INTRODUCTION TO O-MINIMAL STRUCTURES AND AN APPLICATION TO NEURAL NETWORK LEARNING

MARCUS TRESSL

Contents

1. Definition of o-minimality and first examples

We start with some notational conventions. Let $M = (M, \leq)$ be a linearly (or 'totally') ordered set (hence $\leq$ is a binary, reflexive, antisymmetric, transitive and total relation). We will frequently enlarge M by two elements $+\infty$, $-\infty$ with the property $-\infty < M < +\infty$.

- A subset C of M is **convex** if for all $c, c' \in C$ and every $m \in M$ with $c \leq m \leq c'$ we have $m \in C$.
- An **interval** of M is a convex subset of M "with endpoints in $M \cup \{\pm\infty\}$", i.e. an interval is a set of the form (a, b) (open interval), $[a, b]$ (closed interval), where $a, b \notin \{\pm\infty\}$, $[a, +\infty)$, where $a \neq -\infty$ and $(-\infty, b]$, where $b \neq +\infty$.
- M will always be considered as a topological space given by the **interval topology** that has the set of open intervals as a basis of open sets. The product set M^n will also be considered as a topological space carrying the product topology. Hence a subset of M^n is open if it is a (arbitrary) union of open boxes (= n-fold product of open intervals).
- M is called **dense** if there are no "gaps" in M, i.e. for all $x, y \in M$ with $x < y$ there is some $z \in M$ with $x < z < y$. M is called **discrete** if every element $x \in M$ has a successor and a predecessor, unless x is the smallest or the largest element of M. For example, $(\mathbb{N}, \leq)$ is discrete, whereas $(\aleph_1, \leq)$ is not.

Date: August 9, 2015.

1.1. **Definition.** An expansion $M = (M, \leq, ...)$ of a totally ordered set $(M, \leq)$ is called **o-minimal** (an abbreviation for order-minimal), if every definable subset of M is a finite union of intervals. Here and below "definable" always means, "definable with parameters from M". A first order theory T in a language $\{\leq, ...\}$ is o-minimal if all its models are o-minimal.

Recall that an interval has endpoints in $M \cup \{\pm\infty\}$, so if we expand the totally ordered set $\mathbb{Q}$ by the set S of all rational numbers $> \sqrt{2}$ then this structure is not o-minimal, since the definable set S is not a finite union of intervals.

1.2**. Examples of o-minimal orders** Let $M = (M, \leq)$ be a totally ordered set. If M is dense (with or without endpoints), then M has quantifier elimination (after naming the endpoints, i.e. $Th(M)$ has quantifier elimination in the language $\{\leq\}$ enlarged by constant symbols for the endpoints); cf. [Mar02, Thm 3.1.3].

Every discretely ordered set has quantifier elimination after naming the endpoints, the successor function and the predecessor function.

Hence in both cases every definable subset is defined by a boolean combination of formulas of the form $a \leq x$ and $x \leq b$ with $a, b \in M$. It is then clear that M is o-minimal.

Thus for example $(\mathbb{Q}, \leq)$ and $(\mathbb{N}, \leq)$ are o-minimal.

On the other hand, if M is a well ordered set (e.g. an ordinal) and M is uncountable, then M is not o-minimal, witnessed by the set of all elements of M which do not have a predecessor.

In logical terms, o-minimality is equivalent to saying that every definable subset of $M = (M, \leq, ...)$ is defined by a boolean combination of formulas of the form $a \leq x$, $x \leq b$ with $a, b \in M$. In this sense M is indeed "order-minimal". This is in analogy with the notion of a minimal structure, where the definable subsets of that structure are boolean combinations of sets defined by $a = x$.

However there is a fundamental difference between the notions "o-minimal" and "minimal": Let $M = (\mathbb{N}, \leq)$. Then M is o-minimal and minimal (why?). Now every structure N, elementary equivalent to M is again o-minimal (since N is again discrete), whereas N is not minimal if N is not isomorphic to M: in this case there must be some $a \in N$ such that both $(-\infty, a)$ and $(a, +\infty)$ are infinite.

We shall see later that indeed o-minimality of any structure is preserved by elementary equivalence.

1.3. *Remark.* Here some easy consequences of o-minimality used throughout. Let $M = (M, \leq, ...)$ be o-minimal and let $S \subseteq M$ be definable.

(i) S has a supremum and an infimum in $M \cup \{\pm\infty\}$.
(ii) If S is infinite then S contains an infinite open interval.
(iii) The closure $\overline{S}$ and the interior of S are definable. The boundary $\overline{S} \setminus \text{interior}(S)$ of S is finite.
(iv) There is no infinite sequence $a_1 < a_2 < ...$ in M with the property $a_1 \in S$, $a_2 \notin S$, $a_3 \in S$, $a_4 \notin S$,... .
(v) If the ordered set underlying M is the natural order of the reals, then M is o-minimal if and only if every definable subset of M has finitely many connected components.

It should be noted that (i) together with (iv) imply o-minimality. On the other hand, the totally ordered set $\mathbb{Q}$ expanded by the set S of all rational numbers

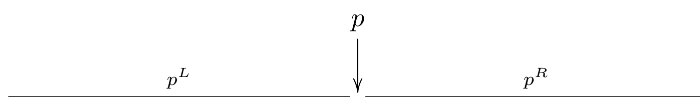
$> \sqrt{2}$, satisfies properties (ii)-(iv), because this structure is **weakly o-minimal**: every definable subset is a finite union of convex sets.

Proof. Exercise. □

If $M = (M, \leq, ...)$ is an expansion of a totally ordered set and $p \in S_1(M)$ is a 1-type of M, then p partitions M into three sets:

$$p^L = \{a \in M \mid a < x \in p\},\ p^= = \{a \in M \mid a = x \in p\},\ p^R = \{a \in M \mid x < a \in p\}.$$

Of course $p^=$ is a singleton if and only if p is realised in M. If p is not realised in M, then p^L and p^R partition M; as $p^L < p^R$ (i.e. $a < b$ for all $a \in p^L, b \in p^R$) and p induces a (Dedekind) cut of M.



Another characterisation of o-minimality of a structure $M = (M, \leq, ...)$ is given by saying that non-realised 1-types of M are determined by the cut they induce on M:

1.4. **Proposition.** *An expansion $M = (M, \leq, ...)$ of a totally ordered set is o-minimal if and only if every 1-type p of M is uniquely determined by the set of formulas $a < x$ and $x < b$ $(a, b \in M)$ contained in p.*

Proof. Exercise. □

We now characterise o-minimal ordered groups.

1.5. **Proposition.** ([PS86, Thm. 2.1])
Let $M = (M, \leq, \cdot, e)$ be a totally ordered group (recall: this means $x \leq y \Rightarrow uxv \leq uyv$). Then M is o-minimal if and only if M is abelian and divisible (i.e. for all $a \in M$ and each $n \in \mathbb{N}$ there is some $x \in M$ with $nx = a$).

Proof. Suppose M is o-minimal. We first show that M does not have non-trivial definable subgroups G: If $[g, +\infty) \subseteq G$ for some $g \in M$, then $G = M$, since for each $m \in M$ with $e \leq m$ we have $mg \in [g, +\infty)$ and so $m = (mg)g^{-1} \in G$.

Otherwise, by o-minimality, G has a supremum s in M. Suppose $s > e$, in particular there is some $g \in G$, $e < g$. Then $g^{-1} < e$ and so $g^{-1}s < s$. By choice of s, there is some $h \in G$ with $g^{-1}s \leq h$. It follows $s \leq gh \in G$. By choice of s this implies $s = gh$. Thus $s \in G$, which is impossible as $s < s^2 \in G$ and s is the supremum of G.

Hence M does not have non-trivial definable subgroups. With this information the rest is easy: For $a \in M$ the commutator subgroup $C := \{b \in M \mid ab = ba\}$ of a is obviously definable. Hence $C = 0$ or $C = M$. As $a \in C$ we get $C = M$ in either case. This shows that M is abelian.

Similarly, for $n \in \mathbb{N}$ the subgroup (note that M is abelian!) $\{a^n \mid a \in M\}$ is definable in M and it is obvious that this group must be M.

It remains to show that M is o-minimal if M is abelian and divisible. Since the theory DOAG of Divisible Ordered Abelian Groups has quantifier elimination in the language $\{+, -, \leq, 0\}$ cf. [Mar02, Cor. 3.1.17], every definable subset of M is a boolean combination of sets defined by formulas of the form $a \pm x \geq 0$. It is clear that each such set is a finite union of intervals. □

In 2.4 we shall see that the o-minimal ordered fields are precisely the real closed fields.

A complete description of pure o-minimal ordered sets can be found in [PS86, (3.12)]. We will from now on focus on o-minimal expansions of densely ordered sets. Observe that by 1.5, this is automatically satisfied if we work with an expansion of a group.

Exercises.

1. Show that a totally ordered set $X = (X, \leq)$ is compact in the interval topology if and only if X is Dedekind complete, i.e. every subset of X has a supremum in X.
2. Let $\sqsubseteq$ be the lexicographic order of $\mathbb{R} \times \mathbb{R}$. Show that there is a subset X of $\mathbb{R} \times \mathbb{R}$, definable with parameters in $(\mathbb{R}, \leq)$ such that X is not a finite union of $\sqsubseteq$-convex subsets. Is the structure $(\mathbb{R} \times \mathbb{R}, \sqsubseteq)$ o-minimal?
   Solution. Take $X := \{(a, b) \mid a \geq 0, b \geq 0\}$. Yes, the structure $(\mathbb{R} \times \mathbb{R}, \sqsubseteq)$ is o-minimal. □
3. Determine the definable 1-types of an o-minimal expansion $M = (M, \leq, ...)$ of a dense order. Recall that an n-type of a structure M is called definable if for each formula $\varphi(x_1, ..., x_n, y_1, ..., y_k)$ there is a formula $\psi(\bar{y})$ (possibly with parameters) such that
$$M \models \psi(\bar{a}) \iff \varphi(\bar{x}, \bar{a}) \in p \ (\bar{a} \in M^k).$$
4. Show that the only connected divisible ordered abelian group is (up to isomorphism) the group $(\mathbb{R}, +, \leq)$. Are there connected densely ordered sets apart from $(\mathbb{R}, \leq)$?
   Solution. If G is a connected divisible ordered abelian group, then G has to be archimedean (i.e. for all $g, h > 0$ there is some $n \in \mathbb{N}$ with $ng > h$) since for each $g > 0$, the set of all h with $h > \mathbb{N}g$ is open and closed. It follows that G can be embedded into $(\mathbb{R}, + \leq)$ and the connectedness of G implies $G = \mathbb{R}$.
   Every Dedekind complete order without jumps is connected. E.g. the set of all non-definable cuts of a densely ordered set is connected. □
5. Show that $(\mathbb{R}, \leq, \exp(x))$ is o-minimal. Is there a total order $\sqsubseteq$ on $\mathbb{R}$ such that $(\mathbb{R}, \sqsubseteq, \sin(x))$ is o-minimal?
   Solution. To show that $(\mathbb{R}, \leq, \exp(x))$ is o-minimal, note first that this structure is a model of the theory T of densely totally ordered structures $(M, \leq)$ without endpoint, expanded by a function $f : M \longrightarrow M$ with the following properties:
   (a) f is an increasing isomorphism $(M, \leq) \longrightarrow ((a, \infty), \leq)$ for some $a \in M$
   (b) $m < f(m)$ for all m.
   Let M be a countable elementary restriction of $(\mathbb{R}, \leq, \exp(x))$. Then M inherits the following property from $(\mathbb{R}, \leq, \exp(x))$:
   (∞) for all $m \in M$, the iteration
$$f^{(0)}(m) = m, f^{(1)}(m) = f(m), f^{(2)}(m) = f(f(m)), ...$$
   is unbounded in M
   Now show that every countable model of T with property (∞) is isomorphic to $(\mathbb{Q}^{>0}, \leq, x \mapsto 1 + x)$: to see this we may replace M by an isomorphic copy expanding $(\mathbb{Q}^{>0}, \leq)$ such that $f : \mathbb{Q}^{>0} \longrightarrow (1, \infty) \subseteq \mathbb{Q}^{>0}$ is an order

preserving isomorphism with property (∞). Then an isomorphism $\varphi : (\mathbb{Q}^{>0}, \leq, x \mapsto 1+x) \longrightarrow (\mathbb{Q}^{>0}, \leq, x \mapsto f)$ is given by

$$\varphi(n+q) = f^{(n)}(q),\ n \in \mathbb{N}_0, q \in (0,1].$$

Observe that φ indeed satisfies $\varphi(1+x) = f(\varphi(x))$ for every $x \in \mathbb{Q}^{>0}$.

