Comparing continuous and discrete versions of Hilbert's thirteenth problem

Lynnelle Ye

1 Introduction

Hilbert's thirteenth problem is the following conjecture: "a solution to the equation $t^7+xt^3+yt^2+zt+1=0$ cannot be represented as a superposition of continuous functions of two variables (out of the three variables x, y, z)". Here a superposition of functions basically means a circuit whose gates compute those functions. (The reason for this particular formulation is that the solution to a polynomial of degree $n \leq 4$ can be written in terms of radicals and arithmetic operations on the coefficients, so is clearly a superposition of continuous functions of two variables; on the other hand, for $n > 4$ there is an arithmetic-operations-plus-radicals transformation called the Tschirnhausen transformation which reduces any degree-n polynomial to one in which the terms $t^{n-1}, t^{n-2}, t^{n-3}$ are all 0 and the constant coefficient is 1, so a solution to a degree-n polynomial is a superposition of continuous functions of $n-4$ variables of the coefficients, and $7-4=3$.) For more background, see [4].

This conjecture was disproven by Kolmogorov and Arnol'd in 1957. In fact, they proved the much stronger theorem that any continuous function of s variables can be written as a superposition of functions of one variable together with the binary function of addition, by explicit construction. There have since been various improvements to the continuity condition on some of the one-variable functions used in this representation, as well as simpler nonconstructive proofs. Working in the other direction, Vitushkin has shown that Kolmogorov's theorem fails if the representing functions are required to be continuously differentiable, or to satisfy various other interesting smoothness conditions.

Recently, there has been significant interest in versions of Hilbert's thirteenth problem in discrete settings. We focus on work of Hansen, Lachish, and Miltersen on word circuits. They propose the natural substitution of functions from $(\{0,1\}^w)^2$ to $\{0,1\}^w$ in place of functions on the unit square, and polynomial-time computability as an analogue for continuity. They ask whether any polynomial-time computable function $f : (\{0,1\}^w)^3 \to \{0,1\}^w$ can be computed by a circuit whose gates compute functions $g : (\{0,1\}^w)^2 \to \{0,1\}^w$, conjecturing a negative answer, and provide a few bounds in both directions.

In this paper, we present a summary of the work of Kolmogorov, Arnol'd, Vitushkin, and Lorentz on both sides of the continuous setting of Hilbert's thirteenth problem, as well as that of Hansen, Lachish, and Miltersen on the discrete setting. We then consider whether and how it might be reasonable to deepen the analogy between the continuous and discrete problems. We will argue that it may be useful to modify Hansen, Lachish, and Miltersen's analogy in a way which provides a closer match between both results and methods.

2 Kolmogorov's theorem

In this section, we state and sketch a proof of Kolmogorov's original theorem disproving Hilbert's conjecture. Let $I = [0, 1]$ and $S = [0, 1]^s$ with coordinates $x_1, \ldots, x_s$.

Theorem 2.1 (Kolmogorov's theorem)**.** *There exist constants* $0 < \lambda_1, \ldots, \lambda_s \leq 1$ *and strictly increasing functions* $\varphi_0, \ldots, \varphi_{2s} : I \to I$ *which are Lipschitz of positive exponent (in particular continuous), so that for each continuous function* f *on* S*, there is a continuous function* g *on* $[0, s]$ *such that*

$$f(x_1, \cdots, x_s) = \sum_{q=0}^{2s} g(\lambda_1 \varphi_q(x_1) + \cdots + \lambda_s \varphi_q(x_s)).$$

Observe that if we completely drop all continuity conditions, Kolmogorov's theorem is trivial: for $s = 2$, consider the standard injection $\varphi : [0, 1]^2 \to [0, 1]$, where if x has binary expansion $0.x_1x_2 \cdots$ and y has binary expansion $0.y_1y_2 \cdots$, we define $\varphi(x, y) = 0.x_1y_1x_2y_2 \cdots$ (if x or y has multiple binary expansions, take the terminating one). Note that this can be written $\varphi(x, y) = \varphi_1(x) + \varphi_2(y)$, where φ_1 takes $0.x_1x_2 \cdots$ to $0.x_10x_20 \cdots$ and φ_2 takes $0.y_1y_2 \cdots$ to $0.0y_10y_2 \cdots$. Then for any function $f(x, y)$ on $[0, 1]^2$, we can write $f(x, y) = g(\varphi_1(x) + \varphi_2(y))$ where $g(a) = f(\varphi^{-1}(a))$ for $a \in \varphi([0, 1]^2)$. It is clear that this can be easily extended to any number of variables, even a countable number.

