THE SEMIALGEBRAIC CASE

Today's Goal

Tarski-Seidenberg Theorem. *For every $n = 1, 2, 3, \ldots$, every $\mathcal{L}_{\mathbf{alg}}$-definable subset of $\mathbb{R}^n$ can be defined by a quantifier-free $\mathcal{L}_{\mathbf{alg}}$-formula.*

Thus every $\mathcal{L}_{\mathbf{alg}}$-definable subset of $\mathbb{R}^n$ is a finite boolean combination (i.e., finitely many intersections, unions, and complements) of sets of the form

$$\{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid p(x_1, \ldots, x_n) > 0\}$$

where $p(x_1, \ldots, x_n)$ is a polynomial with coefficients in $\mathbb{R}$. These are called the *semialgebraic* sets.

A *function* $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ is *semialgebraic* if its graph is a semialgebraic subset of $\mathbb{R}^n \times \mathbb{R}^m$.

As for the semilinear sets, every semialgebraic set can be written as a finite union of the intersection of finitely many sets defined by conditions of the form

$$p(x_1, \ldots, x_n) = 0$$
$$q(x_1, \ldots, x_n) > 0$$

where $p(x_1, \ldots, x_n)$ and $q(x_1, \ldots, x_n)$ are polynomials with coefficients in $\mathbb{R}$.

Proof Strategy

- Prove a geometric structure theorem that shows that any semialgebraic set can be decomposed into finitely many semialgebraic generalized cylinders and graphs.
- Deduce quantifier elimination from this.

Thom's Lemma. *Let $p_1(X), \ldots, p_k(X)$ be polynomials in the variable X with coefficients in $\mathbb{R}$ such that if $p_j'(X) \neq 0$ then $p_j'(X)$ is included among $p_1, \ldots, p_k$. Let $S \subset \mathbb{R}$ have the form*

$$S = \cap_j p_j(X) *_j 0$$

where $*_j$ is one of $<$, $>$, or $=$, then S is either empty, a point, or an open interval. Moreover, the (topological) closure of S is obtained by changing relaxing the sign conditions (changing $<$ to $\leq$ and $>$ to $\geq$.

Note There are 3^k such possible sets, and these form a partition of $\mathbb{R}$.

Some Standard Tricks

- Identify the complex numbers $\mathbb{C}$ with $\mathbb{R}^2$ via

$$a + bi \quad \longleftrightarrow \quad (a, b)$$

  where $a, b \in \mathbb{R}$ and $i = \sqrt{-1}$. With this identification, multiplication of complex numbers is a semialgebraic function from $\mathbb{R}^2 \times \mathbb{R}^2$ to $\mathbb{R}^2$. Also, $\mathbb{C}^n$ is identified with $\mathbb{R}^{2n}$

- The collection of polynomials in the variable X with coefficients in $\mathbb{R}$ of degree not greater than n can be identified with $\mathbb{R}^{n+1}$ via

$$a_0 + a_1 X + \cdots + a_n X^n \leftrightarrow (a_0, a_1, \ldots, a_n).$$

  Similarly for polynomials with coefficients in $\mathbb{C}$. Addition, multiplication, differentiation of polynomials are semialgebraic functions.

Let $B_k^n(\mathbb{R})$ denote (as a subset of $\mathbb{R}^{n+1}$) the collection of polynomials in the variable X with real coefficients of degree not greater than n that have *exactly* k distinct complex roots.

Let $M_k^n(\mathbb{R}) \subset B_k^n(\mathbb{R})$ be those polynomials of degree n with this property.

Lemma. *Suppose that $A \subset M_k^n(\mathbb{R})$ is connected. For each $\bar{a} \in A$ let $r_{\bar{a}}$ be the number of distinct real roots of the polynomial $p_{\bar{a}}(X)$ associated with $\bar{a}$. Then*

a. *$r_{\bar{a}} = r$ is constant on A;*
b. *There are continuous functions*

$$f_1, \ldots, f_r \colon A \to \mathbb{R}$$

such that for all $\bar{a} \in A$ we have $f_i(\bar{a}) < f_{i+1}(\bar{a})$ for $i = 1, \ldots, r-1$ and $p_{\bar{a}}(f_i(\bar{a})) = 0$ for $i = 1, \ldots, r$.

