Lecture 1: An introduction to definability and quantifier elimination

I first want to thank Professor Brown for his kind introduction. My goal is to give everyone a sense of the subject of o-minimality. Ten hours is a lot of time to lecture, but the literature on this topic has grown vast, so I will ignore a lot of things in order to get at issues relevant for this audience. I confess that this is the first time I have spoken to economists.

I want to start with an outline of the talks. The topic is tame topology and o-minimal structures, but today I'm not going to tell you what either one really is. Today I'm going to give you a sense of what's important so that you can understand things better in my subsequent lectures.

Outline of lectures

1. An introduction to definability and quantifier elimination
2. The semialgebraic case
3. O-minimality and some basic properties
4. Examples and some further properties
5. VC dimension and applications

An introduction to definability

For which x do we have

$$\exists y \begin{bmatrix} ax + by > e_1 \\ cx + dy \le e_2 \end{bmatrix} \tag{$*$}$$

where $a, b, c, d, e_1, e_2 \in \mathbb{R}$? We want to know

$$\{x \in \mathbb{R} \,|\, \text{statement } (*) \text{ holds for } x\}.$$

We also can replace the constants a, b, c, d, e_1, e_2 by variables u, v, w, z, s_1, s_2 and ask for which $u, v, w, z, s_1, s_2 \in \mathbb{R}$ do we have

$$\exists x \exists y \begin{bmatrix} ux + vy > s_1 \\ wx + zy \le s_2 \end{bmatrix}. \tag{$**$}$$

That is, we want to know

$$\left\{(u, v, w, z, s_1, s_2) \in \mathbb{R}^6 \mid (u, v, w, z, s_1, s_2) \text{ satisfies } (**)\right\}.$$

These are simple examples of *definable sets*. I want to give a couple more examples before I start to get precise.

Examples

1. $S = \{x \in \mathbb{R} \mid \sin \pi x = 0\}$

   Observe $S = \mathbb{Z}$.

2. Let $L = \{(m,n) \in \mathbb{R}^2 \mid m,n \in \mathbb{Z}\}$ be the integer lattice in $\mathbb{R}^2$. View L as a two-place relation on pairs of real numbers:

$$L(m,n) \iff m,n \in \mathbb{Z}.$$

   Define

$$C = \left\{(x,y) \in R^2 \mid \exists u \exists v \left[L(u,v) \wedge (x-u)^2 + (y-v)^2 \leq 1/16\right]\right\}.$$

   The set C is the set of all closed discs of radius 1/4 whose centers are points in L.

You can see that as you introduce quantifiers, you have to think a bit about what you are defining.

Structures on the real numbers $\mathbb{R}$

Throughout these lectures, I am going to limit my discussion to structures on the real line, for concreteness. Fix a stock

1. $\mathcal{F}$ of *basic functions* $f : \mathbb{R}^k \to \mathbb{R}$, for $k = 1, 2, 3, \ldots$

2. $\mathcal{R}$ of *basic relations* (subsets) $R \subseteq \mathbb{R}^k$, for $k = 1, 2, 3, \ldots$, such that the "less than" relation $< \subset \mathbb{R}^2$ (and equality) always is included in $\mathcal{R}$.

Refer to $\mathcal{F}$ and $\mathcal{R}$ together as a language $\mathcal{L}$. The real numbers $\mathbb{R}$ together with the functions and relations included in $\mathcal{L}$ is called an $\mathcal{L}$-*structure* that in general we denote by $\mathfrak{R}_{\mathcal{L}}$. Special structures will have special names.

Examples

1. $\mathcal{F} = \{+, \cdot, -\}$ where $- : \mathbb{R} \to \mathbb{R}$ is given by $x \longmapsto -x$ and $\mathcal{R} = \{<\}$. This is the language $\mathcal{L}_{\text{alg}}$ of the ordered field of real numbers, the structure denoted by $\mathbb{R}_{\text{alg}}$.

2. $\mathcal{F}$ contains $+$ and, for each $r \in \mathbb{R}$, the scalar multiplication function $\mu_r : \mathbb{R} \to \mathbb{R}$ given by $x \longmapsto rx$, and $\mathcal{R} = \{<\}$. This is the ordered real vector space language $\mathcal{L}_{\text{lin}}$ for the real numbers, whose corresponding structure is denoted by $\mathbb{R}_{\text{lin}}$.

3. $\mathcal{F} = \{+, \cdot, -, \exp\}$ where $\exp : \mathbb{R} \to \mathbb{R}$ is the exponential function $x \longmapsto e^x$ and $\mathcal{R} = \{<\}$. This is the language $\mathcal{L}_{\text{exp}}$ of the ordered exponential field of real numbers; the structure is denoted by $\mathbb{R}_{\text{exp}}$.

Terms

Construct an expanded class of functions by repeated composition starting from the functions in $\mathcal{F}$. These are called *terms.*

Examples

1. The $\mathcal{L}_{\text{alg}}$-terms are all integer coefficient polynomial functions $p(x_1, \ldots, x_k)$ for $k = 1, 2, 3, \ldots$
2. The $\mathcal{L}_{\text{lin}}$-terms are all $\mathbb{R}$-linear functions $L(x_1, \ldots, x_k) = \mu_{r_1} x_1 + \cdots + \mu_{r_k} x_k$.

The obvious kinds of rules apply in these examples, for instance $\mu_{r_1}(x) + \mu_{r_2}(x) = \mu_{r_1+r_2}(x)$.

Formulas

If you think of the definable sets I wrote down at the beginning of the lecture, they were defined in terms of equalities and inequalities. The basic formulas are

1. $t_1 = t_2$ for terms t_1, t_2.
2. $R(t_1, \ldots, t_k)$ where R is a k-place relation and $t_1, \ldots, t_k$ are terms. Note that $t_1 < t_2$ is a special case.

So if you think of $\mathbb{R}_{\text{alg}}$, we can write down polynomial equalities and inequalities, and for $\mathbb{R}_{\text{lin}}$ we can write down linear equalities and inequalities. Next, from the basic formulas recursively construct

1. by boolean operations from formulas φ and ψ :

$$\begin{array}{ll} \varphi \wedge \psi & \text{read as “}\varphi\text{ and }\psi\text{”} \\ \varphi \vee \psi & \text{read as “}\varphi\text{ or }\psi\text{”} \\ \neg\varphi & \text{read as “not }\varphi\text{”.} \end{array}$$

2. by existential quantification over $\mathbb{R}$ from a formula φ:

$$\exists v \varphi \quad \text{read as “there exists } v \text{ such that } \varphi\text{”.}$$

Caution: to get to the point more quickly I have mixed syntax and semantics. When you look at a formal text on model theory you will see a distinction between symbols in a formal language and their interpretations.

Now what I'm going to do is go back to what I was vague about earlier: $\mathcal{L}$-definable sets.

$\mathcal{L}$-definable sets

For our purposes these are subsets of $\mathbb{R}^k$, for $k = 1, 2, 3, ...$, specified as follows. For each $\mathcal{L}$-formula φ, certain variables are bound to quantifiers and others are not. Call the latter *free variables*, and it is for these that we can substitute real numbers.

For an $\mathcal{L}$-formula φ list its free variables as $x_1, ..., x_k, z_1, ..., z_m$. Choose $c_1, ..., c_m \in \mathbb{R}$ and substitute them for $z_1, ..., z_m$, respectively. Write $\overline{c}$ for $(c_1, ..., c_m)$. The set

$$D_{\varphi,\overline{c}} = \left\{(x_1, \ldots, x_k) \in \mathbb{R}^k \,|\, \varphi\,(x_1, \ldots, x_k, \overline{c}) \text{ is true}\right\} \subseteq \mathbb{R}^k$$

is an $\mathcal{L}$-definable subset of $\mathbb{R}^k$. If $\mathcal{L}$ is clear from context, we usually shall drop the $\mathcal{L}$.

Examples

Let's go back to our earlier examples of definable sets. The set

$$\{x \in \mathbb{R} \,|\, \text{statement } (*) \text{ holds for } x\}$$

is $\mathcal{L}_{\text{lin}}$-definable. The set

$$\left\{(u, v, w, z, s_1, s_2) \in \mathbb{R}^6 \mid (u, v, w, z, s_1, s_2) \text{ satisfies } (**)\right\}$$

is $\mathcal{L}_{\text{alg}}$-definable. The set

$$\{x \in \mathbb{R} \mid \sin \pi x = 0\}$$

is $(\mathcal{F}, \mathcal{R})$-definable, where $\mathcal{F} = \{\cdot, \sin\}, \mathcal{R} = \{<\}$. The set

$$\left\{(x, y) \in R^2 \,|\, \exists u \exists v \left[L\,(u, v) \wedge (x-u)^2 + (y-v)^2 \leq 1/16\right]\right\}$$

is $(\mathcal{F}, \mathcal{R})$-definable, where $\mathcal{F} = \{+, \cdot, -\}$, $\mathcal{R} = \{L, <\}$.

