Validating the completeness of the real solution set of a system of polynomial equations

Daniel A. Brake*  
University of Notre Dame  
dbrake@nd.edu

Jonathan D. Hauenstein*  
University of Notre Dame  
hauenstein@nd.edu

Alan C. Liddell, Jr.*  
University of Notre Dame  
aliddel1@nd.edu

ABSTRACT

Computing the real solutions to a system of polynomial equations is a challenging problem, particularly verifying that all solutions have been computed. We describe an approach that combines numerical algebraic geometry and sums of squares programming to test whether a given set is “complete” with respect to the real solution set. Specifically, we test whether the Zariski closure of that given set is indeed equal to the solution set of the real radical of the ideal generated by the given polynomials. Examples with finitely and infinitely many real solutions are provided, along with an example having polynomial inequalities.

1. INTRODUCTION

Local solving methods, such as using Newton’s method, local optimization approaches, and critical point methods as highlighted in Section 5 allow one to compute real solutions to system of equations. The typical drawback of using such local methods is the inability to verify that all real solutions have been computed. This article uses sums of squares programming to validate that a complete real solution set has been computed, that is, the Zariski closure of the given set is equal to the Zariski closure of the set of all real solutions. The method follows in the footsteps of work combining numerical and symbolic methods, particularly [4, 18].

A typical situation where one may need to test the completeness of a real solution set is computing critical points. For example, [5, 6, 7] considers computing the critical points of a potential energy landscape. In such situations, local numerical methods, e.g., [20, 51], exist for locating real critical points. Our approach provides a global stopping criterion for validating that all real solutions have been identified.

A related situation is the computation of the real critical points of a projection of a solution set used in the numerical decomposition of real curves and surfaces [3, 11, 14, 17]. The failure to correctly compute the set of real solutions leads to a failure in the decomposition of the real component. Hence, correct and complete computation of sets of real solutions is paramount to correctly computing the decomposition.

One approach for certifying the existence of real solutions is based on the local analysis of Newton’s method using Smale’s α-theory [60] developed in [33]. Building on α-theory, there are methods for certifying smooth continuous paths for Newton homotopies [29, 30] and general homotopies [10]. For example, if a smooth path is defined by a real system of equations which has a real starting point, then the endpoint of the path must also be real.

From an algebraic viewpoint, the radical of an ideal generated by a given collection of polynomials consists of all polynomials that vanish on the solution set of the given polynomials. There are several algorithms for computing the radical of a zero-dimensional ideal – some numerical, e.g., [35, 12, 43] and some symbolic, e.g., [9, 22]. When there are infinitely many solutions, one can reduce to the zero-dimensional case, for example, via [22, 39].

The real radical of an ideal generated by a given collection of polynomials with real coefficients consists of all polynomials that vanish on the real solution set of the given polynomials. There have been several proposed methods for computing the real radical of an ideal. Some are symbolic, e.g., [8] based on the primary decomposition (see also [54, 64, 66, 67]). Others are numerical, based on moment matrices when the number of real solutions is finite, e.g., [40, 41, 42, 43, 44]. A promising approach for computing the real radical when there are infinitely many real solutions was developed in [49] providing a stopping criterion for verifying that a Pommaret basis has been computed. Other methods for computing real solutions include computing a point on each semi-algebraically connected component of the real solution set, e.g., [1, 2, 27, 57].

As discussed in [49], one key issue related to computing the real radical using semidefinite programming with moment matrices is knowing when the generated polynomials form a basis for the real radical. In our approach, we first compute a set S which is a subset of the Zariski closure of the real solution set. Then, we compute polynomials that vanish on S. Finally, for each of the computed polynomials, we use sums of squares programming to verify that it is indeed in the real radical. Since the polynomials can be validated independently, one could easily parallelize this part of the computation. Since S is contained in the Zariski closure of the real solution set, every polynomial contained in the real radical vanishes on S. Conversely, if every polynomial
that vanishes on $S$ is contained in the real radical, we know that a generating set for the real radical has been computed. Hence, $S$ is complete since the Zariski closure of $S$ is equal to the Zariski closure of the real solution set of the original system of equations, i.e., the solution set of the real radical.

One of the pitfalls of using purely symbolic methods to compute real radicals is the typical computation of field extensions. As an illustrative example, consider the polynomial $f(x) = x^3 - 2$ having rational coefficients, i.e., $f \in \mathbb{Q}[x]$. Since $f = 0$ has one real solution, namely $x = \sqrt[3]{2}$, the real radical of the ideal generated by $f$ is $(x - \sqrt[3]{2})$ which is generated by a polynomial not in $\mathbb{Q}[x]$.

