscr{I}$ *with respect to* g*, then*

$$\sigma(H_g(\mathscr{I})) = N_+ - N_-$$

where $N_+ := \#\{x \in \mathbb{R}^n \mid \mathbf{f}(x) = 0 \text{ and } g(x) > 0\}$ *and* $N_- := \#\{x \in \mathbb{R}^n \mid \mathbf{f}(x) = 0 \text{ and } g(x) < 0\}$.

Proof. Let u be a primitive element as Definition 1.1.15. The elements $1, u, \dots, u^{k-1}$ are linearly independent in $\mathbb{R}[x_1,\dots,x_n]/\mathscr{I}$. $\{1, u, \dots, u^k\}$ can be completed to a basis $B = \{\omega_1 = 1, \omega_2 = u, \dots, \omega_k = u^{k-1}, \omega_k, \dots, \omega_D\}$ of $\mathbb{R}[x_1,\dots,x_n]/\mathscr{I}$. For any $p \in \mathbb{R}[x_1,\dots,x_n]/\mathscr{I}$, $p = p_1 + p_2\omega_2 + \cdots + p_D\omega_D$. By the definition of the Hermite bilinear map and the Stickelberger's theorem (see Theorem 1.5.15),

$$[H_g(\mathscr{I})]_{i,j} = Tr(M_{\omega_i g \omega_j}) = \sum_{\xi \in V_{\mathbb{C}}(\mathscr{I})} \mu(\xi)\omega_i(\xi)g(\xi)\omega_j(\xi) \qquad (4.7)$$

where $\mu(\xi)$ be the multiplicity of ξ as a root of $\mathbf{f}$. Then the same argument used in univariate Hermite theorem (Theorem 4.2.6) applies. Since complex roots appear as conjugates, the trace is a real number. The signature of Hermite matrix only require the summation of signs related to the real roots.

□

Remark 4.2.8. *Let $\mathbf{f} = (f_1,\dots,f_N)$, $f_i, g \in \mathbb{R}[x]$ for all $i = 1,\dots,N$ with a zero dimensional ideal $\mathscr{I} =< f_1,\dots,f_N >$. Consider a Hermite matrix $H_g(\mathscr{I})$ (as in 4.2) where defined by univariate polynomials $\mathbf{f}, g$ with all distinct roots $\xi_1,\dots,\xi_k \in \mathbb{C}$ (i.e., we basically omit the multiplicities) of $\mathbf{f}$.*

- *Let $g(x) = 1$, then the entries of the Hermite matrix are*

$$[H_1(\mathscr{I})]_{i,j} = \sum_{l=1}^{k} \xi_l^{i+j-2}, \qquad (4.8)$$

which is equal to the $(i+j-2)$-th Newton power sum (see 1.5.17).

- *The Hermite matrix $H_g(\mathscr{I})$ is a real Hankel matrix.*
  $H_g(\mathscr{I})$ is a Hankel matrix since $[H_g(\mathscr{I})]_{i,j} = [H_g(\mathscr{I})]_{i+1,j-1}$. $H_g(\mathscr{I})$ is also a real matrix. First, notice that $g(\overline{\xi}) = \overline{g(\xi)}$ since $(\overline{\xi})^n = \overline{\xi^n}$ and $g(x) \in \mathbb{R}[x]$. For any real solution ξ_p, $g(\xi_p)$ will remain real. For a complex solution ξ_p, its conjugate $\overline{\xi_p}$ will also be a solution. Since each entry of the Hermite matrix requires addition of all roots, the sum of a conjugate pair

$$\xi_p^{i+j-2} g(\xi_p) + (\overline{\xi_p})^{i+j-2} g(\overline{\xi_p}) = \xi_p^{i+j-2} g(\xi_p) + \overline{(\xi_p^{i+j-2})}\overline{g(\xi_p)}$$

is also real.

4.3 Root Certification

Let $\mathbf{f} = (f_1,\dots,f_N) \in \mathbb{R}[x_1,\dots,x_n]$ and $\mathscr{I} =< f_1,\dots,f_N >$ is a zero dimensional ideal. We are given $z^* \in \mathbb{R}^n$, we would like to know if there is any exact root of $\mathbf{f}$ within the ϵ neighborhood of $z^* \in \mathbb{R}^n$. In order to answer this question we will use two auxiliary functions, those will allow us to use Hermite matrices for certification of the given approximate solution. First auxiliary function we will use is $g(x) = 1$ for all $x \in \mathbb{R}$.

Proposition 4.3.1. *Let* $\mathbf{f} = (f_1, \ldots, f_N) \in \mathbb{R}[x_1, \ldots, x_n]$, $\mathscr{I} =< f_1, \ldots, f_N >$ *and* $g(x) = 1$, *then*

$$\sigma(H_1(\mathscr{I})) = \#\{z \in \mathbb{R} \mid \mathbf{f}(z) = 0\}$$

where $\sigma(.)$ *is the signature as defined in Definition 4.2.5.*

Proof. By the Hermite theorem, $\sigma(H_1(\mathscr{I})) = N_+ - N_-$. $N_- = 0$ since there is no $z \in \mathbb{R}$ such that $g(z) < 0$. We have $\sigma(H_1(z_1, \ldots, z_k)) = \#\{z \in \mathbb{R}^n \mid \mathbf{f}(z) = 0 \text{ and } g(z) > 0\}$, since $g(z)$ is always positive for all $z \in \mathbb{R}$, we can conclude $\sigma(H_1(\mathscr{I}) = \#\{z \in \mathbb{R}^n \mid \mathbf{f}(z) = 0\}$. □

The second auxiliary function we will use is $g(x) = ||x - z^*||_2^2 - \epsilon^2$, where z^* and ϵ are rational.

Notice that, if $g(x) \geq 0$ for any $x \in V_{\mathbb{R}}(\mathbf{f})$, then $||x - z^*||_2 \geq \epsilon$, which implies that there is no $x \in V_{\mathbb{R}}(\mathbf{f})$ within $\mathscr{B}_\epsilon(z^*)$. Similarly, if $g(x) < 0$ then there is at least one $x \in V_{\mathbb{R}}(\mathbf{f})$ within the given neighborhood $\overline{\mathscr{B}}_\epsilon(z^*)$.

Now we can introduce the certification theorem,

Theorem 4.3.2. *Let* $\mathbf{f} = (f_1, \ldots, f_N)$, $f_i \in \mathbb{R}[x_1, \ldots, x_n]$ *for all* $i = 1, \ldots N$, *and* $\mathscr{I} =< f_1, \ldots, f_N >$ *is a zero dimensional ideal. Given an approximate root* z^*, *distance* $\epsilon \in \mathbb{R}$.
Define $g(x) := \|x - z^*\|_2^2 - \epsilon^2$ *then*

$$\sigma(H_1(\mathscr{I})) = \sigma(H_g(\mathscr{I}))$$

if and only if there is no real root within $\overline{\mathscr{B}}_\epsilon(z^*)$.

