Real Algebraic Geometry

Tutorial lectures during the special week on
"Geometry of polynomial system solving, optimization and topology"
Recent Trends in Computer Algebra
Institut Henri Poincaré

Daniel Plaumann
TU Dortmund

12 October 2023

Introduction

These lecture notes accompany the second half (five hours) of a short course at IHP. The first half was given by Ioannis Emiris.

The aim of the lectures is to give a short introduction to real algebraic geometry and semialgebraic geometry, i.e. the geometry of systems of real polynomial equations and inequalities.

I will not focus on symbolic computation or enter into any details about algorithms, but I will emphasise the parts of the theory that form the basis of algorithms and are relevant for computational and effective questions.

For references, I will often point to these two excellent books:

[BCR98] J. Bochnak, M. Coste, M.-F. Roy: **Real Algebraic Geometry**, Springer 1998. (Translated, and also updated, from the French edition published in 1987)

[BPR06] S. Basu, R. Pollack, M.-F. Roy: **Algorithms in Real Algebraic Geometry**, Springer 2006.

These notes will cover a little more material than I can present at the blackboard. Still a lot of details and most proofs are omitted. Instead, I want to focus on ideas and examples. If you notice any (of the almost unavoidable) mistakes, please let me know.

I would like to thank the organizers of the workshop and special week Carlos D'Andrea, Pierre Lairez, Mohab Safey El Din, Éric Schost, and Lihong Zhi for the invitation, and IHP for the kind hospitality.

Daniel Plaumann Paris, 10 October 2023

Contact:
Daniel.Plaumann@math.tu-dortmund.de

Contents

1 Real root counting

The problem of counting the real roots of a polynomial in one variable is of practical importance but also at the heart of the theory in real algebraic geometry. Let

$$f(T) = T^n + a_{n-1}T^{n-1} + \cdots + a_1 T + a_0$$

be a monic real polynomial of degree n in one variable T. For a quadratic ($n = 2$), the sign of the *discriminant*

$$a_1^2 - 4a_0$$

decides whether f has zero, one or two distinct real roots. Note that there is no need to *compute* the roots. In particular, if the coefficients are rationals or integers, then so is the discriminant. There are no arithmetic difficulties, like field extensions.

There is a similar criterion for $n = 3$, but starting with $n = 4$ the discriminant does not suffice. There are two classical symbolic methods for counting real roots in any degree:

- **Sturm sequences**, based on a modified Euclidean algorithm.
- **Hermite's method**, which we are going to look at now.

For f as above, let $\alpha_1, \ldots, \alpha_n$ be the complex roots (not necessarily distinct). The **Newton sums** are the power sums of the roots:

$$p_k = \alpha_1^k + \cdots + \alpha_n^k.$$

The Newton sums are symmetric polynomials in the roots and therefore polynomials in the coefficients of f, by the Fundamental Theorem on Symmetric Polynomials. Explicitly, they can be computed by a recursive application of Newton's identities:

$$p_r + c_1 p_{r-1} + c_2 p_{r-2} + \cdots + c_{r-1} p_1 + c_r r = 0$$

for all $r \geqslant 1$, where $c_i = a_{n-i}$ (resp. 0 for $i > n$). Alternatively, one can use the companion matrix of f (see below). For example,

$$p_0 = n, \quad p_1 = -c_1, \quad p_2 = -p_1 c_1 - 2c_2 = c_1^2 - 2c_2 \quad \text{etc.}$$

We now let $H(f)$ be the **Hermite matrix**

$$H(f) = (p_{i+j-2})_{i,j=1,\ldots,n}$$

of f. It is a symmetric $n \times n$-matrix whose entries are homogeneous polynomials in $a_0, \ldots, a_{n-1}$ of degree at most $2n - 2$. It is also a *Hankel matrix*, i.e. the entry at position (i, j) depends only on $i + j$.

1.1 Theorem (Hermite criterion).

(1) The rank of $H(f)$ is the number of distinct complex roots of f.

(2) The signature of $H(f)$ is the number of distinct real roots of f.

Here, the signature $\operatorname{sgn}(H(f))$ is the difference between the number of strictly positive and the number of strictly negative eigenvalues of $H(f)$.

1.2 Example. For a monic quadratic polynomial $f(T) = T^2 + a_1 T + a_0$, the relevant Newton sums are

$$p_0 = 2,\ p_1 = -a_1,\ p_2 = a_1^2 - 2a_0$$

and so the Hermite matrix is

$$H(f) = \begin{pmatrix} 2 & -a_1 \\ -a_1 & a_1^2 - 2a_0 \end{pmatrix}.$$

We can find the signature with Sylvester's criterion from linear algebra: The determinant is

$$\det(H(f)) = a_1^2 - 4a_0$$

and therefore agrees with the discriminant. (This is true in any degree.) If $a_1^2 - 4a_0 > 0$, then $H(f)$ is positive definite, so the signature is 2. If $a_1^2 - 4a_0 < 0$, then $H(f)$ is indefinite, so the signature is 0. If $a_1^2 - 4a_0 = 0$, then $H(f)$ has rank 1 and signature 1. So Hermite's criterion agrees with what we know about the quadratic equation. ◇

Proof. We can verify this through a direct computation. We assume that $\alpha_1, \dots, \alpha_p$ are the distinct real roots of f and $\alpha_{p+1}, \dots, \alpha_{p+q}, \overline{\alpha_{p+1}}, \dots, \overline{\alpha_{p+q}}$ the distinct non-real roots, so that there are $r = p + 2q$ distinct roots in total. Each root α_j occurs with a multiplicity $m_j \geqslant 1$. We write

$$v_j = (1, \alpha_j, \dots, \alpha_j^{n-1})^t$$

(column vectors) for $j = 1, \dots, r$, which implies

$$H(f) = \sum_{j=1}^{r} m_j v_j v_j^t$$

by definition. We also note that the vectors $v_1, \dots, v_r$ are linearly independent (by the Vandermonde formula). This shows (1). We can further write

$$\begin{aligned} H(f) &= \sum_{j=1}^{p} m_j v_j v_j^t + \sum_{j=p+1}^{q} m_j (v_j v_j^t + \overline{v_j v_j}^t) \\ &= \sum_{j=1}^{p} m_j v_j v_j^t + \sum_{j=p+1}^{q} 2m_j \mathrm{Re}(v_j)\mathrm{Re}(v_j)^t - \sum_{j=p+1}^{q} 2m_j \mathrm{Im}(v_j)\mathrm{Im}(v_j)^t. \end{aligned}$$

This shows that $H(f)$ has signature $2(p+q) - r = p$ (see also the problem below). ■

Problem 1. Verify that if $w_1, \dots, w_r \in \mathbb{R}^n$ are linearly independent, then the symmetric matrix $\sum_{j=1}^{k} w_j w_j^t - \sum_{j=k+1}^{r} w_j w_j^t$ has rank r and signature $2k - r$.

There is a very useful extension of Hermite's criterion: Given $f \in \mathbb{R}[T]$ as above and $g \in \mathbb{R}[T]$ a second polynomial, we may wish to count only those real roots α of f with $g(\alpha) > 0$. We can define the **generalized Hermite matrix**

$$H(f, g) = (p_{i+j-2}(f, g))_{i,j=1,\dots,n} \qquad \text{where } p_k(f, g) = \alpha_1^k \cdot g(\alpha_1) + \dots + \alpha_n^k \cdot g(\alpha_k).$$

1.3 Theorem (Hermite criterion with sign conditions).

(1) The rank of $H(f,g)$ is the number of distinct complex roots α of f with $g(\alpha) \neq 0$.

(2) If $\alpha_1, \dots, \alpha_r$ are the distinct real roots of f, then $\operatorname{sgn}(H(f,g)) = \sum_{j=1}^{r} \operatorname{sgn}(g(\alpha_j))$.

