The Complexity of the Hausdorff Distance

Paul Jungeblut[1], Linda Kleist[2], and Tillmann Miltzow[3]

1 **Karlsruhe Institute of Technology, Germany**
`paul.jungeblut@kit.edu`

2 **Technische Universität Braunschweig, Germany**
`kleist@ibr.cs.tu-bs.de`

3 **Utrecht University, The Netherlands**
`t.miltzow@uu.nl`

Abstract

We determine the computational complexity of computing the Hausdorff distance. Specifically, we show that the decision problem of whether the Hausdorff distance of two semi-algebraic sets is bounded by a given threshold is complete for the complexity class $\forall\exists_{<}\mathbb{R}$. This implies that the problem is NP-, co-NP-, $\exists\mathbb{R}$- and $\forall\mathbb{R}$-hard.

Related Version Full version available on arXiv: 2112.04343.

1 Introduction

The question of 'how similar are two given objects' occurs in numerous settings. A typical tool to quantify their similarity is the Hausdorff distance. Two sets have a small Hausdorff distance if every point of one set is close to some point of the other set and vice versa. The Hausdorff distance appears in many branches of science. To illustrate the range of use cases, we consider two examples. For illustrations consider Figure 1. In mathematics, the Hausdorff distance provides a metric on sets and henceforth also a topology. This topology can be used to discuss continuous transformations of one set to another [7]. In computer vision and geographical information science, the Hausdorff distance is used to measure the similarity between spacial objects [17, 18], for example the quality of quadrangulations of complex 3D models [20]. In this paper, we study the computational complexity of the Hausdorff distance from a theoretical perspective.

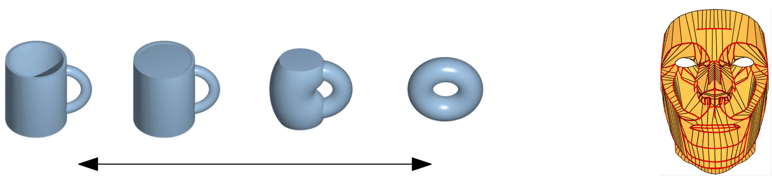

Figure 1 Left: Continuous deformation of a cup into a doughnut [10]. Right: Quadrangulation of a smooth surface used for rendering [20].

Definition. The *directed Hausdorff distance* between a non-empty set $A \subseteq \mathbb{R}^n$ and a non-empty set $B \subseteq \mathbb{R}^n$ is defined as

$$\vec{d}_{\mathrm{H}}(A, B) := \sup_{a \in A} \inf_{b \in B} \|a - b\|.$$

The directed Hausdorff distance between A and B can be interpreted as the smallest value $\varepsilon \geq 0$ such that the (closed) ε-neighborhood of B contains A. Hence, it nicely captures the intuition of how much B has to be blown up to contain A. Note that $\vec{d}_{\mathrm{H}}(A, B)$ and $\vec{d}_{\mathrm{H}}(B, A)$

38th European Workshop on Computational Geometry, Perugia, Italy, March 14–16, 2022.
This is an extended abstract of a presentation given at EuroCG'22. It has been made public for the benefit of the community and should be considered a preprint rather than a formally reviewed paper. Thus, this work is expected to appear eventually in more final form at a conference with formal proceedings and/or in a journal.

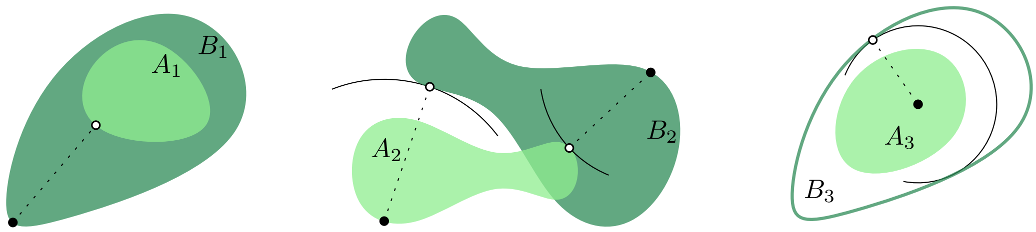

Figure 2 How similar are these sets?

must not be equal. For an example, consider Fig. 2; while $A_1 \subset B_1$ and thus $\vec{d}_{\mathrm{H}}(A_1, B_1) = 0$, it holds that $\vec{d}_{\mathrm{H}}(B_1, A_1) > 0$. The (undirected) *Hausdorff distance* is symmetric and defined as $d_{\mathrm{H}}(A, B) := \max\{\vec{d}_{\mathrm{H}}(A, B), \vec{d}_{\mathrm{H}}(B, A)\}$. In this paper, we investigate the *computational complexity* of deciding whether the Hausdorff distance of two sets is at most a given threshold.