Since $(\mathbb{Q}^{>0}, \leq, x \mapsto 1+x)$ has an o-minimal theory by 1.5, also $(\mathbb{R}, \leq, \exp(x))$ is o-minimal.

Finally, there is no total order $\sqsubseteq$ on $\mathbb{R}$ such that $(\mathbb{R}, \sqsubseteq, \sin(x))$ is o-minimal: If we project the graph S of sin onto the y-axis, the image has dimension 1 and each nonempty fibre has dimension 1. So by the dimension formula, S has dimension 2 (w.r.t. the $\sqsubseteq$-interval topology). But then the projection onto the x-axis also must have an infinite fibre. Contradiction. Is there an easier argument? □

2. Basic real analysis and the monotonicity theorem

Let $M = (M, \leq, ...)$ be an o-minimal expansion of a **dense** linear ordering.

2.1. **Proposition.** *If $f : M \longrightarrow M$ is definable and $x \in M \cup \{\pm\infty\}$, then $\lim_{t \nearrow x} f(t)$ and $\lim_{t \searrow x} f(t)$ exists in $M \cup \{\pm\infty\}$.*

Proof. We show that $\lim_{t \to +\infty} f(t)$ exists, all other cases are analog. By o-minimality, for each $m \in M$, the set $S_m = \{t \in M \mid f(t) \geq m\})$ is a finite union of intervals. hence there is some $a \in M$ such that $(a, +\infty)$ is contained in S_m or disjoint from S_m.

In other words, f is eventually $\geq m$, or f is eventually $< m$. let G be the (definable!) set of all $m \in M$ such that f is eventually $\geq m$. If $G = \emptyset$, then for all $m \in M$, f is eventually $< m$ which means $\lim_{t \to +\infty} f(t) = -\infty$. If $G = M$, then for all $m \in M$, f is eventually $\geq m$ which means $\lim_{t \to +\infty} f(t) = +\infty$. Hence $\emptyset \neq G \subsetneq M$ and every $m \in M \setminus G$ is an upper bound of G. By o-minimality, G has a supremum $s \in M$ and it follows easily that $\lim_{t \to +\infty} f(t) = s$. □

2.2**. Intermediate value theorem**
Continuous, definable functions map intervals onto intervals.

Proof. The proof is as in classical real analysis. Take $a, b, y \in M$ with $f(a) < y < f(b)$. By o-minimality the supremum c of all $x \in [a, b]$ with $f(a) \leq y$ exists. We claim that $f(c) = y$. If $y < f(c)$ then $a < c \leq b$ and by continuity there is some $d \in [a, c)$ with $y < f(x)$ for each $x \in (d, c)$. This contradicts the choice of c.

On the other hand if $f(c) < y$, then $c < b$ and by continuity there are elements $> c$ in $[a, b]$ with $f(x) < y$, which again contradicts the choice of c. □

In fact, 2.2 holds true in greater generality. We say that a definable subset S of M^n is **definably connected**, if S can not be written as the disjoint union of two nonempty definable and open subsets of S. It is then straightforward to see that the definably connected subsets of an o-minimal structure M are precisely the intervals. Moreover, images of definably connected subsets S of M^n under continuous definably maps $S \longrightarrow M^k$ are again definably connected (proof is identical to the topological one).

2.3. **Corollary.** *If $f:(a,b)\longrightarrow M$ is definable, continuous and injective, then f is strictly monotone.*

Proof. Otherwise there are $x<y<z$ such that $f(x)<f(z)<f(y)$, say. By 2.2, there is some $u\in(x,y)$ with $f(u)=f(z)$ which contradicts the injectivity of f. □

The intermediate value property allows us to determine the o-minimal ordered fields. Recall from [Mar02, Thm 3.3.9] that an ordered field $M=(M,+,\cdot,\le)$ is **real closed** if and only if all polynomials in one variable over M satisfy the intermediate value property.

2.4. **Theorem.** *An ordered field M is o-minimal if and only if M is real closed.*

Proof. If M is real closed then M has quantifier elimination in the language $\{\le,+,-,\cdot,0,1\}$ of ordered rings by Tarski's theorem (cf. [Mar02, Thm 3.3.15]). Hence a definable subset of M is a boolean combination of sets defined by polynomial inequalities $P(T)\ge 0$ with $P\in R[T]$, T a single variable. It is therefore enough to show that these sets are finite union of intervals: This follows easily from the intermediate value property of univariate polynomials in real closed fields.

Conversely, if M is an o-minimal ordered field, then every continuous and definable map $M\longrightarrow M$ satisfies the intermediate value property (by 2.2). On the other hand, by copying the proof from real analysis, univariate polynomials are continuous with respect to the order topology in any ordered field. Thus, univariate polynomials of o-minimal ordered fields satisfy the intermediate value property and by [Mar02, Thm 3.3.9], M is real closed. □

2.5. *Remark.* In fact every o-minimal ordered ring M is real closed ([PS86, Thm2.3]). By 2.4 the only thing we need to show is that M is a field. This is left as an exercise (hint: show that the positive elements of M are a group, then use 1.5 to show that this group is abelian).

Sets, definable in a pure real closed field are called **semi-algebraic**.

2.6. **Monotonicity Theorem**
Let $M=(M,\le,...)$ be o-minimal and let $f:M\longrightarrow M$ be definable. There are $a_0=-\infty<a_1<...<a_n=+\infty$ such that for each i, f is constant or continuous and strictly monotone on (a_{i-1},a_i).

Proof. Let

$$X:=\{x\in M \mid f \text{ is constant or strictly monotone and continuous } \underbrace{\text{in an open interval containing } x}_{\text{we say ``near } x\text{''}}\}.$$

and let Y be the complement of X in M. Suppose Y is finite. Then with

$$\begin{aligned} X_= &= \{x\in X \mid f \text{ is constant near } x\}\\ X_\nearrow &= \{x\in X \mid f \text{ is strictly increasing and continuous near } x\}\\ X_\searrow &= \{x\in X \mid f \text{ is strictly decreasing and continuous near } x\},\end{aligned}$$

and using o-minimality, there are $-\infty=a_0<a_1<...<a_n=+\infty$ such that each (a_{i-1},a_i) is either contained in $X_=$, or in $X_\nearrow$ or in $X_\searrow$ (the elements in Y are among the a_i). It is clear that f is continuous in each (a_{i-1},a_i). Moreover if (a_{i-1},a_i) is contained in $X_=$, $X_\nearrow$, $X_\searrow$ respectively, then f is constant, strictly increasing, strictly decreasing in (a_{i-1},a_i): for example say (a_{i-1},a_i) is contained

in $X_{\nearrow}$ and $x \in (a_{i-1}, a_i)$. Then the set D of all $y \in (a_{i-1}, a_i)$ with $x < y$ and $f(x) < f(y)$ has a supremum in M and by choice of $X_{\nearrow}$, this supremum must be a_i.

Hence in order to show the theorem, we only need to show that the set Y is finite. Suppose Y is infinite. By o-minimality, Y contains an open interval I. We will shrink I to reach a contradiction.

Claim. We can shrink I such that f is injective and bounded on I.

Proof. Every $y \in f(I)$ has only finitely many preimages in I: otherwise there would be a subinterval of I contained in a single fibre, which contradicts $Y \cap X_{=} = \emptyset$. Thus $f(I)$ is infinite and there is an interval (u, v) contained in it. Moreover we may define $g : (u, v) \longrightarrow I$ by $g(y) = \min\{x \in I \mid f(x) = y\}$. Obviously g is injective with infinite image and so the image of g contains an interval. On this interval, f is injective (with inverse g) and bounded (by u and v). □

Hence we have an injective function f on I which is not strictly monotone in any open subinterval (as $I \subseteq Y$). Let C be the (definable!) set of all points of I, where f is continuous. If C is infinite, then C contains an open interval and by 2.3 f is strictly monotone on this interval. Thus C is finite and by shrinking I further we may assume that f is not continuous at any point of I. By shrinking I even further and applying o-minimality again we may assume that for each $a \in M$, $\lim\limits_{t \nearrow a} f(t) \neq f(a)$ (note that the limit exists by 2.1).

Since f is bounded on I we may define for all $x \in I$:

$$g(x) := \lim_{t \nearrow x} f(t).$$

We will now make use of the assumption that M expands a group.

Let $h(x) := |f(x) - g(x)|$. Routine checking shows that $\lim_{t \nearrow x} h(t) = 0$ for all x. By assumption h has no zero. Therefore h has an infinite image which again contains an interval (u, v) with $u > 0$. Then also $h^{-1}((u, v))$ is infinite and contains an interval (a, b). Now $h(t) \geq u > 0$ for all $t \in (a, b)$ which contradicts $\lim_{t \nearrow x} h(t) = 0$ for all x. □

If M expands an ordered field, then every definable unary function is in fact differentiable on a cofinite set. For the proof we need

2.7. **Lemma.** *If $F : [a, b] \longrightarrow M$ is a continuous and definable map then there is some $c \in [a, b)$ with*

$$\frac{F(x) - F(c)}{x - c} \leq \frac{F(b) - F(a)}{b - a} \text{ for all } x \in (c, b).$$

Proof. Let $l(x)$ be the line trough $(a, F(a))$ and $(b, F(b))$, thus

$$l(x) = mx + F(a) - ma \text{ with } m = \frac{F(b) - F(a)}{b - a}.$$

We may assume that for some $x \in (a, b)$ we have $\frac{F(x)-F(a)}{x-a} > m$ (otherwise we choose $c = a$). Then $r := F(x) - l(x) > 0$, as

$$F(x) - l(x) > F(a) + m(x - a) - l(x) = F(a) + m(x - a) - (mx + F(a) - ma) = 0.$$

By o-minimality the supremum c of all $x \in [a, b]$ with $F(x) - l(x) \geq r$ exists and we claim that c satisfies (i). Since F is continuous we have $F(c) - l(c) \geq r$. As

$F(b) = l(b)$ and $r > 0$ we get $c < b$ (as required). Pick $y \in (c, b)$. By the choice of c we have $F(y) - l(y) < r \leq F(c) - l(c)$. Hence $F(y) - F(c) < l(y) - l(c) = m(y - c)$ as desired. □

Note that 2.7 applied to $-F$ says that there is some $c \in [a, b)$ with

$$\frac{F(x) - F(c)}{x - c} \geq \frac{F(b) - F(a)}{b - a} \text{ for all } x \in (c, b).$$

2.8. **Theorem.** *If M is an o-minimal expansion of an ordered field, then every definable function $f : M \longrightarrow M$ is differentiable apart from a finite set.*

Proof. Suppose not. By the monotonicity theorem we may assume that there is an interval $[a, b]$ such that $F|[a, b]$ is continuous, a strictly increasing homeomorphism onto $[F(a), F(b)]$ and nowhere differentiable in (a, b).

For $x, y \in M$ with $x \neq y$ let

$$G(x, y) := \frac{F(y) - F(x)}{y - x}.$$

The set $\{x \in (a, b) \mid \lim_{y \searrow x} G(x, y) = +\infty\}$ has to be finite, otherwise it would contain a proper interval which contradicts 2.7. Similarly (but applying 2.7 to $-F(b + a - x)$) shows that the set $\{x \in (a, b) \mid \lim_{y \nearrow x} G(x, y) = +\infty\}$ is finite, too. Hence by shrinking $[a, b]$ (and using the fact that $G > 0$ everywhere) we may assume that for each $x \in (a, b)$,

$$\lim_{y \searrow x} G(x, y) \neq \lim_{y \nearrow x} G(x, y) \text{ and both limits are in } M.$$

By the monotonicity theorem we may shrink $[a, b]$ such that the functions $G^+(x) := \lim_{y \searrow x} G(x, y)$ and $G^-(x) := \lim_{y \nearrow x} G(x, y)$ are continuous on $[a, b]$. As $G^-(x) \neq G^+(x)$ for all $x \in [a, b]$ we know from the intermediate value theorem that $G^- < G^+$ or $G^- > G^+$ on $[a, b]$, say $G^- < G^+$. Further shrinking and using continuity we find some $r \in M$ with $G^- < r < G^+$ on $[a, b]$. As $G^- < r$ we may shrink $[a, b]$ such that $\frac{F(b) - F(a)}{b - a} < r$.