If we require g, φ_q to be continuous, the constructions become more complicated, but the fundamental idea is similar: choose $\varphi_q : [0, 1]^2 \to [0, 2]$ to be "sort of" injective on the majority of $[0, 1]^2$, then choose g to approximate "$f \circ \varphi_q^{-1}$" on those intervals. We will sketch the case $s = 2$. The rest of this section follows Lorentz's exposition of Kolmogorov's proof in [2]. We set $\lambda_1 = 1, \lambda_2 = \lambda$, so that we are looking for a representation of the form

$$f(x, y) = \sum_{q=0}^{4} g(\varphi_q(x) + \lambda \varphi_q(y)).$$

Assume λ and φ_q defined, and for any $g \in C([0, 2])$, let $h_g(x, y) = \sum_{q=0}^{4} g(\varphi_q(x) + \lambda \varphi_q(y))$. We will construct φ_q so that the following lemma holds.

Lemma 2.2. *Fix* $2/3 < \theta < 1$. *For each* $f \in C(S)$ *there is* $g \in C([0, 2])$ *such that* $\|f - h_g\| \leq \theta \|f\|$ *and* $\|g\| \leq \frac{1}{3}\|f\|$ *(where* $\|\cdot\|$ *is the* L^∞ *norm).*

That is, we can find h_g of the desired form very roughly approximating f. It turns out that this is enough.

Proof of Theorem 2.1 from Lemma 2.2. We can construct g_1 such that $\|f - h_{g_1}\| \leq \theta \|f\|$ and $\|g_1\| \leq \frac{1}{3}\|f\|$, followed by g_2 such that $\|f - h_{g_1} - h_{g_2}\| \leq \theta \|f - h_{g_1}\| \leq \theta^2 \|f\|$ and $\|g_2\| \leq \frac{1}{3}\|f - h_{g_1}\| \leq \frac{1}{3}\theta \|f\|$, and so on, producing g_r such that

$$\left\| f - \sum_{a=1}^{r} h_{g_a} \right\| \leq \theta^r \|f\|, \qquad \|g_r\| \leq \frac{1}{3}\theta^{r-1}\|f\|$$

for all r. Then $g = \sum_{r=1}^{\infty} g_r$ and $h_g = \sum_{r=1}^{\infty} h_{g_r}$ converge uniformly and satisfy $h_g(x, y) = \sum_{q=0}^{4} g(\varphi_q(x) + \lambda \varphi_q(y))$, and the above inequalities imply $f = h_g$, giving Kolmogorov's theorem. □

Now we obtain appropriate choices of φ_q so that Lemma 2.2 holds.

Proof sketch of Lemma 2.2. As stated earlier, the goal is to make φ_q "sort of" injective on large portions of $[0,1]^2$, by which we mean not actually injective but at least able to distinguish between values of (x,y) that are not too close together. For $i = 0, \ldots, 10^{k-1}$ and $k = 1, 2, \ldots$, let

$$I_{0i}^k = \left[\frac{i}{10^{k-1}} + \frac{1}{10^k}, \frac{i}{10^{k-1}} + \frac{9}{10^k}\right],$$

and for $q = 1, 2, 3, 4$,

$$I_{qi}^k = I_{0i}^k - \frac{2q}{10^k},$$

that is, the translation of I_{0i}^k to the left by $2q/10^k$, where I_{qi}^k is replaced by $I_{qi}^k \cap I$ if any part of it falls outside I. Note that for each fixed k and q, the intervals I_{qi}^k cover 4/5ths of I, and so for each fixed k, each $x \in I$ falls in at least 4 out of 5 of the intervals I_{qi}^k. Let $S_{qij}^k = I_{qi}^k \times I_{qj}^k$. Then for each fixed k and any $(x,y) \in I^2$, x is missed by at most one I_{qi}^k and y by at most one I_{qj}^k, so (x,y) falls in at least 3 out of 5 of the squares S_{qij}^k.

