# Computing the "shape" of semialgebraic sets

Peter Bürgisser, Felipe Cucker, Josué Tonelli-Cueto

Berlin Mathematical School — CityU 香港城市大學 City University of Hong Kong — Technische Universität Berlin

## Semialgebraic sets

Semialgebraic sets are the class of geometric objects that can be described by real polynomials and inequalities.
A way of describing a semialgebraic set is to use *formulas*. Formulas are expression obtained by combining *atoms* of the form

- $(p(x) < 0)$,
- $(p(x) \geq 0)$,
- $(p(x) \leq 0)$,
- $(p(x) > 0)$ and
- $(p(x) = 0)$,
- $(p(x) \neq 0)$,

which represent the most basic semialgebraic sets; using

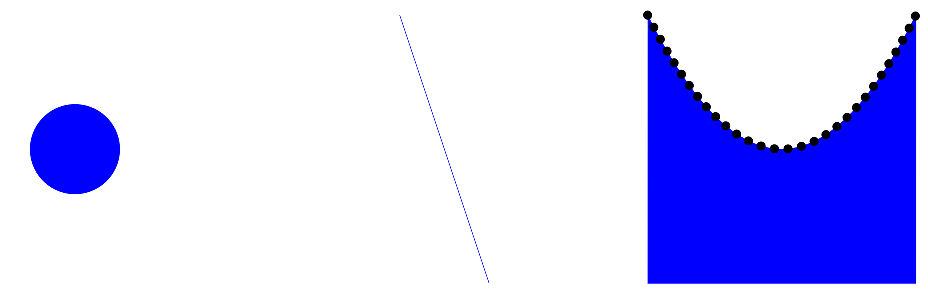
- *negations* $(\neg)$, which represent complements;
- *conjunctions* $(\wedge)$, which represent intersections; and
- *disjunctions* $(\vee)$, which represent unions.

Formulas should be seen as "recipes" telling us how to construct the described set from the most basic ones.

**Example** Consider the formula

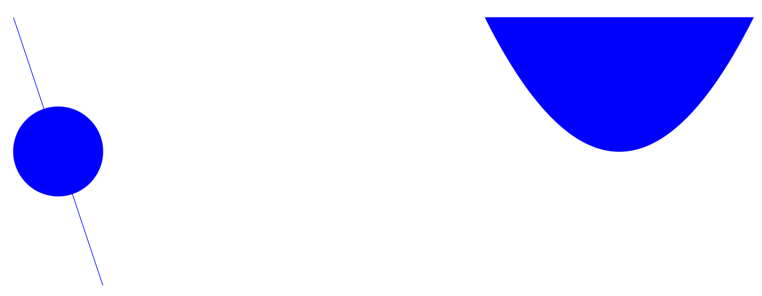
$$((x^2+y^2-1 \leq 0) \vee (3x+y=0)) \wedge (\neg(3y-x^2<0)).$$

In this formula, we have three atoms: $(x^2+y^2-1 \leq 0)$, which is the filled unit circle; $(x+3y=0)$, a line through the origin; and $(3y-x^2<0)$, the points below a parabola. These can be seen below:



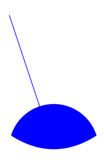
$(x^2+y^2-1 \leq 0)$ $(3x+y=0)$ $(3y-x^2<0)$

Following the formula, on the left side, $((x^2+y^2-1 \leq 0) \vee (3x+y=0))$ tells us to take the points that are given either by $(x^2+y^2-1 \leq 0)$ or $(x+3y=0)$ and, on the right side, $(\neg(y-x^2<0))$ to take the points not in $(y-x^2<0)$. These operations give the sets below:



$((x^2+y^2-1 \leq 0) \vee (3x+y=0))$ $(\neg(3y-x^2<0))$

In the last step, $((x^2+y^2-1 \leq 0) \vee (3x+y=0)) \wedge (\neg(3y-x^2<0))$ tells us to take only those points coming at the same time both from $((x^2+y^2-1 \leq 0) \vee (3x+y=0))$ and $(\neg(3y-x^2<0))$. Below an image of the final set.



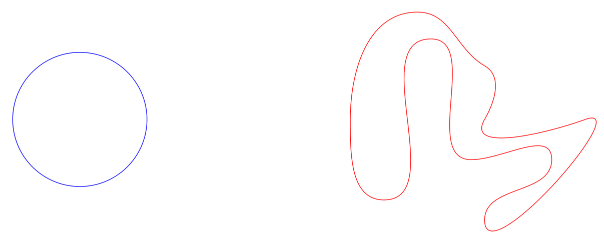
$((x^2+y^2-1 \leq 0) \vee (x+3y=0)) \wedge (\neg(3y-x^2<0))$

## Why do we care?

1. Semialgebraic sets are a large class of geometric objects preserved under many of the usual operations that one can do with sets (intersection, union, complementation, projection,...).
2. Semialgebraic sets can be used to describe:
   - Configuration space of a robotic arm.
   - Realization space of a polytope.
   - Configuration space of a molecule.
   - Regions of behavior of a real algebraic object.

## Homology and shape

The shape of a geometric object $X$ encodes so much geometric information about $X$. Because of this, describing the shape of $X$ is hard.
To deal with this overflow of information, we get rid of part of it so that one can manage what remains. In our case, we do this by focusing on the topological properties of the shape and, more concretely, the so-called homology groups, denoted $H_i(X)$.



In the above example, the red and the blue curves have very different shapes. However, the topological properties of their shapes are the same. This shows graphically how much information we lose about the shape.