By 2.7, there is some $c \in [a, b)$ with $\frac{F(x) - F(c)}{x - c} \leq \frac{F(b) - F(a)}{b - a}$ $(x \in (c, b))$, hence $\frac{F(x) - F(c)}{x - c} < r$ $(x \in (c, b))$. But then $G^+(c) \leq r$, a contradiction. □

Many classical statements from Real Analysis also hold true in o-minimal expansions of arbitrary real closed fields. E.g.:

- Continuous definable functions on intervals (more generally on closed and bounded sets) are uniformly continuous.
- Rolle's theorem and the mean value theorem.
- L'Hospital's rule.
- The implicit function theorem

The proofs are in many cases identical to those from Real Analysis and in expansions of the real field they hold true by Real Analysis.

Exercise. Let R be an o-minimal expansion of a real closed field and let $f : (r, +\infty) \longrightarrow R$ be R-definable and differentiable. If $\lim_{x \to \infty} f(x) = 0$, then also $\lim_{x \to \infty} x \cdot f'(x) = 0$. (hint: use the mean value theorem).

Solution. Suppose not. By o-minimality we may assume that $f(x) > 0$ for all x. As $\lim_{x \to \infty} x \cdot f'(x) \neq 0$, there is some $\delta > 0$ such that $x \cdot f'(x) < -\delta$ for sufficiently large x (as $f > 0$ and $\lim_{x \to \infty} f(x) = 0$, f must be finally decreasing).

Let $r < a < b$. By the mean value theorem, there is some $\xi \in [a, b]$ such that

$$f'(\xi) = \frac{f(b) - f(a)}{b - a}.$$

Since $f'(\xi) < \frac{-\delta}{\xi}$ and $a < b$ we have $f'(\xi) \cdot (b - a) < -\frac{\delta(b-a)}{\xi}$, hence $f(b) < f(a) - \frac{\delta(b-a)}{\xi}$. Since $\xi \leq b$ we have $\frac{\delta(b-a)}{\xi} \geq \frac{\delta(b-a)}{b}$ thus

$$f(b) < f(a) - \frac{\delta(b-a)}{b} = f(a) - \delta + \frac{\delta a}{b} \ (r < a < b).$$

Since $f(x) \to 0$ as $x \to \infty$ we can choose a so that $f(a) < \frac{\delta}{3}$ and $b = 3a$. Then $f(b) < \frac{\delta}{3} - \delta + \frac{\delta}{3} = -\frac{\delta}{3}$ in contradiction to $f > 0$ everywhere. □

3. Definable Skolem functions and elimination of imaginaries

3.1. **Theorem.** *Let $M = (M, \leq, +, ...)$ be an o-minimal expansion of an ordered group and let $1 \in M$ be a positive element. Let $Z \subseteq M^k \times M^n$ be $\emptyset$-definable and let $X \subseteq M^k$ be its projection onto the first k coordinates. For $x \in M^k$ we write $Z_x = \{y \in M^n \mid (x, y) \in Z\} \subseteq M^n$. Then there is a map $f : X \longrightarrow M^n$ definable with parameters from $\{1\}$ such that*

(i) Z contains the graph of f (thus $f(x) \in Z_x$ for all $x \in X$), and
(ii) for all $x_1, x_2 \in X$ with $Z_{x_1} = Z_{x_2}$ we have $f(x_1) = f(x_2)$.

Proof. We first do the case $n = 1$. For a nonempty definable subset W of M we define a point $p(W) \in W$ as follows: Let B be the boundary $\overline{W} \setminus \text{interior}(W)$ of W. By o-minimality, B is finite. If $W = M$ then we take $p(W) = 0$. Otherwise $B \neq \emptyset$ and B has an infimum b; then we take

$$p(W) := \begin{cases} b & \text{if } b \in W \\ b - 1 & \text{if } b \notin W,\ b - 1 \in W \\ b + 1 & \text{if } b \notin W,\ b - 1 \notin W \text{ and } B = \{b\} \\ \frac{b+c}{2} & \text{if } b \notin W,\ b - 1 \notin W,\ B \text{ is not a singleton and } c = \inf(B \setminus \{b\}) \end{cases}$$

We now define for $x \in X$,

$$f(x) := p(Z_x).$$

As $p(Z_x) \in Z_x$ by definition of p, f satisfies (i). By definition the value of f at x only depends on Z_x, hence (ii) holds true. Moreover, the definition of p can be expressed by a formula using the formula defining Z and the parameter 1. This shows the theorem in the case $n = 1$.

We now assume the theorem for n and prove it for $n + 1$. So here $Z \subseteq M^k \times (M^n \times M)$. Let $\hat{Z} \subseteq M^k \times M^n$ be the projection of Z onto $M^k \times M^n$. We apply the case $n = 1$ to Z and $k + n, 1$ and get a map $g : \hat{Z} \longrightarrow M$ satisfying (i) and (ii) with respect to $\hat{Z}$ and $k + n, 1$

Moreover the projection X of Z to M^k is the projection of $\hat{Z}$ to M^k. We apply the induction hypothesis to $\hat{Z}$ and k, n and get a map $\hat{f} : X \longrightarrow M^n$ satisfying (i) and (ii) with respect to X and k, n. In particular $(x, \hat{f}(x)) \in \hat{Z}$ for all $x \in X$. We may therefore define $f : X \longrightarrow M^n \times M$ by

$$f(x) = (\hat{f}(x), g(x, \hat{f}(x))).$$

With $\hat{f}$ and g, also f is $\{1\}$-definable. Finally, routine checking shows that f satisfies (i) and (ii):

(i): As $\hat{f}$ satisfies (i) we know $(x, \hat{f}(x)) \in \hat{Z}$; as g satisfies (i) we get $(x, f(x)) = (x, \hat{f}(x), g(x, \hat{f}(x))) \in Z$.

(ii): Assume $Z_{x_1} = Z_{x_2}$. Then $\hat{Z}_{x_1} = \hat{Z}_{x_2}$ (for $y \in \hat{Z}_{x_1} \setminus \hat{Z}_{x_2}$ pick $m \in M$ with $(x_1, y, m) \in Z$; then $(y, m) \in Z_{x_1} \setminus Z_{x_2}$). As $\hat{f}$ satisfies (ii) we know $\hat{f}(x_1) = \hat{f}(x_2)$. Together with $Z_{x_1} = Z_{x_2}$ we obtain $Z_{(x_1, \hat{f}(x_1))} = Z_{(x_2, \hat{f}(x_2))}$. As g satisfies (ii) we get $f(x_1) = f(x_2)$. □

3.2. **Corollary.** *Let $M = (M, \leq, +, 1, ...)$ be an o-minimal expansion of an ordered group and $1 > 0$. Then M has definable Skolem functions and elimination of imaginaries.*

Proof. Recall: "M has elimination of imaginaries" means that for every $n \in \mathbb{N}$, each $\emptyset$-definable equivalence relation E of M^n is of the form $f(x) = f(y)$ for some $\emptyset$-definable map $M^n \longrightarrow M^k$ and some k. Here we can do even better: Each $\emptyset$-definable equivalence relation E of M^n has a $\emptyset$-definable subset of representatives: Take $Z = E$ in 3.1 and let $g : M^n \longrightarrow M^n$ be the map from 3.1. Then the image of g is a set of representatives of E by 3.1(ii). □

The existence of definable Skolem functions implies that definably closed subsets of a structure are elementary restriction of this structure, as follows easily from the Tarski-Vaught test (note that the converse is also true!). Thus we obtain

3.3. **Corollary.** *Let $M = (M, \leq, +, ...)$ be an o-minimal expansion of an ordered group. If $A \subseteq M$ contains an element different from 0, then the* ***definable closure*** $\mathrm{dcl}(A)$ *of A in M is an elementary substructure of M.*

3.4. *Examples.*

(a) If M is a divisible, ordered abelian group, then M is a $\mathbb{Q}$-vector space. The definable closure of a subset A of M is the subspace generated by A.
(b) If M is a real closed field then the definable closure of a subset A of M is the algebraic closure (in the sense of algebra) of the subfield R generated by A in M.

Another very important consequence of definable Skolem functions in o-minimal structures is the so called

3.5. **Curve Selection Lemma**
Let $M = (M, \leq, +, ...)$ be an o-minimal expansion of an ordered group. If $X \subseteq M^n$ is definable and $y \in \overline{X} \setminus X$, then there is a continuous definable map $\gamma : (0, \delta] \longrightarrow X$ (a "curve") for some $\delta > 0$ such that

$$\lim_{t \to 0} \gamma(t) = y.$$

Proof. Let $Z \subseteq M^n \times M$ be defined by

$$Z = \{(x, \varepsilon) \mid x \in X, \varepsilon > 0 \text{ and } ||y - x|| < \varepsilon\},$$

where $||(a_1, ..., a_n)||$ is defined to be the maximum of all the $|a_i|$. As y is in the closure of X, the projection of Z to the last coordinate is $(0, \infty)$. By 3.2 there is a definable map $\gamma : (0, \infty) \longrightarrow X$ with $||y - \gamma(\varepsilon)|| < \varepsilon$ for all $\varepsilon > 0$. By 2.1 applied to the coordinates of γ the limit $\lim_{t \to 0} \gamma(t)$ exists in M^n. By choice of Z, this limit has to be y. By the monotonicity theorem 2.6, there is some $\delta > 0$ such that γ is continuous on $(0, \delta]$. □

The curve selection lemma 3.5 in the o-minimal context plays the role of convergent sequences in Real Analysis. For example, the compact subsets of $\mathbb{R}^n$ are precisely the closed and bounded sets. However, in a general o-minimal structure the only compact subsets of M are the finite sets. In the o-minimal context, the correct notion is "definably compact": We call a subset S of M^n **definably compact** if every definable curve $\gamma : (0, 1] \longrightarrow S$ can be completed in S, i.e. the limit $\lim_{t\to 0} \gamma(t)$ exists and is in S.

3.6. **Corollary.** *(Characterisation of definably compact sets)*
Let M be an o-minimal expansion of an ordered field. A subset S of M^n is closed and bounded (in M^n) if and only if S is definably compact.

Proof. If S is closed and bounded then S is definably compact as follows easily from 2.1 (which says that one sided limits exist).

If S is unbounded then S has a projection onto some coordinate axis which is unbounded. By o-minimality this projection contains an interval of the form $[a, +\infty)$. Using definable Skolem functions and the field structure we can find a definable map $\gamma : (0, 1] \longrightarrow S$ such that the limit $\lim_{t\to 0} \gamma(t)$ does not exist in M^n. Using the monotonicity theorem and another re-scaling gives a curve as desired.

If S is not closed, then the curve selection lemma 3.5 implies that some curve with values in S can not be completed in S. □

3.7. **Corollary.** *In the situation of 3.6, definable continuous images of closed and bounded sets are closed and bounded: If $f : S \longrightarrow M^k$ is continuous and definable with $S \subseteq M^n$ closed and bounded, then also $f(S) \subseteq M^k$ is closed and bounded.*

Proof. Exercise. □

4. Dimension, part 1

We fix an o-minimal expansion M of an ordered group which has at least 2 definable constants. Recall from 3.3 that in this case, every definably closed subset is an elementary substructure of M.

A combinatorial consequence of the monotonicity theorem 2.6 is the following

4.1. **Exchange principle for** dcl
If $A \subseteq M$ and $b, c \in M$ then

$$c \in \mathrm{dcl}(A \cup \{b\}) \text{ and } c \notin \mathrm{dcl}(A) \;\Rightarrow\; b \in \mathrm{dcl}(A \cup \{c\}).$$

Here dcl is the definable closure in M.