("Continuity of roots")

Lemma. *The subsets* $B_k^n(\mathbb{R})$ *and* $M_k^n(\mathbb{R})$ *of* $\mathbb{R}^{n+1}$ *are semialgebraic.*

Idea

- The polynomial $p(X)$ has a repeated root if and only if $p(X)$ and its derivative $p'(X)$ have a common factor.
- This can be expressed by the condition that the determinant of a matrix constructed from the coefficients of $p(X)$, the so-called *resultant* of p and p' (also called the *discriminant* of p), has value 0.
- This is a semialgebraic condition on the coefficients.
- Then extend this idea to capture $B_k^n(\mathbb{R})$ and $M_k^n(\mathbb{R})$ semialgebraically.

Graphs and Cylinders

The structure theorem shows that a semialgebraic set $S \subseteq \mathbb{R}^n$ can be partitioned into finitely many sets of two kinds, all of which are semialgebraic.

Graphs

Let $A \subset \mathbb{R}^k$ and $f\colon A \to \mathbb{R}$ be continuous. The *graph of f* is the subset of $\mathbb{R}^{k+1}$ given by

$$\text{Graph}(f) = \{(\bar{x}, y) \in \mathbb{R}^{k+1} \mid \bar{x} \in A \text{ and } y = f(\bar{x})\}.$$

Generalized Cylinders

Let $A \subset \mathbb{R}^k$, and let $f, g\colon A \to \mathbb{R}$ be continuous and satisfy $f(\bar{x}) < g(\bar{x})$ for all $\bar{x} \in A$. The *cylinder* determined by f, g, and A is the subset of $\mathbb{R}^{k+1}$ given by

$$(f, g)_A = \{(\bar{x}, y) \in \mathbb{R}^{k+1} \mid \bar{x} \in A \text{ and } f(\bar{x}) < y < g(\bar{x})\}.$$

- If A is connected, then graphs and cylinders based on A are connected.

Structure Theorem. *Let $S \subset \mathbb{R}^n$ be semialgebraic. Then:*

I_n. *S has finitely many connected components and each one is semialgebraic*

II_n. *There is a finite partition $\mathcal{P}$ of $\mathbb{R}^{n-1}$ into connected semialgebraic sets such that for each $A \in \mathcal{P}$ there is $k_A \in \mathbb{N}$ and $f_i^A \colon A \to \mathbb{R} \cup \{\pm\infty\}$ for $i = 0, 1 \ldots, k_A + 1$ satisfying*

a. *$f_0^A = -\infty$, $f_{k_A+1}^A = \infty$, f_i^A is continuous for $1 \leq i \leq k_A$, and $f_{i-1}^A(\bar{x}) < f_i^A(\bar{x})$ for all $1 \leq i \leq k_A + 1$ and $\bar{x} \in A$;*

b. *all graph sets $\mathrm{Graph}(f_i^A)$ for $1 \leq i \leq k_A$ and generalized cylinders $(f_{i-1}^A, f_i^A)_A$ are semialgebraic.*

The graphs and cylinders in (b) for all $A \in \mathcal{P}$ partitions $\mathbb{R}^n$ and S.

The Proof

- The proof is by induction on n, and I shall outline the induction step.
- Most broadly, the argument is as follows: show $\mathrm{I}_{n-1} \Rightarrow \mathrm{II}_n$ and $\mathrm{II}_n \Rightarrow \mathrm{I}_n$.
- $\mathrm{II}_n \Rightarrow \mathrm{I}_n$ is evident; the crux is $\mathrm{I}_{n-1} \Rightarrow \mathrm{II}_n$.
- Split the coordinates of $\mathbb{R}^n$ as

$$(x_1, \ldots, x_{n-1}, t).$$

- Using standard set theory, write S as the union of finitely many finite intersections of polynomial equalities and inequalities.
- Extend the finite collection of polynomials in the given definition of S by including all iterated partial derivatives with respect to t. Let this expanded list of polynomials be $q_1, \ldots, q_r$.