By working over $\mathbb{R}$, one avoids the use of field extensions for computing the generators of the real radical. Thus, the theoretical results for our approach assume an exact computational model over $\mathbb{R}$ as in [12]. The drawback is that the computations in practice are performed using floating-point arithmetic. In the examples, we approximate the computational model by utilizing adaptive precision floating-point computations where each point in the subset $S$ of the solution set is described via a numerical approximation and computations in practice are performed using floating-point arithmetic. In the examples, we approximate the computations where each point in the subset $S$ of the solution set is described via a numerical approximation and computations in practice are performed using floating-point arithmetic. In the examples, we approximate the computations where each point in the subset $S$ of the solution set is described via a numerical approximation and computations in practice are performed using floating-point arithmetic. In the examples, we approximate the computations where each point in the subset $S$ of the solution set is described via a numerical approximation and computations in practice are performed using floating-point arithmetic.

The remainder of the article is as follows. Section 2 focuses on radicals, irreducible decomposition, and Zariski closures. Real radicals, sums of squares, and semidefinite programming are discussed in Section 3. In Section 4 we present the criterion for showing that a set $S$ is complete with respect to the real radical. This approach depends on the ability to compute such an $S$, which is highlighted in Section 5, and the ideal of $S$, which is discussed in Section 6. Section 7 considers the real solution set for collection of equations and inequalities. Several examples are presented in Section 8 with a discussion of the limitations of our approach in Section 9. We conclude in Section 10.

Acknowledgments

The authors thank Mohab Safey El Din, Charles Wampler, and Lihong Zhi for discussions related to real solution sets, as well as the reviewers for their helpful comments.

2. ZARISKI CLOSURE AND RADICALS

Let $f_1, \ldots, f_k \in \mathbb{C}[x_1, \ldots, x_n]$ and consider the ideal generated by these polynomials, namely $I = \langle f_1, \ldots, f_k \rangle$. The polynomials $f = \{f_1, \ldots, f_k\}$ and the corresponding ideal $I = \langle f \rangle$ define the same solution set in $\mathbb{C}^n$, namely $V_C(I) = \{x \in \mathbb{C}^n \mid f_i(x) = 0 \text{ for } i = 1, \ldots, k\}$.

A set $A \subset \mathbb{C}^n$ is called an algebraic set if there is a collection of polynomials $g_1, \ldots, g_m \in \mathbb{C}[x_1, \ldots, x_n]$ such that $A = V_C(g)$. The algebraic set $A$ is irreducible if there does not exist algebraic sets $A_1, A_2 \subseteq A$ with $A = A_1 \cup A_2$. Given an algebraic set $A$, there exists a unique collection (up to relabeling) of irreducible algebraic sets $X_1, \ldots, X_t$ such that $A = \bigcup_{i=1}^t X_i$ and $X_j \not\subseteq \bigcup_{j \neq i} X_i$.

Each $X_i$ is called an irreducible component of $A$.

In numerical algebraic geometry, an irreducible algebraic set is represented by a witness set, see, e.g., [63] Chap. 13. A numerical irreducible decomposition for an algebraic set $A$ is a collection of witness sets for the irreducible components of $A$. Such a decomposition can be computed using various algorithms, e.g., [6], [61], [62].

For any subset $T \subset \mathbb{C}^n$, the ideal defined by $T$ is $I(T) = \{f \in \mathbb{C}[x_1, \ldots, x_n] \mid f(t) = 0 \text{ for all } t \in T\}$.

The Zariski closure of $T$ is the algebraic set $T = \mathcal{V}_C(I(T))$, which is the intersection of all algebraic sets that contain $T$.

For an ideal $I$, the radical of $I$ is $\sqrt{I} = \{p \in \mathbb{C}[x_1, \ldots, x_n] \mid p^\alpha \in I \text{ for some } \alpha \in \mathbb{Z}_{>0}\}$ with $\sqrt{I} = I(\mathcal{V}_C(I))$ following Hilbert’s Nullstellensatz.

3. REAL RADICAL & SUMS OF SQUARES

Many of the topics from §2 have analogous statements over $\mathbb{R}$. Let $f_1, \ldots, f_k \in \mathbb{R}[x_1, \ldots, x_n]$ with $f = \{f_1, \ldots, f_k\}$ and $I = \langle f \rangle$. The set of solutions in $\mathbb{R}^n$ is $V_R(f) = V_C(I) \cap \mathbb{R}^n$.

The real radical of $I$ is

$$\sqrt{I} = \left\{ p \in \mathbb{R}[x] \mid p^{2^\alpha} + \sum_{j=1}^t g_j^2 \in I \text{ for some } \alpha \in \mathbb{Z}_{>0}, g_j \in \mathbb{R}[x] \right\}.$$  

with $\sqrt{I} = I(\mathcal{V}_C(I))$ following the real Nullstellensatz, e.g., see [13] Chap. 4.