Proof. $c \in \mathrm{dcl}(A \cup \{b\})$ says that for some $\emptyset$-definable function $f : M^n \times M \longrightarrow M$ and some n-tuple $\bar{a}$ from A we have $c = f(\bar{a}, b)$. Since $A \prec M$ there are finitely many intervals of A (so they have endpoints in $A \cup \{\pm\infty\}$) such that $f(\bar{a}, x)$ is constant or injective on each of these intervals. Since $c \notin \mathrm{dcl}(A)$, b can not be among the endpoints of any of these intervals. Hence there are $u, v \in A \cup \{\pm\infty\}$ with $u < b < v$ such that $f(\bar{a}, x)$ is constant or injective on (u, v) (note that $A \prec M$, hence the latter statement can be read in A and in M alike). If $f(\bar{a}, x)$ were constant on (u, v), then this constant would be in A, which contradicts $f(\bar{a}, b) = c \notin \mathrm{dcl}(A)$. So $f(\bar{a}, x)$ is injective on (u, v) and the inverse g is an A-definable map satisfying $g(c) = b$. But this means $b \in \mathrm{dcl}(A \cup \{c\})$. □

Observe that in any expansion of a totally ordered set, definable closure is the same as algebraic closure. Hence we will always talk about definable closure here.

4.2. **Definition.** A subset A of M is called **independent** if

$$a \notin \operatorname{dcl}(A \setminus \{a\})$$

for every $a \in A$.

4.3. **Corollary and Definition.** *By Zorn, every subset S of M contains maximal independent subsets and each of these sets is called a* **basis** *of S.*

All bases have the same cardinality, called the **dimension** *of S.*

Proof. The proof is literally the same as for the fact that every vector space has a basis: Replace the notion "linearly independent" with our notion of independence and use obvious properties of dcl together with the exchange principle 4.1. □

As in linear algebra we may also define the **dimension of a subset** B **of** M **over another subset** A of M as smallest size of a subset C of B which has the property $\operatorname{dcl}(A \cup C) = \operatorname{dcl}(A \cup B)$. Of course, then

$$\dim(A \cup B) = \dim A + \dim(B/A).$$

4.4. *Examples.*

(1) If M is a divisible, ordered abelian group, then M is a $\mathbb{Q}$-vector space. The dimension of a subset A of M is the vector space dimension of the subspace generated by A.
(2) If M is a real closed field then the dimension of a subset A of M is the transcendence degree of the subfield generated by A (in M).

5. Cell decomposition

In this section we will discuss a fundamental theorem of o-minimal structures which gives a nice description of definable subsets $S \subseteq M^n$. Topologically and very roughly speaking, the nice description says that S is a finite union of points and definable subsets C of M^n, each one being homeomorphic to an open box of M^k (k depending on C); moreover the homeomorphism can also be chosen to be definable in M.

We start by defining cells of o-minimal structures. Let M be an o-minimal expansion of a densely linearly ordered set.

5.1. **Definition.** (cells)

(i) A **continuous graph** over a definable subset $S \subseteq M^n$ is a graph of a continuous definable function $S \longrightarrow M$.
(ii) A **band** B over a definable subset $S \subseteq M^n$ is the open set between two continuous graphs of S or $\pm\infty$, i.e. there are continuous definable maps $f, g : S \longrightarrow M$ such that $f < g$ everywhere on S and

$$\begin{array}{rcl} B & = & (f,g)_S := \{(x,y) \in M^n \times M \mid f(x) < y < g(x)\}, \\ \text{or } B & = & (f,+\infty)_S := \{(x,y) \in M^n \times M \mid f(x) < y\}, \\ \text{or } B & = & (-\infty,g)_S := \{(x,y) \in M^n \times M \mid y < g(x)\}, \\ \text{or } B & = & S \times M \end{array}$$

A **cell** in M^n is any subset of M^n obtained from iterating continuous graph and band constructions, starting with points and open intervals of M.

5.2. *Warning.* The definition above depends on the order of the coordinates: If $C \subseteq M^{n+1}$ is a cell, then by definition, the projection of C to the **FIRST** n coordinates is a again a cell (and C is by definition a graph or a band over this projection). However the projection of C on other n coordinates is in general **NOT** a cell.

There is a convenient way to denote cells which indicates how they are constructed: We say that a cell C of M^n is an $(\varepsilon_1, ..., \varepsilon_n)$-cell, where $\varepsilon_i \in \{0, 1\}$ and $\varepsilon_i = 0$ if at stage i, the construction is a graph. Hence a (0)-cell is a point of M, a (1)-cell is an open interval of M, a $(0, 0, 0)$-cell is a point in M^3 and a $(1, 1)$-cell is a band over an open interval of M.

Observe that by definition, the projection of an $(\varepsilon_1, ..., \varepsilon_n)$-cell to the first k coordinates is a $(\varepsilon_1, ..., \varepsilon_k)$-cell.

5.3. **Lemma.** *Let $C \subseteq M^n$ be an $(\varepsilon_1, ..., \varepsilon_n)$-cell and let $i_1, ..., i_k$ be those coordinates with $\varepsilon_{i_j} = 1$. Let π be the projection onto these coordinates. Then $\pi(C)$ is an open cell (with respect to the coordinates $i_1, ..., i_k$) and $\pi|_C$ is a homeomorphism onto $\pi(C)$.*

Proof. By induction on n, where the case $n = 1$ is obvious. Assume we know the lemma for n. Let $p : M^n \times M \longrightarrow M^n$ be the projection onto the first n coordinates and let π_0 be the projection from M^n onto the coordinates $i \in \{1, ..., n\}$ with $\varepsilon_i = 1$.

As $p(C)$ is a $(\varepsilon_1, ..., \varepsilon_n)$-cell, by induction, $\pi_0(p(C))$ is an open cell and $\pi_0|_{p(C)}$ is an homeomorphism onto $\pi_0(p(C))$.

If $\varepsilon_{n+1} = 0$, then C is the graph of a definable continuous function $f : p(C) \longrightarrow M$ and $\pi(C) = \pi_0(p(C))$ is open. Clearly, the restriction of p to the graph of f (hence to C) is an homeomorphism onto $p(C)$ (with inverse $\bar{a} \mapsto (\bar{a}, f(\bar{a}))$). Thus $\pi|_C = (\pi_0|_{p(C)}) \circ p|_C$ is the composition of two homeomorphisms.

Now suppose $\varepsilon_{n+1} = 1$, i.e. C is a band over $p(C)$, say $C = (f_1, f_2)_{p(C)}$ with continuous definable maps $f_i : p(C) \longrightarrow M$ (the other cases being similar). Then $\pi = \pi_0 \times p$. We define functions $\hat{f}_i : \pi_0(p(C)) \longrightarrow M$ by

$$\hat{f}_i = f_i \circ (\pi_0|_{p(C)})^{-1}.$$

As $f_1 < f_2$ everywhere, also $\hat{f}_1 < \hat{f}_2$ everywhere and so

$$(\hat{f}_1, \hat{f}_2)_{\pi_0(p(C))} = \{(\bar{a}, b) \in \pi_0(M^n) \times M \mid \bar{a} \in \pi_0(p(C)), \hat{f}_1(\bar{a}) < b < \hat{f}_2(\bar{a})\}$$

is an open cell in $\pi_0(M^n) \times M$. It is now straightforward to see that $(\hat{f}_1, \hat{f}_2)_{\pi_0(p(C))} = (\pi_0 \times p)(C) = \pi(C)$ and $\pi|_C$ is an homeomorphism onto $\pi(C)$ with inverse

$$(\bar{a}, b) \mapsto ((\pi_0|_{p(C)})^{-1}(\bar{a}), b).$$

□

So it seems that cells "are" more or less open cubes. However the situation is not quite so simple, in particular if we want to understand how the cell lies in its ambient space. For example, the paper [BeFo] contains a 2 dimensional semi-algebraic cell C in $\mathbb{R}^4$, whose closure has a "hole" (i.e. $\overline{C}$ is homotopic to a circle).

Exercise. Find an open semi-algebraic cell $C \subseteq \mathbb{R}^2$ and a bounded continuous s.a. function $f : C \longrightarrow [0, 1] \subseteq \mathbb{R}$ which can not be extended continuously to the closure of C.

Find a semi-algebraic $(1,1,0)$-cell in $\mathbb{R}^3$ whose closure in $\mathbb{R}^3$ contains the z-axis $\{0\}^2 \times \mathbb{R}$.
Solution. Let C be the open right half space of $\mathbb{R}^2$ and let $f : C \longrightarrow \mathbb{R}$

$$f(x,y) = \frac{y}{x}$$

Then f is semi-algebraic, continuous, f can not be continuously extended to the origin´ and the graph C of f is a $(1,1,0)$-cell whose closure contains the z-axis. This answers the second question. □

A fundamental property of o-minimal structures says that definable sets in all dimensions are finite unions of cells. This alone implies that every o-minimal structure has an o-minimal theory (cf. 5.6). In fact we have

5.4. **Cell decomposition theorem**
Let M be an o-minimal expansion of a densely ordered set, let $S \subseteq M^n$ and let $f : S \longrightarrow M^m$ be definable. There is a decomposition of M^n into finitely many cells $C_1, ..., C_k$ (i.e. M^n is the disjoint union of the C_i), which is compatible with S (i.e. each C_i is either disjoint from S or contained in S), such that $f|_{C_i}$ is continuous for each i.

Proof. This is a lengthy induction and can be found in [vdD98, Chapter 3, (2.11)] and in [PS86]. □

Obviously, in dimension 1, this specialises to the monotonicity theorem.

There are many variants and improvements of this theorem in the literature; most of them modify the definition of a cell by saying that all functions used in the definition of "cells" have to be of a particular form. Some examples:

- If M expands a field, then the cell decomposition theorem is also true in its C^k-version, where $k \in \mathbb{N}$. That is, if we replace in the theorem and in the definition of "cell" every occurrence of "continuous" by k-times differentiable, then the theorem is still true, cf. [vdD98, Chapter 7, (3.2)].
- One can improve 5.4 to include certain monotonicity assertions about the occurring functions; this is called the regular cell decomposition, cf. [vdD98, Chapter 3, (2.19)]
- If M is a pure real closed field then all data in 5.4 can be chosen to be "Nash", i.e. all occurring functions are definable and infinitely differentiable in an open neighborhood of their domain. Cf. [BCR98]

We state various consequences of cell decomposition. First a notation. For $S \subseteq M^n \times M^k$ and $a \in M^n$ let

$$S_a := \{b \in M^k \mid (a,b) \in S\}.$$

It is good to think of S_a as the fibre of S above a (under the projection $\pi : M^n \times M^k \longrightarrow M^n$), although this is not true, since this fibre actually is $\{a\} \times S_a$. The set S is sometimes thought of as the **definably family** $(S_a)_{a \in M^n}$.

Note that by definition, for a cell $C \subseteq M^n \times M^k$ and $a \in M^n$, the set C_a is a cell of M^k (replace in the inductive definition of C, the variables $x_1, ..., x_n$ by $a_1, ..., a_n$).