Proof. Let $N := \#\{\boldsymbol{x} \in \mathbb{R} | \mathbf{f}(\boldsymbol{x}) = 0\}$. By the definitions, $N \geq N_+$ and $N \geq N_-$. By Proposition 4.3.1, $N = \sigma(H_1(\mathscr{I}))$. Defining $N_0 := \#\{\boldsymbol{x} \in \mathbb{R} | \mathbf{f}(\boldsymbol{x}) = 0 \text{ and } g(\boldsymbol{x}) = 0\}$, it can be seen that $N = N_- + N_0 + N_+$. Now assume that $\sigma(H_1(\mathscr{I})) = \sigma(H_g(\mathscr{I}))$, then by Hermite theorem we have

$$N = N_+ - N_-.$$

Since $N = N_- + N_0 + N_+$ then, $N_- + N_0 + N_+ = N_+ - N_-$. We obtain

$$2N_- + N_0 = 0.$$

N_- and N_0 are both non-negative integer numbers. Sum of two non-negative integers is zero only if they both equal zero. Therefore $N_- = N_0 = 0$, then we have $\{\boldsymbol{x} \in \mathbb{R}^n \mid \mathbf{f}(\boldsymbol{x}) = 0 \text{ and } g(\boldsymbol{x}) \leq 0\}$ is an empty set. This indicates for all $\boldsymbol{x} \in V_{\mathbb{R}}(\mathbf{f})$, $g(\boldsymbol{x}) > 0$. The way we defined $g(\boldsymbol{x})$ tells us, $\|x - \xi\|^2 > \epsilon$ for all real common roots of **f**. Now one can conclude that there is no real common root of **f** within $\mathscr{B}_\epsilon(\xi)$. □

Or equivalently,
If $\sigma(H_1(\mathscr{I})) \neq \sigma(H_g(\mathscr{I}))$ then there is at least one real root within ϵ neighborhood of z^*.

Thus, if we can compute the signatures of $H_g(\mathscr{I})$ and $H_1(\mathscr{I})$, we can decide if there is an exact root within a certain neighborhood of a given point. Now the question is, how we can use Hermite matrices for certification purposes if we do not know the exact roots, only approximates.

4.4 Symbolic-Numeric computation of Hermite matrices

Now we introduce a method using Hermite matrices to certify approximate real solutions of polynomial systems over $\mathbb{Q}$. Both univariate and multivariate cases are covered in the following description of our method.

Let $\mathbf{f} = (f_1, \ldots, f_N) \in \mathbb{Q}[x_1, \ldots, x_n]$ be a system of polynomials and assume $\mathscr{I} =< f_1, \ldots, f_N >$ is a zero dimensional ideal. Let $\xi_1, \ldots, \xi_k \in \mathbb{C}^n$ be common distinct roots of **f** and let $z_1, \ldots, z_k \in \mathbb{C}^n$ be approximations to $\xi_1, \ldots, \xi_k$.

Now we define Hermite matrices with respect to a list of distinct points $z_1, z_2, \ldots, z_k$:

Definition 4.4.1. *Let $g \in \mathbb{R}[x_1, \ldots, x_n]$, then the* **Hermite matrix** *of g associated to the distinct points $z_1, \ldots, z_k \in \mathbb{C}^n$ is*

$$H_g(z_1, \ldots, z_k) := V^T G V \tag{4.9}$$

where $V = V_{\mathscr{B}}(z_1, z_2, \ldots, z_k)$ is the $k \times k$ Vandermonde matrix of $z_1, z_2, \ldots, z_k$ with respect to some monomial basis (normal set) $\mathscr{B}$ for $\mathbb{R}[x_1, \ldots, x_n]/\mathscr{I}(z_1, \ldots, z_k)$ and G is a $k \times k$ diagonal matrix with $[G]_{jj} = g(z_j)$ for $j = 1 \ldots k$. Next, we define

$$\mathscr{B}^+ = \mathscr{B} + \bigcup_i x_i \mathscr{B} = \{b, x_1 b, \ldots, x_n b \mid b \in \mathscr{B}\} \quad \textit{for } i = 1, \ldots, n. \tag{4.10}$$

Then the **extended Hermite matrix** *associated to distinct points* $z_1,\ldots,z_k \in \mathbb{C}^n$ *is*

$$H_g^+(z_1,\ldots,z_k) := V^T G V \tag{4.11}$$

where $V = V_{\mathscr{B}^+}(z_1,z_2,\ldots,z_k)$ *and* G *is a* $k\times k$ *diagonal matrix with* $[G]_{jj} = g(z_j)$ *for* $j = 1\ldots k$.

Note that, in univariate case we will use the following basis extension, let $z_1,\ldots,z_k$ be given distinct points and consider standard monomial basis $\mathscr{B} = \{1,x,\ldots,x^{k-1}\}$. Then one can construct an extended Hermite matrix H_g^+ using the $k\times k$ Vandermonde matrix of $z_1,\ldots,z_k$ with respect to the basis extension

$$\mathscr{B}^+ = \{1,x,\ldots,x^k\}. \tag{4.12}$$

We propose the following method to certify that the point $z^* \in \mathbb{R}^n$ is an approximate root of **f** within a distance ϵ :

Algorithm 1: Real Root Certification

```
Input  $\mathbf{f} = (f_1,\ldots,f_N) \in \mathbb{Q}[x_1,\ldots,x_n]$,
       $\{z_1,\ldots,z_k\}$ a set of all approximate roots of f where $z_i \in \mathbb{C}^n$ for $i = 1,\ldots,k$,
       $z^* \in \mathbb{R}^n$ is the approximate solution we want to certify,
       $E$ an error bound on $\|z_i - \xi_i\|_2$,
       $\epsilon$ is a distance.
Output  True: means $z^* \in \mathbb{R}^n$ is within $\epsilon$ distance from some exact root of f.
        False: The certification failed.
  1: Define the auxiliary function $g(x) := \|x - z^*\|_2^2 - \epsilon^2$.
  2: Compute an extended Hermite matrix with respect to $g(x) = 1$,
     $H_1^+ \leftarrow H_1^+(z_1,\ldots,z_k)$.
  3: Set Rationalize the entries of $H_1^+$ using continued fractions and error bounds for the moments,
     explained below.
  4: Call Algorithm 3 with input f, $g(x)$, $H_1^+$ (in the univariate case call Algorithm 2). If it returns $H_1$
     and $H_g$ then it is certified that $H_1 = H_1(\xi_1,\ldots,\xi_k)$ and $H_g = H_g(\xi_1,\ldots,\xi_k)$ where $\xi_1,\ldots,\xi_k$ are
     the exact roots of f. if it returns Fail then return False.
  5: Compute the signatures using Definition 4.2.5,
     $\sigma_1 \leftarrow \sigma(H_1)$,
     $\sigma_g \leftarrow \sigma(H_g)$.
```

```
6: By Theorem 4.3.2,
   if $\sigma_1 \neq \sigma_g$ then return True,
   else return False.
```

4.4.1 Explanation of the Algorithm

Step 1 First define the auxiliary function $g(x) := \|x - z^*\|_2^2 - \epsilon^2$. The sign of g provides some information on the ϵ neighborhood of z^*. Notice that if $g(x) > 0$, then $\|x - z^*\|_2 > \epsilon$ means there is no x within the ϵ neighborhood of z^*. Similarly if $g(x) \leq 0$, there is at least one x within the ϵ neighborhood of z^*. Note we can assume that the coordinates of z^* are given as floating point numbers, so $z^* \in \mathbb{Q}^n$, thus g is a rational polynomial.

Step 2 Compute the approximate extended Hermite matrix $H_1^+(z_1, \ldots, z_k)$ (as in Definition 4.4.1) using either univariate (see Subsection 4.5.1) or multivariate (see Subsection 4.5.2) basis extension $\mathscr{B}^+$ depending on the given system.

Step 3 Next, we rationalize each entry of the approximate Hermite matrix H_1^+ using rational number reconstruction (see Section 1.4).

The error bound we use is computed as follows. Assume that we want to reconstruct the moment corresponding to x^α for some monomial of degree $|\alpha| = d$. Let E be an upper bound for $\|z_i - \xi_i\|$, given as part of the input, and let C be an upper bound for the coordinates of z_i for all $i = 1, \ldots, k$. Then the error in the moment is bounded by

$$\left| \sum_i (z_i^\alpha - \xi_i^\alpha) \right| \leq k E d C^{d-1},$$

using the same proof as in Proposition 2.2.1 Using this bound, we set

$$B := \left\lceil (2 E d C^{d-1})^{-1/2} \right\rceil \tag{4.13}$$

as the bound for the denominators in the rational number reconstruction algorithm.