Semi-Algebraic Sets. The algorithmic complexity of the Hausdorff distance clearly depends on the type of the considered sets. If we are given the sets in a way that we cannot even decide if they are empty, it seems near impossible to compute their Hausdorff distance. However, if the sets consists of finitely many points, their Hausdorff distance can be easily computed by checking all pairs of points. In practice, we are often somewhere between those two extreme situations. For instance, the sets could be a collection of disks in the plane or cubic splines, describing a surface in three dimensions, see also Fig. 3.

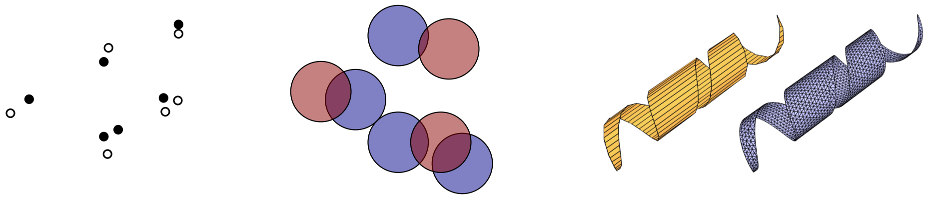

Figure 3 The Hausdorff distance can appear in simpler or more complicated settings. Left: Two finite point sets (black and white) in the plane. Middle: Two sets of blue and red disks in the plane. Right: Two surfaces in 3-space with different meshes, image taken from [20].

In this paper, we focus on semi-algebraic sets, i.e., sets that can be described by polynomial inequalities. Formally, a semi-algebraic set is the finite union of basic semi-algebraic sets. A basic semi-algebraic set S is specified by two families of polynomials $\mathcal{P}$ and $\mathcal{Q}$ such that

$$S = \{x \in \mathbb{R}^n \mid \bigwedge_{P \in \mathcal{P}} P(x) \leq 0 \wedge \bigwedge_{Q \in \mathcal{Q}} Q(x) < 0\}.$$

Semi-algebraic sets cover clearly the vast majority of practical cases and finding efficient algorithms for this problem would be a tremendous contribution. Simultaneously, when considering smooth sets, one is quickly in the situation that one needs to deal with polynomials anyway. So the step to general semi-algebraic sets is not a very big one.

General Decision Algorithm. We consider a situation where we are given two semi-algebraic sets A and B as well as a threshold t; for simplicity, we assume here (only in this paragraph) that A and B are closed. The statement $\vec{d}_{\mathrm{H}}(A, B) \leq t$ can be encoded into a logical sentence

$$\forall a \in A \, . \, \exists b \in B : \|a - b\|^2 \leq t^2,$$

where $\|x\|$ denotes the Euclidean norm of the vector x. We can decide the truth of this sentence by employing sophisticated algorithms from real algebraic geometry that can deal with two blocks of quantifiers [5, Chapter 14]. These algorithms are impractical for all non-trivial instances. Our main result roughly states that in general there is little hope for an improvement. To state this formally, we continue by defining suitable complexity classes.

Algorithmic Complexity. Let φ be a *quantifier-free formula in the first-order theory of the reals*, i.e., a formula formed over the alphabet $\Sigma = \{0, 1, +, \cdot, =, \leq, <, \vee, \wedge, \neg\}$ together with symbols for the variables. The Universal Existential Theory of the Reals (UETR) asks to decide the truth value of a sentence

$$\Phi := \forall X \in \mathbb{R}^n \,.\, \exists Y \in \mathbb{R}^m : \varphi(X, Y).$$