Comments

- We call the $\overline{c}$ *parameters*; the division of the free variables in a formula into those which are parameter variables and those that are not can be made arbitrarily.
- A function $F : A \subseteq \mathbb{R}^m \to \mathbb{R}^n$ is said to be $\mathcal{L}$-definable if its graph, as a subset of $\mathbb{R}^m \times \mathbb{R}^n$, is $\mathcal{L}$-definable.
- If a function $F\,(x_1, \ldots, x_m)$ is definable, we can treat it as if it is a term for the purposes of constructing more complicated definable functions and formulas.
- Similarly, we can treat a definable subset of $\mathbb{R}^k$ as a basic relation in the language for such purposes.

What I want to do now is consider some further examples of definability, to explore the capabilities of the language we are dealing with.

Some further examples

1. Let $f : \mathbb{R} \to \mathbb{R}$ be $\mathcal{L}$-definable, where $\mathcal{L}$ contains $\mathcal{L}_{\text{alg}}$. Then
$$\{x \in \mathbb{R} \,|\, f \text{ is convex in an interval around } x\}$$
is $\mathcal{L}$-definable, since "f is convex in an interval around x" can be written as
$$\exists x_1 \exists x_2 \left(x_1 < x \wedge x < x_2 \wedge \text{"}f \text{ is convex in the interval } (x_1, x_2)\text{"}\right),$$
and "f is convex in the interval (x_1, x_2)" can be written as
$$\forall t \left((x_1 < t \wedge t < x_2) \to f(t) < \frac{x_2 - t}{x_2 - x_1} f(x_1) + \frac{t - x_1}{x_2 - x_1} f(x_2)\right).$$
Note that $\forall t := \neg \exists t \neg$ and $P \to Q := \neg P \vee Q$.

2. Let $f : A \subset \mathbb{R}^k \to \mathbb{R}$ be $\mathcal{L}$-definable, where $\mathcal{L}$ contains $\mathcal{L}_{\text{alg}}$. Then $\{\bar{x} \in A \,|\, f \text{ is continuous at } \bar{x}\}$ is $\mathcal{L}$-definable, since "f is continuous at x" can be written as
$$\forall \varepsilon \left(\varepsilon > 0 \to \exists \delta \left(\delta > 0 \wedge \forall y \left(|y - x| < \delta\right) \to |f(y) - f(x)| < \varepsilon\right)\right).$$
Same for differentiability. Note that $|y - x| < \delta := -\delta < y - x \wedge y - x < \delta$.

3. Let $f : \mathbb{R} \to \mathbb{R}$ be $\mathcal{L}$-definable, where $\mathcal{L}$ contains $\mathcal{L}_{\text{alg}}$. If f is differentiable, then f' is $\mathcal{L}$-definable. Same for functions of several variables.

4. Let $f : \mathbb{R} \to \mathbb{R}$ be $\mathcal{L}$-definable, where $\mathcal{L}$ contains $\mathcal{L}_{\text{alg}}$. Then
$$\{x \in \mathbb{R} \,|\, f \text{ is Lipschitz in an interval around } x\}$$
is $\mathcal{L}$-definable.

5. Let $A \subset \mathbb{R}^k$ be $\mathcal{L}_{\text{alg}}$-definable. Then there is a formula that expresses that "A is convex". Same for any $\mathcal{L}$ containing $\mathcal{L}_{\text{alg}}$.

6. Let $A \subset \mathbb{R}^k$ be $\mathcal{L}$-definable, where $\mathcal{L}$ is arbitrary. Then the topological closure of A in $\mathbb{R}^k$ is $\mathcal{L}$-definable. Many other notions from point-set topology also are definable.

7. Let $A \subset \mathbb{R}^k$ be $\mathcal{L}$-definable. Then all level sets of A are $\mathcal{L}$-definable.

8. Closed polyhedra in $\mathbb{R}^k$ are $\mathcal{L}_{\text{lin}}$-definable.

Some guidelines for which sets are definable:

- Defining formulas cannot be infinitely long. The formula "$X \subseteq \mathbb{R}$ is finite", which we can write as "X contains 1 element $\vee$ X contains 2 elements $\vee \ldots$", is not definable.

- Quantification over real numbers only is allowed. Quantifying over the class of all polynomials is not permitted, but quantifying over all polynomials of degree $\leq n$ is allowed, since such polynomials are defined by $n+1$ coefficients.

Definability set-theoretically

The $\mathcal{L}$-definable subsets of $\mathbb{R}^n$ for $n = 1, 2, 3, \ldots$ is the smallest collection $\mathfrak{D} = \{\mathcal{D}_n \mid n \geq 1\}$ such that

1. Each $D \in \mathcal{D}_n$ is a subset of $\mathbb{R}^n$;
2. $\mathbb{R}^n \in \mathcal{D}_n$;
3. The graph of each $f : \mathbb{R}^n \to \mathbb{R}$ in $\mathcal{F}$ is in $\mathcal{D}_{n+1}$;
4. Each $R \subseteq \mathbb{R}^n$ in $\mathcal{R}$ is in $\mathcal{D}_n$;
5. For all $1 \leq i, j \leq n$, $\{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_i = x_j\} \in \mathcal{D}_n$;
6. Each $\mathcal{D}_n$ is closed under intersection, union, and complement;
7. If $\pi : \mathbb{R}^n \to \mathbb{R}^m$ is a projection map $(x_1, \ldots, x_n) \longmapsto (x_{i_1}, \ldots, x_{i_m})$ and $X \in \mathcal{D}_n$ then $\pi(X) \in \mathcal{D}_m$;
8. If π is as above and $Y \in \mathcal{D}_m$ then $\pi^{-1}(Y) \in \mathcal{D}_n$;
9. If $X \in \mathcal{D}_{n+m}$ and $\bar{b} \in \mathbb{R}^m$, then $\left\{\bar{a} \in \mathbb{R}^n \mid \left(\bar{a}, \bar{b}\right) \in X\right\} \in \mathcal{D}_n$.

Finer structure of definable sets

A definable set typically has several different definitions. We have the $\mathcal{L}_{\text{alg}}$ definitions of the unit interval $[-1, 1]$:

$$\begin{aligned}
&\{x \in \mathbb{R} \mid (-1 < x \wedge x < 1) \vee x = -1 \vee x = 1\} \\
&\left\{x \in \mathbb{R} \mid \exists y \exists z\ x^2 + y^2 + z^2 = 1\right\} \\
&\left\{x \in \mathbb{R} \mid \forall y \exists z\ x^2 + (y - z)^2 = 1\right\}.
\end{aligned}$$

These are three different definitions of the same set. We could construct infinitely many. The maxim here is that quantification generates conceptual complexity. One goal of model theory is to attempt to analyze definable sets in a specific context by showing that definable sets can be defined by simple formulas.

Quantifier elimination for $\mathcal{L}_{\text{lin}}$-definable sets

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

For convenience, write rx instead of $\mu_r(x)$, where $r \in \mathbb{R}$. The theorem tells us that every $\mathcal{L}_{\text{lin}}$-definable subset of $\mathbb{R}^n$ is a finite boolean combination (i.e., finitely many intersections, unions, and complements) of sets of the form

$$\{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid a_1 x_1 + \cdots + a_n x_n > b\}$$

where $a_1, ..., a_n, b$ are fixed but arbitrary real numbers.

The $\mathcal{L}_{\mathrm{lin}}$-definable sets are called the *semilinear* sets. By routine set theoretic manipulation, semilinear sets can be written as a finite union of the intersection of finitely many sets defined by conditions of the form

$$\begin{aligned} a_1x_1 + \cdots + a_nx_n + b &= 0 \\ c_1x_1 + \cdots + c_nx_n + d &> 0. \end{aligned}$$

Thus $\mathcal{L}_{\mathrm{lin}}$-definability reduces (basically) to linear algebra, which we understand well.

$\mathcal{L}_{\mathrm{lin}}$-definable subsets of $\mathbb{R}$

All $\mathcal{L}_{\mathrm{lin}}$-definable subsets of $\mathbb{R}$ are finite boolean combinations of sets of the form $\{x \in \mathbb{R} \mid ax > b\}$. Geometrically these are the union of finitely many (possibly unbounded) open intervals and points. Consequently, neither $\mathbb{Z}$ nor $\mathbb{Q}$ is $\mathcal{L}_{\mathrm{lin}}$-definable.

Idea of the proof

Eliminate quantifiers one at a time (proceed inductively).

Example

Eliminate the quantifier for the $\mathcal{L}_{\mathrm{lin}}$-definable set

$$\{x \in \mathbb{R} \mid \exists y\ [2x - 3y > 2 \wedge 4x - 2y \leq 0]\}\,.$$

High school algebraic elimination leads to $\{x \in \mathbb{R} \mid x < 1/2\}$.

Question: What about $\mathcal{L}_{\mathrm{alg}}$-definability? I will discuss this in my next lecture.

Some general references

D. Marker, *Model Theory: An Introduction.* Graduate Texts in Mathematics, vol. 217, New York: Springer-Verlag, 2002.

Two articles in

D. Haskell, A. Pillay, and C. Steinhorn (editors), *Model Theory, Algebra, and Geometry* (Mathematical Sciences Research Institute Publications, vol. 39), Cambridge: Cambridge University Press, 2000:

Haskell, Pillay, and Steinhorn, "Overview."

Marker, "Introduction to Model Theory."