Proof. See [BPR06, §4.3.2] ∎

The case of polynomials in one variable is the most important in practice because one can often proceed inductively. But I want to mention a natural generalization to polynomial systems in several variables. This starts from the following observation: The Newton sums of a polynomial $f = T^n + a_{n-1}T^{n-1} + \cdots + a_1 T + a_0$ can be computed from the *companion matrix*

$$C(f) = \begin{pmatrix} 0 & & \cdots & & 0 & -a_0 \\ 1 & 0 & & \cdots & 0 & -a_1 \\ 0 & 1 & 0 & & \vdots & \vdots \\ \vdots & & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & 0 & -a_{n-2} \\ 0 & & \cdots & & 1 & -a_{n-1} \end{pmatrix}$$

of f, whose characteristic polynomial is f, so that its eigenvalues are the roots of f and the Newton sums p_k of f are therefore given by

$$p_k = \operatorname{tr}(C(f)^k).$$

In algebraic terms, we can think of $C(f)$ as the matrix describing the multiplication by T (shift operator) in the factor ring $\mathbb{R}[T]/f$. This generalizes to the following:

1.4 Theorem (Multivariate Hermite criterion).
Let $f_1, \dots, f_k \in \mathbb{R}[X_1, \dots, X_n]$ and assume that the system $f_1 \cdots = f_k = 0$ is zero-dimensional, i.e. with finitely many complex solutions. Let $I = (f_1, \dots, f_k)$ be the ideal generated by $f_1, \dots, f_k$ and $A = \mathbb{R}[X_1, \dots, X_n]/I$ the factor ring modulo I. On the finite-dimensional $\mathbb{R}$-vector space A, let

$$\varphi_f \colon A \longrightarrow A,\ h \mapsto fh$$

be multiplication with a fixed element $f \in A$, and consider the bilinear map

$$H \colon A \times A \longrightarrow \mathbb{R},\ (f,g) \mapsto \operatorname{tr}(\varphi_{fg}).$$

(1) The rank of H is the number of distinct complex solutions.

(2) The signature of H is the number of distinct real solutions.

Proof. See [BPR06, §4.6] ∎

Problem 2. Verify that the univariate Hermite criterion 1.1 is a special case of Thm. 1.4.

For example, the multivariate Hermite method has been employed recently for solving parametric systems of zero-dimensional systems over the reals.[1]

In this context, I also want to advertise msolve, a modern and easy-to-use library for solving multivariate polynomial systems developed here at LIP6 by the PolSys Team:

https://msolve.lip6.fr

[1] see H. P. Le and M. Safey El Din: *Solving parametric systems of polynomial equations over the reals through Hermite matrices*. J. Symbolic Comput. 112(2022), 25–61

2 Semialgebraic sets and quantifier elimination

Throughout, we will write $X = (X_1, \dots, X_n)$, and $\mathbb{R}[X]$ as shorthand for the polynomial ring in the variables $X_1, \dots, X_n$.

An **algebraic set** in $\mathbb{R}^n$ is one of the form

$$\mathcal{Z}(f_1, \dots, f_r) = \{p \in \mathbb{R}^n \mid f_1(p) = \cdots = f_r(p) = 0\}$$

for polynomials $f_1, \dots, f_r \in \mathbb{R}[X]$. A set of the form

$$\mathcal{W}(f_1, \dots, f_r) = \{p \in \mathbb{R}^n \mid f_1(p) \geqslant 0, \cdots, f_r(p) \geqslant 0\}$$

is a called a **basic closed (semialgebraic) set**. Likewise, a set of the form

$$\mathcal{U}(f_1, \dots, f_r) = \{p \in \mathbb{R}^n \mid f_1(p) > 0, \cdots, f_r(p) > 0\}$$

is **basic open**.

Definition. A **semialgebraic set** in $\mathbb{R}^n$ is a finite boolean combination of basic closed sets.

A finite boolean combination is made up of a finite number of unions, intersections and complements. Equivalently, we may take finite boolean combinations of basic open sets.

2.1 Examples. (1) The closed unit disk is the semialgebraic set $\mathcal{W}(1 - X_1^2 - X_2^2)$. Half of the disk is $\mathcal{W}(1 - X_1^2 - X_2^2, X_2)$. Three quarters of the disk (midnight to nine o'clock) is the set $\mathcal{W}(1 - X_1^2 - X_2^2, -X_2) \cup \mathcal{W}(1 - X_1^2 - X_2^2, X_1, X_2)$.

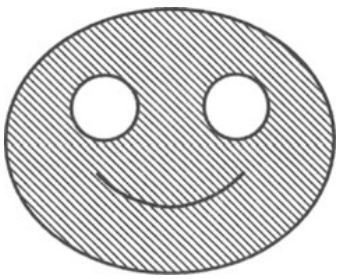
(2) The semialgebraic subsets of $\mathbb{R}$ are precisely the finite unions of points and intervals, i.e.

$$[a, b], \quad (a, b), \quad [a, b), \quad (a, b], \quad \{c\}, \quad \emptyset$$

for $a, b \in \mathbb{R} \cup \{\infty\}$ with $a < b$ and $c \in \mathbb{R}$ (Exercise).

(3) Hence $\mathbb{Z} \subset \mathbb{R}$ is not semialgebraic.

(4) I just had to steal this one from [BCR98, Fig. 2.1]:



$$\{(x, y) \in \mathbb{R}^2 \mid x^2/25 + y^2/16 < 1 \text{ and } x^2 + 4x + y^2 - 2y > -4 \\ \text{and } x^2 - 4x + y^2 - 2y > -4 \text{ and } (x^2 + y^2 - 2y \neq 8 \text{ or } y > -1)\}$$

(5) The epigraph of the exponential $\{(x_1, x_2) \in \mathbb{R}^2 \mid x_2 \geqslant e^{x_1}\}$ is not semialgebraic. Neither is its graph. This is easy to believe, but how would you prove it?

(6) Easier: The epigraph $X = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_2 \geqslant \sin(x_1)\}$ of sine is not semialgebraic. For if it were, then the intersection with the horizontal axis $\mathcal{Z}(X_2)$ would also be semialgebraic. But this is an infinite union of intervals. ◇

The first example already brings up an interesting point: Not every subset of $\mathbb{R}^n$ that is closed and semialgebraic is basic closed.

Problem 3. Show that the sector $\mathcal{W}(1-X_1^2-X_2^2,-X_2)\cup\mathcal{W}(1-X_1^2-X_2^2,X_1,X_2)$ in Example 2.1(1) is not basic closed.

A peculiar fact in real algebraic geometry is that every real algebraic set can be defined by a single equation, since $\mathcal{Z}(f_1,\dots,f_r)=\mathcal{Z}(f_1^2+\cdots+f_r^2)$. There is no such simple trick for inequalities. Also, we can take finite boolean combinations of basic sets in any order. Nevertheless, semialgebraic sets admit certain normal forms:

2.2 Proposition. *Every semialgebraic set in $\mathbb{R}^n$ can be expressed as a finite union of sets of the form*

$$\mathcal{Z}(f)\cap\mathcal{U}(g_1,\dots,g_l)$$

for polynomials $f,g_1,\dots,g_l\in\mathbb{R}[X]$.

Proof. Since these sets are semialgebraic and include all basic open sets, it is enough to check that this class is closed unter finite boolean operations (Exercise). ∎

The following looks similar at a first glance but is much more difficult to prove:

2.3 Theorem (Finiteness theorem). *Every closed (resp. open) semialgebraic set is a finite union of basic closed (resp. basic open) sets.*

Proof. We will not need this, but see [BCR98, Thm. 2.7.2]. ∎

The foundation of semialgebraic geometry is the projection theorem:

2.4 Theorem (Projection Theorem). *Let $S\subset\mathbb{R}^m\times\mathbb{R}^n$ be semialgebraic and let $\pi\colon\mathbb{R}^m\times\mathbb{R}^n\to\mathbb{R}^n$ be the projection onto the second factor. Then $\pi(S)$ is again a semialgebraic set.*

The proof of the projection theorem is an application of root counting. Let us sketch the simplest case: Let $f\in\mathbb{R}[T,X_1,\dots,X_n]$ be a polynomial in $n+1$ variables with zero set $S=\mathcal{Z}(f)$ in $\mathbb{R}\times\mathbb{R}^n$. Let π be the projection onto $\mathbb{R}^n$. Then

$$\pi(S)=\big\{p\in\mathbb{R}^n\mid\exists a\in\mathbb{R}\colon f(a,p)=0\big\},$$

in other words, $\pi(S)$ is the set of all $p\in\mathbb{R}^n$ for which the polynomial $f(T,p)\in\mathbb{R}[T]$ has at least one real root. We can set up the Hermite matrix $H(f)$ of f with respect to T. Its entries are polynomials in the coefficients of f and therefore polynomials in $X_1,\dots,X_n$. By Hermite's criterion 1.1, the set $\pi(S)$ consists of those points p for which the signature of $H(f)$ is strictly positive. This condition can be expressed by polynomial inequalities in the entries of $H(f)$ and hence in $X_1,\dots,X_n$ (see below). These inequalities describe $\pi(S)$ as a semialgebraic set.