Now we let λ be any fixed irrational number in $(0,1)$, and choose φ_q strictly increasing and Lipschitz with exponent $\log_{10} 2$ so that $\psi_q(x,y) = \varphi_q(x) + \lambda\varphi_q(y)$ satisfies the following property: for each fixed k, the intervals $\Delta_{qij}^k = \psi_q(S_{qij}^k)$ are all disjoint as q, i, j vary. We will not discuss exactly how to do this, except to say that it consists of successively choosing the values of φ_q on the endpoints of I_{qi}^k as k increases while enforcing the listed properties.

Having fixed such φ_q, we prove Lemma 2.2. Let k be sufficiently large so that the oscillation of f on each S_{qij}^k is less than $(\theta - 2/3)\|f\|$. Let f_{qij}^k be the value of f at the center of S_{qij}^k. Let g on Δ_{qij}^k be the constant $\frac{1}{3}f_{qij}^k$, and extend g linearly in the gaps between the Δ_{qij}^k; this clearly gives $\|g\| \leq \frac{1}{3}\|f\|$ as required. Then for any (x,y), three of the values $g(\varphi_q(x) + \lambda\varphi_q(y)) = g(\psi_q(x,y))$ are exactly $\frac{1}{3}f_{qij}^k$, which is off from $\frac{1}{3}f(x,y)$ by no more than $\frac{1}{3}(\theta - 2/3)\|f\|$. The other two values of $g(\psi_q(x,y))$ each have magnitude at most $\frac{1}{3}\|f\|$. Therefore

$$|f(x,y) - h_g(x,y)| \leq 3 \cdot \frac{1}{3}(\theta - 2/3)\|f\| + \frac{2}{3}\|f\| = \theta\|f\|$$

for all (x,y), which is what we wanted. $\square$

3 Vitushkin's work

In this section, we state and sketch a proof of a theorem in the vein of Vitushkin's work on conditions under which Hilbert's conjecture holds. It goes as follows.

Theorem 3.1. *Let G be the closure of an open connected bounded set in $\mathbb{R}^n$, and $C(G)$ the set of continuous functions on G. Let F be the set of functions $f \in G$ such that for some continuous functions $f_1, \ldots, f_m$ on k variables, we have*

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{m} p_i(x_1, \ldots, x_n) f_i(q_{i1}(x_1, \ldots, x_n), \ldots, q_{ik}(x_1, \ldots, x_n))$$

where m and $k < n$ are fixed integers, the p_i are fixed continuous functions on G, and the q_{ij} are fixed continuously differentiable functions on G. Then F is a "set of first category" in $C(G)$, meaning a countable union of nowhere dense sets.

By the Baire Category Theorem, this implies that the complement of F is dense (in particular nonempty). Note that this is basically a counting argument: we conclude that there exists f which cannot be decomposed in the proposed manner from showing that the set of f which can be so decomposed is small. It is possible to choose f to be continuously differentiable, but we will not show this.

We proceed to handwave through Lorentz's proof of this theorem in [3], starting with a bunch of definitions.

Definition 3.1. Let $\|f\|_\delta = \sup_{Q_\delta \subset G} \frac{1}{\delta^n} \left| \int_{Q_\delta} f(x)dx \right|$. This is a seminorm. Note that δ^n is the volume of Q_δ.

Definition 3.2. For a subset F of $C(G)$, the entropy $H_\epsilon^\delta(F)$ is $\log N_\epsilon^\delta(F)$ where $N_\epsilon^\delta(F)$ is the minimum number of balls of radius ϵ (in the $\|\cdot\|_\delta$ seminorm) in a set of such balls which covers F.

Definition 3.3. For a subset F of $C(G)$, the capacity $C_\epsilon^\delta(F)$ is $\log M_\epsilon^\delta(F)$ where $M_\epsilon^\delta(F)$ is the maximum number of functions $f_1, \ldots, f_{M_\epsilon^\delta(F)} \in F$ such that $\|f_i - f_j\|_\delta > \epsilon$ for all $i \neq j$.

We can see that $C_\epsilon^\delta(F) \leq H_{\epsilon/2}^\delta(F)$. We now prove Theorem 3.1 by approximating the entropy of F.

Proof sketch of Theorem 3.1. Because of an upper bound on the entropy of a direct sum of compact subsets in terms of the entropies of those subsets, we may assume $m = 1$, so that F consists of functions $p(x_1, \ldots, x_n)g(q_1(x_1, \ldots, x_n), \ldots, q_k(x_1, \ldots, x_n))$ where p, q_i are fixed. Next, by applying the Implicit Function Theorem and possibly restricting G (this is where the condition that the q_i are continuously differentiable comes in), we may assume that in fact F consists of functions $p(x_1, \ldots, x_n)g(x_1, \ldots, x_l)$, where l is the maximum rank of the Jacobian of the function $(q_1, \ldots, q_k)$ on G (obviously this is at most k).