Lecture 2: The semialgebraic case

Today I'm going to be discussing the semialgebraic case, that is, the case of $\mathcal{L}_{\text{alg}}$. To construct this language, you start with polynomials with real coefficients, and build up through the operations of $+, \cdot, -$. The focus of today's lecture is the Tarski-Seidenberg theorem.

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

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

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

where $p(x_1, ..., x_n)$ 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 I did yesterday, I want to draw special attention to the definable sets in one variable: the $\mathcal{L}_{\text{alg}}$-definable subsets of $\mathbb{R}$. You start off with sets of the form $p(x) > 0$, where $p(x)$ is a polynomial. This is nothing other than finitely many open intervals. Consider the negation $p(x) \leq 0$; this defines finitely many closed intervals. So, the $\mathcal{L}_{\text{alg}}$-definable subsets of $\mathbb{R}$ are simply the union of finitely many intervals and points. In particular, $\mathbb{Z}$ and $\mathbb{Q}$ are not $\mathcal{L}_{\text{alg}}$-definable. This is the same as in the case of $\mathcal{L}_{\text{lin}}$, which is surprising, because the structure of $\mathcal{L}_{\text{alg}}$ seems more complex. This complexity manifests itself in higher dimensions.

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

$$\begin{aligned} p(x_1, ..., x_n) &= 0 \\ q(x_1, ..., x_n) &> 0 \end{aligned}$$

where $p(x_1, ..., x_n)$ and $q(x_1, ..., x_n)$ are polynomials with coefficients in $\mathbb{R}$. This is just standard set theoretic manipulation of the kind I talked about yesterday.

Let me tell you now about the proof, which is the substance of what today's lecture is about. There are two steps. First we prove a geometric structure theorem that shows that any semialgebraic set can be decomposed into finitely many semialgebraic generalized cylinders and graphs. Then we deduce quantifier elimination from this. I will define what I mean by cylinders and graphs a little later on. We start with a nice result called Thom's lemma.

Thom's Lemma: *Let $p_1(X), ..., p_k(X)$ 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 $p_1, ..., p_k$. 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 the sign conditions (changing $<$ to $\leq$ and $>$ to $\geq$).

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

I'll draw a picture (*see figure 2.1*) to give you some idea of what the derivatives do for you. You can see that, by imposing conditions on derivatives, you can work your way down to a connected set.

Proof

The proof is by induction on k. When $k = 1$, $p_1(x)$ is a constant polynomial, so $*_j$ gives either $\mathbb{R}$ or $\emptyset$. Assume the theorem is true for $k-1$; we will show that it must be true for k. Without loss of generality, suppose $p_k(x)$ has the largest degree of the polynomials. Then $\{p_1, \ldots, p_{k-1}\}$ must also satisfy the conditions of Thom's lemma. Let $*_1, \ldots, *_k$ be given, and form the set $S' = \cap_{j=1}^{k-1} p_j(X) *_j 0$. If $S' = \emptyset$ or $\{r\}$, then it clear that $S = S' \cap p_k(X) *_k 0$ has the right form. Now suppose S' is an interval I. Note that p_k' is among $p_1, \ldots, p_{k-1}$. On I, $p_k'(X) > 0$ or $p_k'(X) < 0$ or $p_k'(X) = 0$. So p_k is either monotone or constant on I, and so S has the right form.

We now need some "tricks" to continue with our proof. First, we identify the complex numbers $\mathbb{C}$ with $\mathbb{R}^2$ via

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

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 $\mathbb{R}^2$. More generally, $\mathbb{C}^n$ is identified with $\mathbb{R}^{2n}$. Second, 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

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

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.

I'm now going to state a lemma that is vital for our proof.

Lemma ("continuity of roots"): *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*

1. *$r_{\bar{a}} = r$ is constant on A;*
2. *There are continuous functions $f_1, \ldots, f_r : A \to \mathbb{R}$ 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 $i = 1, \ldots, r$.*

I'm not going to prove this result, but I'll try to give you some intuition by going back to the quadratic case. Consider the polynomial $p(x) = \underline{a}x^2 +$

$\underline{b}x + \underline{c}$. The determining factor here is the discriminant $\underline{b}^2 - 4\underline{a}\underline{c}$. Consider the sets obtained by setting the discriminant equal to zero and fixing $\underline{a}$. This is a parabola in the plane $x = \underline{a}$ in $\mathbb{R}^3$ (with $\underline{b}$ and $\underline{c}$ serving as the y and z coordinates). We have zero distinct real roots below the parabola, two distinct real roots above the parabola, and one distinct real root on the parabola. So these three cases constitute connected sets.

Continuity of roots takes a little while to prove. You first prove local continuity, then use connectedness to show that the number of roots is constant across a set. This proof is involved, and I will not show it here.

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

The idea here is that 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 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. We ca extend this idea to capture $B_k^n(\mathbb{R})$ and $M_k^n(\mathbb{R})$ semialgebraically.

I'm going to go back to the case of the quadratic equation again. Let $p(x) = ax^2 + bx + c$, so that $p'(x) = 2ax + b$. The discriminant can be written as

$$\begin{aligned} D(p, p') &= \begin{vmatrix} c & b & a \\ b & 2a & 0 \\ 0 & b & 2a \end{vmatrix} \\ &= a\left(4ac - b^2\right). \end{aligned}$$

So we ask that $a\left(4ac - b^2\right) = 0$. This condition ensures that p has a repeated root. So we have a semialgebraic condition for p to have a repeated root.

What we've seen is that the sets $B_k^n(\mathbb{R})$ and $M_k^n(\mathbb{R})$ are semialgebraic, and as long as we stay on a connected subset of $M_k^n(\mathbb{R})$, we get the same number of real roots. And these root functions are continuous.

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

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

Generalized cylinders

Let $A \subset \mathbb{R}^k$, and let $f, g : A \to \mathbb{R}$ 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

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

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

Structure Theorem: *Let S 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 $f_i^A : A \to \mathbb{R} \cup \{\pm\infty\}$ 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 $\left(f_{i-1}^A, f_i^A\right)_A$ are semialgebraic.*

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

I want to draw a picture so you can see what this looks like (*see figure 2.2*). The essence of the theorem is that S can be constructed as a finite union of semialgebraic cylinders and graphs.

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, because the graphs are connected since the functions are semialgebraic, and the cylinders are connected since the base sets are connected. 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 $q_1, \ldots, q_r$. The nice thing about polynomials is that we only have to do this finitely many times to obtain closure under differentiation.

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 $l \leq \mathrm{degree}\, Q_{\mathcal{S},\bar{x}}(t)$ and $k \leq l$, let

$$M_{\mathcal{S},k}^l = \left\{\bar{x} \in \mathbb{R}^{n-1} \,|\, \mathrm{degree}\, Q_{\mathcal{S},\bar{x}}(t) = l \text{ and it has exactly } k \text{ distinct real roots}\right\}.$$

It should come as no surprise that you can in fact show that this subspace of $\mathbb{R}^{n-1}$ is semialgebraic.

Next we partition $\mathbb{R}^{n-1}$ by taking all intersections of all $M^l_{\mathcal{S},k}$. 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.

That's the rough idea of the proof of the structure theorem. Now we want to use it to deduce the Tarski-Seidenberg theorem.

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

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

is semialgebraic.

Proof: Let $\pi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ be the projection map onto the last m coordinates. Then $f(x) = \pi(\text{graph}(f))$. It thus suffices to show that the image under π of a semialgebraic set $S \subseteq \mathbb{R}^{m+n}$ is semialgebraic. We will show this by induction on n. When $n = 1$, our desired result follows directly from the structure theorem. Suppose the result is true for $n-1$; we will show that it holds for n. Let $\pi_1 : \mathbb{R}^{n-1} \times \mathbb{R}^{1+m} \to \mathbb{R}^{1+m}$ be the projection onto the last $m+1$ coordinates, and let $\pi_2 : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}^m$ be the projection onto the last m coordinates. Then $\pi(S) = \pi_2(\pi_1(S))$. But $\pi_1(S)$ is semialgebraic by the induction hypothesis, so $\pi(S)$ is semialgebraic by the structure theorem.■

The Tarski-Seidenberg theorem gives quantifier elimination because if we take the formula $\exists y\, \varphi(x_1, \ldots, x_n, y)$, where φ is a semialgebraic quantifier-free formula, then this is just the projection to $\mathbb{R}^n$ of a semialgebraic subset of $\mathbb{R}^{n+1}$, and is thus semialgebraic.

Algorithmic cruelty

It turns out that everything I've explained can be done algorithmically. That is, we can take any definable set and algorithmically give it a semialgebraic description. The proof I've given you today relates to work done by George Collins on cylindrical algebraic decomposition. 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. Computational 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. So, quantifier elimination is something that is do-able in principle, but not by any computer that you and I are ever likely to see. Well, I'll retract that last statement because it's probably false.