An instance of UETR belongs to Strict-UETR if the corresponding formula φ is over the alphabet $\Sigma = \{0, 1, +, \cdot, <, \vee, \wedge\}$, i.e., if every atom is a strict inequality and there are no negations. The complexity classes $\forall\exists\mathbb{R}$ and $\forall\exists_<\mathbb{R}$ contain all decision problems for which there exists a polynomial-time many-one reduction to UETR and Strict-UETR, respectively. We propose to pronounce the complexity classe $\forall\exists\mathbb{R}$ as 'UER' or 'forall exists R' and $\forall\exists_<\mathbb{R}$ as 'Strict-UER' or 'strict forall exists R'. To the best of our knowledge, $\forall\exists\mathbb{R}$ was first introduced by Bürgisser and Cucker [9, Section 9] under the name $\mathsf{BP}^0(\forall\exists)$ (in the constant-free Blum-Shub-Smale-model [6]). The notation $\forall\exists\mathbb{R}$ arised later in [13] extending the notation from Schaefer and Števankovič [19]. The sister class co-$\forall\exists_<\mathbb{R} = \exists\forall_\leq\mathbb{R}$ was first studied by D'Costa, Lefaucheux, Neumann, Ouaknine and Worrel [12].

Problem and Results. We now have all ingredients to state our problem and main results. Let $\Phi_A(X)$ and $\Phi_B(X)$ be two quantifier-free formulas defining the semi-algebraic sets $A = \{x \in \mathbb{R}^n \mid \Phi_A(x)\}$ and $B = \{x \in \mathbb{R}^n \mid \Phi_B(x)\}$, and let $t \in \mathbb{Q}$ be a rational number. The Hausdorff problem asks whether $d_{\mathrm{H}}(A, B) \leq t$. Here the dimension n of the ambient space of A and B is part of the input (there is a polynomial-time algorithm for every fixed n, see the related work in Section 2). The computational complexity of this problem was posed as an open question by Dobbins, Kleist, Miltzow and Rzążewski [13].

▶ **Theorem 1.1.** *The* Hausdorff *problem is* $\forall\exists_<\mathbb{R}$*-complete.*

Note that prior to our result, it was not even known if computing the Hausdorff distance was NP-hard. As $\forall\exists_<\mathbb{R}$ contains, NP, co-NP, $\exists\mathbb{R}$ and $\forall\mathbb{R}$, we also get hardness for all of these complexity classes. In the proof of $\forall\exists_<\mathbb{R}$-hardness for Theorem 1.1, we create instances with some additional properties. In particular, we can guarantee a gap, i.e., the Hausdorff distance is either below the threshold t or at least $t \cdot 2^{2^{\Omega(d)}}$, where d denotes the number of variables of Φ_A and Φ_B. Thus our result also rules out approximation algorithms.

▶ **Corollary 1.2.** *Let A and B be two semi-algebraic sets in $\mathbb{R}^d$ and $f(d) = 2^{2^{o(d)}}$. Then there is no polynomial time $f(d)$-approximation algorithm to compute $d_{\mathrm{H}}(A, B)$, unless* $\mathsf{P} = \forall\exists_<\mathbb{R}$.

2 Related Work

This section reviews previous work concerning two directions. First, we discuss the complexity of computing the Hausdorff distance for special sets. Afterwards, we investigate previous work on the complexity class $\forall\exists\mathbb{R}$.

Computing the Hausdorff Distance. The notion of the Hausdorff distance was introduced by Felix Hausdorff in 1914 [16]. Most of the early works focused on the Hausdorff distance for finite point sets. For a set of n points and a set of m points in any fixed dimension, the Hausdorff distance can be easily computed by checking all pairs, i.e., in time $O(mn)$. In the plane, this can be improved to $O((n+m)\log(m+n))$ by using Voronoi diagrams [1]. In fact, this method can be extended to sets consisting of pairwise non-crossing line segments in the plane, e.g., simple polygons and polygonal chains fulfill this property. If the polygons are additionally convex, their Hausdorff distance can even be computed in linear time [4].

More generally, the Hausdorff distance can be computed in polynomial time whenever the two sets can be described by a simplicial complex of fixed dimension. Based on the PhD thesis of Godau [15], Alt et al. [2, Theorem 3.3] show how to compute the directed Hausdorff distance between two sets in $\mathbb{R}^d$ consisting of n and m k-dimensional simplices in time $O(nm^{k+2})$ (assuming d is constant). Using a Las Vegas algorithm for computing the vertices of the lower envelope, similar ideas yield an approach with randomized expected time in $O(nm^{k+\varepsilon})$ for $k > 1$ and every $\varepsilon > 0$ [2, Theorem 3.4]. They additionally present algorithms with better randomized expected running times for sets of triangles in $\mathbb{R}^3$ and point sets in $\mathbb{R}^d$.