Let F_r be the subset of F with norm at most r. We claim that $\log H_\epsilon^\delta(F_r) < (l+\sigma)\log(1/\delta)$ for each $\sigma > 0$ and all sufficiently small $\delta > 0$. To show this, we construct cubes $Q_1, \ldots, Q_J$ with the property that for each $Q_\delta \subset G$, there is some Q_j such that

$$\frac{1}{\delta^n} \left| \int_{Q_\delta} f(x)dx - \int_{Q_j} f(x)dx \right| < \frac{\epsilon}{3}$$

for all $f \in F_r$. This is done in the only reasonable way, by spreading the Q_j evenly and closely enough across G. Importantly, they can be chosen so that J is proportional to δ^{-l}.

For $f \in F_r$, the values of $\frac{1}{\delta^n} \int_{Q_\delta} f(x)dx$ lie in $[-r, r]$. Cover $[-r, r]$ with $O(r/\epsilon)$ intervals $K_1, \ldots, K_{O(r/\epsilon)}$ of length $\epsilon/3$. For $I_1, \ldots, I_J \in \{K_i\}$, let $A(I_1, \ldots, I_J)$ be the set of functions $f \in F_r$ such that $\frac{1}{\delta^n} \int_{Q_j} f(x)dx \in I_j$ for all $j = 1, \ldots, J$. Then for any Q_δ and $f_1, f_2 \in A(I_1, \ldots, I_J)$, we can choose j to get

$$\frac{1}{\delta^n} \left| \int_{Q_\delta} (f_1 - f_2) \right| \leq \frac{1}{\delta^n} \left(\left| \int_{Q_\delta} f_1 - \int_{Q_j} f_1 \right| + \left| \int_{Q_j} f_1 - \int_{Q_j} f_2 \right| + \left| \int_{Q_j} f_2 - \int_{Q_\delta} f_2 \right| \right) \leq \epsilon.$$

So $A(I_1, \ldots, I_J)$ is contained in a ball of radius ϵ. There are $(O(r/\epsilon))^J = (O(r/\epsilon))^{O(1/\delta^l)}$ distinct sets $A(I_1, \ldots, I_J)$ and they cover F_r. Hence we have

$$N_\epsilon^\delta(F_r) \leq (O(r/\epsilon))^{O(1/\delta^l)},$$

$$H_\epsilon^\delta(F_r) \leq O(1/\delta^l) \log O(r/\epsilon),$$

$$\log H_\epsilon^\delta(F_r) \leq l \log(1/\delta) + C \leq (l + \sigma) \log(1/\delta).$$

Finally, we claim that for any ball U_ρ of radius ρ in $C(G)$, we have $\log C_{\epsilon,\delta}(U_\rho) \geq (n - \sigma) \log(1/\delta)$ for each $\sigma > 0$ and all sufficiently small $\delta > 0$. Assume U_ρ is centered at 0. We obtain $Q'_1, \ldots, Q'_K$ disjoint closed cubes of side length δ, where $K = \Omega(\delta^{-n})$. Let $\epsilon_j \in \{\pm 1\}$ for $j = 1, \ldots, K$. Then we can construct a continuous function f such that $\|f\| \leq \rho$ and $f(x) = \epsilon_j \rho$ on Q_j. This gives a list of $2^K = 2^{\Omega(\delta^{-n})}$ functions in U_ρ, any two of which are at least ϵ apart. So

$$M_\epsilon^\delta(U_\rho) \geq 2^{\Omega(\delta^{-n})},$$

$$C_\epsilon^\delta(U_\rho) \geq \Omega(\delta^{-n}),$$

$$\log C_\epsilon^\delta(U_\rho) \geq n \log(1/\delta) + C \geq (n - \sigma) \log(1/\delta).$$

From $C_\epsilon^\delta(F) \leq H_{\epsilon/2}^\delta(F)$, we conclude that $\log H_\epsilon^\delta(U_\rho) \geq (n-\sigma) \log(1/\delta) > (l+\sigma) \log(1/\delta) \geq \log H_\epsilon^\delta(F_r)$. Therefore F_r cannot be dense in U_ρ (otherwise any set of balls of radius ϵ covering F_r would also cover U_ρ and the two entropies would be equal). Therefore F_r is nowhere dense in $C(G)$, and $F = \cup_{r=1}^\infty F_r$ is a countable union of nowhere dense sets, as desired. □

4 Word circuits

In this section, we present Hansen, Lachish, and Miltersen's concept of a word circuit along with some of their results from [1].