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)

Lecture 3: O-minimal structures

An $\mathcal{L}$-structure $\mathfrak{R}$ is *o-minimal* if every definable subset of $\mathbb{R}$ is the union of finitely many points and open intervals (a,b), where $a<b$ and $a,b\in\mathbb{R}\cup\{\pm\infty\}$. Thus far we have seen two examples of o-minimal structures: $\mathbb{R}_{\mathrm{lin}}$, the semilinear context, and $\mathbb{R}_{\mathrm{alg}}$, the semialgebraic context. We showed by quantifier elimination that these structures are o-minimal. O-minimal is short for ordered minimal. We use this name because, for an o-minimal structure, the definable subsets of $\mathbb{R}$ are exactly those that must be there because of the presence of $<$. The hypothesis of o-minimality combined with the power of definability have remarkable consequences.

Minimal structures

Consider the field of complex numbers $(\mathbb{C},\cdot,+)$. There is a theorem due to Chevalley which says that $(\mathbb{C},\cdot,+)$ has the quantifier elimination property. All definable subsets of $\mathbb{C}$ can be defined by finite boolean combinations of polynomials equalities $p(x)=0$ or inequalities $p(x)\neq 0$, where p has coefficients in $\mathbb{C}$. As long as p is not the zero polynomial, $p(x)$ will consist of finitely many points, and $p(x)\neq 0$ will consist of cofinitely many points. So all definable subsets of $\mathbb{C}$ are either finite or cofinite. Notice that these are the sets you *have* to be able to define using equality. So $(\mathbb{C},\cdot,+)$ is minimal.

Minimal structures were first know by model theorists. One of my contributions was to look at the properties of minimal structures when we have an ordering. The theme of what I'm going to be talking about today and in the remaining lectures is o-minimal structures.

O-minimal structures

The first result, which I'm going to spend quite a bit of time on, is one of the most important results on o-minimality.

Monotonicity Theorem: *Let $\mathfrak{R}$ be an $\mathcal{L}$-structure that is o-minimal. Suppose that $f:\mathbb{R}\to\mathbb{R}$ is $\mathfrak{R}$-definable. Then there are $-\infty=a_0<a_1<\cdots<a_{k-1}<a_k=\infty$ in $\mathbb{R}\cup\{\pm\infty\}$ such that for each $j<k$ either $f\restriction(a_j,a_{j+1})$ is constant or is a strictly monotone bijection of (possibly unbounded) open intervals in $\mathbb{R}$.*

In particular, all definable $f:\mathbb{R}\to\mathbb{R}$ are piecewise continuous. Now we can see the power of o-minimality. We are not confined to looking at semialgebraic sets.

I'm going to give you some idea of what's in the proof of the monotonicity theorem. We need to show that for every open interval $I\subseteq\mathbb{R}$, there is an open interval $I^*\subseteq I$ on which f is constant or strictly montone. This will give us a local version of the theorem, which we need to globalize. For now, assume the local version and consider the task of globalization.

Let $\theta(x)$ say "x is the left endpoint of an interval on which f is constant or strictly monotone and this interval cannot be extended properly on the left while preserving this property." You should check that this is a meaningful expression in the context of our structure. Clearly $\theta(x)$ defines a finite subset of $\mathbb{R}$, say $b_1, \ldots, b_p$. We know that $f \restriction (b_i, b_{i+1})$ is strictly monotone or constant, because otherwise the local version of the theorem implies that $(b_i, b_{i+1}) \cap \theta(x) \neq \emptyset$. To finish working from local to global, we note that if $f \restriction (b_i, b_{i+1})$ is strictly monotone or constant then $f((b_i, b_{i+1}))$ consists of finitely many points and intervals, and so we can partition each (b_i, b_{i+1}) into a finite collection of points and intervals in such a way that the image under f of each interval in the partition is itself an open interval.

Now I'll try to give you some idea of what's in the proof of the local version of the theorem. Define the following formulas:

$\varphi_0(x)$ says "$\exists$ open interval J with x as an endpoint on which f is constantly equal to $f(x)$."

$\varphi_1(x)$ says "$\exists$ open interval J containing x such that $f(y) < f(x)$ for $y \in J$, $y < x$ and $f(y) > f(x)$ for $y \in J$, $y > x$."

$\varphi_2(x)$ says "$\exists$ open interval J containing x such that $f(y) < f(x)$ for $y \in J$, $y > x$ and $f(y) > f(x)$ for $y \in J$, $y < x$."

$\varphi_3(x)$ says "$\exists$ open interval J containing x such that $f(y) < f(x)$ for $y \in J$, $y \neq x$."

$\varphi_4(x)$ says "$\exists$ open interval J containing x such that $f(y) > f(x)$ for $y \in J$, $y \neq x$."

These five formulas exhaust all possibilities, by o-minimality. Also by o-minimality, possibilities 3 and 4 cannot hold, so $\exists I_1 \subseteq I$ contained in the set defined by one of $\varphi_0, \ldots, \varphi_2$. If I_1 is contained in the set defined by φ_0, then f is constant on I_1. If I_1 is contained in the set defined by φ_1, then f is strictly monotone increasing on I_1. And if I_1 is contained in the set defined by φ_2, then f is strictly monotone decreasing on I_1.

Cells

What I want to do next is talk about a particular kind of definable set: cells. Cells in $\mathbb{R}$ are either points (0-cells) or intervals (1-cells). Cells in $\mathbb{R}^2$ can be constructed as follows: begin with cells in $\mathbb{R}$, and take the graph of a continuous function defined on those cells, as well as the cylinders defined by pairs of continuous functions. Let me draw a picture (*see figure 3.1.*) This is the same kind of picture I drew for the structure theorem, but now we are working with a structure more general than semialgebraic.

More formally, let $\mathfrak{R}$ be an $\mathcal{L}$-structure. The collection of $\mathfrak{R}$-cells is a subcollection $\mathcal{C} = \cup_{n=1}^{\infty} \mathcal{C}_n$ of the $\mathfrak{R}$-definable subsets of $\mathbb{R}^n$ for $n = 1, 2, 3, \ldots$ defined recursively as follows.

Cells in $\mathbb{R}$

The collection of cells $\mathcal{C}_1$ in $\mathbb{R}$ consists of all single point sets $\{r\} \subset \mathbb{R}$ and all open intervals $(a, b) \subseteq \mathbb{R}$, where $a < b$ and $a, b \in \mathbb{R} \cup \{\pm\infty\}$.

Cells in $\mathbb{R}^{n+1}$

Assume the collection of cells $\mathcal{C}_n$ in $\mathbb{R}_n$ have been defined. The collection $\mathcal{C}_{n+1}$ of cells in $\mathbb{R}_{n+1}$ consist of two different kinds: graphs and cylinders.

Graphs Let $C \in \mathcal{C}_n$ and let $f : C \subseteq \mathbb{R}^n \to \mathbb{R}$ be $\mathfrak{R}$-definable and continuous. Then Graph$(f) \subseteq \mathbb{R}^{n+1}$ is a cell.

Generalized Cylinders Let $C \in \mathcal{C}_n$. Let $f, g : C \subseteq \mathbb{R}^n \to \mathbb{R}$ be $\mathfrak{R}$-definable and continuous such that $f(\bar{x}) < g(\bar{x})$ for all $\bar{x} \in C$. Then the cyclinder set $(f, g)_C \subset \mathbb{R}^{n+1}$ is a cell.

Cells are $\mathfrak{R}$-definable and connected. There is a concept of dimension for cells: for each cell $C \subseteq \mathbb{R}^n$ there is a largest $k \leq n$ and $i_1, \ldots, i_k \in \{1, 2, \ldots, n\}$ such that if $\pi : \mathbb{R}^n \to \mathbb{R}^k$ is the projection mapping given by $\pi(x_1, \ldots, x_n) = (x_{i_1}, \ldots, x_{i_k})$, then $\pi(C) \subseteq \mathbb{R}^k$ is an open cell in $\mathbb{R}^k$. This value of k we call the *dimension* of C. Basically speaking, the dimension of a cell is the number of times we construct cylinders in the "bottom-up" construction of a cell.

Cell decomposition

An $\mathfrak{R}$-*decomposition* $\mathcal{D}$ of $\mathbb{R}^n$ is a partition of $\mathbb{R}^n$ into finitely many $\mathfrak{R}$-cells satisfying:

1. If $n = 1$, then $\mathcal{D}$ consists of finitely many open intervals and points.
2. If $n > 1$ and $\pi_n : \mathbb{R}^n \to \mathbb{R}^{n-1}$ denotes projection onto the first $n - 1$ coordinates, then $\{\pi_n(C) : C \in \mathcal{D}\}$ is a decomposition of $\mathbb{R}^{n-1}$.

This is just a generalized version of the cylindrical algebraic decomposition.

Cell Decomposition Theorem: *Let $\mathfrak{R}$ be o-minimal and let $S \subset \mathbb{R}^n$ be definable. Then there is a decomposition $\mathcal{D}$ of $\mathbb{R}^n$ that partitions S into finitely many cells. In particular, if $f : A \subset \mathbb{R}^n \to \mathbb{R}$ is definable, then there is a partition of A into cells such that the restriction of f to each cell is continuous.*

Some obvious consequences

- Using the definition of a cell as defined above, we obtain a good geometric definition of the dimension of a definable set. Namely, the dimension of a set is the maximium dimension of the cells in a decomposition that partitions it.
- Since cells are connected, it follows that every definable set has finitely many connected components. This is one aspect of the "tameness" alluded to in the name "tame topology."
- The topological closure of a definable set consists of finitely many connected components; same for the the interior and the frontier (or boundary). Even if you start with a connected set, this won't necessarily be

true unless the set is definable. Consider a comb with infinitely many teeth, where the points of those teeth are not in the set; the frontier of this connected set consists of infinitely many connected components.