**Example 1.** For $f(x) = x^3 - 2$ and $I = \langle f \rangle$, we have:

- $V_C(I) = \{\sqrt[3]{2}, \sqrt[3]{2} \omega, \sqrt[3]{2} \omega^2\}$ and $V_R(I) = \{\sqrt[3]{2}\}$.
- $\sqrt{I} = I$ and $I = (x - \sqrt[3]{2})$.

where $\omega$ is the primitive cube root of unity. In particular,

$$(x - \sqrt[3]{2})^4 + (\sqrt[3]{3}x^2 - \sqrt[3]{3} \sqrt{2})^2 = 4(x^3 - 2)(x - \sqrt[3]{2}) \in I.$$  

The algebraic description of the real radical $\sqrt{I}$ presented in [4] shows that this definition depends on sums of squares. A polynomial $s \in \mathbb{R}[x_1, \ldots, x_n]$ is called a sum of squares if $s = \sum_{j=1}^t g_j^2$ for some $g_1, \ldots, g_t \in \mathbb{R}[x_1, \ldots, x_n]$. Clearly, every polynomial that is a sum of squares has even degree.

The polynomials of even degree that are sums of squares are characterized by positive semidefinite matrices. A symmetric matrix $M \in \mathbb{R}^{m \times m}$ is positive semidefinite if, for all $y \in \mathbb{R}^m$, $y^T M y \geq 0$. This condition is equivalent to all eigenvalues of $M$ being nonnegative. We will write $M \succeq 0$ if $M$ is positive semidefinite.

Let $s \in \mathbb{R}[x_1, \ldots, x_n]$ be a polynomial of degree $2d$ and $X_d$ the vector of all monomials in $x_1, \ldots, x_n$ of degree at most $d$. Hence, there exists a symmetric matrix $C$ such that

$$s(x) = X_d^T \cdot C \cdot X_d.$$  

The polynomial $s$ is a sum of squares if and only if there is a positive semidefinite matrix $C$ such that [2] holds, e.g., [16].

**Example 2.** As shown in Ex. 1 the quartic polynomial $s(x) = 4(x^3 - 2)(x - \sqrt[3]{2})$ is a sum of squares. Let

$$X_2 = \begin{bmatrix} 1 & x \\ x^2 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 8 \sqrt[3]{2} & -4 & -2 \sqrt[3]{2} \\ -4 & 4 \sqrt[3]{2} & -2 \sqrt[3]{2} \\ -2 \sqrt[3]{2} & -2 \sqrt[3]{2} & 4 \end{bmatrix}.$$  

It is easy to verify that $C \succeq 0$ and $s(x) = X_2^T \cdot C \cdot X_2$. 
For a given polynomial \( s \) of degree \( 2d \), the set of symmetric matrices \( C \) such that \( [2] \) holds is a linear space. Hence, testing that a polynomial is a sum of squares can be accomplished by solving a semidefinite feasibility problem.

**Example 3.** Continuing with \( s(x) = 4(x^3 - 2)(x - \sqrt{2}) \) from Ex. 2, consider the linear space

\[
\mathcal{L} = \left\{ \begin{bmatrix} s_{00} & s_{01} & s_{02} \\ s_{10} & s_{11} & s_{12} \\ s_{20} & s_{21} & s_{22} \end{bmatrix} \right| \begin{array}{c} s_{00} = 8 \sqrt{2} \\ 2s_{01} = -8 \\ 2s_{02} + s_{11} = 0 \\ 2s_{12} = -4 \sqrt{2} \\ s_{22} = 4 \end{array} \right\}.
\]

Since \( s(x) = X^2 \cdot C \cdot X \) if and only if \( \alpha \in \mathcal{L} \), it follows that \( s \) is a sum of squares if and only if there exists \( C \in \mathcal{L} \) such that \( \alpha \geq 0 \), which is a semidefinite feasibility problem.

To facilitate the task of converting between sums of squares problems and semidefinite programming problems in the examples, we utilize the software package **SOSTools** [5].

Given a polynomial \( p \in \mathbb{R}[x_1, \ldots, x_n] \), we can decide if \( p \in \sqrt{T} \) using [1]. That is, \( p \in \sqrt{T} \) if and only if there exists \( \alpha \in \mathbb{Z}_{\geq 0} \) and \( h_1, \ldots, h_k, g_1, \ldots, g_l \in \mathbb{R}[x_1, \ldots, x_n] \) such that

\[
p^{2\alpha} + \sum_{j=1}^l g_j^2 = \sum_{i=1}^k h_if_i,
\]

which is equivalent to requiring that

\[-p^{2\alpha} + \sum_{i=1}^k h_if_i \text{ is a sum of squares.} \quad (3)\]

Thus, given a polynomial \( p \in \mathbb{R}[x_1, \ldots, x_n] \), one can test if \( p \in \sqrt{T} \) by solving a semidefinite feasibility problem. We use this observation in our method, which is described next.

### 4. VALIDATION

Given an ideal \( I \subset \mathbb{R}[x_1, \ldots, x_n] \), \( S \subset \mathcal{V}_\mathbb{R}(I) = \mathcal{V}_C(\sqrt{T}) \), and \( I(S) = (g_1, \ldots, g_l) \), we describe an approach for validating that \( S \) is complete, i.e., \( S = \mathcal{V}_\mathbb{R}(I) \) which occurs if and only if \( \sqrt{T} = (g_1, \ldots, g_l) = I(S) \).