For the general case, one needs to use the signed version of Hermite's criterion 1.3. The proofs in [BCR98, Thm. 2.2.1] and [BPR06, §1.3] are based on Sturm sequences. (The latter also describes the algorithm much more explicitly.)

Problem 4. Show that the set of real symmetric $n\times n$-matrices of a fixed signature k is a semialgebraic subset of the space $\mathbb{S}_n$ of real symmetric matrices, defined by polynomials with integer coefficients. (*Suggestion:* Use induction on n. Show that there are finitely many invertible matrices $T_1,\dots,T_k$ such that for every $A\in\mathbb{S}_n$ there is an index j for which $T_j^tAT_j$ is block-diagonal of the form $\begin{pmatrix}a'_{11}&0\\0&A'\end{pmatrix}$.)

2.5 Example. We will see several consequences and reformulations of the Projection Theorem for semialgebraic geometry.

A simple application in geometry is something like the following: Suppose that, for some reason, we wish to consider the set of all polynomials $f \in \mathbb{R}[X, Y]$ of degree 2 for which the zero set (conic) $\mathcal{Z}(f)$ in $\mathbb{R}^2$ is fully contained in the upper half plane. We can write down an *incidence correpondence*

$$\{(f, (a_1, a_2)) \in \mathbb{R}[X, Y]_{\leqslant 2} \times \mathbb{R}^2 \mid a_2 < 0,\ f(a_1, a_2) = 0\}.$$

Under the identification of $\mathbb{R}[X, Y]_{\leqslant 2}$ with $\mathbb{R}^6$ by taking coefficients, this is a semialgebraic set in $\mathbb{R}^6 \times \mathbb{R}^2 = \mathbb{R}^8$. Its projection onto the first factor is the set of quadratics having some root in the lower half plane. It is semialgebraic by the Projection Theorem. Its complement is also semialgebraic, and it is the set we want. So we know that there is *some* semialgebraic description of this set of polynomials. ◇

Problem 5. Compute inequalities describing the set in the above example.

3 Real closed fields

Many statements of real algebraic geometry hold not only over the real numbers but over every **real closed field**. These are the fields that behave *algebraically* like the real numbers. There are several equivalent definitions (see [BCR98, §1.2] or [KS22, Ch. 1]), such as:

Definition. A real closed field is an ordered field $(R, \leqslant)$ such that

(1) every positive element in R is a square;
(2) every polynomial of odd degree over R has a root in R.

An equivalent characterization, it turns out, is that R is not algebraically closed, but $R(\sqrt{-1})$ is (Artin-Schreier, 1927; see [KS22, Thm. 1.5.4 and Thm. 1.6.1]).

Why is this relevant? First, there are some important examples of real closed fields:
(a) The **real numbers** $\mathbb{R}$, of course.
(b) The field $\mathbb{R}_{\text{alg}} = \overline{\mathbb{Q}} \cap \mathbb{R}$ of **real algebraic numbers**. These are the numbers we typically use for symbolic computations, without all the "dead wood" of real transcendental numbers.

There is a notion of **real closure** for an ordered field, analogous to the algebraic closure.

3.1 Theorem. *Let K be an ordered field. There exists a real closed field R containing K as a subfield such that R/K is algebraic and the ordering of R extends that of K. The field R is unique up to a unique order-preserving isomorphism.*

For example, the real algebraic numbers $\mathbb{R}_{\text{alg}}$ are the real closure of $\mathbb{Q}$.

3.2 Example. The field $\mathbb{R}((T))$ or real *Laurent series* is the field of fractions of the formal power series ring $\mathbb{R}[[T]]$. It can be ordered by comparing *initial cofficients*, i.e. the non-zero coefficients with lowest exponents. In other words, we have $f > 0$ for $f(T) \in \mathbb{R}((T))$ if the first non-zero (formal) derivative of f at $T = 0$ is positive.

The real closure of $\mathbb{R}((T))$ is the field $\mathbb{R}\{\{T\}\}$ of **real Puiseux series**. Explicitly, a Puiseux series is a formal series of the form

$$\sum_{i \geqslant k} a_i T^{i/q}$$

where $k \in \mathbb{Z}$, $a_i \in \mathbb{R}$, i runs over integers $\geqslant k$, and q is a positive integer. (In other words, T may occur with an infinite number of rational exponents, but starting from a minimum and over a common denominator; it is not obvious that $\mathbb{R}\{\{T\}\}$ is indeed a field.)

This description of the real closure is a consequence of the **Newton-Puiseux theorem** (see e.g. [BPR06, Thm. 2.91] for a proof). Likewise, the field $\mathbb{C}\{\{T\}\}$ is the algebraic closure of $\mathbb{C}((T))$.

Puiseux series occur naturally in real algebraic geometry, because we can think of varieties and semialgebraic sets over $\mathbb{R}\{\{T\}\}$ as families of such objects over $\mathbb{R}$ parametrized in T, while still having all the benefits of working over a real closed field. This also makes them a useful algebraic tool in **real tropical geometry** (see [MS15] for a general reference).

Note that the variable T in the fields $\mathbb{R}((T))$ and $\mathbb{R}\{\{T\}\}$ is *infinitesimal*, i.e., it is smaller than $1/n$ for every natural number n. Consequently, $1/T$ is larger than any natural number. The ordered fields $\mathbb{R}((T))$ and $\mathbb{R}\{\{T\}\}$ are therefore **non-archimedean**. ◇

4 Formulas

We have to take a short detour and introduce a few basic notions from mathematical logic. Let A be a ring (commutative with 1), e.g. $A = \mathbb{Z}$ or $A = \mathbb{R}$.

- An A-**prime formula** is a formula of the form $f(X) > 0$ for a polynomial $f \in A[X]$.
- An A-**formula** arises by iteration as follows: Any A-prime formula is an A-formula. If φ and ψ are two A-formulas, then so are

$$\varphi \vee \psi, \quad \neg\varphi, \quad \exists x_i \varphi.$$

- Using this, we can also express the remaning standard logical operators via $\varphi \wedge \psi = \neg(\neg\varphi \vee \neg\psi)$, $\varphi \Longrightarrow \psi = \neg(\varphi \wedge \neg\psi)$ and $\forall X_i \varphi = \neg\exists X_i(\neg\varphi)$. Also, the formula $f(X) \geqslant 0$ is defined as $\neg(-f(X) > 0)$; the formula $f(X) = 0$ is similarly defined. Also, $f(X) > g(X)$ is defined by $f(X) - g(X) > 0$, etc.
- A variable in a formula is called **free**, if it occurs (at least once) outside the scope of any quantifier. Otherwise, the variable is called **bound**. An A-formula, all of whose variables are bound, is called an A-**sentence**.
- A formula is **quantifier-free**, if it does not involve any quantifiers.

We will usually take $A = \mathbb{Z}$ or $A = R$ a real closed field. Whether an A-sentence is true may of course depend on what values we are allowed to substitute for the variables. For example, the $\mathbb{Z}$-sentence $\exists X(X^2 = 2)$ is true in $\mathbb{R}$ but false in $\mathbb{Z}$. On the other hand, a formula containing a free variable has no truth value: It is meaningless to say that $\exists X_1(X_1 = X_2)$ is true or false.

Definition. Let R be a ring containing A as a subring. If φ is an A-formula and $X_1, \ldots, X_m$ are free variables in φ (or not occuring in φ at all), we can form the **set of satisfying assignments** $\mathrm{SAT}_R(\varphi)$ as the set of all points $a = (a_1, \ldots, a_n) \in R^n$ such that φ becomes a true sentence (in R) if a_i is substituted for X_i, for all $i = 1, \ldots, m$.