Definable families

Let $S \subset \mathbb{R}^{n+p}$ be a definable set in the o-minimal structure $\mathfrak{R}$. For each $\bar{b} \in \mathbb{R}^n$ define $S_{\bar{b}} := \left\{\bar{y} \in \mathbb{R}^p \mid (\bar{b}, \bar{y}) \in S\right\}$. (Note that some $S_{\bar{b}}$ may be empty.) The family $\left\{S_{\bar{b}} \mid \bar{b} \in \mathbb{R}^n\right\}$ of subsets of $\mathbb{R}^p$ is called a *definable family.*

The next result is quite surprising; we did not expect to find it.

Uniform Bounds Theorem: *Let $\mathfrak{R}$ be o-minimal and let $S \subset \mathbb{R}^{n+1}$ be a definable set such that $S_{\bar{b}}$ is finite for all $\bar{b} \in \mathbb{R}^n$. Then there is a fixed $K \in \mathbb{N}$ satisfying $|S_{\bar{b}}| \leq K$ for all $\bar{b} \in M^n$.*

Note that a definable subset of $\mathbb{R}$ in an o-minimal structure $\mathfrak{R}$ is infinite if and only if it contains an interval. So the theorem actually is stronger. It can be thought of as a generalization of our earlier result on the roots of polynomial equations.

Recall our discussion of minimality and the field of complex numbers. We can also obtain a version of the uniform bounds theorem for subsets of $\mathbb{C}^{n+1}$.

Sketch proof of the cell decomposition theorem

Next I want to give you a quick sketch of the proof of the cell decomposition theorem. The proof is by induction on n. Define the following induction hypotheses:

I_n :	There exists a cell decomposition that respects finitely many definable subsets of $\mathbb{R}^n$.
II_n :	Let $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$ be definable. Then there is a decomposition $\mathcal{D}$ of $\mathbb{R}^n$ that partitions A such that $f \restriction C$ is continuous for all $C \in \mathcal{D}$, $C \subseteq A$.
III_n :	The uniform finiteness property holds for all definable families $\left\{S_{\bar{b}} \mid \bar{b} \in \mathbb{R}^n\right\}$, where $S \subseteq \mathbb{R}^{n+1}$ is definable.

I_1 is true by the definition of o-minimality, II_1 is just the monotonicity theorem, and III_1 requires an intricate direct argument that we will not go into. By assuming I_m, II_m, III_m for all $m < n$, we can use the monotonicity theorem to show that I_n holds, and then that II_n holds, and finally that III_n holds.

Quantifier elimination

Theorem (van den Dries): *Let I be an index set and for each $i \in I$ let $f_i : \mathbb{R}^{n_i} \to \mathbb{R}$ be (total) analytic functions. Then the structure $(\mathbb{R}_{\mathrm{alg}}, \{f_i : i \in I\})$ admits quantifier elimination if and only if each f_i is semialgebraic.*

So, e.g., $\mathbb{R}_{\exp}$, the real exponential field does not have quantifier elimination. Quantifier elimination is not easy to come by.

Partial Elimination

Suppose that a structure $\mathfrak{R}$ has the property that every definable set is definable by an *existential formula*, that is, a formula having the form

$$\exists x_1 \exists x_2 \cdots \exists x_k \, \varphi$$

where φ is a quantifier-free $\mathcal{L}$-formula. How can this help? Suppose that the $\mathfrak{R}$-definable sets that are definable using quantifier-free formulas can be analyzed, and that all such have finitely many connected components. The continuous image of a connected set is connected (elementary topology). Existential quantification corresponds to projection, and projection is a continuous map. Thus all $\mathfrak{R}$-definable subsets of $\mathbb{R}$ have finitely many connected components, that is, all such are the union of finitely many points and open intervals. We conclude that $\mathfrak{R}$ is o-minimal, and so all the geometric and topological properties available as consequences of o-minimality apply. We will talk more about this tomorrow. Tomorrow I want to give you a deeper understanding of what tame topology means, and take you on a tour of structures that we now know to be o-minimal.

Some References

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

J. Knight, A. Pillay, and C. Steinhorn, Definable sets in ordered structures II, *Transactions of the A. M. S.*, **295** (1986) 593-605.

A. Pillay and C. Steinhorn, Definable sets in ordered structures I, *Transactions of the A. M. S.*, **295** (1986) 565-592.

Lecture 4: Examples and some further properties

I'm going to start off by giving you a tour of some o-minimal structures, and I'll finish today by giving you some of the finer tame topological results.

Examples of o-minimal structures

So far we have two examples of o-minimal structures: $\mathbb{R}_{\text{lin}}$, the semilinear context, and $\mathbb{R}_{\text{alg}}$, the semialgebraic context. As I shall discuss later, o-minimality implies a wealth of good analytic and topological properties. This provided ample motivation to seek out o-minimal structures that expand $\mathbb{R}_{\text{alg}}$ to include transcendental data.

We now survey some of the remarkable results that have been obtained beginning in the mid-1980s.

Expansions of $\mathbb{R}_{\text{alg}}$

1. van den Dries 1986

Consider the class of *restricted analytic functions*, **an**, where $g : \mathbb{R}^n \to \mathbb{R} \in \mathbf{an}$ if there is some analytic $f : U^{\text{open}} \subset \mathbb{R}^n \to \mathbb{R}$ such that $[0,1]^n \subset U$, $g \restriction [0,1]^n = f \restriction [0,1]^n$, and $g(\bar{x}) = 0$ otherwise. (The point of doing something like this is that you can avoid considering behaviour at $\pm\infty$ by appropriate restriction of the functions' domain.) Let $\mathbb{R}_{\text{an}}$ be the expansion of $\mathbb{R}_{\text{alg}}$ by adding as basic functions all $g \in \mathbf{an}$. Then the structure $\mathbb{R}_{\text{an}}$ admits elimination down to existential formulas and is o-minimal.

Let me give you a sense of what kind of functions you can actually get here. You can not only get bounded functions, but also functions that live on all of $\mathbb{R}$. For instance, we can obtain $\arctan x$ from the restriction of $\sin x$ and $\cos x$ to $(-\pi, \pi)$. Everything I said about definable functions - monotonicity and the cell decomposition - works for these functions. This result depends on the work of Lojasiewicz and Gabrielov (in the 1960's). Another useful fact about $\mathbb{R}_{\text{an}}$ is that it is polynomially bounded.

Polynomial growth: Let $f : (a, \infty) \to \mathbb{R}$ be definable in $\mathcal{R}_{\text{an}}$. Then there is some $N \in \mathbb{N}$ such that $|f(x)| < x^N$ for sufficiently large x.

2. Denef-van den Dries 1988

Adjoin to $\mathbb{R}_{\text{an}}$ the function $^{-1}$ given by $x \longmapsto 1/x$ for $x \neq 0$ and $0^{-1} = 0$.

Theorem (Denef-van den Dries '88): $(\mathbb{R}_{\text{an}}, ^{-1})$ *admits elimination of quantifiers.*

Whereas in the semialgebraic context we know all the basic functions (they are just polynomials), in this case we have a much larger collection of basic functions, so that our descriptive language is much richer. The languages in (1) and (2) are large, but nonetheless natural. Quantifier elimination always can be achieved by enlarging the language, but no advantage is gained: in general, the quantifier-free sets thus obtained can be horribly badly behaved.

3. Wilkie 1991

The next theorem is really quite spectacular, and was a breakthrough for the subject.

Theorem (Wilkie '91): $\mathbb{R}_{\exp}$ *admits elimination down to existential formulas.*

O-minimality then follows by a result of Khovanskii 1980 (which Wilkie also uses in his proof). Recall that yesterday I showed that quantifier elimination is not possible in $\mathbb{R}_{\exp}$. This result tells us "the best that we can do." This was the first o-minimal structure with functions growing faster than polynomials.

This theorem addresses a question posed originally by Tarski. He asked if his results on $\mathbb{R}_{\mathrm{alg}}$ could be extended to $\mathbb{R}_{\exp}$. Wilkie's result from the syntactic and topological points of view is the best possible.

Macintyre and Wilkie link decidability of the theory of the real exponential field to the following conjecture.

Schanuel's Conjecture: *Let* $r_1, \ldots, r_n \in \mathbb{R}$ *be linearly independent over* $\mathbb{Q}$. *Then the transcendence degree over* $\mathbb{Q}$ *of* $\mathbb{Q}(r_1, \ldots, r_n, e^{r_1}, \ldots, e^{r_n})$ *is at least* n.

This has implications about the transcendence of various things, which I might talk more about tomorrow. Schanuel's conjecture is now regarded by mathematicians as being intractable, so I don't know if we will ever see it verified. But most mathematicians seem to believe that it is true.