Given two semi-algebraic sets $A, B \subseteq \mathbb{R}^n$, the HAUSDORFF problem can be encoded as a sentence of the form $\forall X \in \mathbb{R}^n . \exists Y \in \mathbb{R}^n : \varphi(X,Y)$ with $\Theta(n)$ variables, where φ is quantifier-free. Such a sentence can be decided in time roughly equal to $(sd)^{O(n^2)}$ [5, Theorem 14.14] where d denotes the maximum degree of any polynomial in φ and s denotes the number of atoms.

In other contexts the two sets are allowed to undergo certain transformations (e.g. translations) such that the Hausdorff distance is minimized [8]. See Alt [3] for a survey.

Universal Existential Theory of the Reals. As mentioned above, the complexity class $\forall\exists\mathbb{R}$ was first studied by Bürgisser and Cucker who prove complexity results for many decision problems involving circuits [9]. Dobbins, Kleist, Miltzow, and Rzążewski [14, 13] consider $\forall\exists\mathbb{R}$ in the context of area-universality of graphs. A plane graph is *area-universal* if for every assignment of reals to the inner faces of a plane graph, there exists a straight-line drawing such that the area of each inner face equals the assigned number. Dobbins et al. conjecture that the decision problem whether a given plane graph is area-universal is complete for $\forall\exists\mathbb{R}$. They support this conjecture by proving hardness for several related notions [13]. Additionally, for future research directions, they present a number of candidates for potentially $\forall\exists\mathbb{R}$-hard problems. Among them, they stated a question motivating this paper as an open problem, namely whether the HAUSDORFF problem is $\forall\exists\mathbb{R}$-complete. The other candidates exhibit intrinsic connections to imprecision, robustness and extendability.

The sister class $\exists\forall\mathbb{R}$ was recently investigated by D'Costa et al. [12]. They show that it is $\exists\forall_{\leq}\mathbb{R}$-complete to decide for a given rational matrix A and a compact semi-algebraic set $K \subseteq R^n$, whether there exists a starting point $x \in K$ such that $x_n := A^n x$ is contained in K for all $n \in N$.

3 Techniques and Proof Overview

In this section, we present the general idea behind the hardness reduction for the HAUSDORFF problem. The goal is to convey the intuition and to motivate the technical intermediate steps needed. The sketched reduction is oversimplified and thus neither in polynomial time nor fully correct. We point out both of these issues and give first ideas on how to solve them.

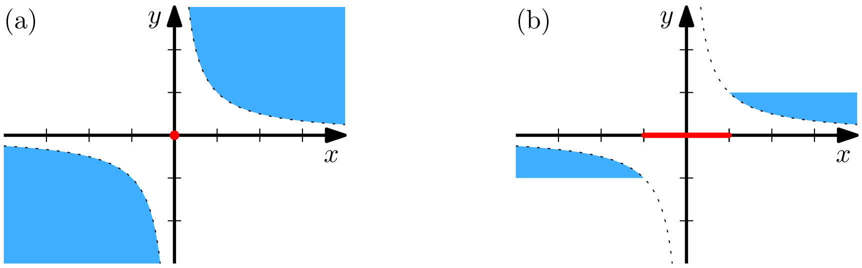

Figure 4 Consider the formula $\forall X \in \mathbb{R} . \exists Y \in \mathbb{R} : XY > 1$. (a) Each point $(x, y) \in \mathbb{R}^2$ in the blue open region satisfies $xy > 1$. Only for $x = 0$ (in red) no suitable $y \in \mathbb{R}$ exists. (b) Restricting the range of Y to $[-1, 1]$, then for all $x \in [-1, 1]$ (in red) no y with $xy > 1$ exists.

Let $\Phi := \forall X \in \mathbb{R}^n . \exists Y \in \mathbb{R}^m : \varphi(X, Y)$ be a Strict-UETR instance. We define two sets

$$A := \{x \in \mathbb{R}^n \mid \exists Y \in \mathbb{R}^m : \varphi(x, Y)\} \quad \text{and}$$
$$B := \mathbb{R}^n$$