5.5. **Corollary.** (*Uniform finiteness property*)
If M is an o-minimal $\mathscr{L}$-structure and $\varphi(\bar{x}, y)$ is an $\mathscr{L}$-formula, then there is some

$K \in \mathbb{N}$ *such that for all* $\bar{x}$*-tuples* $\bar{a}$ *from* M*, the subset of* M *defined by* $\varphi(\bar{a}, y)$ *is a union of at most* K *intervals.*

Proof. Let $Z \subseteq M^n \times M$ (where n is the length of $\bar{x}$) be the set defined by $\varphi(\bar{x}, y)$. By cell decomposition, Z is a union of K cells C of $M^n \times M$. Each C is a band or a graph over its projection onto the first n coordinates and so for $\bar{a} \in M^n$, the set $C_{\bar{a}} = \{b \in M \mid (\bar{a}, b) \in C\}$ is empty, a singleton or an open interval. Since the set defined by $\varphi(\bar{a}, y)$ (which is $Z_{\bar{a}}$) is the union of the $C_{\bar{a}}$'s, this set consists of at most K intervals. □

5.6. **Corollary.** *If* M *is an o-minimal structure and* N *is elementary equivalent to* M*, then also* N *is o-minimal. This property is referred as "o-minimal structures are* ***strongly o-minimal****".*

Proof. We have to show that every (parametrically!) definable subset $S \subseteq N$ is a finite union of intervals. Let S be defined by a formula $\varphi(\bar{a}, y)$ with an $\mathscr{L}$-formula $\varphi(\bar{x}, y)$ and some $\bar{x}$-tuples $\bar{a}$ from N. Choose $K \in \mathbb{N}$ according to uniform finiteness 5.5. The property described in 5.5 is clearly expressible in a first order statement. Since N is elementary equivalent to M, this statement also holds true in N. Thus S is a union of at most K intervals. □

5.7. **Corollary.** *O-minimal structures are* ***geometric****. Recall that a structure* M *is geometric if model theoretic algebraic closure has the exchange property in all* $N \equiv M$ *and if "*$\exists^\infty$*" is definable, i.e. for each* $\mathscr{L}$*-formula* $\varphi(\bar{x}, y)$ *there is some* $K \in \mathbb{N}$ *such that for all* $\bar{x}$*-tuples* $\bar{a}$ *from* M*, the subset of* M *defined by* $\varphi(\bar{a}, y)$ *is infinite or os size at most* K*.*

Proof. By 5.6, all $N \equiv M$ are again o-minimal and so by the exchange principle 4.1, acl (which is equal to dcl here) has the exchange property. By 5.5, M defines "$\exists^\infty$". □

Remark. The interest in geometric structures is the presence of a definable dimension function: As in 4.3 we can associate a dimension to subsets of the geometric structure M and in all of its elementary extensions. We then define the dimension of a definable subset S of M^n as the maximum dimension over M of an n-tuple $\bar{\alpha}$ from an elementary extension $N \succ M$ with $N \models S(\bar{\alpha})$. Using the second requirement of geometric structures one can then show that $\dim S \geq k$ if and only there is a projection π of S onto k coordinates such that in M,

$$\exists^\infty x_1 \ \dots \ \exists^\infty x_k \ (x_1, \dots, x_k) \in \pi(S)$$

holds true.

5.8. **Proposition.** *Definable connectedness is definable, i.e. for a definable sets* $S \subseteq M^n \times M^k$*, the set of all* $a \in M^n$ *for which* S_a *is definably connected is itself definable.*

Proof. Sketch: Firstly, it is easy to see by induction on d that cells of M^d are definably connected.

Now a finite union $C_1 \cup \dots \cup C_K$ of cells is definably connected if and only if there is some sequence $i_1, \dots, i_N$ of indices from $\{1, \dots, K\}$, enumerating $\{1, \dots, K\}$ with $N \leq K^2$ such that for each $j < N$, $\overline{C}_j \cap C_{j+1} \neq \emptyset$ or $C_j \cap \overline{C}_{j+1} \neq \emptyset$.

In order to see the proposition, choose finitely many cells $C_1, \dots, C_K$ of $M^n \times M^k$ whose union is S. Then for each $a \in M^n$, $S_a = C_{1,a} \cup \dots \cup C_{K,a}$ and the $C_{i,a}$ are again cells (in M^k). Now use the characterisation above to write down a formula which expresses that $C_{1,a} \cup ... \cup C_{K,a}$ is definably connected. □

Exercise. Show that for every cell C of an o-minimal expansion of a group, the closure is **definably path connected**, i.e. for all $a, b \in \overline{C}$ there is a continuous definable map $[0,1] \longrightarrow C$ starting at a and ending in b.
Exercise. Show that for an o-minimal structure on the set of real numbers, every definably connected set is connected.

6. Dimension, part 2

We will now describe the dimension of o-minimal structures (cf. 4.3) in geometric terms:

6.1. **Definition.** For a nonempty definable subset $S \subseteq M^n$ we define its **dimension** as

$$\dim S := \max\{d \in \mathbb{N}_0 \quad | \quad \text{there is a projection } \pi \text{ onto } d \text{ coordinates such that } \pi(S) \text{ has nonempty interior in } \pi(M^n)\}$$

Here M^0 stands for a singleton. We extend this to $\dim \emptyset := -1$

Important observation. The dimension of definable sets is definable in the following sense: for a definable set $S \subseteq M^n \times M^k$, the set of all $a \in M^n$ for which the fibre $S_a = \{b \in R^k \mid (a,b) \in S\}$ has dimension d, is again definable. This property is called "**definability of dimension**".

6.2. **Proposition.** *Let M be o-minimal.*

(i) If $S \subseteq M^n$ is definable then

$$\dim S = \max\{\dim(\bar{\alpha}/M) \mid \bar{\alpha} \in N^n, N \succ M \text{ and } N \models S(\bar{\alpha})\}.$$

(ii) If $N \succ M$ and $\bar{\alpha} \in N^n$, then

$$\dim(\bar{\alpha}/M) = \min\{\dim S \mid S \subseteq M^n \text{ definable and } N \models S(\bar{\alpha})\}.$$

Proof. (i). Let $d := \dim S$.
"$\leq$": We may assume that the projection π onto the first d coordinates has nonempty interior in M^d. Let $a_1 < b_1, ..., a_d < b_d \in M$ such that $\pi(S)$ contains the open box $B = \prod_{i=1}^d (a_i, b_i)$. Observe that $\dim(\alpha_1/M) = 1$ as $\alpha_1 \notin M$.

Take α_1 from an elementary extension N_1 of M such that $a_1 < \alpha_1 < c$ for every $c \in M$ with $a_1 < c$.

Take α_2 from an elementary extension N_2 of N_1 such that $a_2 < \alpha_2 < c$ for every $c \in N_1$ with $a_2 < c$. Observe that $\dim(\alpha_2/M\alpha_1) = 1$ as $\alpha_2 \notin \operatorname{dcl} M\alpha_1 \subseteq N_1$. From the additivity of dimension we get $\dim(\alpha_1, \alpha_2/M) = 2$.

Continuing in this way we see that we can produce a d-tuple $\bar{\alpha}$ in some elementary extension of M with $a_i < \alpha_i < b_i$ and $\dim(\bar{\alpha}/M) = d$. Since $B \subseteq \pi(S)$ is an elementary statement and $N \models B(\bar{\alpha})$ we get $N \models \pi(S)(\bar{\alpha})$. Now this means we can extend the d-tuple $\bar{\alpha}$ to an n-tuple which satisfies $N \models S(\bar{\alpha})$.
"$\geq$": Take $\bar{\alpha} \in N^n, N \succ M$ and $N \models S(\bar{\alpha})$. After decomposing S into finitely many cells C, we see that $N \models C(\bar{\alpha})$ for one such cell. Let $\varepsilon_1, ..., \varepsilon_n \in \{0,1\}$ such that C is a $(\varepsilon_1, ..., \varepsilon_n)$-cell and let $i_1, ..., i_k$ be those indices for which $\varepsilon_{i_j} = 1$. Let π be the projection onto these coordinates. By 5.3, $\pi(C)$ is open. Thus $d \geq k$ and it suffices to show $k \geq \dim(\bar{\alpha}/M)$. By 5.3, $\pi|_C$ is an homeomorphism onto $\pi(C)$, hence

for each $s \in C$, $(\pi|_C)^{-1}(\pi(s)) = s$. As $N \models C(\bar{\alpha})$ we also have $(\pi|_C)^{-1}(\pi(\bar{\alpha})) = \bar{\alpha}$. As $(\pi|_C)^{-1}$ is definable in M, we get

$$\bar{\alpha} \subseteq \operatorname{dcl}(M \cup \{\pi(\bar{\alpha})\})$$

and so $\dim(\bar{\alpha}/M) = \dim(\pi(\bar{\alpha})/M) \leq k$ (note that $\pi(\bar{\alpha})$ is a k-tuple).

(ii). By (i), the inequality $\leq$ of (ii) holds true. To prove "$\geq$", let $d = \dim(\bar{\alpha}/M)$. We may assume that $\alpha_{d+1}, ..., \alpha_n \subseteq \operatorname{dcl}(M \cup \{\alpha_1, ..., \alpha_d\})$, in other words there is a definable function $F : M^d \longrightarrow M^{n-d}$ such that $F(\alpha_1, ..., \alpha_d) = (\alpha_{d+1}, ..., \alpha_n)$. Let $S \subseteq M^n$ be the graph of F. Clearly $N \models S(\bar{\alpha})$ and it suffices to show that $\dim S \leq d$. We use (i) again: We know that $\dim S = \dim(\bar{\beta}/M)$ for some n-tuple $\bar{\beta}$ from some elementary extension N' of M with $N' \models S(\bar{\beta})$; thus $F(\beta_1, ..., \beta_d) = (\beta_{d+1}, ..., \beta_n)$ and $\bar{\beta} \subseteq \operatorname{dcl}(M \cup \{\beta_1, ..., \beta_d\})$. It follows $\dim S = \dim(\bar{\beta}/M) = \dim(\beta_1, ..., \beta_d/M) \leq d$. □

Let us examine how the dimension of definable sets behave under definable maps. The proof is basically an application of 6.2:

6.3. **Proposition.** *Let $S \subseteq M^n$ and $F : S \longrightarrow M^k$ be definable. Then*

$$\dim F(S) + \min_{y \in F(S)} \dim F^{-1}(y) \quad \leq \quad \dim S \quad \leq \quad \dim F(S) + \max_{y \in F(S)} \dim F^{-1}(y)$$

Proof. To see the first inequality, assume that all fibres of F have dimension $\geq d$. Take $\bar{\beta} \in N^k$ for some $N \succ M$ with $N \models F(S)(\bar{\beta})$ such that $\dim(\bar{\beta}/M) = \dim F(S)$ (cf. 6.2(i)). Since all fibres of F have dimension $\geq d$ and this is an elementary statement about M, the fibre $F^{-1}(\bar{\beta}) \subseteq N^n$ also has dimension $\geq d$. Hence there are $N_1 \succ N$ and $\bar{\alpha} \in N_1^n$ with $N_1 \models S(\bar{\alpha})$ and $F(\bar{\alpha}) = \bar{\beta}$ such that $\dim(\bar{\alpha}/N) \geq d$, in particular $\dim(\bar{\alpha}/M\bar{\beta}) \geq d$. It follows

$$\dim S \geq \dim(\bar{\alpha}/M) = \dim(\bar{\alpha}, \bar{\beta}/M) = \dim(\bar{\beta}/M) + \dim(\bar{\alpha}/M\bar{\beta}) = \dim F(S) + d.$$

To see the second inequality take $N \succ M$ and some $\bar{\alpha} \in N^n$ with $N \models S(\bar{\alpha})$ and $\dim(\bar{\alpha}/M) = \dim S$ (cf. 6.2(i)). Let $\bar{\beta} = F(\bar{\alpha}) \in N^k$ and let $d := \dim(\bar{\alpha}/M\bar{\beta})$. Then

$$\dim S = \dim(\bar{\alpha}/M) = \dim(\bar{\alpha}, \bar{\beta}/M) = \dim(\bar{\beta}/M) + \dim(\bar{\alpha}/M\bar{\beta}) \leq \dim F(S) + d$$

and it suffices to show that F has a fibre of dimension $\geq d$.

By 3.3, $N_0 := \operatorname{dcl}(M\bar{\beta})$ is an elementary substructure of N (provided M has definable Skolem functions, e.g. if M expands a group; we assume this here for simplicity). We apply 6.2(i) to the fibre of $F : S(N_0^n) \longrightarrow N_0^k$ above $\bar{\beta}$ and see that the dimension of this fibre is at least d. Hence F considered in N_0 has a fibre of dimension $\geq d$. Since this is an elementary statement, also the original map F has a fibre of dimension d. □

Here a list of obvious consequences of 6.3:

6.4. **Corollary.** *Let $S \subseteq M^n$ be definable.*

(1) If $F : S \longrightarrow M^k$ is a definable map then $\dim S \geq \dim F(S)$ and if the dimension of the fibres is constant d, then $\dim S = \dim F(S) + d$. In particular, if F is injective then $\dim S = \dim F(S)$.
(2) $\dim(S \times T) = \dim S + \dim T$ for all definable $T \subseteq M^k$.
(3) $\dim(S \cup T) = \max\{\dim S, \dim T\}$ for all definable $T \subseteq M^n$.