**Procedure 1** Validating Real Solution Sets

**Input:** Polynomials \( f = \{f_1, \ldots, f_k\} \subset \mathbb{R}[x_1, \ldots, x_n] \), a set \( S \subset \mathcal{V}_\mathbb{R}(I) \) with \( I(S) = (g_1, \ldots, g_l), \) integer \( \alpha_{\max} \in \mathbb{Z}_{\geq 0} \).

**Output:** A boolean which is True if \( I(S) = \sqrt{T}(f) \) can be validated up to \( \alpha_{\max} \), otherwise False.

1. (Optional) Replace \( f \) with a Gröbner basis for \( \sqrt{T}(f) \).
2. for \( m = 1, \ldots, k \) do
3.    Initialize \( \alpha := 1 \) and \( \text{success} := \text{FALSE} \).
4.    while \( \text{success} := \text{FALSE} \) do
5.      if there exists \( h_i \in \mathbb{R}[x_1, \ldots, x_n] \) such that the polynomial \( q = -g_m^{2\alpha} + \sum h_if_i \) is a sum of squares then
6.        Set \( \text{success} := \text{TRUE} \).
7.      else
8.        Increment \( \alpha := \alpha + 1 \).
9.        if \( \alpha > \alpha_{\max} \) then
10.         return \( (S, \text{FALSE}) \).
11.     return \( (S, \text{TRUE}) \).

**Theorem 4.** Procedure 1 is a correct algorithm.

**Proof.** Let \( I = (f_1, \ldots, f_k) \). Since \( \mathcal{V}_\mathbb{R}(I) = \mathcal{V}_\mathbb{R}(\sqrt{T}) \), replacing \( f \) by a Gröbner basis for \( \sqrt{T} \) does not impact the set of real solutions.

For a given \( p \in \mathbb{R}[x_1, \ldots, x_n] \), we know \( p \in \sqrt{T} \) if and only if there exists \( \alpha \in \mathbb{Z}_{\geq 0} \) and \( h_1, \ldots, h_k \in \mathbb{R}[x_1, \ldots, x_n] \) such that \( [3] \) holds. In particular, since \( S \subset \mathcal{V}_\mathbb{R}(I) \), we know \( \sqrt{T} \subset I(S) = (g_1, \ldots, g_l) \) so that \( \sqrt{T} = I(S) \) if and only if \( [2] \) holds for each \( g_i \). If all of the corresponding \( \alpha \)'s for each \( g_i \) are at most \( \alpha_{\max} \), then Procedure 1 will correctly determine \( I(S) = \sqrt{T} \).

If \( p \not\in \sqrt{T} \), then for every \( \alpha \in \mathbb{Z}_{\geq 0} \), \( [3] \) does not hold. In Procedure 1 we use the upper bound \( \alpha_{\max} \) so that Procedure 1 always terminates.

If one could compute an a priori upper bound on the largest possible value for \( \alpha \), then we could replace \( \alpha_{\max} \) with this bound. However, without such a bound, we simply keep searching for new points to add to \( S \), which is described next. If \( p \not\in \sqrt{T} \), then there must exist a point \( x \in \mathcal{V}_\mathbb{R}(I) \) such that \( p(x) \neq 0 \). In fact, there is an irreducible component \( X \subset \mathcal{V}_\mathbb{R}(I) = \mathcal{V}_C(\sqrt{T}) \) such that \( p(x) \neq 0 \) for every \( x \) in a dense open subset of \( X \).

We have implemented this procedure using floating-point arithmetic computations for several example systems while a full general-purpose implementation is left for future work.

### 5. GENERATING A CANDIDATE SET

**Input** for Procedure 1 is a set \( S \) and generators for \( I(S) \). Hence, in this section we provide techniques for generating \( S \) with the next focusing on computing \( I(S) \).

#### 5.1 Approaches for locating real solutions

A classical approach for attempting to find a real solution is to use Newton’s method or related variants, see, e.g., [28]. For a polynomial system with real coefficients, if the initial point is real, then every solution obtained from Newton’s method is also real. Of course, there are many challenges associated with finding real solutions using Newton’s method, particularly when \( \mathcal{V}_C(f) \) is not a complete intersection or the real solutions are singular with respect to \( f \). That is, problems can occur with Newton’s method, e.g., divergence, if the dimension of the solution set is less than dimension of the null space of the Jacobian at the solution [23, 24]. Nonetheless, heuristic techniques such as damping methods, reusing Jacobians for several iterations, or using chord or secant methods can be utilized [28].

Another approach for computing real solutions is to utilize numerical optimization techniques. Standard iterative techniques include those based on nonlinear least squares approaches such as the Levenberg-Marquardt algorithm and alternating least squares [37]. Other standard methods in optimization include the worker bees method, genetic algorithms, and the Nelder-Mead method, see, e.g., [17].