Definition 4.1. A word is a bit string of some length w. A word circuit is a circuit where the wires pass words, the input gates take words (of the same length), and the internal gates compute binary functions $(\{0,1\}^w)^2 \to \{0,1\}^w$, defined for all word lengths w.

Definition 4.2. A function $f : (\{0,1\}^w)^3 \to \{0,1\}^w$ (defined for all w) is decomposable if f can be computed by a single constant-sized word circuit.

The first point to make here is that there exist nondecomposable ternary functions, in fact functions whose restrictions to a given word length w require word circuits of size exponential in w to compute.

Proposition 4.1 (Hansen, Lachish, Miltersen 2009)**.** *There exists a function $f : (\{0,1\}^w)^3 \to \{0,1\}^w$ so that no word circuit of size smaller than $(1 - o(1))2^w$ computes f. In particular, f is not decomposable.*

Proof. This is a simple counting argument. A word circuit of size s can be described with $2s\log_2(s+3) + sw2^{2w}$ bits, where the first term specifies the structure of the circuit and the second term specifies the binary function on each gate. There are $2^{w2^{3w}}$ ternary functions. So if s is an upper bound on the word circuit size of all functions, we have

$$s(2\log_2(s+3) + sw2^{2w} \geq w2^{3w}$$

which implies $s = (1-o(1))2^w$. □

The second point is that this lower bound is tight within a constant factor.

Theorem 4.2 (Hansen, Lachish, Miltersen 2009). *Every function $f : (\{0,1\}^w)^3 \to \{0,1\}^w$ is computed by a word circuit of size at most $(2+o(1))2^w$.*

Proof. This is an explicit construction, obtained essentially by separately computing each restriction of f where the first word is fixed, using the corresponding binary function. □

Finally, Hansen, Lachish, and Miltersen present a lower bound for a certain polynomial-time computable function, showing that any word circuit for it must have size at least 5. This is a very small lower bound, but already far from trivial.

Theorem 4.3 (Hansen, Lachish, Miltersen 2009). *There exists a polynomial-time computable function $F : (\{0,1\}^w)^3 \to \{0,1\}^w$ so that for all $w \geq 8$, F_w cannot be computed by a word circuit of size 4.*

Proof. F is obtained by an explicit, complicated construction for which I can offer no simplifying intuition. To prove the lower bound, Hansen, Lachish, and Miltersen enumerate all seven word circuits of size 4 which cannot be further simplified, and show that they all satisfy the property of "weakness", meaning that one of the inputs is connected to the output by a unique path, along with some other properties. They then show that a circuit which satisfies these properties and computes F must have depth at least 2^{w-5}, by counting the number of distinct inputs on which such a circuit may produce a given output and comparing this to how often F may produce the same output. This gives the desired lower bound. □

(The above paragraph conceals many pages of calculations, which we will not provide because they are not enlightening.)

5 Comparisons

To simplify terminology, if we are trying to represent a function f on n variables by a superposition of functions g_i on fewer than n variables, we will call f the "target function" and g_i the "representing functions". We can summarize the results of the previous sections as follows:

1. For functions on $[0,1]^n$, Hilbert's conjecture is trivially false if the target and representing functions are arbitrary, false but difficult if both are required to be continuous, and true if both are required to be continuously differentiable (or satisfy other such smoothness properties).

2. For word functions $(\{0,1\}^w)^3 \to \{0,1\}^w$, Hilbert's conjecture is easily seen to be true if the target and representing functions are arbitrary, and conjectured to be still true if the target function is required to be polynomial-time but the representing functions may be arbitrary.

The substitution of word functions for functions on $[0,1]^n$ in a given circuit is clean and elegant, but the proposed analogy between continuity and polytime computability does not seem to fit very well superficially. There is not much parallelism between the corresponding results.