We can define semialgebraic sets over a real closed field R in exactly the same way as over the real numbers. Everything we have proved so far (root counting, projection theorem,…) will work just as before. Using the language of formal logic, we can now say:

4.1 Theorem. *Semialgebraic sets in R^n are the sets of satisfying assignments of R-formulas in at most n free variables.*

Proof. The description of a semialgebraic set as a boolean combination of basic sets is easily rewritten as an R-formula. The converse is proved by "induction on the recursive construction of the formula": It's clear that $\vee$ corresponds to union and $\neg$ to taking complements. The crucial point is that if we know that the set of satisfying assignments of φ is semialgebraic, then the set of satisfying assignments of $\exists X_i \varphi$ in R^{n-1} is also semialgebraic by the Projection Theorem. ■

4.2 Example. The closure of a semialgebraic set S in $\mathbb{R}^n$ is again semialgebraic. To see this, write $S = \mathrm{SAT}_{\mathbb{R}}(\varphi)$ for some $\mathbb{R}$-formula φ and note that the closure of S is the set of satisfying assignments of the formula

$$\forall \varepsilon \left(\varepsilon > 0 \Rightarrow \exists Y \left(\varphi(Y) \wedge \sum (X_i - Y_i)^2 < \varepsilon \right) \right)$$

in the free variables $X = (X_1, \ldots, X_n)$. The interior of S is likewise algebraic, by a similar argument or by taking complements. ◇

4.3 Example. The closure of a basic open set $\mathcal{U}(f_1, \ldots, f_r)$ can be strictly smaller than the basic closed set $\mathcal{W}(f_1, \ldots, f_r)$. In other words, relaxing inequalities may not give the closure. For instance, $\mathcal{U}(f_1, \ldots, f_r)$ could be empty and $\mathcal{W}(f_1, \ldots, f_r)$ non-empty (Example?). For computational purposes, this can be extremely annoying (see also CAD below). ◇

Problem 6. It's even worse: Can you find an example of a basic open set whose closure is not basic closed at all? (Maybe not. But it's somewhere in [ABR96].)

4.4 Theorem (Quantifier elimination). *Let A be a ring and φ an A-formula. Then there exists a quantifier-free A-formula ψ with the same free variables as φ and such that $\mathrm{SAT}_R(\varphi)$ and $\mathrm{SAT}_R(\psi)$ agree for any real closed field that contains A.*

This is another reformulation of the Projection Theorem, with one additional detail: The quantifier-free formula ψ does not depend on R. This independence does not follow from the statement of the Projection Theorem, but it does follow from its proof, with minimal extra effort:

Sketch of proof. The proof is the same as that of Thm. 4.1 if one verifies that the formulas obtained from A-formulas are again A-formulas. This is true, because the signature of a matrix whose entries are polynomial in A is determined by polynomial inequalities with coefficients in A. ■

Elimination is a very versatile tool, as can be already seen in Examples 2.5 and 4.3. Many, if not most, operations we perform in algebraic geometry can be set up as elimination problems. Unfortunately, elimination is often too slow in practice. For algebraic geometry over $\mathbb{C}$ or algebraically closed fields, Gröbner bases are the main tool. For real algebraic geometry, the situation is more complicated and in practice often requires a mix of different methods.

Algorithms for quantifier elimination (or the Projection Theorem) are based on the cylidrical algebraic decomposition, to be discussed later. The complexity is double exponential in n.

For now, we will concentrate on some more abstract consequences of quantifier elimination, in particular the important Tarski principle:

4.5 Corollary (Transfer principle of Tarski-Seidenberg). *Let K be a field and let R_1 and R_2 be two real closed fields containing K and inducing the same ordering on K. Then the true K-sentences over R_1 are exactly the same as over R_2.*

Proof. For any K-sentence φ, there is a quantifier-free K-sentence that is equivalent to φ over R_1 and over R_2. But whether such a sentence is true or false is decided in K and does not involve R_1 and R_2 at all. (Just think about what a "quantifier-free K-sentence" really is.) ■

In the language of model theory, this is usually stated as: *The theory of real closed fields in the language of ordered rings is model complete*.

4.6 Examples. For any system of equations and inequalities defined over $\mathbb{Z}$ or $\mathbb{Q}$, the solvability in real algebraic numbers (over $\mathbb{R}_{\text{alg}}$) is equivalent to the solvability over $\mathbb{R}$. The *set* of solutions will in general be larger, but one will be non-empty if and only if the other is non-empty. So symbolic computations over $\mathbb{Q}$ can capture the full picture, in principle.

Even going to something like $\mathbb{R}\{\{T\}\}$, introducing the infinitesimal element T, changes nothing. This is closely related to the fact that the Archimedean axiom itself

$$\forall r \in \mathbb{R} \exists n \in \mathbb{N}\colon n > r$$

is not an $\mathbb{R}$-formula, since it specifically asks for the existence of a *natural* number. So it may (and does) hold in some real closed fields but not in others. In this sense, algebra cannot detect this property.

Likewise, we can use any fixed polynomial expression over a real closed field R in an R-formula, but we cannot express

There exists a polynomial such that…

in an R-formula. Only once a degree d is fixed, we can encode

There exists a polynomial of degree at most d such that…

in an R-formula. Therefore, bounding degrees in some construction often has great theoretical consequences, in addition to questions of complexity or practical runtime. ◇

5 Digression: Hilbert's 17th Problem

At the International Congress of Mathematicians in Paris in 1900, David Hilbert presented a list of 23 Problems he considered to be important and fruitful. Most Hilbert problems have been solved, but a few are still open (e.g. the Riemann hypothesis). Two of Hilbert's problems belong to the area of real algebraic geometry, namely number 16 (still largely open and again included by Smale in his list of problems) and number 17:

Question (Hilbert's 17th Problem). Let $f \in \mathbb{R}(X)$ be a rational function. Assume that f is positive on its domain, i.e. $f(a) \geqslant 0$ for all $a \in \mathbb{R}^n$ at which f does not have a pole. Are there rational functions $g_1, \ldots, g_r \in \mathbb{R}(X)$ such that

$$f = g_1^2 + \cdots + g_r^2 \qquad ?$$

In other words, is every positive rational function a sum of squares? This condition is obviously sufficient for positivity, but is it necessary?

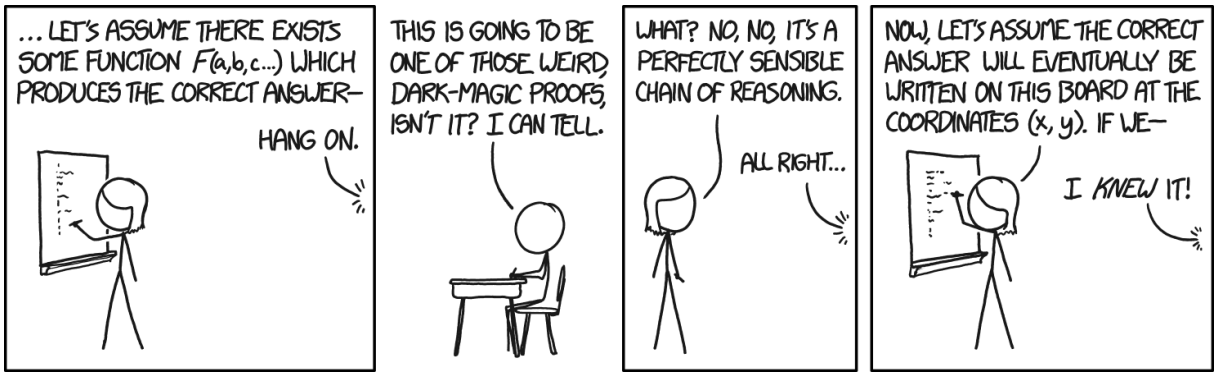
Hilbert's 17th problem has a history going back a few years to earlier work of Hilbert on sums of squares of homogeneous polynomials. It was solved 27 years later by Emil Artin. Its solution is the starting point for most of the following theory.

Using the Tarski principle, Hilbert's 17th problem has a short and ingenious solution.

5.1 Theorem (E. Artin 1927). *Every positive rational function is a sum of squares.*

Proof. Let $f \in \mathbb{R}(X)$ be a rational function and assume that f is not a sum of squares. Then there is some ordering of the field $\mathbb{R}(X)$ in which f is a negative element. In other words, sums of squares are the only elements in a field that are necessarily positive with respect to *any* ordering. This is not very hard to show using Zorn's lemma; see below for the proof. The field $\mathbb{R}(X)$ possesses a real closure $\widetilde{R}$ with respect to this ordering. Since $\mathbb{R}(X)$ is a subfield of $\widetilde{R}$, the variables $X = (X_1, \dots, X_n)$ are just elements of $\widetilde{R}$. That f is negative in $\mathbb{R}(X)$ and therefore in $\widetilde{R}$ just means that the polynomial $f(T)$ (in a new set of variables $T = (T_1, \dots, T_n)$, if we want) evaluates negatively at the point $X \in \widetilde{R}^n$. For fixed f, the existence of such a point is an $\mathbb{R}$-sentence. By the transfer principle, it must also hold over $\mathbb{R}$. So f is not positive. ∎

This is one of those proofs…



https://xkcd.com/1724/

This is not Artin's original proof, since Tarski's principle came later (1940). Instead, Artin used a more algebraic statement later developed into the *Artin-Lang Theorem* on real places (see [BCR98, §4.1]). The basic idea, however, is the same.