Step 4 Since we use approximate roots, it is not always true that the rationalized Hermite matrix actually correspond to the exact roots. Rational number reconstruction of entires may not converge to the exact ones with the computed bounds. One can use the method described in Section 4.5 and certify $H_1(\xi_1, \xi_2, \ldots, \xi_k)$ and $H_g(\xi_1, \xi_2, \ldots, \xi_k)$ without any knowledge of the exact roots. If the input system **f** is univariate use Algorithm 2, if it is multivariate use equivalent Algorithm 3. In case Algorithm 2 (or Algorithm 3) fails, then Algorithm 1 returns false.

Step 5 Once the rational Hermite matrices are certified, Algorithm 2 or Algorithm 3 will return $H_1(\xi_1,\dots,\xi_k)$ and $H_g(xi_1,\dots,\xi_k)$. Then we can use Definition 4.2.5 to obtain the signatures. One can investigate the number of sign changes in the characteristic polynomials of $H_1(\xi_1,\dots,\xi_k)$ and $H_g(\xi_1,\dots,\xi_k)$ and yield the signatures using various root counting methods, see Section 1.6.

Step 6 Then one can simply use Theorem 4.3.2 and compare the signatures $\sigma(H_1(\xi_1,\dots,\xi_k))$ and $\sigma(H_g(\xi_1,\dots,\xi_k))$. If the signatures are not equal then we conclude that there exists at least one root within the ϵ neighborhood of z^*. If not we conclude that our method failed to certify the given approximate solution z^*.

Remark 4.4.2. *At the initialization of our algorithm we assume that we have approximation for all the roots of* $\mathbf{f}$. *Certifying that we have all distinct roots of $\mathscr{I} =< \mathbf{f} >$ is a separate question that we do not address in this thesis. One can use homotopy method to get all approximate solutions, as explained in Subsection 2.2.1.*

Also, note that Algorithm 1 can be used as long as the roots $\xi_1,..,\xi_k \in \mathbb{C}^n$ form a rational component of $V(\mathscr{I})$.

As stated at **Step 3**, the Hermite matrices are constructed using approximate roots and then rational number reconstruction of each approximate entry. As it is stated in Section 1.4, rational reconstruction may not exist, or depending on the denominator bound, it may not converge to the right value. Therefore it is not always true that the rationalized Hermite matrix actually correspond to the exact roots (see Section 1.4). In case we obtain a rational Hermite matrix, it is still needed to be certified that it is actually corresponding to the exact roots.

4.5 Certification of Hermite matrices

In this section, we describe a method to certify a given Hermite matrix over rationals.

4.5.1 Univariate Case

Let $\mathbf{f} = (f_1,\dots,f_N) \in \mathbb{Q}[x]$ be a system of univariate polynomials with distinct approximate solutions $z_1,\dots,z_k \in \mathbb{C}$. Fix a basis $\mathscr{B} = \{1, x,\dots,x^{k-1}\}$ and assume that the Vandermonde matrix of $z_1,\dots,z_k$

with respect to $\mathscr{B}$ is nonsingular, which is guaranteed in the univariate case if $z_i \neq z_j$ for all $i, j = 1, \dots, k$.

Let V be the Vandermonde matrix of $z_1, \dots, z_k \in \mathbb{C}$ with respect to $\mathscr{B}^+ := \{1, x, \dots, x^k\}$. Then the $(k+1) \times (k+1)$ extended Hermite matrix is $H_1^+(z_1, \dots, z_k) = V^T G V$, where G is a $k \times k$ diagonal matrix with $[G]_{ii} = g(z_i)$ for $i = 1 \dots k$.

Algorithm 2: Univariate Hermite Matrix Certification

```
Input: f = $(f_1, \dots, f_N) \in \mathbb{Q}[x]$,
       $g(x) = \|x - z^*\|_2^2 - \epsilon^2 \in \mathbb{Q}[x]$ ,
       $H_1^+ \in \mathbb{Q}^{(k+1)\times(k+1)}$ extended Hermite matrix with respect to $\mathscr{B}^+$.
Output  Return $H_1$ and $H_g$, or Fail.
  1: $H_1 \leftarrow k \times k$ submatrix of $H_1^+$, consisting the first $k$ rows and the first $k$ columns.
     $H_1^k \leftarrow k \times k$ submatrix of $H_1^+$, consists of the first $k$ rows and the last $k$ columns.
  2: Check if $H_1^+$ has Hankel structure and $\mathrm{rank\ H_1 = rank\ H_1^+ = k}$,
     then $M \leftarrow H_1^{-1} \cdot H_1^k$.
  3: Check if $M$ has a companion shape and $f(M) = 0$,
     then $p(x) \leftarrow$ characteristic polynomial to the companion matrix $M$,
     check if it is square free, which certifies that $M$ as the multiplication matrix of $x$,
     $M_x \leftarrow M$.
  4: Using coefficients of $p$, yield Newton sums (see Remark 4.2.8) and check if each one matches
     to the corresponding entry of $H_1$, which certifies $H_1$.
  5: Once $H_1$ and $M_x$ are certified,
     $H_g \leftarrow H_1 \cdot g(M_x)$.
  6: Return $H_1$ and $H_g$.
```

4.5.1.1 Explanation of Algorithm 2

At Step 2, ranks of H_1 and H_1^+ can be checked using exact arithmetic since they are rational matrices. Then, by the construction of M, it has a companion matrix structure as defined in 1.5.10. If M has a companion shape, next we check if its characteristic polynomial p is square free (i.e. $\gcd(p, p') = 1$)

and $f(M)=0$. If that is the case, M is similar to a diagonal matrix D with diagonal entries $\xi_1,\dots,\xi_k \in \mathbb{C}$ which are the roots of $\mathbf{f}$, more specifically

$$M = V^{-1} D V$$

where V is the Vandermonde of $\xi_1,\dots\xi_k$ with respect to the basis $\mathscr{B}=\{1,x,\dots,x^{k-1}\}$. Since M_x has companion matrix shape, the last column contains coefficients of its characteristic polynomial. These coefficients are the elementary symmetric polynomials of the k common roots, those coefficients can be used to find power sums (i.e., Newton identities) as described in Definition 1.5.17. Once we certify H_1 and M_x, we can compute H_g by taking

$$H_g := H_1 \cdot g(M_x)$$

which is $H_1 \cdot M_g = (V^T V)\cdot(V^{-1} G V) = V^T G V = H_g$.

4.5.2 Multivariate Case

Now we show how to certify a Hermite matrix defined by multivariate polynomials.