A very useful and more complicated property of the dimension function is the following:

6.5. **Theorem.** ([vdD98, chapter 4, (1.8)])
*If $S \subseteq M^n$ is definable then the **frontier** $\overline{S} \setminus S$ of S has dimension strictly less than* $\dim S$.

7. Restricted analytic functions and global exponentiation

So far we have seen real closed fields and divisible ordered abelian groups as examples of o-minimal structures. In the mid 1980's the first proper o-minimal expansions of the real field $\mathbb{R}$ was discovered by van den Dries (the expansion of $\mathbb{R}$ by the restricted exponential function $\exp|[0,1]$). In the meantime a huge class of non-algebraic functions defined on subsets of $\mathbb{R}^n$ is known. We give here a brief introduction to two types of such functions which were and are studied in many areas of mathematics and for which the discovery of their o-minimality had big impact outside (and of course also inside) model theory.

Recall that a function $f : U \longrightarrow \mathbb{R}$, $U \subseteq \mathbb{R}^n$ open, is **analytic** if for each $a \in U$, the Taylor series expansion of f in U converges in an open neighborhood of a and it converges to f in that neighborhood. Observe that, in contrast to the complex case, for analytic $f : \mathbb{R} \longrightarrow \mathbb{R}$ the power series expansion of f at 0 in general does not converge everywhere: for example the function $f(x) = \frac{1}{1+x^2}$ is real analytic on $\mathbb{R}$, but its Taylor series about 0 is $1 - x^2 + x^4 - x^6 + \ldots$.
If U is an open subset of $\mathbb{R}^n$ with $[-1,1]^n \subseteq U$ and $f : U \longrightarrow \mathbb{R}$ is analytic, we define $\hat{f} : \mathbb{R}^n \longrightarrow \mathbb{R}$ by

$$\hat{f}(x) := \begin{cases} f(x) & \text{if } x \in [-1,1]^n, \\ 0 & \text{otherwise.} \end{cases}$$

Let $\mathbb{R}_{\mathrm{an}}$ be the expansion of the real field together with all functions $\hat{f}$ for every analytic function f defined on an open superset of $[-1,1]^n$ and every $n \in \mathbb{N}$.

So $\mathbb{R}_{\mathrm{an}}$ in a $\mathscr{L}_{\mathrm{an}}$-structure, where $\mathscr{L}_{\mathrm{an}}$ is the language of ordered rings together with an n-ary function symbol $\hat{f}$ for every analytic function f defined on an open superset of $[-1,1]^n$.

7.1. **Theorem.** (*Denef, van den Dries,* [DvdD88])
$\mathbb{R}_{\mathrm{an}}$ *is o-minimal and admits quantifier elimination in the language $\mathscr{L}_{\mathrm{an}}$ extended by a name for the function $x \mapsto \frac{1}{x}$.*

Comments on the proof: After re-scaling, we can restrict the universe to $[-1,1]$. Quantifier elimination is achieved by first applying the Weierstrass preparation theorem to analytic functions in n variables to write them in terms of polynomials with respect to one of the variables (here some work is involved as the Weierstrass preparation theorem is only applicable to regular germs and we have to reduce to this case first). Then quantifier elimination for the real field is used to eliminate this variable locally. By compactness, the variable is then eliminable globally.

Observe that every function defined on a compact subset K of $\mathbb{R}^n$ which has an analytic extension on some open neighborhood of K, has an $\mathbb{R}_{\mathrm{an}}$-definable analytic extension on some open neighborhood of K (Exercise!). For example, the complex exponential function viewed as a map $\mathbb{R}^2 \longrightarrow \mathbb{R}^2$ restricted to every ball is definable in $\mathbb{R}_{\text{an}}$. On the other hand, global complex exponentiation $\mathbb{R}^2 \longrightarrow \mathbb{R}^2$ is not o-minimal (why?) and therefore not definable in $\mathbb{R}_{\text{an}}$.

Real exponentiation is also not definable in $\mathbb{R}_{\text{an}}$ since $\mathbb{R}_{\text{an}}$ is **polynomially bounded**, i.e. for every unary definable function $f : M \longrightarrow M$, there is some $d \in \mathbb{N}$ such that $f(x) \leq x^d$ for all sufficiently large x (cf. [vdD86]); in fact in this paper van den Dries shows that $f(x)$ is asymptotic to $c \cdot x^q$ for some $c \in M$ and a rational number q (so $\lim_{x\to\infty} \frac{f(x)}{cx^q} = 1$). In particular, the function $x \mapsto x^{\sqrt{2}}$ is not definable in $\mathbb{R}_{\text{an}}$. Of course, $x \mapsto x^{\sqrt{2}}$ would be definable if global exponentiation $\mathbb{R} \longrightarrow \mathbb{R}$ were available.

Let $\mathbb{R}_{\text{exp}}$ be the expansion of the real field by the real exponential function in the language $\mathscr{L}_{\text{exp}}$ of ordered rings expanded by a unary function symbol exp.

7.2. **Theorem.** (*Wilkie*) $\mathbb{R}_{\text{exp}}$ *is model complete and o-minimal.*

Comments on the proof (cf. [Wil96]). Recall that a structure M is model complete if all $\emptyset$-definable subsets are projections of quantifier free $\emptyset$-definable sets (these are precisely the existentially definable sets). Wilkie uses the Robinson test for model completeness and shows that for any extension $M \subseteq N$ of models of $\text{Th}(\mathbb{R}_{\text{exp}})$, M is existentially closed in N (the language here is $\mathscr{L}_{\text{exp}}$).

It was shown by Khovanskii earlier that o-minimality follows from model completeness: Khovanskii proved that every **exponential variety**, i.e. zero set in $\mathbb{R}^n$ of a system of equations

$$\begin{gathered} P_1(x_1, ..., x_n, e^{x_1}, ..., e^{x_n}) = 0 \\ \vdots \\ P_k(x_1, ..., x_n, e^{x_1}, ..., e^{x_n}) = 0, \end{gathered}$$

with polynomials $P_i \in \mathbb{R}[X_1, ..., X_n, Y_1, ..., Y_n]$, has only finitely many connected components. Now it is an easy exercise to show that every set, existentially definable in $\mathbb{R}_{\text{exp}}$, is the projection of an exponential variety. Hence model completeness implies that all $\mathbb{R}_{\text{exp}}$-definable sets are projections of sets with only finitely many connected components. So these sets also have only finitely many connected components, which implies o-minimality.

We can also merge $\mathbb{R}_{\text{an}}$ and $\mathbb{R}_{\text{exp}}$: Let $\mathbb{R}_{\text{an,exp}}$ be the expansion of $\mathbb{R}_{\text{an}}$ by the real exponential function in the language $\mathscr{L}_{\text{an,exp}} = \mathscr{L}_{\text{an}}(\text{exp})$.

7.3. **Theorem.** ([vdDMM94])
$\mathbb{R}_{\text{an,exp}}$ *has quantifier elimination after naming the logarithm (i.e., in the language* $\mathscr{L}_{\text{an,exp}}(\log)$*) and* $\mathbb{R}_{\text{an,exp}}$ *is o-minimal.*

There are much more o-minimal expansions of the real field known, e.g. if M is an o-minimal expansion of the real field, $I \subseteq \mathbb{R}$ is an open interval and $a \in I$ then for every continuous function $f : I \longrightarrow \mathbb{R}$, definable in M, we can add the function $x \mapsto \int_a^x f(t)\mathrm{d}t$ (defined on I) to M and get again an o-minimal structure. This is a special case of the construction of the so-called Pfaffian closure of an o-minimal expansion of the real field by Speissegger (cf. [Spe99]). The integration of functions in this way indeed is not possible in $\mathbb{R}_{\text{an,exp}}$: In [vdDMM97] the error function $x \mapsto \int_0^x e^{-t^2}\mathrm{d}t$ is shown to be not definable in $\mathbb{R}_{\text{an,exp}}$.

A surprising result by Miller says that any o-minimal expansion of the real field which is not polynomially bounded must define global exponentiation $\mathbb{R} \longrightarrow \mathbb{R}$. This is called the **growth dichotomy** (cf. [Mil94]).

The first order theory of $\mathbb{R}_{\exp}$.

Beside the geometric interest in producing o-minimal structures of large classes of functions (and therefore providing the o-minimal machinery for these functions), decidability questions were a strong motivation for the development of the theory.

Recall that Tarski proved the decidability of the real field $\mathbb{R}$ by showing effective quantifier elimination in the language $\mathscr{L} = \{\le, +, -, \cdot, 0, 1\}$. Hence an algorithm for testing truth of an $\mathscr{L}$-sentence φ in $\mathbb{R}$ explicitly transforms φ into a quantifier free $\mathscr{L}$-sentence χ, equivalent to φ in $\mathbb{R}$. In a second step it is then trivial to decide whether χ holds in $\mathbb{R}$, since χ is only a boolean combination of expressions of the form $P(\bar{a}) \ge 0$, where P is a polynomial with coefficients in $\mathbb{Z}$ and $\bar{a} \subseteq \mathbb{Z}$.

Tarski asked whether the decidability can be extended to $\mathbb{R}_{\exp}$. Both steps in Tarski's approach are not directly accessible: Firstly, $\mathbb{R}_{\exp}$ does not have quantifier elimination in the language $\mathscr{L}_{\exp}$ (a counterexample can be found in [vdD84]). Secondly, how to decide existential sentences of $\mathbb{R}_{\exp}$? Note that we at least have to decide quantifier free sentences like

$$P(e, e^2, e^e) = 0 \ (P \in \mathbb{Z}[x_1, x_2, x_3])$$

in $\mathbb{R}_{\exp}$.

Here some highlights of what is known about Tarski's problem: Ressayre, van den Dries, Marker and Macintyre have shown that the elementary theory of $\mathbb{R}_{\exp}$ is decidable provided the theory of the restricted exponential $\exp(\frac{1}{1+x^2})$ is decidable. Wilkie has shown that the theory of $\mathbb{R}_{\exp(\frac{1}{1+x^2})}$ is effectively model complete. Wilkie and Macintyre in [MW96] have shown that the theory of $\mathbb{R}_{\exp}$ is decidable provided the following number theoretic conjecture holds true:

7.4. **Schanuel's Conjecture** If $\lambda_1, ..., \lambda_n \in \mathbb{C}$ are linearly independent over $\mathbb{Q}$, then the field $\mathbb{Q}(\lambda_1, ..., \lambda_n, e^{\lambda_1}, ..., e^{\lambda_n})$ has transcendence degree at least n.

Note that this conjecture is known to hold true if $\lambda_1, ..., \lambda_n$ are algebraic over $\mathbb{Q}$ by the Lindemann-Weierstrass Theorem. Also note that Schanuel's Conjecture would imply the algebraic independence of e and π (choose $\lambda_1 = 1, \lambda_2 = 2\pi i$).

8. NIP and neural networks

8.1. **Vapnik-Chervonenkis dimension.**

8.1.1. **Definition.** Let X be a set and let $\mathcal{S}$ be a collection of subsets of X. We say that $\mathcal{S}$ **shatters** a subset $B \subseteq X$ if every subset of B is of the form $B \cap S$ for some $S \in \mathcal{S}$.

If there is some $d \in \mathbb{N}$ such that $\mathcal{S}$ does not shatter any subset of size d of X, then the smallest such d is called the **VC-dimension**, or **VC-index**, of $\mathcal{S}$. 'VC' stands for Vapnik-Chervonenkis. In this case $\mathcal{S}$ is called a **VC-class**.

If there is no such d, then $\mathrm{VC}(\mathcal{S}) := \infty$.

Let $\mathcal{S}$ be a collection of subsets of a set X. For $B \subseteq X$, let $B \cap \mathcal{S} = \{B \cap S \mid S \in \mathcal{S}\}$. For $n \in \mathbb{N}$ let

$$f_{\mathcal{S}}(n) = \max\{|B \cap \mathcal{S}| \mid B \subseteq X \text{ and } |B| = n\}.$$

Thus $f_{\mathcal{S}}(n) = 2^n$ if and only if $\mathcal{S}$ shatters a set of size n. Surprisingly, $f_{\mathcal{S}}(n)$ is polynomially bounded for large n, if $\mathcal{S}$ has finite VC-dimension:

8.1.2. **Theorem.** [vdD98, chapter 5, (1.6)] *Suppose $\mathcal{S}$ does not shatter any subset of X of size d. Then for all $n \geq d$, $f_{\mathcal{S}}(n)$ is at most the number of subsets of an n-element set of size $< d$, given by*

$$p_d(n) = \sum_{i<d} \binom{n}{i}.$$

Observe that $p_d(n)$ is a polynomial of degree $d-1$.