Theorem (Macintyre-Wilkie 1993): *Schanuel's conjecture implies that the theory of the real exponential field is decidable.*

What I want to do now is talk about things which are *not* definable. For instance, we know that the exponential function is not definable in $\mathbb{R}_{\mathrm{alg}}$. Using elimination down to existential formulas, some commutative algebra, and a functional theoretic version of Schanuel due to Ax 1971, Bianconi proved the following.

Theorem (Bianconi '95): *No arc of the sine function is definable in* $\mathbb{R}_{\exp}$ *and conversely no restriction of the exponential function is definable in* $(\mathbb{R}_{\exp}, \sin \restriction [-\pi, \pi])$.

4. van den Dries-Miller 1992

A natural question is to ask what happens if we combine the restricted analytic and the exponential functions in our basic functions. Van den Dries and Miller adapt Wilkie's techniques to prove the following.

Theorem (van den Dries-Miller '92): $\mathbb{R}_{\mathrm{an},\exp}$ *admits elimination down to existential formulas and (by Khovanskii) is o-minimal.*

5. van den Dries-Macintyre-Marker 1992

Inspired by the work of Ressayre 1992, these authors analyze $\mathbb{R}_{\text{an,exp}}$ further.

Theorem (D-M-M 1992): $\mathbb{R}_{\text{an,exp,log}}$ *admits elimination of quantifiers.*

Their analysis further shows that every definable function in one variable is bounded by an iterated exponential. Macintyre-Marker 1996 show that the logarithm is necessary for the quantifier elimination. In a second paper (1995) D-M-M develop tools that enable them to obtain several further results.

What I want to do now is talk about results of definability and undefinability. Let $f(x) = (\log x)(\log\log x)$ and let $g(x)$ be a compositional inverse to f defined on some interval (a, ∞). Hardy conjectured in 1912 that g is not asymptotic to a composition of exp, log, and semialgebraic functions.

Theorem (D-M-M 1995): *Hardy's conjecture is true.*

Let me talk about some other undefinability results which people will perhaps find more down to earth. Building on some remarkable ideas and results of Mourges-Ressayre 1993, D-M-M derive some "undefinability" results also.

Theorem (D-M-M 1995): None of the following functions is definable in $\mathbb{R}_{\text{an,exp}}$.

i. the restriction of the gamma function $\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} \mathrm{d}t$ to $(0, \infty)$

ii. the error function $\int_0^x e^{-t^2} \mathrm{d}t$

iii. the logarithmic integral $\int_x^\infty t^{-1} e^{-t} \mathrm{d}t$

iv. the restriction of the Riemann zeta function $\zeta(s) = \sum_{n=1}^\infty n^{-s}$ to $(1, \infty)$.

These are functions that we deal with all the time, but are not definable in the context of $\mathbb{R}_{\text{an,exp}}$. So our work is not yet complete.

6. Miller 1994

For $r \in \mathbb{R}$ let x^r denote the real power function

$$x^r = \begin{cases} x^r & \text{if } x > 0 \\ 0 & \text{if } x \leq 0. \end{cases}$$

Let $\mathbb{R}_{\text{an}}^{\mathbb{R}}$ denote the expansion of $\mathbb{R}_{\text{an}}$ by all power functions x^r for $r \in \mathbb{R}$.

Theorem (Miller 1994): $\mathbb{R}_{\text{an}}^{\mathbb{R}}$ *has elimination of quantifiers.*

Again, this is a structure with functions whose growth at infinity is bounded by polynomials. An expansion $\mathfrak{R}$ of $\mathbb{R}_{\text{alg}}$ is *polynomially bounded* if for every definable $f : \mathbb{R} \to \mathbb{R}$ there is some $N \in \mathbb{N}$ so that $|f(x)| < x^N$ for sufficiently large x. I've given you examples of functions that are polynomially bounded (**an**) and those that are not (exp).

Growth Dichotomy Theorem (Miller 1992): *Let* $\mathfrak{R}$ *be an o-minimal expansion of* $\mathbb{R}_{\text{alg}}$*. Then either the exponential function* e^x *is definable in* $\mathfrak{R}$ *or*

$\mathfrak{R}$ is polynomially bounded. In the second case, for every definable $f : \mathbb{R} \to \mathbb{R}$ in $\mathfrak{R}$ not ultimately identically zero, there are $c \in \mathbb{R} \setminus \{0\}$ and $r \in \mathbb{R}$ such that $f(x) = cx^r + o(x^r)$ as $x \to \infty$.

This amazing theorem shows that there is no "middle ground": if functions are not bounded by polynomial growth, then exponential growth must be possible.

7. van den Dries-Gabrielov 1993

Let $\partial\Phi$ be a collection of restricted analytic functions that is closed under differentiation. Since derivatives are definable in $\mathbb{R}_{\text{an}}$ (definability of derivatives was mentioned in lecture 1), all of the functions in $\partial\Phi$ are definable in this structure.

Theorem: *The structure $(\mathbb{R}_{\text{alg}}, f)_{f \in \partial\Phi}$ has elimination down to existential formulas.*

This result does not give us new functions, but tells us how we can think of what we have in a nicer way. The next result gives us new functions.

8. van den Dries-Speissegger 1996

Using (delicate) generalized power series methods new expansions of $\mathbb{R}_{\text{alg}}$ are constructed. There are two polynomially bounded versions that have elimination down to existential formulas: generalized convergent power series (using real, rather than integer, powers) and multisummable series. Moreover, the exponential function can be added while preserving o-minimality. If, in addition the logarithmic function is adjoined as a basic function, these expansions admit quantifier elimination. In one of these expansions, the gamma function on $(0, \infty)$ is definable, and in the second, the Riemann zeta function on $(1, \infty)$ is definable.

9. Wilkie 1996

Now we come to some really beautiful results of Wilkie. I have to introduce another class of functions, which again is quite natural. A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be *Pfaffian* if there are functions $f_1, \dots, f_k : \mathbb{R}^n \to \mathbb{R}$ and polynomials $p_{ij} : \mathbb{R}^{n+i} \to \mathbb{R}$ such that

$$\frac{\partial f_i}{\partial x_j}(\bar{x}) = p_{ij}(\bar{x}, f_1(\bar{x}), \dots, f_i(\bar{x}))$$

for all $i = 1, \dots, k$, $j = 1, \dots, n$, and $\bar{x} \in \mathbb{R}^n$. So this is a chaining procedure, going one step at a time to generate more complicated functions. Wilkie proved (by quite different methods than used previously) that the expansion of $\mathbb{R}_{\text{alg}}$ by all Pfaffian functions is o-minimal. The introduction of Pfaffian functions allows us to integrate and retain o-minimality. Our class of functions is now even richer.

10. Speissegger 2000

Here Wilkie's methods are extended to obtain the "Pfaffian closure" of an o-minimal expansion of $\mathbb{R}_{\text{alg}}$. In particular, such a structure is closed under integration (antidifferentiation) of functions in one variable.

I have one more comment before I move on to the second part of this talk. What about quantifier elimination in these expansions? Unfortunately, this is completely unknown. We have seen that many geometric results are obtainable for these o-minimal strcutures, but results on quantifier elimination are not available.

Finer analytic and topological consequences of o-minimality

For this section, assume throughout that we work in some o-minimal expansion $\mathfrak{R}$ of $\mathbb{R}_{\text{alg}}$. We showed earlier in our discussion of the cell decomposition that our functions are continuous on each cell. A natural question to ask is: can we do better than this? The next theorem shows that the answer is yes.

$\mathcal{C}^k$ **Cell Decomposition Theorem:** *For each definable set $X \subset \mathbb{R}^m$ and $k = 1, 2, \ldots$, there is a decomposition of $\mathbb{R}^m$ that respects X and for which the data in the decomposition are $\mathcal{C}^k$.*

In the next theorem, "definably homeomorphic" means that the homeomorphism between structures is itself a definable function.

Triangulation Theorem: *Every definable set $X \subset \mathbb{R}^m$ is definably homeomorphic to a semilinear set. More precisely, X is definably homeomorphic to a union of simplices of a finite simplicial complex in $\mathbb{R}^m$.*

Note that these results are "nice" topological results, in the spirit of the term "tame topology" coined by Grothendieck. The next theorem is another nice result. It says that if we look at the fibers of some set, then the fibers corresponding to a given connected component of that set are homeomorphic.

Theorem (Number of Homeomorphism Types): *Let $S \subset \mathbb{R}^{m+n}$ be definable, so that $\{S_{\bar{a}} \,|\, \bar{a} \in \mathbb{R}^m\}$ is a definable family of subsets of $\mathbb{R}^n$. Then there is a definable partition $\{B_1, \ldots, B_p\}$ of $\mathbb{R}^m$ such that for all $\bar{a}_1, \bar{a}_2 \in \mathbb{R}^m$, the sets $S_{\bar{a}_1}$ and $S_{\bar{a}_2}$ are homeomorphic if and only if there is some $j = 1, \ldots, p$ such that $\bar{a}_1, \bar{a}_2 \in B_j$.*

Uniform finiteness combined with Wilkie's theorem yields Khovanskii's theorem. You all know the theorem that says that a polynomial of degree k has no more than k distinct real roots. One of the implications of Khovanskii's theorem is that there is a uniform bound on the number of distinct real roots of $p_{l,k}(x) = ax^l + bx^k$ as l, k vary.