Critical point methods combine optimization and polynomial system solving techniques. For example, Seidenberg [58] considered the critical points of the distance function between the set of real solutions and a given real point \( y^* \) that was not a solution. The set of all such critical points contains a point on every connected component of the real solution set [1, 58, 58]. By utilizing homotopy continuation, one can compute a finite subset of critical points containing

1Available at \texttt{www.nd.edu/~aliddell1/validate-reals}
a point on every connected component \([27]\). Moreover, one can then sample more real points by moving \(y^*\).

 Rather than compute all critical points, one can attempt to compute the closest critical point to the given \(y^*\). This can be accomplished using a classical optimization approach such as the gradient descent method or a homotopy-based approach called gradient descent homotopies \([25]\). By testing at many values of \(y^*\), one aims to quickly generate many real solutions, e.g., as shown in \([25, \text{Fig. } 3]\).

 Other so-called “local” solving methods exist for finding real solutions, which have been used in various disciplines. Some examples include techniques in theoretical chemistry, e.g., \([20, 51, 52]\) and solving power-flow equations in electrical engineering, e.g., \([45, 48]\).

### 5.2 Real solutions and isosingular sets

To extract additional information about the geometry of the solution set near this point. One approach is to compute a local irreducible decomposition using local witness sets \([15]\) to improve the conditioning of interpolation. Moreover, for positive-dimensional components, sampling points that are spread out over the component using numerical algebraic geometry as in \([52]\) also helps to improve conditioning.

**Example 5.** The solution set of the polynomial system

\[ f = \{x^2 + y^2 + z^2 - 1, \, x^2 + y^2 + z - 1, \, x\} \]  

consists of the three points

\[ \mathcal{V}_C(f) = \mathcal{V}_F(f) = \{(0, 1, 0), (0, -1, 0), (0, 0, 1)\} \]

where the point \((0, 0, 1)\) has multiplicity two with respect to \(f\).

To illustrate, for \(d = 2\), we choose the monomial basis

\[ B = \{1, x, y, z, x^2, yx, zy, y^2, yz, z^2\} \]

for \(\mathbb{R}[x, y, z]_{\leq 2}\) with \(S = T = \mathcal{V}_F(f)\) where \(M\) is

\[
\begin{array}{cccccccccccc}
1 & x & y & z & x^2 & xy & xz & y^2 & yz & z^2 \\
(0, 1, 0) & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
(0, -1, 0) & 1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 \\
(0, 0, 1) & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
\end{array}
\]

A basis for \(M\) is given by the columns of the matrix

\[
\begin{bmatrix}
0 & 0 & 0 & 0 & -1 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & -1 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 \\
\end{bmatrix}
\]

corresponding to the polynomials

\[x, \, x^2, \, xy, \, xz, \, y^2 + z - 1, \, yz, \, z^2 - z\]
which form a basis for the linear space $\sqrt{I}\leq 2$. Note that since each polynomial $f_i$ has degree at most 2, each $f_i$ is contained in the linear span of these polynomials.

For illustrative purposes, we selected a monomial basis. In practice, the choice of basis should be made based on numerical conditioning.

For $d \gg 0$, we know $I(S) = (I(S)_{\leq d})$. If $S$ is a finite set, then one can determine an upper bound on $d$ such that $I(S)$ is generated by $I(S)_{\leq d}$. In particular, the function

$$c \mapsto \dim \mathbb{R}[x_1, \ldots, x_n]_{\leq c} - \dim I(S)_{\leq c}$$

is the Hilbert function of $I(S)$. If $r$ is the minimum such that $|S| = \dim \mathbb{R}[x_1, \ldots, x_n]_{\leq r} - \dim I(S)_{\leq r}$, i.e., the index of regularity, then one knows that $I(S)$ is either generated by $I(S)_{\leq r}$ or $I(S)_{\leq r+1}$. In fact, $I(S)_{\leq r}$ generates $I(S)$ if and only if $(I(S)_{\leq r})_{\leq r+1} = I(S)_{\leq r+1}$, i.e., the Hilbert function of $J = (I(S)_{\leq r})$ in degree $r+1$ is also equal to $|S|$.

**Example 6.** Continuing with Ex. 2 since

$$\dim \mathbb{R}[x, y, z]_{\leq 2} - \dim I(S)_{\leq 2} = 10 - 7 = 3 = |S|,$$

one can easily verify that $I(S)$ is generated by $I(S)_{\leq 2}$, i.e.,

$$\sqrt{I} = \langle x, y^2 + z, 1, yz, z^2 - z \rangle.$$

**Example 7.** The Hilbert function for the ideal $I(S)$ where $S = \{(0, 0), (0, 1), (1, 0)\}$ is 1, 3, 5, 6, . . . so that $I(S)$ is either generated by $I(S)_{\leq 1}$ or $I(S)_{\leq 2}$. Since $I(S)_{\leq 1} = \{0\}$, we know that $I(S)_{\leq 2}$ must generate $I(S)$.