Let $z_1,\dots,z_k \in \mathbb{C}^n$ with $z_i=(z_{i,1},\dots,z_{i,n})$ for $i=1,\dots,k$ be the distinct approximate solutions of $\mathbf{f}=(f_1,\dots,f_N)\in\mathbb{Q}[x_1,\dots,x_n]$. If $z_{i,1}\neq z_{j,1}$ for $i\neq j$, then $\mathscr{B}=\{1,x_1,\dots,x_1^{k-1}\}$ is a basis for $\mathbb{R}[x_1,\dots,x_n]/\mathscr{I}(z_1,\dots,z_k)$. And we use $\mathscr{B}^+$ as described in (4.10). Then the corresponding Hermite matrix can be written in the moment matrix form (e.g., see [54]) as

$$H_1^+ = [y_{b\cdot b'}]_{b,b'\in\mathscr{B}^+}.$$

Algorithm 3: Multivariate Hermite Matrix Certification

```
Input: $\mathbf{f}=(f_1,\dots,f_N)\in\mathbb{Q}[x_1,\dots,x_n]$,
    $g(x)=\|x-z^*\|_2^2-\epsilon^2\in\mathbb{Q}[x_1,\dots,x_n]$,
    $H_1^+\in\mathbb{Q}^{l\times l}$ extended Hermite matrix with respect to $\mathscr{B}^+$, with some $l$.
Output: Return $H_1$ and $H_g$, or Fail.
  1: $H_1 \leftarrow [y_{b\cdot b'}]_{b,b'\in\mathscr{B}}$ which is the $k\times k$ submatrix of $H_1^+$ consisting the first $k$ rows and columns.
     $H_1^{x_s} \leftarrow k\times k$ submatrix of $H_1^+$ with rows corresponding to $\mathscr{B}$ and columns corresponding to $x_s\mathscr{B}$ for $s=1,\dots,n$.
  2: Check if $H_1^+$ has Hankel (or moment matrix) structure and $\mathrm{rank\ H_1} = \mathrm{rank\ H_1^+} = \mathrm{k}$,
     then $M_s \leftarrow H_1^{-1}\cdot H_1^{x_s}$ for $s=1,\dots,n$.
```

```
3: Check $M_1$ has a companion matrix structure,
   then $p \leftarrow$ characteristic polynomial polynomial to $M_1$.
   Check if $\gcd(p, p') = 1$.
4: Check if $\{M_i,\ i = 1, \dots, n\}$ commute pairwise and
       $f_i(M_1, M_2, \dots, M_n) = 0$ for $i = 1, \dots, N$,
   then $M_{x_i} \leftarrow M_i$ for all $i = 1, \dots, n$.
5: Compute coefficients of $p(x_1)$, yield Newton sums (see Remark 4.2.8) and check that what we
   get is equal to the corresponding entry of $H_1$, which certifies $H_1$.
6: Once $H_1$ and $M_i$ for all $i = 1, \dots, n$ are certified,
   $H_g \leftarrow H_1 \cdot g(M_1, \dots, M_n)$.
7: Return $H_1$ and $H_g$.
```

4.5.2.1 Explanation of Algorithm 3

At step 2, rank $H_1 =$ rank $H_1^+ = k$ can be checked by exact arithmetics. Then we verify that H_1 and H_1^+ has the correct Hankel or moment matrix structure. Because of the Hankel structure of $H_1^{x_1}$, M_1 has a companion matrix structure, we the characteristic polynomial $p(x) \in \mathbb{Q}[x]$. We check p is square free, i.e. $\gcd(p, p') = 1$. Then if $\{M_i\ :\ i = 1, \dots n\}$ are pairwise commuting, then they are simultaneously diagonalizable. Then, $\{M_i\ :\ i = 1, \dots n\}$ are certified if they vanish on the given system, i.e., $f_i(M_1, \dots, M_n) = 0$ for all $i = 1, \dots, N$.

The coefficients of $p(x)$ are the elementary symmetric polynomials of the k common roots $\xi_1, \dots, \xi_k \in \mathbb{C}$, those coefficients can be used to find power sums of roots using Newton identities, as described in Definition 1.5.17. H_1 is certified if the power sums are equal to the corresponding entry of H_1.

Once we certify H_1 and M_1, we can compute H_g by taking

$$H_g := H_1 \cdot g(M_1, \dots, M_n)$$

which is $H_1 \cdot M_g = (V^T V) \cdot (V^{-1} G V) = V^T G V = H_g$.

4.6 Examples

4.6.1 A toy example

We start with an obvious example to illustrate the steps of our algorithms:

Consider $f(x) = 16x^4 - 10x^2 + 1$, the exact roots of f are $1/\sqrt{2}, -1/\sqrt{2}, 1/2\sqrt{2}, -1/2\sqrt{2}$. Let say we want to show that there is no root of f near zero. We set the distance $\epsilon = 1/10$.

We can get the following approximate solutions using homotopy method on Maple:
$z_1 = 0.7071067810, z_2 = -0.7071067810, z_3 = 0.3535533905, z_4 = -0.3535533905$.
This solution has error bound $E := 10^{-8}$.

Algorithm 1: Real Root Certification

Input: $f(x) = 16x^4 - 10x^2 + 1$,
$z_1 = 0.7071067810, z_2 = -0.7071067810, z_3 = 0.3535533905, z_4 = -0.3535533905$,
$z^* = 0$
$\epsilon = 1/10$.

Step 1: $g(x) := x^2 - (\frac{1}{10})^2$.

Step 2: First we need the Vandermonde matrix with respect to the following extension of the standard monomial basis $\mathscr{B}^+ := \{1, x, x^2, x^3, x^4\}$,

$$V := V_{\mathscr{B}^+}(z_1, z_2, z_3, z_4) = \begin{bmatrix} 1.0 & 0.7071067832 & 0.5000000028 & 0.3535533936 & 0.2500000028 \\ 1.0 & -0.7071067832 & 0.5000000028 & -0.3535533936 & 0.2500000028 \\ 1.0 & 0.353553390 & 0.1249999996 & 0.04419417360 & 0.01562499990 \\ 1.0 & -0.3535533907 & 0.1250000001 & -0.04419417386 & 0.01562500002 \end{bmatrix}$$

Then by Definition 4.4.1, $H_1^+(z_1, z_2, z_3, z_4) = V^T V$ for $g(x) = 1$

$$H_1^+ = \begin{bmatrix} 4.0 & -0.000000000700000002407108468 & 1.25000000529999999 & -0.000000000260000000695814748 & 0.531250005520000013 \\ -0.000000000700000002407108468 & 1.25000000535086242 & -0.000000000264276704686761121 & 0.531250005587795671 & -5.33639070432467832 \times 10^{-11} \\ 1.25000000529999999 & -0.000000000264276704686761121 & 0.531250005525000013 & -5.45970881352109139 \times 10^{-11} & 0.253906254185312541 \\ -0.000000000260000000695814748 & 0.531250005587795671 & -5.45970881352109139 \times 10^{-11} & 0.253906254235507889 & -9.36580088657656962 \times 10^{-12} \\ 0.531250005520000013 & -5.33639070432467832 \times 10^{-11} & 0.253906254185312541 & -9.36580088657656962 \times 10^{-12} & 0.125488284047500037 \end{bmatrix}.$$

Step 3: Rationalize H_1^+ using an upper bound $B = 4083$ on denominators as defined in 4.13 with

$E = 10^{-8}$, $C =$ and $d = 4$.

$$H_1^+ = \begin{bmatrix} 4 & 0 & \frac{5}{4} & 0 & \frac{17}{32} \\ 0 & \frac{5}{4} & 0 & \frac{17}{32} & 0 \\ \frac{5}{4} & 0 & \frac{17}{32} & 0 & \frac{65}{256} \\ 0 & \frac{17}{32} & 0 & \frac{65}{256} & 0 \\ \frac{17}{32} & 0 & \frac{65}{256} & 0 & \frac{257}{2048} \end{bmatrix}$$

Step 4: Use **Algorithm 2** to obtain H_1, H_g (see below for details).