5.2 Example. Motzkin's polynomial $f = X^4Y^2 + X^2Y^4 - 3X^2Y^2 + 1$ is the best-known example of a polynomial that is positive ($\geqslant 0$) on $\mathbb{R}^2$ but cannot be written as a sum of squares of polynomials. By Artin's theorem, it can be written as a sum of squares of rational functions, though the proof is not constructive. Explicitly, one can verify that

$$f = \frac{X^2Y^2(X^2+Y^2+1)(X^2+Y^2-2)^2 + (X^2-Y^2)^2}{(X^2+Y^2)^2}$$

is a representation of f as a sum of squares in $\mathbb{R}(X,Y)$ (see for example [Mar08] and the lectures of Didier Henrion for more on positive polynomials). ◇

The proof of Artin's theorem made use of the following lemma, which we include here for reference (see also [BCR98, §1.1] or [KS22, §1.1]).

5.3 Lemma. *Let K be a field of characteristic 0. If $c \in K$ is not a sum of squares in K, then there exists a field ordering $\leqslant$ of K with $c < 0$.*

Proof. A field ordering $\leqslant$ of K can be identified with its positive cone $P_{\geqslant} = \{a \in K \mid a \geqslant 0\}$, since $a \geqslant b$ is equivalent to $a - b \geqslant 0$ and therefore to $a - b \in P_{\geqslant}$. A *preordering* of K is a subset $P \subsetneq K$ satisfying the conditions

$$P + P \subset P, \quad P \cdot P \subset P, \quad \forall a \in K\colon a^2 \in P, \quad -1 \notin P.$$

Clearly, the positive cone $P_{\geqslant}$ of a field ordering $\leqslant$ is a preordering. Also, $\Sigma = \{a_1^2 + \cdots + a_k^2 \mid a_1, \ldots, a_k \in K,\ k \in \mathbb{N}\}$, the set of all sums of squares is a preordering, provided that $-1 \notin \Sigma$. (If -1 is a sum of squares in K, then every element is a sum squares, since we can write $a = (\frac{a+1}{2})^2 - (\frac{a-1}{2})^2 = (\frac{a+1}{2})^2 + (-1) \cdot (\frac{a-1}{2})^2$ for any $a \in K$. So the statement of the lemma is empty in this case.) For a preordering to be an actual ordering, it must contain either a or $-a$ for every $a \in K \setminus \{0\}$. Thus Σ is, in general, not an ordering.

The crucial step is the following: If P is a preordering of K and a an element of $K \setminus P$, then the set $P_{-a} = \{p - qa \mid p, q \in P\}$ is again a preordering. The first three properties are checked at once. To see $-1 \notin P_{-a}$, suppose to the contrary that $-1 = p - qa$ for some $p, q \in P$. Then $q \neq 0$, since $-1 \notin P$, hence $a = q^{-2} \cdot q \cdot (1 + p) \in P$, a contradiction.

We now look at the set of all preorderings of K containing the element $-c$, ordered by inclusion. It is not empty, since it contains Σ_{-c}. Ascending unions of preorderings are obviously preorderings, hence Zorn's lemma applies and produces a maximal preordering P of K containing $-c$. It is easy to check that $a \leqslant b \Leftrightarrow b - a \in P$ is an ordering of K: If there were $a \in K$ with neither $a \in P$ nor $-a \in P$, then one of P_{-a} or P_a would be a preordering containg $-c$ that is larger than P. ∎

Note that the lemma also implies that a field K possesses some field ordering if and only if not every element in K is a sum of squares or, equivalently, if -1 is not a sum of squares.

6 Brief Summary: Semialgebraic maps and topology of semialgebraic sets

Let always R be a real closed field.

Definition. Let $A \subset R^m$ and $B \subset R^n$ be semialgebraic sets. A map $\varphi\colon A \to B$ is called **semialgebraic**, if its graph $\Gamma(\varphi) = \{(p, q) \in \mathbb{R}^m \times \mathbb{R}^n \mid p \in A,\ \varphi(p) = q\}$ is a semialgebraic set.

6.1 Examples. Every polynomial function is semialgebraic. The square root $\mathbb{R} \to \mathbb{R}$, $x \mapsto \sqrt{x}$ is semialgebraic. The absolute value is semialgebraic. Any step function with a finite number of steps is semialgebraic. ◇

6.2 Example. For a more interesting example, let $S \subset \mathbb{R}^n$ be a non-empty semialgebraic set. The map

$$d_S\colon \mathbb{R}^n \to \mathbb{R}_{\geqslant 0},\ d_S(x) = \inf\big\{\|x - y\| \ \mid\ y \in S\big\}$$

is semialgebraic with $d_S^{-1}(0) = \overline{S}$ (Exercise). ◇

6.3 Proposition. *The image of a semialgebraic set under a semialgebraic map is semialgebraic.*

Proof. By the Projection Theorem, since the image is a projection of the graph. ∎

Now for the topology: Let R be a real closed field, for example $\mathbb{R}_{\text{alg}}$, the real algebraic numbers. The field R carries a topology induced by its ordering, which is exactly the usual open-ball topology on R^n and its subsets. We have already seen that the closure of a semialgebraic set in $\mathbb{R}^n$ in again semialgebraic. This is true over any real closed field: With the order topology on R^n, the closure is described by the same R-formulas as over $\mathbb{R}$. Topology is great, but we have to be careful when working over general real closed fields.

The order topology does not have very good properties: For $R \neq \mathbb{R}$, the real line is always totally disconnected. However, this problem (mostly) disappears as long as we restrict ourselves to semialgebraic sets only. For example:

Definition. A semialgebraic set $S \subset R^n$ is **semialgebraically connected** if the following holds: If $S_1, S_2 \subset S$ are two disjoint open subsets of S with $S = S_1 \cup S_2$, then $S_1 = \emptyset$ or $S_2 = \emptyset$.

With this definition, the real line R is easily seen to be connected once again, as are all intervals in R.

6.4 Theorem. *Every semialgebraic set S is a finite union of disjoint semialgebraically connected subsets, which are both closed and open in S, called the (semialgebraic) **connected components** of S.*

Proof. See [BCR98, Thm. 2.4.4]. ∎

One difference is that, due to the disconnectedness of the topology, we cannot argue with sequences to describe convergence and closures, etc.

Definition. Let $S \subset R^n$ be a semialgebraic set. A **semialgebraic path** in S is a continuous semialgebraic map $\alpha\colon I \to S$ defined on an interval $I \subset R$.

6.5 Theorem (Curve selection lemma). *Let $S \subset R^n$ be a semialgebraic set, and let $x \in \overline{S}$. Then there is a semialgebraic path $\alpha\colon (0,1] \to S$ such that $\alpha(t) \in S$ for every $t \in (0,1)$ and $\alpha(1) = x$.*

The curve selection lemma can be applied in many cases to replace arguments with convergent sequences. Two different proofs are given in [BCR98, §2.5] and [BPR06, Thm. 3.19], but neither is exactly short.

It should be mentioned that connected semialgebraic sets are always (semialgebraically) path-connected, which follows from the curve selection lemma.

We also have to be careful with the definition of compactness.

Problem 7. Show that the unit interval $[0,1]$ in $\mathbb{R}_{\text{alg}}$ is not compact (with the usual open-cover definition).

Compactness can be replaced with the property of being closed and bounded, as in the following statements.