Theorem (Khovanskii): *There exists a bound in terms of m and n for the number of connected components of a system of n polynomial inequalities with no more than m monomials.*

There is a trick to proving this theorem. Replace x^m by $e^{m \log x}$, and let m vary over $\mathbb{R}$. The set of (a, b, m, n, x) such that $ae^{m \log x} + be^{n \log x} = 0$ is in

$\mathbb{R}_{\exp}$. Now fix x and let the other parameters vary. Uniform finiteness gives us a bound on the size of the fibers.

The next theorem gives an o-minimal improvement of the previous result.

Theorem: *There is a bound in terms of m and n for the number of homeomorphism types of the zero sets in $\mathbb{R}^n$ of polynomials $p(x_1, \ldots, x_n)$ over $\mathbb{R}$ with no more than m monomials.*

Theorem (Marker-CS 1994): *Let $\mathfrak{R}$ be an o-minimal expansion of $\mathbb{R}_{a\,\mathrm{lg}}$, and let $g_{\bar{a}} : B \subseteq \mathbb{R}^m \to \mathbb{R}$ for $\bar{a} \in A \subset \mathbb{R}^m$ be an $\mathfrak{R}$-definable family $\mathcal{G}$ of functions. Then every $f : B \subseteq \mathbb{R}^m \to \mathbb{R}$ which is in the closure of $\mathcal{G}$ is definable in $\mathfrak{R}$.*

Here, closure refers to closure in the product topology $\mathbb{R}^B$. This result is very surprising. Consider it in the case of semialgebraic functions: the pointwise limit of semialgebraic functions is semialgebraic.

The Euler characteristic

Now what I want to talk about is how we can begin to do algebraic topology from the viewpoint of o-minimality. Let me draw a picture (*see figure 4.1*). Take this set and cut it up into triangles. I claim that the number of vertices, minus the number of edges, plus the number of faces, is equal to minus one. In fact, you can see that we have 21 vertices, 45 edges, and 23 faces, so this claim is true. This is an example of the *Euler characteristic*. We would like to consider this characteristic in an o-minimal context.

Let $S \subset \mathbb{R}^n$ be definable and $\mathcal{P}$ be a partition of S into cells. Let $n(\mathcal{P}, k)$ be the number of cells of dimension k in $\mathcal{P}$, and define $E_{\mathcal{P}}(S) = \sum (-1)^k n(\mathcal{P}, k)$.

Proposition: *If $\mathcal{P}$ and $\mathcal{P}'$ are partitions of S into cells, then $E_{\mathcal{P}}(S) = E_{\mathcal{P}'}(S)$.*

So we define $E(S) = E_{\mathcal{P}}(S)$ for any partition $\mathcal{P}$. The Euler characteristic has played a very interesting role in o-minimal theory generally. Its properties include the following:

1. Let A and B be disjoint definable subsets of $\mathbb{R}^n$. Then $E(A \cup B) = E(A) + E(B)$.

2. Let $A \subset \mathbb{R}^m$ and $B \subset \mathbb{R}^n$ be definable. Then $E(A \times B) = E(A)\, E(B)$.

3. Let $f : A \subset \mathbb{R}^m \to \mathbb{R}^n$ be definable and injective. Then $E(A) = E(f(A))$.

Some References

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

L. van den Dries, O-minimal Structures and Real Analytic Geometry, in *Current Developments in Mathematics 1998*, International Press, 2000.

Chapters 3 (van den Dries) and 6 (Macpherson), in D. Haskell, A. Pillay, and C. Steinhorn (editors), *Model Theory, Algebra, and Geometry* (Mathematical Sciences Research Institute Publications, vol. 39), Cambridge: Cambridge University Press, 2000.

P. Speissegger, O-minimal Expansions of the Real Field, preprint, 2003. (Available on his website: www.math.wisc.edu/~speisseg/.)

Lecture 5: VC dimension and applications

Before I begin, I want to thank the organizer, Don. I know it was talked about on Wednesday night - the wonderful attention you pay to graduate students - and I can see that this week. And for me too, not a graduate student, it has been a wonderful week. So I just want to thank Don for that.

Vapnik-Chervonenkis dimension

A collection $\mathcal{C}$ of subsets of a set X *shatters* a finite subset F if $\{F \cap C \mid C \in \mathcal{C}\} = \mathcal{P}(F)$, where $\mathcal{P}(F)$ is the set of all subsets of F. The collection $\mathcal{C}$ is a *VC-class* if there is some $n \in \mathbb{N}$ such that no set F containing n elements is shattered by $\mathcal{C}$, and the least such n is the VC-dimension, $\mathcal{V}(\mathcal{C})$, of $\mathcal{C}$.

Let $\mathcal{C} \cap F := \{\mathcal{C} \cap F \mid C \in \mathcal{C}\}$ and for $n = 1, 2, \ldots$ let

$$f_{\mathcal{C}}(n) := \max\{\, |\mathcal{C} \cap F| \mid F \subset X \text{ and } |F| = n\}.$$

Also, let $p_d(n) = \sum_{i<d} \binom{n}{i}$. The next theorem gives us a polynomial bound on the growth of $f_{\mathcal{C}}$ for VC-classes.

Theorem (Sauer): *Suppose that $f_{\mathcal{C}}(d) < 2^d$ for some d. Then $f_{\mathcal{C}}(n) \leq p_d(n)$ for all n.*

To prove Sauer's theorem we need the following proposition.

Proposition: *Let $|F| = n$, and let $\mathcal{D}$ be a collection of subsets of F such that $|\mathcal{D}| > p_d(n)$, where $d \leq n$. Then there is $E \subseteq F$, $|E| = d$, so that $\mathcal{D}$ shatters E.*

Proof: The proof is by induction on n. The result is clear when $d = 0$ or $d = n$, so assume $0 < d < n$. Fix $x \in F$ and let $F' = F \backslash \{x\}$, $\mathcal{D}' = \{D \backslash \{x\} \mid D \in \mathcal{D}\}$. Consider the map $\pi(D) = D \backslash \{x\}$. Note that $\pi^{-1}(D')$ has either one or two elements (i.e. D', $D' \cup \{x\}$), depending on whether or not $x \in D$. Write $\mathcal{D}' = \mathcal{D}_1' \cup \mathcal{D}_2'$, where $\mathcal{D}_1'$ is the class of all sets with one preimage under π, and $\mathcal{D}_2'$ is the class of all sets with two preimages under π. If $|\mathcal{D}'| > p_d(n-1)$, then by the induction hypothesis we have $E' \subseteq F'$, $|E'| = d$, so that $\mathcal{D}'$ shatters E'. It follows that $\mathcal{D}$ shatters E'. If $|\mathcal{D}'| \leq p_d(n-1)$, then we have $|\mathcal{D}| = |\mathcal{D}_1'| + 2|\mathcal{D}_2'| = |\mathcal{D}'| + |\mathcal{D}_2'|$. But $|\mathcal{D}| > p_d(n) = p_d(n-1) + p_{d-1}(n-1)$, and so $|\mathcal{D}_2'| > p_{d-1}(n-1)$. Thus by the induction hypothesis we have $E' \subseteq F'$, $|E'| = d-1$ so that $\mathcal{D}_2'$ shatters E'. It follows that $\mathcal{D}$ shatters $E' \cup \{x\}$.■

Now we can use our proposition to prove Sauer's theorem.

Proof: If $d > n$ then $p_d(n) = 2^n$, and the inequality holds trivially. So let $d \leq n$. Consider an arbitrary set $F \subset X$ with $|F| = n$. If $|\mathcal{C} \cap F| > p_d(n)$, then by our proposition there exists $E \subseteq F$, $|E| = d$ such that $\mathcal{C}$ shatters E. But this contradicts $f_{\mathcal{C}}(d) < 2^d$. Thus for all F we must have $|\mathcal{C} \cap F| \leq p_d(n)$, implying that $f_{\mathcal{C}}(d) \leq p_d(n)$.■

Now I'm going to try to connect back to the logical issues we have been discussing through the week. An $\mathcal{L}$-formula $\varphi(x_1, \ldots, x_k; y_1, \ldots, y_m)$ has the independence property with respect to the $\mathcal{L}$-structure $\mathfrak{R}$ if for every $n = 1, 2, \ldots$ there are $\bar{b}_1, \ldots, \bar{b}_n \in \mathbb{R}^m$ such that for every $X \subseteq \{1, \ldots, n\}$, there is some $\bar{a}_X \in \mathbb{R}^k$ satisfying

$$\varphi\left(\bar{a}_X; \bar{b}_i\right) \text{ is true in } \mathfrak{R} \Leftrightarrow i \in X.$$

If φ does not have the independence property with respect to $\mathfrak{R}$, we let $\mathcal{I}(\varphi)$ be the least n for which the property above fails.

For an $\mathcal{L}$-formula $\varphi(\bar{x}; \bar{y})$ and a structure $\mathfrak{R}$, let $S \subseteq \mathbb{R}^{k+m}$ be the set defined by φ. We let $C_\varphi := \left\{S_{\bar{b}} \mid \bar{b} \in \mathbb{R}^m\right\}$ denote the family of subsets of $\mathbb{R}^k$ determined by S.