When $S$ is infinite, we aim to reduce our computations to standard computations performed over $\mathbb{C}$ as summarized in § 2. In particular, by using isosingular sets as discussed in § 2.2, we can actually assume that $S = \mathbb{R}$ and that we have a numerical irreducible decomposition of $S$. Hence, we simply need to compute $d$ large enough so that $S$ and $\mathcal{V}(\mathcal{I}(S)_{\leq d})$ have the same irreducible components so that $S \subseteq \mathcal{V}(\mathcal{I}(S)_{\leq d})$. Hence, $I(S) = \sqrt{I(S)_{\leq d}}$.

### 7. Equalities and Inequalities

One can naturally generalize from real radicals of systems of polynomial equations to $A$-radicals of systems of polynomial equations and inequalities. In particular, let $f_1, \ldots, f_k, r_1, \ldots, r_s \in \mathbb{R}[x_1, \ldots, x_n]$ with

$I = (f_1, \ldots, f_k)$ and $A = \{ x \in \mathbb{R}^n \mid r_i(x) \geq 0 \text{ for all } i = 1, \ldots, s \}$.

The $A$-radical of $I$ is $\sqrt{A} = I(\mathcal{V}_A(I) \cap A)$. Algebraically, one can characterize $\sqrt{A}$ using sums of squares $\lbrack 50 \rbrack$ $\lbrack 65 \rbrack$.

$$\sqrt{A} = \left\{ p \in \mathbb{R}[x] \mid \begin{array}{l}
p^{2^\circ} + \sum_{\sigma \in \Sigma} \prod_{i=1}^r r_i^{\sigma_i} \in I \\
\text{for some } \sigma \in \Sigma_{\geq 0}, \text{sum of squares } \sigma \sigma_i \in \mathbb{R}[x]
\end{array} \right\}. \quad (5)$$

Rather than try to locate sample points that satisfy equalities and inequalities, we will instead reduce to equations by introducing “slack” variables. That is, we consider the ideal

$$J = \langle f_1(x), \ldots, f_k(x), r_1(x) - y_1, \ldots, r_s(x) - y_s \rangle.$$ 

Since $\mathcal{V}_A(I) \cap A = \pi(\mathcal{V}_A(J))$ where $\pi(x,y) = x$, we know

$$\sqrt{A} = \sqrt{J} \cap \mathbb{R}[x_1, \ldots, x_n]. \quad (6)$$

Thus, we compute $S \subseteq \mathcal{V}_A(J)$ but only perform interpolation on $\pi(S)$. If $(g_1, \ldots, g_l) = I(\pi(S)) \subseteq \mathbb{R}[x_1, \ldots, x_n]$ and each $g_i \in \sqrt{J}$, then $I(\pi(S)) = \sqrt{J}$ by (6).

### 8. Examples

As mentioned in the Introduction, the following examples were computed using floating-point arithmetic. By using approaches described in Section 5, each numerical approximation of a solution naturally comes with an algorithm which can be used to refine the approximation to arbitrary accuracy. The scripts needed to run these examples are available at [www.nd.edu/~aliddel1/validate-reals](http://www.nd.edu/~aliddel1/validate-reals).

#### 8.1 An illustrative example

To illustrate our approach, we consider the intersection of a circle and a bivariate cubic, namely

$$f = \{x^2 + y^2 - 2, 2xy^2 - x + 1\}.$$

The system $f = 0$ has six solutions, all of which are real:

$$\mathcal{V}(f) = \{(-1, \pm 1), (1.366, \pm 0.366), (-0.366, \pm 1.366)\}.$$

In our first test, we simply take $S = \mathcal{V}(f)$. Since the Hilbert function of $I(S)$ is 1, 3, 5, 6, . . . we can show that $I(S)$ is generated by $I(S)_{\leq 2}$. A basis for the linear space $I(S)_{\leq 3}$, computed as in § 6 is:

$$G = \left\{ y^3 - 2xyy - 2y, xy - x^2 + 2, x^3 - 2x^2 - x, x^2 + y^2 - 2 \right\}.$$ 

Using either $f$ or a Gröbner basis for $f$, e.g.,

$${x^2 + y^2 - 2, 2xy^2 - x + 1, 2y^4 - 5y^2 - x + 2}, \quad (7)$$

every $g \in G$ was found to be in $\sqrt{\langle f \rangle}$ showing that $S$ is indeed equal to $\mathcal{V}(f)$.

#### Incomplete solution set

Suppose that we take $R = \mathcal{V}(f) \cap \{y \geq 0\}$. Since the Hilbert function of $I(R)$ is 1, 3, 3, . . . and $I(R)_{\leq 1} = \{0\}$, we know that $I(R)$ is generated by three quadratics, approximately

$$G = \left\{ y^2 - 2.049y - 0.86603, xy - 0.18301y + 0.68301x + 1/2, x^2 + 0.18301x + 2.049y - 2.866 \right\}. \quad (8)$$

Using $\alpha_{\max} = 5$, we were unable to validate that any of the polynomials in $G$ where in $\sqrt{\langle f \rangle}$. In fact, we can show that this is indeed correct since each polynomial in $G$ is nonzero at each of the three points in $\mathcal{V}(f) \setminus R$.