Proof. First note (by counting subsets of size $< d$ of an n-element set) that $p_d(n) = p_{d-1}(n-1) + p_d(n-1)$.

We proceed by induction on n. If $n = d$ then $f_{\mathcal{S}}(n) < 2^n = p_n(n) - 1$. Now let $n > d$ and let $B \subseteq X$ be of size n. We must show that $|B \cap \mathcal{S}| \leq p_d(n)$ and of course we may replace $\mathcal{S}$ by $B \cap \mathcal{S}$. Fix $x \in B$ and define

$\mathcal{S}_0 = \{S \in \mathcal{S} \mid x \notin S \text{ and } S \cup \{x\} \in \mathcal{S}\}$
$\mathcal{S}_1 = \{S \in \mathcal{S} \mid x \in S \text{ or } S \cup \{x\} \notin \mathcal{S}\}$

Since $\mathcal{S}$ does not shatter any subset of X of size d, $\mathcal{S}_0$ does not shatter any subset of $X \setminus \{x\}$ of size $d-1$.

Hence the induction hypothesis says $|(B \setminus \{x\}) \cap \mathcal{S}_0| \leq p_{d-1}(n-1)$. As $x \notin S$ for any $S \in \mathcal{S}_0$, $(B \setminus \{x\}) \cap \mathcal{S}_0 = \mathcal{S}_0$ and $|\mathcal{S}_0| \leq p_{d-1}(n-1)$.

On the other hand $|\mathcal{S}_1| \leq |(B \setminus \{x\}) \cap \mathcal{S}_1|$ since the map $\mathcal{S}_1 \longrightarrow (B \setminus \{x\}) \cap \mathcal{S}_1$ which removes x is injective (by definition of $\mathcal{S}_1$ and since all $S \in \mathcal{S}_1$ are assumed to be a subset of B).

By the induction hypothesis we have $|\mathcal{S}_1| \leq p_d(n-1)$. Thus $|\mathcal{S}| = |\mathcal{S}_0| + |\mathcal{S}_1| \leq p_{d-1}(n-1) + p_d(n-1) = p_d(n)$. □

8.2. **O-minimal structures have NIP.**

First a reminder: Let T be an $\mathscr{L}$-theory and let $\varphi(\bar{x}, \bar{y})$ be an $\mathscr{L}$-formula. We say that φ has the **independence property** (w.r.t. $\bar{x}$,$\bar{y}$) if in some model M of T, the set of fibres $(S_{\bar{a}})_{\bar{a} \in M^{\bar{x}}}$, where S is defined by φ, shatters some infinite subset of $M^{\bar{y}}$. Hence there are

$$\bar{a}_I \in M^{\bar{x}} \text{ and } \bar{b}_i \in M^{\bar{y}} \ (i \in \omega, I \subseteq \omega)$$

such that

$$M \models \varphi(\bar{a}_I, \bar{b}_i) \iff i \in I.$$

This is equivalent to saying that the collection $(\varphi[M^{\bar{x}}, \bar{b}_i])_{i\in\omega}$ is independent in the sense of boolean algebras.

The theory T **has the independence property** if some formula has the independence property. If T does not have the independence property, then T is called **dependent**, or T is said to have the **NIP**.

A structure M has (N)IP, if its theory has (N)IP. By compactness, M has the independence property if and only if there is a definable family $(S_{\bar{a}})_{\bar{a}\in M^n} \subseteq M^k$ which shatters finite subsets of M^k of arbitrary size (Exercise!); in other words if $(S_{\bar{a}})_{\bar{a}\in M^n}$ has infinite VC-dimension.

In [vdD98, chapter 5], it is explicitly proved that no formula in an o-minimal structure has the independence property. We shall use the following general model theoretic criterion instead:

8.2.1. **Theorem.** ([Poi85, théorème 12.28])
An arbitrary $\mathscr{L}$-theory T has the independence property if and only if there is a 1-type p over some model M of T with $\operatorname{card} M \geq \operatorname{card} \mathscr{L}$ *and some $N \succ M$ such that p has at most $2^{2^{\operatorname{card} M}}$* **coheirs** *on N.*

In order to apply 8.2.1 we describe coheirs of 1-types of o-minimal structures: Recall that for an elementary extension $M \prec N$ of $\mathscr{L}$-structures, an n-type q of N is called a **coheir** over M if q is finitely realisable in M. In this case q is a coheir of its restriction p to M. Since M, N are o-minimal, p and q are determined by the cuts they induce on M.

8.2.2. **Proposition.** *Every 1-type of an o-minimal structure M has at most 2 coheirs on any $N \succ M$.*

Proof. Suppose there are 3 distinct coheirs q_1, q_2, q_3 of p on N. By 5.6, also N is o-minimal and the q_i are determined by the cuts they induce on N. Up to permutation we may therefore assume that we have $\alpha, \beta \in N$ such that the position of q_1, q_2, q_3 and α, β looks as follows:

$$N \text{ ———————— } q_1 \text{ — } \alpha \text{ — } q_2 \text{ — } \beta \text{ — } q_3 \text{ ———————— }$$

$$M \underset{p^L}{\text{ ———————— }} \qquad\qquad p \qquad\qquad \underset{p^R}{\text{ ———————— }}$$

As q_1 and q_3 lie over p, α and β realise p. But then the interval (α, β) of N does not contain elements of M, in other words, the formula $\alpha < x < \beta$ is contained in q_2 and not realisable in M. Thus q_2 cannot be a coheir of p. □

8.2.3. **Corollary.** *Every o-minimal structure has* NIP.

Proof. By 5.6 every model of the theory of M is o-minimal. Hence the result follows from 8.2.1 and 8.2.2. □

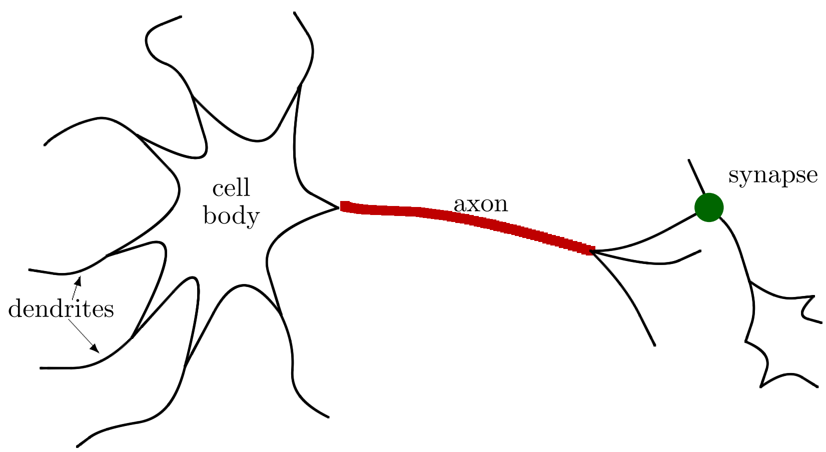
Hence in an o-minimal structure M for every definable family $S \subseteq M^n \times M^k$, there is some $d \in \mathbb{N}$ such that $(S_{\bar{a}})_{\bar{a}\in M^n}$ does not shatter any subset of M^k of size d.

8.2.4. **Corollary.** *If M is o-minimal and $S \subseteq M^n \times M^k$ is definable, then there is some $d \in \mathbb{N}$ such that for all sufficiently large $n \in \mathbb{N}$ and every subset $X \subseteq M^k$ of size n, there are at most n^d sets of the form $X \cap S_a$ where a varies in M^n.*

Proof. By 8.2.3, the collection $\{S_a \mid a \in M^n\}$ has finite VC-dimension d. Now apply 8.1.2 and notice that $p_d(n)$ is a polynomial of degree $d-1$. □

8.3. **An application to neural network learning.** In this section I describe a main instance of how the NIP property is applied in neural network learning. The first two subsections are an attempt to motivate the architecture of a neural network and to develop the intuition behind the mathematical theory of neural network learning. I'm introducing some terminology from the neural network literature to help the reader approaching these texts. The last subsection gives a formal definition of neural network learning for binary output networks and shows in which way the NIP property is fundamental to the subject (cf. 8.3.3.2). Most of the material here is from [AB99]. Further reading and computations of the VC-dimension of various families, definable in o-minimal structures can be found in [MS93] and [KM97].

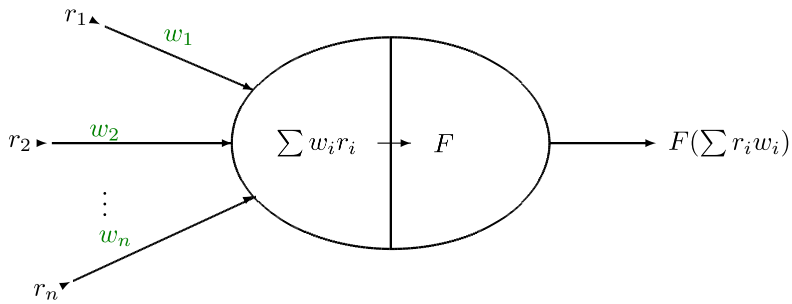
8.3.1. *A neuron.* I want to start with a rough description of a biological neuron and what can be observed when the neuron (or better its host) is learning.



A neuron is a cell consisting of a cell body, several input filaments called **dendrites** and a single long output filament called an **axon**. The cell body receives inhibitory or excitatory electrical impulses through the dendrites and sends an impulse down the axon depending on these inputs. The axon splits into thousands of branches which end in a **terminal button**. Each terminal button is connected to dendrites of other neurons across a small gap called a **synapse**. The synapses convert the activity from the axon into inhibitory or excitatory impulses and transmit them into the dendrites of other neurons.

During learning the synapses change their effect on the dendrites. Now this does not explain how the brain learns, it is merely what can be observed form the neuron during the learning process. However, it is this behavior of the neuron which is rebuilt in artificial neural networks.

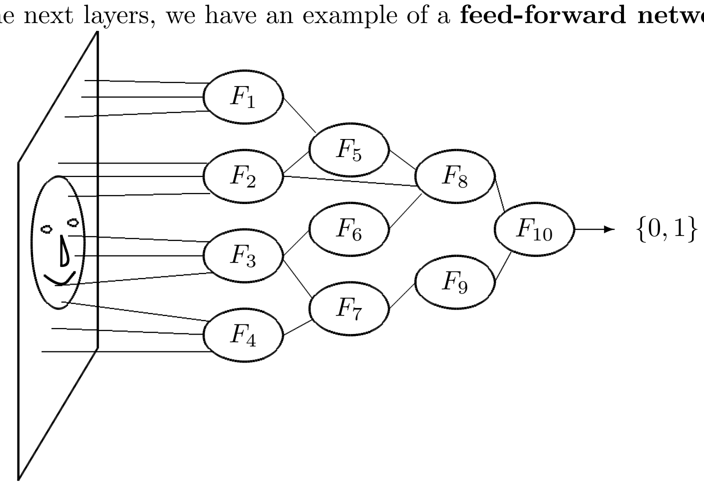
8.3.2. *Artificial neural networks and simulation of learning.*



An artificial neuron receives real numbers r_i as inputs, computes the weighted sum $\sum r_i w_i$ and applies an **activation function** F to this sum. The weights w_i are adjustable and play the role of the synaptic activity during learning of a biological neuron. Hence, in mathematical terms, a neuron is a family H of functions $\bar{r} \mapsto F(\sum r_i w_i)$, where the w_i vary in some parameter space.

Typical activation functions used in practice are characteristic functions of an interval $(a, +\infty)$, piecewise linear functions or the **sigmoid function** $F(t) = \frac{1}{1+e^{-t}}$.