- For each subset $\mathcal{S} \subset \{1, \ldots, r\}$, form the polynomial

$$Q_{\mathcal{S}}(\bar{x}, t) = \prod_{j \in \mathcal{S}} q_j(\bar{x}, t).$$

  View $\bar{x}$ as parameter variables and consider the polynomial as $Q_{\mathcal{S},\bar{x}}(t)$, a polynomial in the variable t whose coefficients are polynomials in $\bar{x}$.

- For each $\ell < \text{degree}\, Q_{\mathcal{S},\bar{x}}(t)$ and $k \leq \ell$, let

$$M^{\ell}_{\mathcal{S},k} = \{\bar{x} \in \mathbb{R}^{n-1} \mid \text{degree}\, Q_{\mathcal{S},\bar{x}}(t) = \ell$$

and it has exactly k distinct real roots$\}$.

  It can be shown that $M^{\ell}_{\mathcal{S},k}$ is semialgebraic.

- Partition $\mathbb{R}^{n-1}$ by taking all intersections of all $M^{\ell}_{\mathcal{S},k}$. This still is a finite semialgebraic partition of $\mathbb{R}^{n-1}$.

- Refine this partition further to obtain a partition $\mathcal{P}_0$ by taking the connected components of the sets in the partition above. By I_{n-1} this again is a finite semialgebraic partition of $\mathbb{R}^{n-1}$.
- For $A \in \mathcal{P}_0$, let $Q_{A,\bar{x}}(t)$ be the product of those $q_j(\bar{x}, t)$ which are nonzero for (all) $\bar{x} \in A$. It can be shown that the number of roots of $Q_{A,\bar{x}}(t)$ is uniform as $\bar{x}$ ranges over A and that the 1^{st}, 2^{nd}, ... root functions are continuous on A, as A is connected.
- Form the corresponding graph and generalized cylinder sets above each set $A \in \mathcal{P}_0$.
- It can be shown that each such set has the form

$$\cap_{j=1}^{r} \{(\bar{x}, t) \mid \bar{x} \in A \text{ and } q_j(\bar{x}, t) *_j 0\}$$

where $*_j$ is one of $<$, $>$, or $=$. This step uses Thom's Lemma.

Tarski-Seidenberg Theorem redux.
Let $f\colon X \subset \mathbb{R}^n \to \mathbb{R}^m$ *be semialgebraic. Then the image of* f*,*

$$f(X) = \{\bar{y} \in \mathbb{R}^m | \bar{y} = f(\bar{x}) \text{ for some } \bar{x} \in X\},$$

is semialgebraic.

Algorithmic Cruelty

Tarski's original proof gives an algorithm for quantifier-elimination: given an $\mathcal{L}_{\text{alg}}$-formula as input, it outputs a quantifier-free formula that defines the same set as the input formula.

Computationial efficiency of a quantifier elimination algorithm thus becomes important for applications (e.g., robot motion planning). Cylindrical algebraic decomposition-based quantifier elimination, such as described above and developed in 1973 by Collins, has played an important role.

Quantifier elimination for $\mathbb{R}_{\text{alg}}$ is, unfortunately, an inherently computationally intensive problem. It is known that there is a doubly exponential lower bound in the number of quantifiers for worst-case time complexity.

Some References

R. Benedetti and J.-J. Risler, *Real algebraic and semi-algebraic sets*, Paris: Hermann, 1990. (particularly Chapters 1 and 2)

J. Bochnak, M. Coste, and M.-F. Roy, *Real algebraic geometry*, Berlin: Springer-Verlag, 1998.

B.F. Caviness and J. R. Johnson (eds.), *Quantifier Elimination and Cylindrical Algebraic Decomposition*, Vienna: Springer-Verlag, 1998.

L. van den Dries, *Tame Topology and O-minimal Structures* (London Mathematical Society Lecture Note Series, vol. 248), Cambridge: Cambridge University Press, 1998. (Chapter 2)