#### Semialgebraic condition

We now validate that $R = \mathcal{V}(f) \cap \{y \geq 0\}$ is the complete solution set for the $A$-radical of $f$ where $A = \{y \geq 0\}$. To that end, we add a slack variable $z$ and consider the system

$$F = \{x^2 + y^2 - 2, 2xy^2 - x + 1, y - z^2\}.$$

As described in § 7.2, we just need to show that each polynomial in $G$ from (8) is contained in $\sqrt{F}$. Using either $F$ or a Gröbner basis for $F$, namely (7) together with $y - z^2$, we validated that $G \subseteq \sqrt{F}$ showing that $R$ is indeed equal to $\mathcal{V}(f) \cap \{y \geq 0\}$, i.e., $\sqrt{\langle f \rangle} = I(R)$.

#### 8.2 Positive-dimensional components

To illustrate the approach on a system such that the real radical ideal is positive-dimensional, consider the system

$$f = \{xyz, z(x^2 + y^2 + z^2 + y), y(y + z)\}.$$
The set $V_\mathfrak{c}(f)$ consists of three lines, two of which are complex conjugates of each other that intersect at the origin and the other is a double line with respect to $f$, and an isolated point. In particular, $V_\mathfrak{b}(f)$ is the line $y = z = 0$ and the isolated point $(0, -1/2, 1/2)$. So, we take

$$S = \{(x, 0, 0) \mid x \in \mathbb{C}\} \cup \{(0, -1/2, 1/2)\} \subset V_\mathfrak{b}(f).$$

To simplify the real computations later, we first replace $f$ with a Gröbner basis for the radical $\sqrt{f}$, namely

$$f = \{2yz - y, 2y^2 + y, xy, 4x^3z + 4y^3 + y\}.$$

With the isolated solution, sampling 3 points on the line is enough to compute a basis for $I(S) \leq 2$ which generates $I(S)$:

$$G = \{z^2 + y/2, yz - y/2, y^2 + y/2, xz, xy, y + z\}.$$

Each element in $G$ was shown to belong to $\sqrt{f}$ with $\alpha \leq 2$.

### 8.3 Katsura-5 system

As an illustration of our approach on a problem which was solved using the semidefinite characterization of the real radical in [41], we consider the Katsura-5 system as in [41 Ex. 5.4]. The system consists of a linear, say $f_1$, and five quadratics, say $f_2, \ldots, f_6$, in six variables, namely

$$f = \begin{cases}
    x_1 + 2(x_2 + x_3 + x_4 + x_5 + x_6) - 1, \\
    x_2^2 + 2(x_2^2 + x_3^2 + x_4^2 + x_5^2 + x_6^2), \\
    2(x_1 x_2 + x_1 x_3 + x_1 x_4 + x_1 x_5 + x_1 x_6 - x_2), \\
    x_3^2 + 2(x_1 x_3 + x_2 x_3 + x_3 x_5 + x_3 x_6 - x_3), \\
    2(x_2 x_4 + x_2 x_5 + x_2 x_6 + x_3 x_4 - x_4), \\
    x_4^2 + 2(x_1 x_4 + x_1 x_5 + x_1 x_6 - x_5),
\end{cases}$$

The set $V_\mathfrak{c}(f)$ consists of 32 points, 12 of which lie in $\mathbb{R}^6$. The set of real solutions, say $S$, is readily computed using homotopy continuation.

The Hilbert function is $1, 6, 12, 12, \ldots$ with $I(S)$ being generated by $I(S) \leq 2$. In particular, $I(S) \leq 2$ is a linear space spanned by the linear $f_1$ and 15 quadratics.$^7$

Trivially, $f_1 \in \sqrt{\langle f_1, \ldots, f_6 \rangle}$ and the quadratics are shown to be in the real radical using $\alpha \leq 2$. This computation validates that $V_\mathfrak{b}(f)$ consists of 12 points. Moreover, this data matches that displayed in [41 Table 4].

### 8.4 High degree

Our next example was also presented in [41] which was modeled after a system originally from [36] Sec 4.3:

$$f = \left\{ \begin{array}{l}
5x_1^2 - 6x_1^2x_2 + x_1x_2^2 + 2x_1x_3 \\
-2x_1^2x_2 + 2x_1^2x_3 + 2x_2x_3 \\
x_1^2 + x_3^2 - 0.265625
\end{array} \right\}.$$

We note that the original system in [36] had the term $-6x_1^2x_2^2$ while the system from [41] has the term $-6x_1x_2x_3$. The system $f = 0$ has 8 real solutions among the 20 complex solutions. Our computations generate a set consisting of 2 quadratics and 10 cubic polynomials for the ideal defined the 8 real solutions which are shown to be in the real radical using $\alpha \leq 2$.