and ask whether $d_{\mathrm{H}}(A, B) = 0$. If Φ is true, then $A = \mathbb{R}^n$ and we have $d_{\mathrm{H}}(A, B) = 0$ because both sets are equal. Otherwise, if Φ is false, then there exists some $x \in \mathbb{R}^n$ for which there is no $y \in \mathbb{R}^m$ satisfying $\varphi(x, y)$ and we conclude that $A \neq \mathbb{R}^n$. In general we call the set of all $x \in \mathbb{R}^n$ for which there is no $y \in \mathbb{R}^m$ satisfying $\varphi(x, y)$ the *counterexamples* $\perp(\Phi)$ of Φ. One might hope that $\perp(\Phi) \neq \emptyset$ is enough to obtain $d_{\mathrm{H}}(A, B) > 0$, but this is not the case. To this end, consider the formula $\Psi := \forall X \in \mathbb{R} . \exists Y \in \mathbb{R} : XY > 0$, which is false. The set $\perp(\Psi) = \{0\}$ contains only a single element, so we have $A = \mathbb{R} \setminus \{0\}$ and $B = \mathbb{R}$. However, their Hausdorff distance also evaluates to $d_{\mathrm{H}}(A, B) = 0$. We conclude that above reduction does not (yet completely) work, because it maps a yes- and a no-instances of Strict-UETR to a yes-instance of Hausdorff.

We solve this issue by blowing up the set of counterexamples. Specifically, Theorem 12 (in the full version) establishes a polynomial time algorithm to transform a Strict-UETR instance Φ into an equivalent formula Φ' such that the set of counterexamples is either empty (if Φ' is true) or contains an open ball of positive radius (if Φ' is false). The radius of the ball serves as a lower bound on the Hausdorff distance $d_{\mathrm{H}}(A, B)$. Thus a reduction starting with Φ' is correct. As a key tool for this step, we restrict the variable ranges from $\mathbb{R}^n$ and $\mathbb{R}^m$ to small and compact intervals. Fig. 4 presents an an example on how such a range restriction may enlarge the set of counterexamples from a single point to an interval. We think that the special property of blown up counterexamples can prove useful in future reductions to show $\forall\exists_{<}\mathbb{R}$-hardness of other problems because it makes handling the no-instances easier.

A further challenge is given by the definition of the sets A and B. While the description complexity of B depends only on n, the definition of A contains an existential quantifier. This is troublesome because our definition of the Hausdorff problem requires quantifier-free formulas as its input, and in general there is no equivalent quantifier-free formula of polynomial length which describes the set A [11]. We overcome this issue by taking the existentially quantified variables as additional dimensions into account. We cannot know their precise values for each possible choice of the universally quantified variables. But by scaling them to a tiny range, their influence on the Hausdorff distance becomes negligible. Therefore instead of the above we work with sets similar to

$$A := \{(x, y) \in [-1, 1]^n \times [-\varepsilon, \varepsilon]^m \mid \varphi(x, y)\} \quad \text{and}$$
$$B := [-1, 1]^n \times \{0\}^m$$

for some tiny value ε depending on the radius r (of the ball contained in the counterexamples). This definition of A and B introduces the new issue that even if Φ is true, the Hausdorff distance $d_{\mathrm{H}}(A,B)$ might be strictly positive. However, we manage to identify a threshold t, such that $d_{\mathrm{H}}(A,B) \leq t$ if and only if Φ is true. This completes the proof of $\forall\exists_{<}\mathbb{R}$-hardness.

$\forall\exists_{<}\mathbb{R}$-membership is shown by formulating the Hausdorff problems as an equivalent Strict-UETR instance (see Section 6 of the full version).