Step 5: Then we find characteristic polynomial of H_1,

$$c_1(\lambda) = \lambda^4 - \frac{1545\,\lambda^3}{256} + \frac{60721\,\lambda^2}{8192} - \frac{8235\,\lambda}{8192} + \frac{81}{4096},$$

$$c_1(-\lambda) = \lambda^4 + \frac{1545\,\lambda^3}{256} + \frac{60721\,\lambda^2}{8192} + \frac{8235\,\lambda}{8192} + \frac{81}{4096}.$$

Then $N_+ = \#$ sign change in $c_1(\lambda)$ is 4 and $N_- = \#$ sign change in $p(-\lambda)$ is 0, thus

$$\sigma_1 := \sigma(H_1) = 4.$$

Characteristic polynomial of H_g

$$c_g(\lambda) = \lambda^4 - \frac{21507\,\lambda^3}{10240} + \frac{317737009\,\lambda^2}{327680000} - \frac{152236287\,\lambda}{6553600000} + \frac{102880449}{1638400000000}$$

$$c_g(-\lambda) = \lambda^4 + \frac{21507\,\lambda^3}{10240} + \frac{317737009\,\lambda^2}{327680000} + \frac{152236287\,\lambda}{6553600000} + \frac{102880449}{1638400000000}.$$

And $N_+ = \#$ sign change in $p(\lambda)$ is 4 and $N_- = \#$ sign change in $p(-\lambda)$ is 0, we obtain

$$\sigma_g := \sigma(H_g) = 4.$$

Step 6: Finally, by Theorem 4.3.2, there is no exact root near (within the distance 1/10) zero since $\sigma_1 = \sigma_g$.

Now we show how Algorithm 2 is used at **Step 4**.

Algorithm 2: Univariate Hermite Matrix Certification

Input: $f(x) = 16x^4 - 10x^2 + 1$,

$g(x) = x^2 - (\frac{1}{10})^2$,
$H_1^+ \in \mathbb{Q}^{5\times 5}$.

Step 1: We get the submatrices:

$$H_1 = \begin{bmatrix} 4 & 0 & \frac{5}{4} & 0 \\ 0 & \frac{5}{4} & 0 & \frac{17}{32} \\ \frac{5}{4} & 0 & \frac{17}{32} & 0 \\ 0 & \frac{17}{32} & 0 & \frac{65}{256} \end{bmatrix}, H_1^x = \begin{bmatrix} 0 & \frac{5}{4} & 0 & \frac{17}{32} \\ \frac{5}{4} & 0 & \frac{17}{32} & 0 \\ 0 & \frac{17}{32} & 0 & \frac{65}{256} \\ \frac{17}{32} & 0 & \frac{65}{256} & 0 \end{bmatrix}$$

Step 2: H_1^+ has Hankel structure and $\mathrm{rank\,H_1^+} = \mathrm{rank\,H_1} = 4$. Then

$$M = H_1^{-1} \cdot H_1^x = \begin{bmatrix} 0 & 0 & 0 & -\frac{1}{16} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & \frac{5}{8} \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Step 3: M has companion matrix structure and $f(M) = 0$ then $p(x) := \frac{5}{8}x^2 - \frac{1}{16}$ with $\gcd(p, p') = 1$. Therefore M is the multiplication matrix M_x.

Step 4: As defined in 1.5.17, and described in Remark 4.2.8, elementary symmetric functions: $e_0 = \frac{5}{8}, e_1 = 0, e_2 = -\frac{1}{16}$, which yields

$$\sum_{i=1}^{4} \xi_i^0 = 4, \sum_{i=1}^{4} \xi_i = 0, \sum_{i=1}^{4} \xi_i^2 = 5/4,$$

$$\sum_{i=1}^{4} \xi_i^3 = 0, \sum_{i=1}^{4} \xi_i^4 = 17/32, \sum_{i=1}^{4} \xi_i^5 = 0, \sum_{i=1}^{4} \xi_i^6 = 65/256.$$

Each entry matches the Newton sums, thus H_1 is indeed the exact Hermite Matrix.

Step 5: H_1 and M_x are certified, then

$$H_g = H_1 \cdot g(M_x) = \begin{bmatrix} \frac{121}{100} & 0 & \frac{83}{160} & 0 \\ 0 & \frac{83}{160} & 0 & \frac{1591}{6400} \\ \frac{83}{160} & 0 & \frac{1591}{6400} & 0 \\ 0 & \frac{1591}{6400} & 0 & \frac{1259}{10240} \end{bmatrix}.$$

Step 5: Return H_1 and H_g.

4.6.2 An illustrative example

A classical example $f_1 = f_0(x_1), f_2 = x_2 - x_1^2, f_3 = x_3 - x_2^2, \ldots, f_n = x_n - x_{n-1}^2$ can be used to show usefulness of our method when $f_0(x_1)$ has roots less than 1. In that case, the last coordinate of the solutions will get exponentially small. If we set $g = x_n$ in our algorithm, then we can demonstrate that there are no roots close to zero with last coordinate.
Let $n = 4$ and $f_0(x_1) = 10{x_1}^3 - 1$, then consider the polynomial system $\mathbf{f} = (f_1, f_2, f_3, f_4)$ with

$$\begin{aligned} f_1 &= 10{x_1}^3 - 1, \\ f_2 &= x_2 - x_1^2, \\ f_3 &= x_3 - x_2^2, \\ f_4 &= x_4 - x_3^2. \end{aligned}$$

$f_1(x_1)$ has 3 exact roots

$$\begin{aligned} \xi_{11} &= \frac{1}{10}10^{(2/3)}, \\ \xi_{21} &= -\frac{1}{20}10^{(2/3)} + \frac{i\sqrt{3}}{20}10^{(2/3)}, \\ \xi_{31} &= -\frac{1}{20}10^{(2/3)} - \frac{i\sqrt{3}}{20}10^{(2/3)}. \end{aligned}$$

The system $\mathbf{f}$ has the following 3 exact roots

$$\begin{aligned} \xi_1 &= (\xi_{11}, \xi_{11}^2, \frac{1}{10}\xi_{11}, \frac{1}{100}\xi_{11}^2), \\ \xi_2 &= (\xi_{21}, \xi_{21}^2, \frac{1}{10}\xi_{21}, \frac{1}{100}\xi_{21}^2), \\ \xi_3 &= (\xi_{31}, \xi_{31}^2, \frac{1}{10}\xi_{31}, \frac{1}{100}\xi_{31}^2). \end{aligned}$$

Maple is used to obtain all the following computation:
Using homotopy method we obtain the approximate solutions z_1, z_2, z_3 which are the approximate

solutions associated to the exact solutions ξ_1, ξ_2, ξ_3 of $\mathbf{f}$ with the error bound $E = 10^{-10}$.

$$
\begin{aligned}
z_1 &= [0.4641588834, 0.2154434690, 0.4641588834e-1, 0.2154434691e-2],\\
z_2 &= [-0.2320794417 + i0.4019733844, -0.01077217345 - i0.1865795172,\\
&\quad -0.2320794417e-1 + i0.4019733844e-1, -0.1077217345e-2 - i0.1865795172e-2],\\
z_3 &= [-0.2320794414 - i0.4019733843, -0.1077217345 + i0.1865795171,\\
&\quad -0.2320794414e-1 - i0.4019733843e-1, -0.1077217354e-2 + i0.1865795169e-2].
\end{aligned}
$$

We set $g := x_4$ and use Algorithm 1:
Since the first coordinates of the solutions are distinct, we can set the monomial basis $\mathscr{B} := \{1, x_1, x_1^2\}$, and construct the following extension of $\mathscr{B}$

$$\mathscr{B}^+ := \{1, x_1, x_1^2, x_1^3, x_2, x_2x_1, x_2x_1^2, \ldots, x_4, x_4x_1, x_4x_1^2\}.$$