6.6 Corollary. *A semialgebraic $S \subset R^n$ is closed and bounded if and only if the following holds: Any semialgebraic path $\alpha\colon (0,1) \to S$ extends to a semialgebraic path $\widetilde{\alpha}\colon (0,1] \to S$ with $\widetilde{\alpha}|_{(0,1)} = \alpha$.*

Proof. Exercise. ∎

6.7 Theorem. *If $S \subset R^m$ is semialgebraic, closed and bounded and $\varphi\colon R^m \to R^n$ is a semialgebraic map, then $\varphi(S)$ is again closed and bounded.*

Proof. [BCR98, Thm. 2.5.8] ∎

7 The real Nullstellensatz

Given polynomials $f_1, \dots, f_k \in R[X]$, they define both the real algebraic set

$$Z_R = \mathcal{Z}(f_1, \dots, f_k) = \{p \in R^n \mid f_1(p) = \cdots = f_k(p) = 0\}$$

and the complex algebraic set (affine variety)

$$Z_C = \mathcal{Z}_C(f_1, \dots, f_k) = \{p \in C^n \mid f_1(p) = \cdots = f_k(p) = 0\}$$

where $C = R(\sqrt{-1})$ is the algebraic closure. If

$$I = (f_1, \dots, f_k) = \{a_1 f_1 + \cdots + a_k f_k \mid a_1, \dots, a_k \in R[X]\}$$

is the ideal generated by $f_1, \dots, f_k$ in the polynomial ring $R[X]$, then it is clear that every element of I also vanishes on Z and Z_C:

$$f \in I \implies f \equiv 0 \text{ on } Z_R$$
$$f \in I \implies f \equiv 0 \text{ on } Z_C$$

(In fact, we have $Z_R = \mathcal{Z}(I)$ and $Z_C = \mathcal{Z}_C(I)$.) The converse implications are not clear, and in general not true. For Z_C this is described by:

7.1 Theorem (Hilbert's Nullstellensatz). *For $h \in R[X]$, we have $h \equiv 0$ on Z_C if and only if there exists $r > 0$ such that*

$$h^r \in I.$$

(In most textbooks, this is only proved for polynomials over an algebraically closed field, for instance in [CLO15, §4.1], but it really holds over any field as long as the points (i.e. solutions) are considered over an algebraically closed field.) The real analogue is as follows:

7.2 Theorem (Real Nullstellensatz). *For $h \in R[X]$, we have $h \equiv 0$ on Z_R if and only if there exist $r > 0$ and $s_1, \dots, s_l \in R[X]$ such that*

$$h^{2r} + s_1^2 + \cdots + s_l^2 \in I.$$

In both the real and the complex theorem, the reverse implication is easy to see, only the other requires work. See [BCR98, Thm. 4.1.4] for a proof. An early version of the real Nullstellensatz goes back to Krivine (1964). The modern version was first proved by Risler (1976).

7.3 Example. (1) If $f = x_1^2 + x_2^2$, then $Z_R = \{(0,0)\}$. The polynomials that vanish in this point are just those without constant term, which is the same as the maximal ideal generated by x_1 and x_2. The real Nullstellensatz immediately reflects this, since

$$x_1^2 + x_2^2 = f \in I$$

so this is the statement with $h = x_1$, $r = l = 1$ and $s_1 = x_2$, and likewise for $h = x_2$.

(2) More generally, whenever we apply the trick to rewrite $Z_R = Z(f_1, \dots, f_k) = Z(f_1^2 + \cdots f_k^2)$, the real Nullstellensatz applies to the (principal) ideal generated by $f_1^2 + \cdots + f_k^2$ and certifies that $f_1, \dots, f_k$ indeed vanish on Z_R. ◇

7.4 Corollary (Weak Nullstellensatz).

(1) If $Z_C = \emptyset$, there exist $a_1, \dots, a_k \in R[X]$ such that $a_1 f_1 + \cdots + a_k f_k = 1$.

(2) If $Z_R = \emptyset$, there exist $a_1, \dots, a_k, s_1, \dots, s_l \in R[X]$ such that $a_1 f_1 + \cdots + a_k f_k + s_1^2 + \cdots + s_l^2 = 1$.

Proof. (1) If $Z_C = \emptyset$, then 1 "vanishes" on Z_C, so $1^r \in I$ for some r by the Nullstellensatz, hence $1 \in I$. This proof is slightly bogus in the sense that Hilbert's Nullstellensatz is almost invariably proved by showing this weak version first. (2) follows in the same way. ■

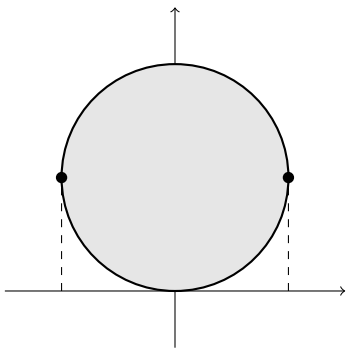
There are semialgebraic versions of the real Nullstellensatz that characterize (strict or non-strict) positivity on basic closed semialgebraic sets. We will skip them here for lack of time. We will see another application of the real Nullstellensatz at the end of the lecture.

8 Decomposition theorems and semialgebraic dimension

Always let R denote a real closed field. There are a number of very strong decomposition theorems for semialgebraic sets and maps. The most basic version is called the *cylindrical algebraic decomposition* (CAD).

8.1 Example. Let $S \subset \mathbb{R}^2$ be the closed disk of radius 1 centered at the point $(0, 1)$. When we project onto the horizontal axis, we can think of S as bounded by graphs of semialgebraic functions. More precisely, we may decompose S into the two single points $(-1, 1)$ and $(1, 1)$ together with the region

$$\left\{(x_1, x_2) \in \mathbb{R}^2 \mid -1 < x_1 < 1 \text{ and } 1 - \sqrt{1 - x_1^2} \leqslant x_2 \leqslant 1 + \sqrt{1 - x_1^2}\right\}.$$



Definition. Let $S \subset R^n$ be semialgebraic and let $f, g \colon S \to R \cup \{\pm\infty\}$ be continuous semialgebraic functions. The set

$$\operatorname{Band}(f, g) = \big\{(x, t) \in S \times R \mid f(x) < t < g(x)\big\}$$

is called the **band between** f **and** g. As usual, let $\Gamma(f) = \{(x, y) \in S \times R \mid y = f(x)\}$ be the graph of f.

8.2 Theorem. *Let $f_1, \dots, f_r \in R[X, T]$. Then there is a decomposition*

$$R^n = S_1 \cup \dots \cup S_k$$

into disjoint semialgebraic sets and for every $i \in \{1, \dots, k\}$ a finite number of continuous semialgebraic functions

$$z_{ij} \colon S_i \to R \qquad (j = 1, \dots, m_i,\ m_i \geqslant 0)$$

with the following properties:

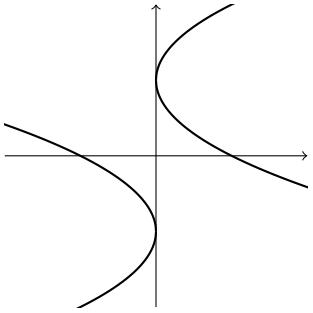
(1) We have $z_{i1} < z_{i2} < \dots < z_{im_i}$ on S_i.

(2) For every $x \in S_i$, the following holds:

$$\big\{z_{i1}(x), \dots, z_{im_i}(x)\big\} = \big\{t_0 \in R \mid \exists j \colon f_j(x, t_0) = 0 \text{ and } f_j(x, t) \not\equiv 0\big\}.$$

(3) The polynomials $f_1, \dots, f_r$ have constant sign on every graph $\Gamma(z_{ij})$ and every $\operatorname{Band}(z_{ij}, z_{i,j+1})$ (for $j = 0, \dots, m_i + 1$), where we let $z_{i0} = -\infty$ and $z_{i,m_i+1} = \infty$.

We refer to [BCR98, §2.3] and [BPR06, §5.1] for the proof, but just point out one technical difficulty. If we think of the example with the disk above, it might seem that it is enough to parametrize all the roots of the polynomials $f_1, \dots, f_k$ in T as a function of X and examine where the different roots collide, but we might run into something like in the following example: Let $f(X, T) = (X + (T - 1)^2)^2 \cdot (X - (T + 1)^2)^2$.



As X varies, this polynomial always has exactly two distinct real roots in T, but there is a discontinuity. This problem is avoided by adding the derivatives (with respect to T) to the description (see Thom's Lemma [BCR98, Prop. 2.5.4]).

Definition. Let $S \subset R^{n+1}$ be a semialgebraic set. A **cylindrical algebraic decomposition (CAD)** of S with respect to the projection $\pi \colon R^{n+1} \to R^n$ is a decomposition of R^n as in Thm. 8.2, such that S is a union of graphs and bands.