Now what I want to do is come to the connection between the concept I have just defined, and VC dimension.

Theorem (Laskowski): *The definable family $\mathcal{C}_\varphi$ is a VC-class if and only if φ does not have the independence property. Moreover, if $\mathcal{V}(\mathcal{C}_\varphi) = d$ and $\mathcal{I}(\varphi) = n$, then $n \leq 2^d$ and $d \leq 2^n$ (and these bounds are sharp).*

Let $\psi(\bar{y}; \bar{x}) := \varphi(\bar{x}; \bar{y})$ be the *dual formula* of φ. That is, ψ and φ are the same formula (and so define the same set) with the roles of $\bar{x}$ and $\bar{y}$ reversed. The theorem follows from the next two lemmas.

Lemma 1: *With the notation as above, $\mathcal{V}(\mathcal{C}_\varphi) = d$ if and only if $\mathcal{I}(\psi) \geq d$.*

Proof: By definition, $\mathcal{V}(\mathcal{C}_\varphi) \geq d$ if and only if there exist $\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_d \in \mathbb{R}^k$ such that for every $X \subseteq \{1, \ldots, d\}$ there is $b_X \in \mathbb{R}^m$ for which $\varphi\left(\bar{a}_j, \bar{b}_X\right)$ true $\Leftrightarrow j \in X$. This exactly says: $\mathcal{I}(\psi) \geq d$.■

Understanding Lemma 1 is just a matter of understanding the definitions of VC dimension and the independence number.

Lemma 2: *Let the notation be as above. Then $\mathcal{I}(\varphi) = n$ implies $\mathcal{I}(\psi) \leq 2^n$.*

Proof: Suppose $\mathcal{I}(\psi) > 2^n$. By definition, there are $\bar{b}_s \in \mathbb{R}^k$ for each $s \subseteq \{1, \ldots, n\}$ so that for every $X \subseteq \{s \mid s \subseteq \{1, \ldots, n\}\}$ there is $\bar{a}_X \in \mathbb{R}^m$ such that $\varphi\left(\bar{a}_X, \bar{b}_s\right)$ true $\Leftrightarrow s \in X$. For $i = 1, 2, \ldots, n$, let $X_i = \{s \subseteq \{1, \ldots, n\} \mid i \in s\}$. We have $\bar{a}_{X_1}, \bar{a}_{X_2}, \ldots, \bar{a}_{X_n} \in \mathbb{R}^m$. Now for each $s \subseteq \{1, \ldots, n\}$, we have $\varphi\left(\bar{b}_s; \bar{a}_{X_i}\right)$ true $\Leftrightarrow s \in X_i \Leftrightarrow i \in s$.■

Now we can use these two lemmas to do a quick proof of Laskowski's theorem.

Proof:

$$\begin{aligned}
\mathcal{C}_\varphi \text{ is a VC-class} \quad &\Leftrightarrow \quad \mathcal{V}(\mathcal{C}_\varphi) = d \text{ for some } d \in \mathbb{N} \\
&\Leftrightarrow \quad \mathcal{I}(\psi) = d \quad \text{(by Lemma 1)} \\
&\Rightarrow \quad \mathcal{I}(\varphi) \leq 2^d \quad \text{(by Lemma 2)} \\
&\Rightarrow \quad \varphi \text{ does not have the independence property.}
\end{aligned}$$

Conversely,

$$
\begin{aligned}
\varphi \text{ does not have the independence property} &\Leftrightarrow \mathcal{I}(\varphi) = n \text{ for some } n \in \mathbb{N} \\
&\Rightarrow \mathcal{I}(\varphi) \leq 2^n \quad \text{(by Lemma 2)} \\
&\Leftrightarrow \mathcal{V}(\mathcal{C}_\varphi) \leq 2^n \quad \text{(by Lemma 1)} \\
&\Rightarrow \mathcal{C}_\varphi \text{ is a VC-class.}
\end{aligned}
$$

■

We say that the $\mathcal{L}$-structure $\mathfrak{R}$ has the *independence property* if there is a formula $\varphi(x; \bar{y})$ *with just the single variable* x that has the independence property with respect to $\mathfrak{R}$.

Applying model theoretic methods, Laskowski gives a clear combinatorial proof of the following theorem due to Shelah.

Theorem (Shelah 1971): *An $\mathcal{L}$-structure $\mathfrak{R}$ has the independence property if and only if there is a formula $\varphi(\bar{x}; \bar{y})$ (in any number of x variables) that has the independence property with respect to $\mathfrak{R}$.*

This is a very hard theorem, and attests to Shelah's incredible combinatorial genius. Interestingly, if you look at the paper with Sauer's theorem, there is a footnote that quotes a referee as saying that these results had been proved earlier by Shellah.

Laskowski's theorem combined with the following result provides the link between o-minimality and VC-classes.

Proposition (Pillay-CS 1986): *O-minimal structures do not have the independence property.*

Theorem (Laskowski 1992): *Let $\mathfrak{R} = (\mathbb{R}, <, \ldots)$ be o-minimal and let $S \subset \mathbb{R}^{k+m}$ be definable. Then the collection $\mathcal{C} = \{S_{\bar{x}} \mid \bar{x} \in \mathbb{R}^m\}$ is a VC-class.*

Thus any definable family of subsets in an o-minimal structure constitutes a VC-class. This theorem gives us a "black box" to generate an enormous variety of VC-classes.

Note that many structures are known not to have the independence property (by work of Shelah), and thus Laskowski's theorem provides significantly more examples of VC-classes. To illustrate, the field of complex numbers, $(\mathbb{C}, +, \cdot)$ does not have the independence property, and thus any definable family of sets in this structure is a VC-class.

Probably approximately correct (PAC) learning

Begin with an *instance space* X that is supposed to represent all instances (or objects) in a learner's world. A *concept* c is a subset of X, which we can identify with a function $c : X \to \{0, 1\}$. A *concept class* $\mathcal{C}$ is a collection of concepts.

A *learning algorithm* for the concept class $\mathcal{C}$ is a function L which takes as input m-tuples $((x_1, c(x_1)), \ldots, (x_m, c(x_m)))$ for $m = 1, 2, \ldots$ and outputs hypothesis concepts $h \in \mathcal{C}$ that are consistent with the input. If X comes equipped with a probability distribution, then we can define the *error* of h to be $\text{err}(h) = P(h\Delta c)$.

The learning algorithm L is said to be PAC if for every $\epsilon, \delta \in (0,1)$ there is $m_L(\epsilon,\delta)$ so that for *any* probability distribution P on X and any concept $c \in \mathcal{C}$, we have for all $m \geq m_{L(\epsilon,\delta)}$ that

$$P\left(\left\{\bar{x} \in X^m \mid \text{err}\left(L\left((x_i, c(x_i))_{i\leq m}\right)\right) \leq \epsilon\right\}\right) \geq 1-\delta.$$

It can be shown that an algorithm that outputs a hypothesis concept h consistent with the sample data is PAC provided that $\mathcal{C}$ is a VC-class. Moreover, for given ϵ and δ, the number of sample points needed is, roughly speaking, proportional to the VC-dimension $\mathcal{V}(\mathcal{C})$.

Neural networks

Macintyre-Sontag 1993 and Karpinski-Macintyre 1994 apply Laskowski's result and the uniform bounds available in o-minimal structures to answer questions about neural networks. The output in a sigmoidal neural network is the result of computing a quantifier-free formula whose atomic formulas have the form $\tau(\bar{x},\bar{w}) > 0$ or $\tau(\bar{x},\bar{w}) = 0$, where τ is built from polynomials and exp, $\bar{x}$ are input values, and $\bar{w}$ represent a tuple of programmable parameters. Varying the parameters gives rise to a definable family in an o-minimal structure and hence Laskowski's theorem applies, which tells us that it is possible to PAC learn the architecture of such a network.

The first results of Macintyre and Sontag applied Laskowski's theorem to prove finite VC-dimension. Using quantitative results of Khovanskii, Karpinski and Macintyre give an upper bound for the VC-dimension that is $O(m^4)$, where m is the number of weights. Koiran and Sontag 1997 have established a quadratic lower bound (in the number of weights) for the VC-dimension.

Some references

M. Anthony, Probabilistic Generalization of Functions and Dimension-based Uniform Convergence Results, *Statistics and Computing*, **8** (1998), 5-14. (available - with much more - on his website: www.maths.lse.ac.uk/Personal/martin/.)

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

M. Karpinski and A. Macintyre, Polynomial Bounds for VC Dimension of Sigmoidal and General Pfaffian Neural Networks, *J. Computing and System Sciences*, **54** (1997), 169-176.

M. Kearns and U. Vazirani, *An Introduction to Computational Learning Theory*, Cambridge, MA and London: The MIT Press, 1994.

P. Koiran and E. Sontag, Neural Networks with Quadratic VC Dimension, *J.C.S.S.*, **54** (1997), 190-198.

M. C. Laskowski, Vapnik-Chervonenkis Classes of Definable Sets, *J. London Math. Soc.*, **245** (1992), 377-384.