### 8.5 Seiler system

As an illustration of our approach on a problem discussed in [49] Ex. 5, namely the Seiler system [59]

$$f = \begin{cases}
    x_1^4 + x_2x_3 - x_1^2, \\
x_1x_3 + x_1x_2 - x_3, \\
x_2x_3 + x_2^2 + x_3 - x_1
\end{cases}.$$

$^7$Available at [www.nd.edu/~aliddel1/validate-reals](http://www.nd.edu/~aliddel1/validate-reals)

---

**Figure 1:** Nearest-neighbor coupling for a $3 \times 3$ grid of nodes.

This system does not have a Pommaret basis with respect to the total degree ordering defined by $x_1 < x_2 < x_3$ [59]. Thus, [19] uses a change of coordinates to overcome this.

Even though $f$ consists of 3 polynomials in 3 variables, $V_\mathfrak{c}(f)$ is actually a curve. In particular, $I = \langle f \rangle$ is a one-dimensional prime ideal, i.e., $I = \sqrt{T}$ and $V_\mathfrak{c}(I)$ is an irreducible curve. Hence, we know that $I = \sqrt{T}$ if we can compute a real point $x \in V_\mathfrak{b}(I)$ which is smooth with respect to $f$, i.e., the rank of $Jf(x)$ is 2.

To that end, we utilize a gradient descent homotopy [25]. We took $y = (1, -3/2, 4)$ and considered the homotopy

$$H(x, \lambda, t) = \left[\lambda_0(x - y) + \lambda_1 \nabla f_1(x) + \lambda_2 \nabla f_2(x) + \lambda_3 \nabla f_3(x)\right].$$

where $\lambda \in \mathbb{P}^3$. Starting at $x = y$ and $\lambda = [1, 0, 0, 0] \in \mathbb{P}^3$ when $t = 1$, we obtain a point, which is approximately $(0.7009, -0.2504, -0.5868)$, that lies on $V_\mathfrak{b}(f)$ and is indeed a smooth point on $V_\mathfrak{c}(f)$. Hence, the isosingular set of this point with respect to $f$ is $V_\mathfrak{c}(f)$ showing that $I = \sqrt{T}$.

### 8.6 An energy landscape

Our last example aims to compute the real critical points of the energy landscape of the two-dimensional nearest-neighbor $\varphi^4$ model on a $3 \times 3$ grid as in [21] [53]. We label the nodes 1, \ldots, 9 with Figure 1 showing the coupling between the nodes. Let $N(i)$ denote the four nearest neighbors of node $i$, e.g., $N(1) = \{2, 3, 4, 7\}$. After selecting various parameters for this model, we consider the potential energy

$$V(x) = \sum_{i=1}^9 \left[ \frac{1}{40} x_i^4 - x_i^2 + \frac{1}{4} \sum_{j \in N(i)} (x_i - x_j)^2 \right].$$

The system defining the critical points is $f = \nabla V$ so that

$$f_i = \frac{1}{10} x_i^3 - 2x_i + \sum_{j \in N(i)} (x_i - x_j).$$

The system $f$ is a Gröbner basis and the set $V_\mathfrak{c}(f)$ consists of $3^9 = 19,683$ points. However, when searching for real stationary points, one only obtains 3 points, namely

$$S = \{(0, 0, 0, 0, 0, 0, 0, 0, 0)\} \pm (w, w, w, w, w, w, w, w, w)$$

where $w = \sqrt{30} \approx 4.4721$. Hence, $I(S)$ is generated by

$$G = \{x_1(x_1^2 - 20), x_2 - x_3, \ldots, x_9 - x_1\}.$$

All nine basis elements were found to be in $\sqrt{\langle f \rangle}$ with $\alpha = 1, 2, 3, \ldots, 2$, respectively. Therefore, $S = V_\mathfrak{b}(f)$, i.e., the energy landscape $V$ has exactly three real critical points.
9. KNOWN LIMITATIONS

As a theoretical approach to computing real radicals using sums of squares, the major practical limitation is based on the value of ρ needed in (3). To prevent prevent arbitrarily long runtimes, Procedure 1 uses an upper bound $\alpha_{\text{max}}$ which we hope will be replaced in the future by an a priori upper bound based on the input system $f$.

When using floating-point computations as in the examples above, one needs to be cognizant of the effects of round-off error and conditioning in the computations. We can aim to control this using adaptive precision computations with numerical approximations of solutions that can be refined to arbitrary accuracy.

Systems phrased with numerical approximations of exact numbers can present problems in practice. Fundamentally, these systems violate the model of computation as exact numbers should be input exactly. Hence, representing 1/3 as 0.3333 means that the user is solving a different system from using 1/3. Nonetheless, the robustness and aspects of conditioning of this approach will be explored in future work.

10. CONCLUSION

By combining numerical algebraic geometry with sums of squares programming, we have produced a method for certifying that a set of polynomials generate the real radical. The set of polynomials arises from the generators of a set of polynomials that generate the real radical. The conditioning of this approach will be explored in future work.

11. REFERENCES