8.3 Corollary. *Every semialgebraic set possesses a CAD.*

Proof. Apply Thm. 8.2 to all polynomials occuring in the description of the input set. ∎

8.4 Corollary. *Every semialgebraic set $S \subset R^n$ can be written as the disjoint union of semialgebraic sets $S = S_1 \cup \dots S_k$, where each S_i is semialgebraically homeomorphic to $(0, 1)^{n_i}$, for some $n_i \geqslant 0$. (Here, $(0, 1)^0$ is a single point).*

Proof. Over a semialgebraic set $S_i \subset R^k$, every graph is semialgebraically homeomorphic to S_i and every band to $S_i \times (0,1)$. Applying CAD successively for every projection $R^n \to R^{n-1} \to \cdots \to R^2 \to R$ yields the claim. ∎

A typical example, where CAD can often be applied in practice, is to determine the topology of a curve in the real (or real projective) plane; see [BPR06, §11.6] for a detailed analysis. Some examples will follow below.

There are many natural follow-up questions. For example, one would very much like to have the closure of each cell to be a union of cells. Also, it can be useful to have not just semialgebraic homeomorphisms, but $\mathbb{C}^\infty$-diffeomorphisms. (For a real closed field other than R, one first has to introduce *Nash functions* to make sense of this.) All these additional requirements can be satisfied for semialgebraic sets.

There are also decomposition results for semialgebraic maps: Let S and B be two semialgebraic sets and let $\varphi\colon S \to B$ be a semialgebraic map. We want to think of φ as a family of semialgebraic sets $\varphi^{-1}(\{x\})$ as x varies over the *base* B.

8.5 Theorem (Hardt-trivialization). *Let $\varphi\colon S \to B$ be a semialgebraic map. There exists a semialgebraic partition $B = \bigcup_{i=1}^r B_i$ and for each $i = 1, \dots, r$ a semialgebraic homeomorphism*

$$\Theta_i\colon B_i \times F_i \to \varphi^{-1}(B_i)$$

for some semialgebraic set F_i which is a local trivialization of φ, i.e. $\varphi \circ \Theta_i$ equals the first projection $B_i \times F_i \to B_i$. In particular, all fibres of φ over B_i are semialgebraically homeomorphic to F_i.

Further refinements are possible. See [BCR98, Thm. 9.3.2].

8.6 Example. Let $R[X]_{\leqslant d}$ be the space of polynomials of degree at most d in n variables, which we may identify with R^N, $N = \binom{n+d}{n}$. The zero set $\mathcal{Z}(f)$ for $f \in R[X]_{\leqslant d}$ is a **real hypersurface** of degree at most d in R^n. We let

$$\mathcal{X} = \{(f, a) \in R[X]_{\leqslant d} \times R^n \mid f(a) = 0\}$$

which is a semialgebraic subset of $R^N \times R^n$. The fibres of the projection $\pi\colon \mathcal{X} \to R[X]_{\leqslant d}$ onto the first factor are exactly the hypersurfaces of degree at most d, i.e. $\pi^{-1}(\{f\}) = \{f\} \times \mathcal{Z}(f)$. A Hardt trivialization of this projection immediately implies that there can be only finitely many semialgebraic homeomorphism classes of such hypersurfaces. ◇

Problem 8. Discuss the possibilities for extending this argument to larger classes of semialgebraic sets than just hypersurfaces.

8.7 Example. Let $C \subset \mathbb{R}^n$ be a convex body (compact, convex with non-empty interior). We may wish to approximate C by a polytope, so we pick points on the boundary $x_1, \dots, x_r \in \partial C$ and form the convex hull $P_r = \mathrm{Conv}(x_1, \dots, x_r)$. By choosing more and more points distributed over the boundary, we can approximate C to arbitary precision. We may even turns this into a continuous family $\varphi\colon S \to [1, \infty)$ with $\varphi^{-1}(\{k\}) \cong P_k$ for every positive integer k with a linear interpolation between P_k and P_{k+1}. The set C is the *limit* of this family (e.g. in the Hausdorff metric).

But even though each individual P_k is a semialgebraic set, this cannot possibly be a semialgebraic family, unless the set $\{x_1, x_2, \dots\}$ is finite. In fact, the trivialization theorem above can be used to show that the number of vertices must be bounded in every semialgebraic family of polytopes. (How would you show this?) ◇

Decomposition theorems are also useful to characterize the dimension of a semialgebraic set.

Definition. The **dimension** of a semialgebraic set $S \neq \emptyset$ is the largest integer d such that there exists a non-empty open semialgebraic set $U \subset R^d$ and an injective semialgebraic map $U \to S$. Also, $\dim(\emptyset) = -\infty$.

This definition is geometrically plausible and it is not too difficult to prove (although not trivial!) that $\dim(R^n) = n$ and to establish some further natural properties (see [BPR06, §5.3]).

Also, it is clear that the dimension can be computed using CAD: Since any band has non-empty interior, while all graphs have empty interior, just keep applying CAD to S and its projections until you encounter a band (c.f. Cor. 8.4). Of course, this may be a rather expensive computation. A similar but more optimized algorithm is described together with a full complexity analysis in [BPR06, §14.5].

There is another useful characterization of dimension that can be computationally more tractable, provided that the semialgebraic set is presented in a good way: We know that every semialgebraic set is a finite union of sets of the form

$$S_i = \mathcal{Z}(f_1, \dots, f_k) \cap \mathcal{U}(g_1, \dots, g_l).$$

This is an open subset of the real algebraic set $\mathcal{Z}(f_1, \dots, f_k)$. If S_i is non-empty, its dimension will *hopefully* agree with that of $\mathcal{Z}(f_1, \dots, f_k)$. The "dimension" of an algebraic set is easier to compute using commutative algebra: It is the Krull dimension of the factor ring $R[X]/(f_1, \dots, f_k)$, which can be computed, for example, using Hilbert polynomials (see [CLO15, Ch. 9]).

However, there is a big catch: This computes the **algebraic dimension** $\dim_{\text{alg}}(Z_C)$ of the complex algebraic variety $Z_C = \mathcal{Z}_C(f_1, \dots, f_k) \subset C^n$, where $C = R(\sqrt{-1})$. We are only interested in the real points, and that dimension may well be smaller.

A simple example in the plane shows what can go wrong.

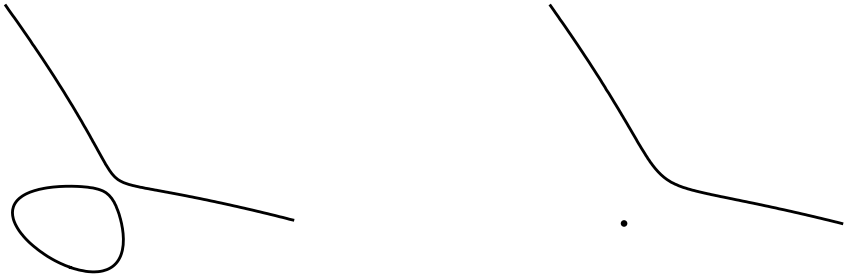
8.8 Example. We will take the cubic polynomials

$$f_1 = (x_1 - x_2)^2 - \left(x_1 + \tfrac{3}{2}y - 1\right)\left((x_1 - \tfrac{3}{2}x_2)^2 - \tfrac{1}{2}\right)$$

and

$$f_2 = (x_1 - x_2)^2 - \left(x_1 + \tfrac{3}{2}y - 1\right)\left(x_1 + \tfrac{3}{2}x_2\right)\Big)^2.$$

(Of course, I could have made these simpler. But I do not want to think about the equations, only the pictures.) Close to the origin, the real algebraic curves $\mathcal{Z}(f_1)$ and $\mathcal{Z}(f_2)$ look like this:

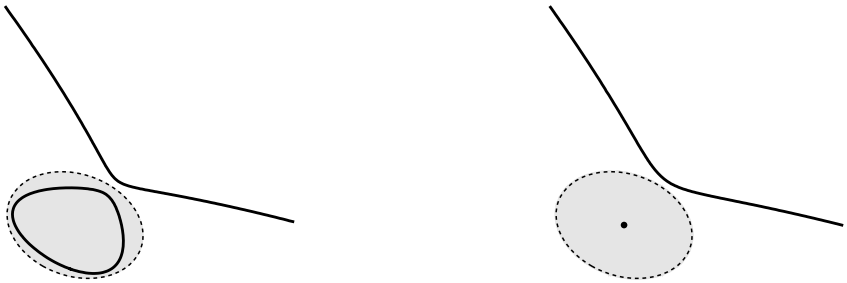


$\mathcal{Z}(f_1)$ $\mathcal{Z}(f_2)$

Both are one-dimensional. (Also, we could easily compute a CAD to determine the topology exactly.) If we now add a quadratic inequality $g \geqslant 0$ with

$$g = 3 - 8x_1^2 - 4x_1x_2 - 13x_2^2$$

defining an oval region around the origin, we obtain the semialgebraic sets



$\mathcal{Z}(f_1) \cap \mathcal{U}(g)$ $\mathcal{Z}(f_2) \cap \mathcal{U}(g)$

If we determine the dimension in the naive way indicated above, we see that we will get it wrong in the second example: The set $\mathcal{Z}(f_2) \cap \mathcal{U}(g)$ is only a point, even though $\mathcal{Z}(f_2)$ is one-dimensional. This happens, because the point we selected is a singularity of the curve. ◇

8.9 Example. Note that rewriting a real algebraic set $\mathcal{Z}(f_1, \dots, f_k)$ into the single equation $\mathcal{Z}(f_1^2 + \cdots + f_k^2)$ will usually change the algebraic dimension. The algebraic dimension of the complex variety $\mathcal{Z}_{\mathbb{C}}(f_1^2 + \cdots + f_k^2)$ is always $n-1$, no matter what $f_1, \dots, f_k$ are (except if they are all constant). It is a complex hypersurface. All real points on this hypersurface are (usually) singular. The simplest example is replacing the point $\mathcal{Z}(x_1, x_2)$ by the singular conic $\mathcal{Z}(x_1^2 + x_2^2)$. ◇

So what does this mean for the dimension of the semialgebraic set

$$S = \mathcal{Z}(f_1, \dots, f_k) \cap \mathcal{U}(g_1, \dots, g_l)$$

in general? In theory, we may simply replace the real algebraic set $\mathcal{Z}(f_1, \dots, f_k)$ by the **real Zariski closure** of S, which is the smallest real algebraic set containing S. If we do that, the semialgebraic and the algebraic dimension will always agree (see the following theorem). In fact, in [BCR98, §2.8] this is taken as the *definition* of the semialgebraic dimension. In practice, computing the real Zariski closure is another difficult computational task, relying on a version of the real Nullstellensatz (Thm. 7.2).

Let us summarize our findings for real algebraic sets in a final theorem, together with some additional characterizations.

8.10 Theorem. *Let $Z = \mathcal{Z}(f_1, \dots, f_k) \subset R^n$ be an algebraic set, let $Z_C = \mathcal{Z}_C(f_1, \dots, f_k)$ be the complex variety defined by the same equations, and assume that Z_C is irreducible (not a finite union of smaller varieties defined over R).*

(1) We have $\dim(Z) \leqslant \dim_{\mathrm{alg}}(Z_C)$.

(2) The equality

$$\dim(Z) = \dim_{\mathrm{alg}}(Z_C)$$

holds if and only if the real algebraic set Z contains a regular point of Z_C.

(3) Assuming that the ideal $I = (f_1, \dots, f_k)$ generated by $f_1, \dots, f_k$ in $R[X]$ is a prime ideal, the following statements are equivalent:

(i) The algebraic set Z contains a regular point of Z_C.

(ii) There exists a point $x \in Z$ such that the Jacobi matrix of the system $f_1, \dots, f_k$ evaluated at x has rank $n - \dim_{\mathrm{alg}}(Z_C)$.

*(iii) The ideal I is **real radical**, which means that if $f \in R[X]$ vanishes at all points of Z, then it is contained in I.*

(iv) The element -1 is not a sum of squares in the function field $R(Z)$, which is the field of fractions of the factor ring $R[X]/I$. (Equivalently, $R(Z)$ admits an ordering; Lemma 5.3.)

Sketch of proof and references. Most of what is needed for a proof can be found in [BCR98], but some results are stated differently

(2) follows from [BCR98, Prop. 7.6.2]. (1) is a special case of (2), since the singular locus of Z_C is an algebraic subvariety and therefore of strictly smaller algebraic dimension.

(3) The equivalence of (i) and (ii) is a standard result in Algebraic Geometry, or may be taken as the definition of *regular*; see for example [Har77, §I.5].

If (i) does not hold, then Z contains only singular points of Z_C. The singular locus is a proper subvariety of Z_C, defined over R. Hence there are polynomials in $R[X]$ vanishing on Z but not on all of Z_C. This shows the implication (iii)⇒(i).

(i),(ii)⇒(iii) can be seen more or less directly for $R = \mathbb{R}$ from the implicit function theorem: Around a real regular point $z \in Z$, the set Z will be an embedded real submanifold of $\mathbb{R}^n$ of dimension $\dim_{\mathrm{alg}}(Z_C)$. Therefore, it cannot be contained in any subvariety of Z_C of smaller dimension, which implies (iii). Alternatively, one can again use [BCR98, Prop. 7.6.2].

The equivalence of (iii) and (iv) is a special case of the real Nullstellensatz (Thm. 7.2). If I is not real radical, then there exists some $h \in R[X]$ which vanishes on all of Z but is not contained in I. By the real Nullstellensatz, there is an identity $h^{2r} + s_1^2 + \cdots + s_l^2 \in I$ for some polynomials $s_1, \dots, s_l \in R[X]$. Modulo I, this reads $-h^{2r} \equiv s_1^2 + \cdots + s_l^2$, where h is non-zero. Dividing by h^{2r} gives $-1 = (s_1/h^r)^2 + \cdots + (s_l/h^r)^2 \in R(Z)$.

The other implication is the easy direction of the Nullstellensatz: If $-1 = (s_1/h_1)^2 + \cdots + (s_l/h_l)^2$ in $R(Z)$, then we can multiply with h^2 for $h = h_1 \cdots h_l$ and find $-h^2 = s_1^2 + \cdots + s_l^2$ modulo I. This implies that h vanishes at all points of Z. But h is not in I, because I is prime and $h_1, \dots, h_l \notin I$. Hence I is not real radical. ∎

In conclusion, using all of the above, we may suggest the following practical procedure to compute the dimension of a semialgebraic set S given in the standard form

$$S = \mathcal{Z}(f_1, \dots, f_k) \cap \mathcal{U}(g_1, \dots, g_l).$$

Step 1. Decompose the algebraic set $\mathcal{Z}(f_1, \dots, f_k)$ into its irreducible components and find prime ideal defining each component. This amounts to a *primary decomposition* of the ideal $(f_1, \dots, f_k)$ in the real polynomial ring $R[X]$.

Step 2. Check if $\mathcal{U}(g_1, \dots, g_l)$ contains a real regular point for each component, using one of the criteria in the Thm. 8.10. If yes, then $\dim(S)$ is the algebraic dimension of $\mathcal{Z}(f_1, \dots, f_k)$.

References

[ABR96] Andradas, C., Bröcker, L., and Ruiz, J. **Constructible sets in real geometry**. Vol. 33. Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. Springer, 1996.

[BCR98] Bochnak, J., Coste, M., and Roy, M.-F. **Real algebraic geometry**. Vol. 36. Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. Springer, 1998.

[BPR06] Basu, S., Pollack, R., and Roy, M.-F. **Algorithms in real algebraic geometry**. 2nd ed. Vol. 10. Algorithms and Computation in Mathematics. Springer, 2006.

[CLO15] Cox, D. A., Little, J., and O'Shea, D. **Ideals, varieties, and algorithms. An introduction to computational algebraic geometry and commutative algebra**. English. 4th revised ed. Undergraduate Texts in Mathematics. Springer, 2015.

[Har77] Hartshorne, R. **Algebraic geometry**. Vol. 52. Graduate Texts in Mathematics. Springer, 1977.

[KS22] Knebusch, M. and Scheiderer, C. **Real algebra. A first course. Translated from the German and with contributions by Thomas Unger**. Universitext. Springer, 2022.

[Mar08] Marshall, M. **Positive polynomials and sums of squares**. Vol. 146. Mathematical Surveys and Monographs. American Mathematical Society, 2008.

[MS15] Maclagan, D. and Sturmfels, B. **Introduction to tropical geometry**. Vol. 161. Graduate Studies in Mathematics. American Mathematical Society, 2015.