We use this to compute the Vandermonde matrix V of z_1, z_2, z_3 with respect to $\mathscr{B}^+$. Then we get H_1^+ using Definition 4.4.1. H_1^+ is a 13×13 matrix that cannot be listed here. Next, we use the bound $B = 1118034$ on denominators and rationalize the entries of H_1^+ (see 2.2.2.2 and (4.13)). Algorithm 3 certifies H_1^+, (see below) and returns rational matrices, H_1 and H_g as follows

$$H_1 = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 0 & 3/10 \\ 0 & 3/10 & 0 \end{bmatrix}, H_g = \begin{bmatrix} 0 & \frac{3}{1000} & 0 \\ \frac{3}{1000} & 0 & 0 \\ 0 & 0 & \frac{3}{10000} \end{bmatrix}.$$

The characteristic polynomial of H_1 is

$$c_1(\lambda) = -\lambda^3 - 3\lambda^2 + \frac{9\lambda}{100} + \frac{27}{100}.$$

The number of sign change in $c_1(\lambda)$ is 2, that is corresponding to the number of positive eigenvalues of H_1. Similarly, the number of negative eigenvalues is and the number of sign change in $c_1(-\lambda)$, which is 1. Thus, the signature σ_1 is 1. Now we compute the signature of H_g. The characteristic polynomial of H_g is

$$c_g(\lambda) = \lambda^3 - \frac{3\lambda^2}{10000} - \frac{9\lambda}{1000000} + \frac{27}{10000000000}.$$

The number of sign change in $c_g(\lambda)$ is 2 and the number of sign change in $c_g(-\lambda)$, which is 1. Then, the signature σ_g is 1.

Therefore, by the Theorem 4.3.2, $\sigma_1 = \sigma_g$ implies that there is no root close to zero with last coordinate.

Next, we show how Algorithm 3 is used to certify approximate Hermite matrices. The input matrix is a 13 × 13 rational matrix which has a Hankel matrix structure. We get its submatrices as described in the algorithm:

$$H_1 := \begin{bmatrix} 3 & 0 & 0 \\ 0 & 0 & 3/10 \\ 0 & 3/10 & 0 \end{bmatrix}, H_1^{x_1} := \begin{bmatrix} 0 & 0 & 3/10 \\ 0 & 3/10 & 0 \\ 3/10 & 0 & 0 \end{bmatrix},$$

$$H_1^{x_2} := \begin{bmatrix} 0 & 3/10 & 0 \\ 3/10 & 0 & 0 \\ 0 & 0 & \frac{3}{100} \end{bmatrix}, H_1^{x_3} := \begin{bmatrix} 0 & 0 & \frac{3}{100} \\ 0 & \frac{3}{100} & 0 \\ \frac{3}{100} & 0 & 0 \end{bmatrix}, H_1^{x_4} := \begin{bmatrix} 0 & \frac{3}{1000} & 0 \\ \frac{3}{1000} & 0 & 0 \\ 0 & 0 & \frac{3}{10000} \end{bmatrix}.$$

H_1^+ and H_1 have rank 3. Next we compute the matrices $M_i = H_1^{-1} \cdot H_1^{x_i}$ for $i = 1,2,3,4$.

$$M_1 := \begin{bmatrix} 0 & 0 & 1/10 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, M_2 := \begin{bmatrix} 0 & 1/10 & 0 \\ 0 & 0 & 1/10 \\ 1 & 0 & 0 \end{bmatrix},$$

$$M_3 := \begin{bmatrix} 0 & 0 & \frac{1}{100} \\ 1/10 & 0 & 0 \\ 0 & 1/10 & 0 \end{bmatrix}, M_4 := \begin{bmatrix} 0 & \frac{1}{1000} & 0 \\ 0 & 0 & \frac{1}{1000} \\ \frac{1}{100} & 0 & 0 \end{bmatrix}.$$

We see that M_1 has a companion matrix structure, and its characteristic polynomial $p(x) := -\frac{1}{10}$ is square free. Since

$$f_i(M_1, M_2, M_3, M_4) = 0, \quad i = 1,2,3,4$$

M_i are the multiplication matrices M_{x_i} for $i = 1,2,3,4$. Now as it is shown in the previous example, at Step 4 of Algorithm 2, the coefficients of the characteristic polynomial $p(x) = 0 \cdot x^2 + 0 \cdot x - 1/10$

provides the following power sums without knowing the exact roots (see 1.5.17)

$$\sum_{i=1}^{3}\xi_{i1}^{0}=3,\sum_{i=1}^{3}\xi_{i1}=0,\sum_{i=1}^{3}\xi_{i1}^{2}=0,\sum_{i=1}^{3}\xi_{i1}^{3}=3/10,\sum_{i=1}^{3}\xi_{i1}^{4}=0.$$

Each entry of H_1 matches to the corresponding power sum, which certifies H_1.
Since H_1 and the multiplication matrices are certified, we can obtain the matrix H_g with respect to $g=x_4$, therefore $g(M_1,M_2,M_3,M_4)=M_4$.

$$H_g := H_1\cdot g(M_1,M_2,M_3,M_4)=H_1\cdot M_4=\begin{bmatrix} 0 & \frac{3}{1000} & 0 \\ \frac{3}{1000} & 0 & 0 \\ 0 & 0 & \frac{3}{10000} \end{bmatrix}.$$

Thus, Algorithm 3 returns H_1 and H_g.

REFERENCES

[1] J. Abbott, *Bounds on factors in $\mathbb{Z}[x]$*, J. Symbolic Comput., 50 (2013), pp. 532–563.

[2] T. A. Akoglu, J. D. Hauenstein, and A. Szanto, *Certifying solutions to overdetermined and singular polynomial systems over $\mathbb{Q}$*. manuscript, 2014.

[3] M.-E. Alonso, E. Becker, M.-F. Roy, and T. Wörmann, *Zeros, multiplicities, and idempotents for zero-dimensional systems*, in Algorithms in algebraic geometry and applications, Springer, 1996, pp. 1–15.

[4] E. A. Arnold, *Modular algorithms for computing gröbner bases*, Journal of Symbolic Computation, 35 (2003), pp. 403 – 419.

[5] S. Basu, R. Pollack, and M.-F. Roy, *Algorithms in real algebraic geometry*, vol. 20033, Springer, 2005.

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

[7] ——, *Numerically solving polynomial systems with Bertini*, vol. 25 of Software, Environments, and Tools, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2013.

[8] R. Bellman, *Introduction to Matrix Analysis (2Nd Ed.)*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1997.

[9] C. Beltrán and A. Leykin, *Certified numerical homotopy tracking*, Exp. Math., 21 (2012), pp. 69–83.

[10] ——, *Robust certified numerical homotopy tracking*, Found. Comput. Math., 13 (2013), pp. 253–295.

[11] G. Björck, *Functions of modulus 1 on Z_n whose Fourier transforms have constant modulus, and "cyclic n-roots"*, in Recent advances in Fourier analysis and its applications (Il Ciocco, 1989), vol. 315 of NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., Kluwer Acad. Publ., Dordrecht, 1990, pp. 131–140.

[12] L. Blum, F. Cucker, M. Shub, and S. Smale, *Complexity and real computation*, Springer-Verlag, New York, 1998. With a foreword by Richard M. Karp.

[13] L. Brand, *The companion matrix and its properties*, The American Mathematical Monthly, 71 (1964), pp. 629–634.

[14] W. D. Brownawell and C. K. Yap, *Lower bounds for zero-dimensional projections*, in ISSAC 2009—Proceedings of the 2009 International Symposium on Symbolic and Algebraic Computation, ACM, New York, 2009, pp. 79–85.

[15] J. CANNY, *Generalised characteristic polynomials*, J. Symbolic Comput., 9 (1990), pp. 241–250.