We suggest that a possibly better analogy for continuity on the discrete side is a condition much weaker than, and not necessarily much related to, polytime computability. To motivate this, consider what would happen if we tried to obtain a circuit for a ternary word function f by embedding $\cup_w \{0,1\}^w$ into $[0,1]$, for example by the naive method of interpreting $x_1 \cdots x_w \in \{0,1\}^w$ as a terminating binary expansion $0.x_1 \ldots x_w$, obtaining a matching function $f_1 : [0,1]^3 \to [0,1]$, writing $f_1(x,y,z) = g(\varphi_1(x) + \varphi_2(y) + \varphi_3(z))$ using the simple construction at the beginning of Section 2 (where g, φ_i are allowed to be arbitrary), and restricting g, φ_i back to $\cup_w \{0,1\}^w$. This would give a word circuit for f, except that the resulting φ_i are not functions $\{0,1\}^w \to \{0,1\}^w$, but rather $\{0,1\}^w \to \{0,1\}^{3w}$. Of course, as a word circuit this is rather silly, since now g is no simpler than f for a given input word length.

So just as we have to restrict the ranges of the word functions in the word circuit in order for Hilbert's conjecture to be interesting, we have to restrict the representing functions in the $[0,1]^n$ case to continuous ones in order for Hilbert's conjecture to be interesting (though still false). And there is a vague conceptual similarity between restricting the range of a word function in terms of the input length and restricting the image of a function of real numbers based on continuity. The point is that instead of polytime computability as an analogue for continuity, it might be reasonable to consider the basic requirement that functions have outputs in $\{0,1\}^w$ as a condition already as strict as continuity.

But if we identify the two, the results still don't match: now we have Kolmogorov's theorem on one hand and Proposition 4.1 on the other. So it might actually be more accurate to compare the output restriction for word functions with continuous differentiability, or one of Vitushkin's smoothness conditions. If we do this, not only the conclusion but also the method of proof matches: both Theorem 3.1 and Proposition 4.1 are obtained by nonconstructive counting arguments. Along these lines, Kolmogorov's theorem might plausibly be compared to Theorem 4.2: both are statements that under less restrictive situations than continuous differentiability/decomposability, functions can be computed in terms of explicitly chosen functions on fewer variables.

Now what happens if we restrict the target (word) function to be polytime computable? In the regime we are suggesting, this corresponds to a condition on a target function of $[0,1]^n$ much more onerous than continuous differentiability—perhaps it is like, for example, requiring the function to be algebraic, as in Hilbert's original formulation in terms of seventh-degree polynomials. Some of Vitushkin's results apply to extremely nice target functions, even polynomials. So there seems to be some hope that Hansen, Lachish, and Miltersen's conjecture could be ported to the $[0,1]^n$ case in a way which provides indirect evidence for it.

In the other direction, perhaps the correct analogue of continuity in the discrete setting is a restriction of the range to $\{0,1\}^{a(w)}$ rather than $\{0,1\}^w$, where $a(w)$ grows is very slightly larger than w (say, $a(w) = w + \log w$), so that the range cannot be pushed back up to $\{0,1\}^{3w}$ in constant depth by repeated applications of such functions. We could then ask which ternary functions can be computed with a constant-sized circuit of this form.

There are a number of other issues here. We could ask, for an indecomposable word function, whether the growth rate of its word circuits has a continuous counterpart—perhaps something about dimensionality, for example. Hansen, Lachish, and Miltersen bring up the possibility of adding a polytime restriction for the representing functions, making the requirements for the representation even stricter; we could ask whether this situation would fit better with Vitushkin's work. There is also always the question of how small we can make the circuit of a decomposable word function, or analogously the representation of a function of real numbers.

In summary, I believe that the analogy between continuous and discrete versions of Hilbert's thirteenth problem is worth exploring more closely, and that it should be pursued with the goal of finding closer connections between results, methods, and conjectures in the two settings, with the hope of making better predictions about which conjectures are likely to be true and how they should be proven.

References

[1] K. A. Hansen, O. Lachish, and P. B. Miltersen, *Hilbert's thirteenth problem and circuit complexity*, ISAAC 2009.

[2] G. G. Lorentz, *Approximation of Functions*, Holt, Rinehart, and Winston, Inc. (Athena Series), 1966.

[3] G. G. Lorentz, *The 13-th problem of Hilbert*, Proceedings of Symposia in Pure Mathematics 28 (1976), 419-430.

[4] A. G. Vitushkin, *On Hilberts thirteenth problem and related questions*, Russian Mathematical Surveys 59 (1) (2004) 11-25.