**Which information about shape do the homology groups give?** In general, it is not easy to tell exactly which topological features are captured by the $H_i(X)$. However, in the case of $H_0(X)$, for example, one can see that it tells us the number of connected components, i.e. regions of $X$ in which you can "walk" from one point to the other without leaving $X$.

## Condition and weak complexity

Our algorithm is numeric. Therefore, its performance (i.e. how much precision it needs and how many arithmetic operations it uses) is dominated by a condition number.

**Ill-posed inputs** Those inputs for which the condition is infinite are called *ill-posed*. For them, an arbitrarily small perturbation of the coefficients of the polynomials changes what we want to compute (the homology groups). Using a numerical algorithm means accepting that for these inputs, there will be no numerical algorithm computing the answer.

**How do we get rid of the condition?** The condition number allows us to understand the performance of the algorithm for each input. However, can we get such an understanding that is not input-dependent?
To do so, we endow the inputs with a probability distribution and we compute probabilistic estimates of the condition. One of them, which explains performance in practice, it's the so called *weak complexity*. This means that we get a "good" bound on the performance which hold for all inputs excepts for those (called *black swans* as they are very improbable) in an exceptional set with exponentially small probability.

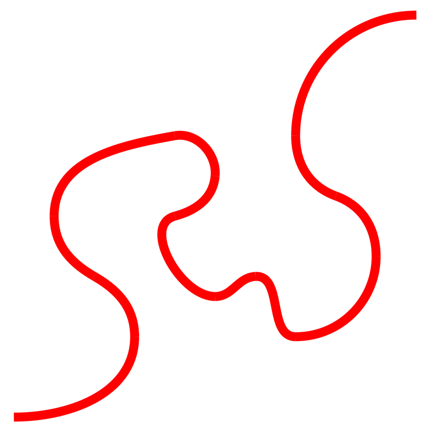
## Our current result

**Theorem.** *There is a numerical algorithm, numerically stable, with input polynomial q-tuples f and lax formulas Φ (i.e. without "<", ">", "≠" and "¬") using the polynomial in f that computes the homology groups of the semialgebraic set described by Φ in weak exponential time for f uniformly distributed on the sphere.*

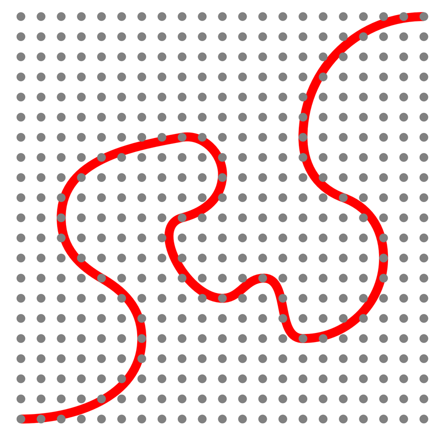
This result is good, because we expect such an algorithm (numeric or not) to take exponential time at least.

## Core of the algorithm

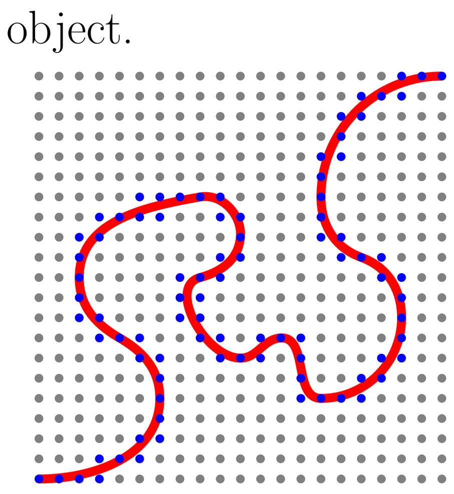
1. Take the geometric object.



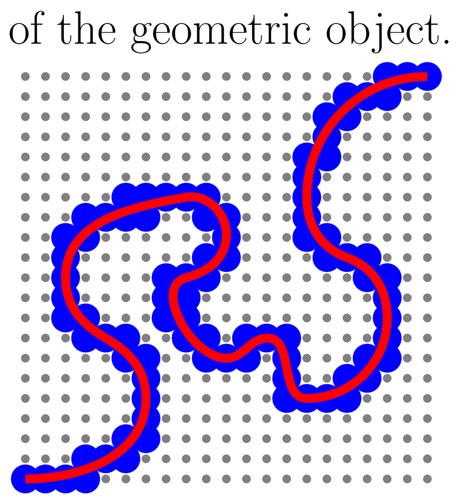
2. Construct a sufficiently fine grid. The size of the grid's mesh is controlled by the condition number.



3. Sample from the grid a cloud of points near enough the geometric object.



4. A thickening of the sampled cloud of points will capture the shape of the geometric object.



5. Standard techniques of algebraic topology allow the computation of the homology groups.

## Open challenges

1. Can we extend the algorithm to work on general formulas and not only on lax formulas?
2. Can we extend the probabilistic analysis of the condition number for more general probability distributions of $f$?
3. Implement the algorithm and evaluate its performance in practice.

## References for further reading

F. Cucker, T. Krick, M. Shub. Computing the Homology of Real Projective Sets. *Found Comput Math*, 2017.
P. Bürgisser, F. Cucker, P. Lairez. Computing the homology of basic semialgebraic sets in weak exponential time. To appear in *Journal of the ACM*, 2018.
P. Bürgisser, F. Cucker, J. Tonelli-Cueto. Computing the Homology of Semialgebraic Sets I: The Lax Case. To appear.

The research has been funded by Felipe Cucker's Einstein Visiting Fellowship "Complexity and accuracy of numerical algorithms in algebra and geometry" funded by the EINSTEIN Foundation.de Einstein Stiftung Berlin Einstein Foundation Berlin

LaTeX TikZposter