[16] D. CASTRO, L. M. PARDO, K. HÄGELE, AND J. E. MORAIS, *Kronecker's and Newton's approaches to solving: a first comparison*, J. Complexity, 17 (2001), pp. 212–303.

[17] T. CHEN, T. LEE, AND T. LI, *Hom4ps-3: an numerical solver for polynomial systems using homotopy continuation methods*.

[18] D. COX, J. LITTLE, AND D. O'SHEA, *Ideals, varieties, and algorithms*, vol. 3, Springer, 1992.

[19] ——, *Using algebraic geometry*, vol. 185, Springer, 2006.

[20] R. E. CURTO AND L. FIALKOW, *Flat extensions of positive moment matrices: recursively generated relations*, Mem. Amer. Math. Soc, 136 (1998), p. 648.

[21] X. DAHAN AND É. SCHOST, *Sharp estimates for triangular sets*, in Proceedings of the 2004 international symposium on Symbolic and algebraic computation, ACM, 2004, pp. 103–110.

[22] B. H. DAYTON AND Z. ZENG, *Computing the multiplicity structure in solving polynomial systems*, in ISSAC'05, ACM, New York, 2005, pp. 116–123 (electronic).

[23] D. DECKER, H. KELLER, AND C. KELLEY, *Convergence rates for newton's method at singular points*, SIAM Journal on Numerical Analysis, 20 (1983), pp. 296–314.

[24] J. P. DEDIEU AND M. SHUB, *Newton's method for overdetermined systems of equations*, Math. Comp., 69 (2000), pp. 1099–1115.

[25] M. S. EL DIN AND L. ZHI, *Computing rational points in convex semialgebraic sets and sum of squares decompositions*, SIAM J. on Optimization, 20 (2010), pp. 2876–2889.

[26] H. R. P. FERGUSON, D. H. BAILEY, AND S. ARNO, *Analysis of PSLQ, an integer relation finding algorithm*, Math. Comp., 68 (1999), pp. 351–369.

[27] J. V. Z. GATHEN AND J. GERHARD, *Modern Computer Algebra*, Cambridge University Press, New York, NY, USA, 2 ed., 2003.

[28] M. GIUSTI, J. HEINTZ, K. HÄGELE, J. E. MORAIS, L. M. PARDO, AND J. L. MONTAÑA, *Lower bounds for Diophantine approximations*, J. Pure Appl. Algebra, 117/118 (1997), pp. 277–317. Algorithms for algebra (Eindhoven, 1996).

[29] M. GIUSTI, J. HEINTZ, J. E. MORAIS, J. MORGENSTERN, AND L. M. PARDO, *Straight-line programs in geometric elimination theory*, J. Pure Appl. Algebra, 124 (1998), pp. 101–146.

[30] M. GIUSTI, G. LECERF, AND B. SALVY, *A Gröbner free alternative for polynomial system solving*, J. Complexity, 17 (2001), pp. 154–211.

[31] A. GRIEWANK AND M. R. OSBORNE, *Analysis of Newton's method at irregular singularities*, SIAM J. Numer. Anal., 20 (1983), pp. 747–773.

[32] B. HASSETT, *Introduction to algebraic geometry*, Cambridge University Press, 2007.

[33] J. D. HAUENSTEIN, I. HAYWOOD, AND A. C. LIDDELL, JR., *An a posteriori certification algorithm for newton homotopies*, in Proceedings of the 39th International Symposium on Symbolic and Algebraic Computation, ISSAC '14, New York, NY, USA, 2014, ACM, pp. 248–255.

[34] J. D. HAUENSTEIN AND A. C. LIDDELL, *Certified predictor-corrector tracking for newton homotopies*, J. Symb. Comput., 74 (2016), pp. 239–254.

[35] J. D. HAUENSTEIN, B. MOURRAIN, AND A. SZANTO, *Certifying isolated singular points and their multiplicity structure*, in Proceedings of the 2015 ACM on International Symposium on Symbolic and Algebraic Computation, ACM, 2015, pp. 213–220.

[36] J. D. HAUENSTEIN, V. Y. PAN, AND A. SZANTO, *Global newton iteration over archimedian and non-archimedian fields*, in In: Proceedings of Computer Algebra in Scientific Computing (CASC 2014), Lecture Notes in Computer Science, Springer-Verlag, 2014.

[37] J. D. HAUENSTEIN AND F. SOTTILE, *Algorithm 921: alphaCertified: certifying solutions to polynomial systems*, ACM Trans. Math. Software, 38 (2012), pp. Art. ID 28, 20.

[38] J. D. HAUENSTEIN AND C. W. WAMPLER, *Isosingular sets and deflation*, Found. Comput. Math., 13 (2013), pp. 371–403.

[39] J. HEINTZ, T. KRICK, S. PUDDU, J. SABIA, AND A. WAISSBEIN, *Deformation techniques for efficient polynomial equation solving*, J. Complexity, 16 (2000), pp. 70–109. Real computation and complexity (Schloss Dagstuhl, 1998).

[40] J. HEINTZ, M.-F. ROY, AND P. SOLERNÓ, *Sur la complexité du principe de tarski-seidenberg*, Bulletin de la Société mathématique de France, 118 (1990), pp. 101–126.

[41] C. HERMITE, *Extrait d'une lettre de mr. ch. hermite de paris Ãă mr. borchardt de berlin sur le nombre des racines d'une Ãl'quation algÃl'brique comprises entre des limites donnÃl'es.*, Journal fÃijr die reine und angewandte Mathematik, 52 (1856), pp. 39–51.

[42] M. HINDRY AND J. H. SILVERMAN, *Diophantine geometry*, vol. 201 of Graduate Texts in Mathematics, Springer-Verlag, New York, 2000. An introduction.

[43] G. JERONIMO AND D. PERRUCCI, *On the minimum of a positive polynomial over the standard simplex*, J. Symbolic Comput., 45 (2010), pp. 434–442.

[44] G. JERONIMO, D. PERRUCCI, AND E. TSIGARIDAS, *On the minimum of a polynomial function on a basic closed semialgebraic set and applications*, SIAM J. Optim., 23 (2013), pp. 241–255.

[45] E. KALTOFEN, *Polynomial-time reductions from multivariate to bi- and univariate integral polynomial factorization*, SIAM J. Comput., 14 (1985), pp. 469–489.

[46] E. KALTOFEN, B. LI, Z. YANG, AND L. ZHI, *Exact certification of global optimality of approximate factorizations via rationalizing sums-of-squares with floating point scalars*, in Proceedings of the Twenty-first International Symposium on Symbolic and Algebraic Computation, ISSAC '08, New York, NY, USA, 2008, ACM, pp. 155–164.

[47] E. L. KALTOFEN, B. LI, Z. YANG, AND L. ZHI, *Exact certification in global polynomial optimization via sums-of-squares of rational functions with rational coefficients*, J. Symb. Comput., 47 (2012), pp. 1–15.

[48] R. KANNAN, A. K. LENSTRA, AND L. LOVÁSZ, *Polynomial factorization and nonrandomness of bits of algebraic and some transcendental numbers*, Math. Comp., 50 (1988), pp. 235–250.

[49] Y. KANZAWA, M. KASHIWAGI, AND S. OISHI, *An algorithm for finding all solutions of parameter-dependent nonlinear equations with guaranteed accuracy*, Electronics and Communications in Japan (Part III: Fundamental Electronic Science), 82 (1999), pp. 33–39.

[50] Y. KANZAWA AND S. OISHI, *Approximate singular solutions of nonlinear equations and a numerical method of proving their existence*, Sūrikaisekikenkyūsho Kōkyūroku, (1997), pp. 216–223. Theory and application of numerical calculation in science and technology, II (Japanese) (Kyoto, 1996).