References

1. Helmut Alt, Bernd Behrends, and Johannes Blömer. Approximate Matching of Polygonal Shapes. *Annals of Mathematics and Artificial Intelligence*, 13(3):251–265, 1995. doi:10.1007/BF01530830.
2. Helmut Alt, Peter Braß, Michael Godau, Christian Knauer, and Carola Wenk. Computing the Hausdorff Distance of Geometric Patterns and Shapes. In Boris Aronov, Saugata Basu, János Pach, and Micha Sharir, editors, *Discrete and Computational Geometry: The Goodman-Pollack Festschrift*, volume 25 of *Algorithms and Combinatorics*, pages 65–76. Springer, 2003. doi:10.1007/978-3-642-55566-4_4.
3. Helmut Alt and Leonidas J. Guibas. Discrete Geometric Shapes: Matching, Interpolation, and Approximation. In Jörg-Rüdiger Sack and Jorge Urrutia, editors, *Handbook of Computational Geometry*, pages 121–153. Elsevier, 2000. doi:B978-044482537-7/50004-8.
4. Mikhail J. Atallah. A Linear Time Algorithm for the Hausdorff Distance Between Convex Polygons. *Information Processing Letters*, 17(4):207–209, 1983. doi:10.1016/0020-0190(83)90042-X.
5. Sauguta Basu, Richard Pollack, and Marie-Françoise Roy. *Algorithms in Real Algebraic Geometry*, volume 10 of *Algorithms and Computation in Mathematics*. Springer, 2006. doi:10.1007/3-540-33099-2.
6. Lenore Blum, Mike Shub, and Steve Smale. On a Theory of Computation and Complexity over the Real Numbers: NP-Completeness, Recursive Functions and Universal Machines. *Bulletin of the American Mathematical Society*, 21:1–46, 1989. doi:10.1090/S0273-0979-1989-15750-9.
7. Glen E. Bredon. *Topology and Geometry*, volume 139 of *Graduate Texts in Mathematics*. Springer Science & Business Media, 1st edition, 2013. doi:10.1007/978-1-4757-6848-0.
8. Karl Bringmann and André Nusser. Translating Hausdorff Is Hard: Fine-Grained Lower Bounds for Hausdorff Distance Under Translation. arXiv preprint, 2021. arXiv:2101.07696.
9. Peter Bürgisser and Felipe Cucker. Exotic Quantifiers, Complexity Classes, and Complete Problems. *Foundations of Computational Mathematics*, 9:135–170, April 2009. doi:10.1007/s10208-007-9006-9.
10. Wiki Community. Homotopy. accessed 2021 November. URL: https://en.wikipedia.org/wiki/Homotopy.
11. James H. Davenport and Joos Heintz. Real Quantifier Elimination is Doubly Exponential. *Journal of Symbolic Computation*, 5(1–2):29–35, 1988. doi:10.1016/S0747-7171(88)80004-X.
12. Julian D'Costa, Engel Lefaucheux, Eike Neumann, Joël Ouaknine, and James Worrel. On the Complexity of the Escape Problem for Linear Dynamical Systems over Compact Semialgebraic Sets. In Filippo Bonchi and Simon J. Puglisi, editors, *International Symposium on Mathematical Foundations of Computer Science (MFCS)*, volume 202 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 33:1–33:21, 2021. doi:10.4230/LIPIcs.MFCS.2021.33.

13 Michael G. Dobbins, Linda Kleist, Tillmann Miltzow, and Paweł Rzążewski. $\forall\exists\mathbb{R}$-Completeness and Area-Universality. In Andreas Brandstädt, Ekkehard Köhler, and Klaus Meer, editors, *Graph-Theoretic Concepts in Computer Science (WG)*, volume 11159 of *Lecture Notes in Computer Science*, pages 164–175. Springer, 2018. doi:10.1007/978-3-030-00256-5_14.

14 Michael G. Dobbins, Linda Kleist, Tillmann Miltzow, and Paweł Rzążewski. Completeness for the Complexity Class $\forall\exists\mathbb{R}$ and Area-Universality. ArXiv preprint, 2021. arXiv:1712.05142v3.

15 Michael Godau. *On the complexity of measuring the similarity between geometric objects in higher dimensions*. PhD thesis, Freie Universtät Berlin, 1999. doi:10.17169/refubium-7780.

16 Felix Hausdorff. *Grundzüge der Mengenlehre*. Von Veit & Company, 1914.

17 Deng Min, Li Zhilin, and Chen Xiaoyong. Extended Hausdorff distance for spatial objects in GIS. *International Journal of Geographical Information Science*, 21(4):459–475, 2007. doi:10.1080/13658810601073315.

18 William Rucklidge. *Efficient Visual Recognition Using the Hausdorff Distance*, volume 1173 of *Lecture Notes in Computer Science*. Springer, 1996. doi:10.1007/BFb0015091.

19 Marcus Schaefer and Daniel Štefankovič. Fixed Points, Nash Equilibria, and the Existential Theory of the Reals. *Theory of Computing Systems*, 60:172–193, 2017. doi:10.1007/s00224-015-9662-0.

20 Floor Verhoeven, Amir Vaxman, Tim Hoffmann, and Olga Sorkine-Hornung. Dev2PQ: Planar Quadrilateral Strip Remeshing of Developable Surfaces. arXiv preprint, 2021. arXiv:2103.00239.