An (artificial) **neural network** consists of a number of artificial neurons, whose input and output wires are connected in some way. Here is an example of a network with 10 neurons and 4 **layers**; since each neuron is only sending output to neurons in the next layers, we have an example of a **feed-forward network**:



Let X be the input space of our network. X can be a finite set or in our example, the network receives a tuple of points from $\mathbb{R}^2 \times \{0, ..., 255\}$ (describing the coordinates and the color value of a point), say of l points. So $X = (\mathbb{R}^2 \times \{0, ..., 255\})^l$; $l = 12$ in the picture above. The network is supposed to recognise patterns and has $Y = \{0, 1\}$ as output space. Each of the artificial neurons is equipped with an activation function $F_i(\bar{x}, \bar{w})$, and so the network is capable of computing a class H of functions $X \longrightarrow Y$.

8.3.2.1. *Remark.* For our discussion here, in particular for the result 8.3.3.2, the only thing that matters is that an artificial neural network is modeled as a family of functions $F(\bar{x}, \bar{w})$, where the $\bar{w}$ vary in some parameter space. In the quantitative analysis of a neural network it is important though how the neurons are implemented, e.g. in [AB99, Section 8.4] a two-layer feed-forward network using sigmoid functions is analysed.

Mathematically, the network's physical configuration is modeled by an input space X (an arbitrary but fixed set), the output space $Y = \{0, 1\}$ and a class H of functions $X \longrightarrow Y$. This is the model for a **binary output network** and there are many others discussed in the literature, e.g. **real valued output networks**.

We want to formalise the way in which the weights in the network are adapted during the learning process so that the network is able to "learn" a pattern which is presented to it. Let us make this process more precise:

The learning cycle:

- At the beginning of the cycle the network is in a certain state, given by the weights, or simply given by a function $h \in H$ coded by the weights.
- The network receives a **training sample** $(x, y) \in X \times Y$, which is randomly chosen by a "teacher". It is a good idea to think of a training sample as a "correct sample".
- We then compute $h(x)$ using the function $h \in H$ of the network's current state.
- Finally we adapt the weights w_{ij} of the $F_i(\bar{x}, \bar{w})$ depending on whether $h(x) = y$ or $h(x) \neq y$. In this last step of the cycle we also want to take into account the training samples of previous cycles, so that we can make use of the information $h(x) = y$.

Goal:

After a finite number of training samples the network has "learned" a function $h \in H$ which approximates the pattern in a best possible way and can now recognise the pattern; so we can give the network an input $x \in X$ and hopefully get a good approximation to an answer of the question on whether x is an instance of the correct pattern.

More formally, the learning cycles indicate that we are looking for an algorithm L receiving training samples, or better sequences of training samples, and returning a weight vector (w_{ij}), or simply: a function $h \in H$ (coded by the weights). In other words the learning cycles are mimicked by a single function

$$L : \bigcup_{m=1}^{\infty} Z^m \longrightarrow H, \text{ where } Z = X \times Y.$$

We consider the goal above to be achieved by L if for large m, and almost all $z \in Z^m$, the function $L(z)$ computes the pattern as good as possible. The mathematical definition expressing this goal can be found in 8.3.3.1 below and we want to motivate this definition a bit further. The evaluation of a state $h \in H$ of the network for a particular pattern and training method, is mimicked with the aid of a **probability measure** p on the set of samples $Z = X \times Y$. At this stage it is important to note that during the design of the network and the implementation of the algorithm L, we do not have any information about the pattern or how the training samples are presented to us; both information are captured in the measure p. Recall that the network only consists of the class of functions H and we have to design an algorithm independent of p.

8.3.3. *A mathematical formulation of neural network learning.*

Here we describe the mathematical formulation of what we have motivated in the previous section. Let X be a set, $Y = \{0,1\}$ and let H be a set of functions $X \longrightarrow Y$. Of course we may also think of H as a collection of subsets of X.

Let $Z = X \times Y$ be the sample space and let p be a probability measure on Z. We define the error of $h \in H$, given p by

$$\mathrm{er}_p(h) = p\{(x,y) \in Z \mid h(x) \neq y\}.$$

p measures the probability that a sample is chosen as a training sample. The best approximation in H for given p is defined as

$$\mathrm{opt}_p(H) = \inf_{h \in H} \mathrm{er}_p(h).$$

8.3.3.1. **Formal definition of learning**

A **learning algorithm** L for the network given by H (and X, $Y = \{0,1\}$) is a map

$$L : \bigcup_{m=1}^{\infty} Z^m \longrightarrow H$$

with the following property:

$\forall\ \varepsilon, \delta \in (0,1)\ \exists\ m_0(\varepsilon,\delta) \in \mathbb{N}\ \forall\ m \geq m_0(\varepsilon,\delta)$:

for every probability measure p on $Z = X \times Y$ and each m-tuple of samples $z \in Z^m$ we have

$$p^m\{\mathrm{er}_p(L(z)) < \mathrm{opt}_p(H) + \varepsilon\} \geq 1 - \delta,$$

where p^m is the product measure on Z^m.

The class H is called **learnable** if there is a learning algorithm L for the network (X, Y, H) and in this case the minimum of all the bounds $m_0(\varepsilon,\delta)$, were L varies through the learning algorithms is called the **inherent sample complexity**, denoted by $m_H(\varepsilon,\delta)$.

8.3.3.2. **Theorem.** ([AB99, Thm 5.5])

H is learnable if and only if H has finite VC-dimension, by which we mean that the collection of all subsets of X of the form $h^{-1}(1)$ has finite VC-dimension.

In this case the inherent sample complexity of H is asymptotic to

$$\frac{1}{\varepsilon^2} \cdot \log(\frac{1}{\delta}),$$

i.e. there is some $K \in \mathbb{N}$ such that for sufficiently small ε, δ we have $m_H(\varepsilon,\delta) \leq K \cdot \frac{1}{\varepsilon^2} \cdot \log(\frac{1}{\delta})$ and $\frac{1}{\varepsilon^2} \cdot \log(\frac{1}{\delta}) \leq K \cdot m_H(\varepsilon,\delta)$.

Observe that the estimating term $\frac{1}{\varepsilon^2} \cdot \log(\frac{1}{\delta})$ of the inherent sample complexity in 8.3.3.2 is independent of H (!)

Summing up, if H is a definable family of functions from an o-minimal expansion of $\mathbb{R}$ (e.g. if in our network picture above, $F_1 - F_{10}$ are definable in $\mathbb{R}_{\mathrm{an,exp}}$), then H is learnable.

If H is learnable then the following **SEM**-algorithm ("Sample Error Minimisation") is also a learning algorithm for H: We define for $z = (x, y) \in Z^m$ and $h \in H$, the **observed error**

$$\widehat{\mathrm{er}}_z(h) = \frac{1}{m} \cdot |\{i \in \{1,...,m\} \mid h(x_i) \neq y_i\}|.$$

Then every map $L : \bigcup_{m=1}^{\infty} Z^m \longrightarrow H$ satisfying

$$\widehat{\mathrm{er}}_z(L(z)) = \min_{h \in H} \widehat{\mathrm{er}}_z(h) \ (z \in Z^m, m \in \mathbb{N})$$

is called a SEM-algorithm.

This describes the principal use of NIP for binary-output networks. The formalism for learning with other neural network architectures can be found in [AB99, parts II, III]. For example [AB99, Def 16.1] gives the formal definition of learning real valued functions. The VC-dimension then is slightly altered (cf. [AB99, section 11]) and results comparable to 8.3.3.2 are available; an example is [AB99, Thm 19.1].

References

[AB99] Martin Anthony and Peter L. Bartlett. **Neural network learning: theoretical foundations**. Cambridge University Press, Cambridge, 1999. 23, 25, 26, 27

[BeFo] Alessandro Berarducci, Antongiulio Fornasiero **O-minimal cohomology: finiteness and invariance results** Preprint, May 2007: http://www.maths.manchester.ac.uk/raag/index.php?preprint=0227 13

[BCR98] Jacek Bochnak, Michel Coste, and Marie-Françoise Roy. **Real algebraic geometry**, volume 36 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1998. Translated from the 1987 French original, Revised by the authors. 14

[DvdD88] J. Denef and L. van den Dries. ***p*-adic and real subanalytic sets.** *Ann. of Math. (2)*, 128(1):79–138, 1988. 18

[Cos99] Michel Coste. **An Introduction to O-minimal Geometry** http://perso.univ-rennes1.fr/michel.coste/polyens/OMIN.pdf

[vdD98] Lou van den Dries. **Tame topology and o-minimal structures**, volume 248 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 1998. 14, 18, 21, 22

[vdDMM97] Lou van den Dries, Angus Macintyre, and David Marker. **Logarithmic-exponential power series.** *J. London Math. Soc. (2)*, 56(3):417–434, 1997. 19

[vdDMM94] Lou van den Dries, Angus Macintyre, and David Marker. **The elementary theory of restricted analytic fields with exponentiation.** *Ann. of Math. (2)*, 140(1):183–205, 1994. 19

[vdD86] Lou van den Dries. **A generalization of the Tarski-Seidenberg theorem, and some nondefinability results.** *Bull. Amer. Math. Soc. (N.S.)*, 15(2):189–193, 1986. 19

[vdD84] Lou van den Dries. **Remarks on Tarski's problem concerning** ($\mathbf{R}, +, \cdot, \exp$). In *Logic colloquium '82 (Florence, 1982)*, volume 112 of *Stud. Logic Found. Math.*, pages 97–121. North-Holland, Amsterdam, 1984. 20

[KM97] Marek Karpinski and Angus Macintyre. **Polynomial bounds for VC dimension of sigmoidal and general Pfaffian neural networks.** *J. Comput. System Sci.*, 54(1, part 2):169–176, 1997. 1st Annual Dagstuhl Seminar on Neural Computing (1994). 23

[KPS86] Julia F. Knight, Anand Pillay, and Charles Steinhorn. **Definable sets in ordered structures. II.** *Trans. Amer. Math. Soc.*, 295(2):593–605, 1986.

[MS93] Angus Macintyre and Eduardo D. Sontag. **Finiteness results for sigmoidal "neural" networks.** Annual ACM Symposium on Theory of Computing. Proceedings of the twenty-fifth annual ACM symposium on Theory of computing table of contents San Diego, California, United States; pages 325 - 334, 1993. 23

[MW96] Angus Macintyre and A. J. Wilkie. **On the decidability of the real exponential field.** In *Kreiseliana*, pages 441–467. A K Peters, Wellesley, MA, 1996. 20

[Mac00] Dugald Macpherson. **Notes on o-minimality and variations.** In *Model theory, algebra, and geometry*, volume 39 of *Math. Sci. Res. Inst. Publ.*, pages 97–130. Cambridge Univ. Press, Cambridge, 2000.

[Mar02] David Marker. **Model theory**, volume 217 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002. An introduction. 2, 3, 6

[Mil94] Chris Miller. **Exponentiation is hard to avoid.** *Proc. Amer. Math. Soc.*, 122(1):257–259, 1994. 19

[PS86] Anand Pillay and Charles Steinhorn. **Definable sets in ordered structures. I.** *Trans. Amer. Math. Soc.*, 295(2):565–592, 1986. 3, 4, 6, 14

[PS88] Anand Pillay and Charles Steinhorn. **Definable sets in ordered structures. III.** *Trans. Amer. Math. Soc.*, 309(2):469–476, 1988.

[Poi85] Bruno Poizat. **Cours de théorie des modèles**. *Bruno Poizat, Lyon, 1985.* Une introduction à la logique mathématique contemporaine. [An introduction to contemporary mathematical logic]. 22

[Spe99] Patrick Speissegger. **The Pfaffian closure of an o-minimal structure.** *J. Reine Angew. Math.*, 508:189–211, 1999. 19

[Tar51] Alfred Tarski. **A decision method for elementary algebra and geometry**. *University of California Press, Berkeley and Los Angeles, Calif.*, 1951. 2nd ed.

[Wil96] A. J. Wilkie. **Model completeness results for expansions of the ordered field of real numbers by restricted Pfaffian functions and the exponential function.** *J. Amer. Math. Soc.*, 9(4):1051–1094, 1996. 19

The University of Manchester, School of Mathematics, Oxford Road, Manchester M13 9PL, UK

Homepage: http://personalpages.manchester.ac.uk/staff/Marcus.Tressl/index.php

E-mail address: `marcus.tressl@manchester.ac.uk`