[51] R. KRAWCZYK, *Newton-algorithmen zur bestimmung von nullstellen mit fehlerschranken*, Computing, 4 (1969), pp. 187–201.

[52] T. KRICK, L. M. PARDO, AND M. SOMBRA, *Sharp estimates for the arithmetic Nullstellensatz*, Duke Math. J., 109 (2001), pp. 521–598.

[53] L. KRONECKER, *Grundzüge einer arithmetischen theorie der algebraischen grössen... von L. Kronecker*, G. Reimer, 1882.

[54] M. LAURENT AND B. MOURRAIN, *A generalized flat extension theorem for moment matrices*, Archiv der Mathematik, 93 (2009), pp. 87–98.

[55] G. LECERF, *Quadratic newton iteration for systems with multiplicity*, Foundations of Computational Mathematics, 2 (2002), pp. 247–293.

[56] A. K. LENSTRA, H. W. LENSTRA, JR., AND L. LOVÁSZ, *Factoring polynomials with rational coefficients*, Math. Ann., 261 (1982), pp. 515–534.

[57] A. LEYKIN, J. VERSCHELDE, AND A. ZHAO, *Newton's method with deflation for isolated singularities of polynomial systems*, Theoret. Comput. Sci., 359 (2006), pp. 111–122.

[58] ——, *Higher-order deflation for polynomial systems with isolated singular solutions*, in Algorithms in algebraic geometry, vol. 146 of IMA Vol. Math. Appl., Springer, New York, 2008, pp. 79–97.

[59] N. LI AND L. ZHI, *Verified error bounds for isolated singular solutions of polynomial systems: Case of breadth one*, Theor. Comput. Sci., 479 (2013), pp. 163–173.

[60] N. LI AND L. ZHI, *Verified error bounds for isolated singular solutions of polynomial systems*, SIAM Journal on Numerical Analysis, 52 (2014), pp. 1623–1640.

[61] V. MAGRON, X. ALLAMIGEON, S. GAUBERT, AND B. WERNER, *Formal proofs for nonlinear optimization*, J. Formaliz. Reason., 8 (2015), pp. 1–24.

[62] A. MANTZAFLARIS AND B. MOURRAIN, *Deflation and certified isolation of singular zeros of polynomial systems*, in Proceedings of the 36th International Symposium on Symbolic and Algebraic Computation, ISSAC '11, New York, NY, USA, 2011, ACM, pp. 249–256.

[63] C. D. MEYER, *Matrix analysis and applied linear algebra*, vol. 2, Siam, 2000.

[64] D. MONNIAUX AND P. CORBINEAU, *On the generation of positivstellensatz witnesses in degenerate cases*, in Interactive Theorem Proving, M. van Eekelen, H. Geuvers, J. Schmaltz, and F. Wiedijk, eds., vol. 6898 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2011, pp. 249–264.

[65] R. MOORE, *A test for existence of solutions to nonlinear systems*, SIAM Journal on Numerical Analysis, 14 (1977), pp. 611–615.

[66] S. MORITSUGU AND K. KURIYAMA, *On multiple zeros of systems of algebraic equations*, in Proceedings of the 1999 International Symposium on Symbolic and Algebraic Computation, ISSAC '99, New York, NY, USA, 1999, ACM, pp. 23–30.

[67] Y. NAKAYA, S. OISHI, M. KASHIWAGI, AND Y. KANZAWA, *Numerical verification of nonexistence of solutions for separable nonlinear equations and its application to all solutions algorithm*, Electronics and Communications in Japan (Part III: Fundamental Electronic Science), 86 (2003), pp. 45–53.

[68] T. OJIKA, *Modified deflation algorithm for the solution of singular problems. I. A system of nonlinear algebraic equations*, J. Math. Anal. Appl., 123 (1987), pp. 199–221.

[69] T. OJIKA, S. WATANABE, AND T. MITSUI, *Deflation algorithm for the multiple roots of a system of nonlinear equations*, J. Math. Anal. Appl., 96 (1983), pp. 463–479.

[70] P. PEDERSEN, M.-F. ROY, AND A. SZPIRGLAS, *Computational Algebraic Geometry*, Birkhäuser Boston, Boston, MA, 1993, ch. Counting real zeros in the multivariate case, pp. 203–224.

[71] H. PEYRL AND P. A. PARRILO, *A macaulay 2 package for computing sum of squares decompositions of polynomials with rational coefficients.*, in SNC, 2007, pp. 207–208.

[72] H. PEYRL AND P. A. PARRILO, *Computing sum of squares decompositions with rational coefficients*, Theor. Comput. Sci., 409 (2008), pp. 269–281.

[73] K. H. ROSEN, *Elementary number theory and its applications*, Reading, Mass., 1993.

[74] F. ROUILLIER, *Solving zero-dimensional systems through the rational univariate representation*, Journal of Applicable Algebra in Engineering, Communication and Computing, 9 (1999), pp. 433–461.

[75] S. RUMP AND S. GRAILLAT, *Verified error bounds for multiple roots of systems of nonlinear equations*, Numerical Algorithms, 54 (2010), pp. 359–377.

[76] S. M. RUMP, *Solving algebraic problems with high accuracy*, in Proc. Of the Symposium on A New Approach to Scientific Computation, San Diego, CA, USA, 1983, Academic Press Professional, Inc., pp. 51–120.

[77] E. SCHOST, *Sur la résolution des systèmes polynomiaux à paramètres*, PhD thesis, 2000.

[78] A. SCHRIJVER, *Theory of linear and integer programming*, John Wiley & Sons, 1998.

[79] S. SMALE, *Newton's method estimates from data at one point*, Springer, 1986.

[80] A. J. SOMMESE AND C. W. WAMPLER, *The Numerical solution of systems of polynomials arising in engineering and science*, vol. 99, World Scientific, 2005.

[81] A. SZANTO, *Solving over-determined systems by the subresultant method*, J. Symbolic Comput., 43 (2008), pp. 46–74. With an appendix by Marc Chardin.

[82] A. SZANTO, *Lecture notes in ma722 computer algebra*, 2012.

[83] W. TRINKS, *On improving approximate results of buchberger's algorithm by newton's method*, in EUROCAL '85, B. Caviness, ed., vol. 204 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 1985, pp. 608–612.

[84] J. VAN DER HOEVEN, *Reliable homotopy continuation*, PhD thesis, LIX, Ecole polytechnique, 2015.

[85] J. VERSCHELDE, *Algorithm 795: Phcpack: A general-purpose solver for polynomial systems by homotopy continuation*, ACM Transactions on Mathematical Software (TOMS), 25 (1999), pp. 251–276.

[86] J. VON ZUR GATHEN AND J. GERHARD, *Modern computer algebra*, Cambridge University Press, New York, 1999.

[87] X. H. WANG AND D. F. HAN, *On dominating sequence method in the point estimate and smale theorem*, SCIENCE IN CHINA SERIES A-MATHEMATICS PHYSICS ASTRONOMY, 33 (1990), pp. 135–144.

[88] F. WINKLER, *A p-adic approach to the computation of Gröbner bases*, J. Symbolic Comput., 6 (1988), pp. 287–304. Computational aspects of commutative algebra.

[89] K. YAMAMURA, H. KAWATA, AND A. TOKUE, *Interval solution of nonlinear equations using linear programming*, BIT Numerical Mathematics, 38 (1998), pp. 186–199.

[90] Z. YANG, L. ZHI, AND Y. ZHU, *Verified error bounds for real solutions of positive-dimensional polynomial systems*, in Proceedings of the 38th International Symposium on International Symposium on Symbolic and Algebraic Computation, ISSAC '13, New York, NY, USA, 2013, ACM, pp. 371–378.