# DEVIATION THEOREMS FOR SOLUTIONS OF LINEAR ORDINARY DIFFERENTIAL EQUATIONS AND APPLICATIONS TO PARALLEL COMPLEXITY OF SIGMOIDS

D. Yu. Grigoriev

**Abstract.** By a sigmoid with a depth $d$ we mean a computational circuit with $d$ layers in which rational operations are admitted at each layer, and to jump to the next layer the substitution of a function computed at the previous layer into an arbitrary real solution of a linear ordinary differential equation with the polynomial coefficients is admitted. Sigmoids appear as a computational model for neural networks. The deviation theorem is proved which states that for a (real) function $0 \not\equiv f$ computed by a sigmoid with a depth (or parallel complexity) $d$ there exist $c > 0$ and an integer $n$ such that the inequalities $(\exp(\cdots(\exp(c|x|^n)\cdots)^{-1} \leq |f(x)| \leq \exp(\cdots(\exp(c|x|^n)\cdots))))$ hold everywhere on the real line except for a finite measure set, where the iteration of the exponential function is taken $d$ times. One can treat the deviation theorem as an analogue of Liouvillean theorem (on the bound of the difference of algebraic numbers) for solutions of ordinary differential equations. Also we estimate the numbers of zeroes of $f$ in the intervals.

## Introduction

Sigmoids (they appear as a computational model for neural networks, see [MSS] and the references there, also the section 1 below) could be treated as the computational circuits with the gate functions being not necessary rational functions as for the usual circuits. In the so-called "standard sigmoid" one takes a function $(1 + \exp(-x))^{-1}$ as a gate function. One of the most interesting questions in this area, how to approximate the functions computed by sigmoids by means of the "simpler" functions?

We consider this question for the class of sigmoids in which as the gate functions could occur arbitrary (real) functions being solutions of the linear ordinary differential equations with the polynomial coefficients (see the section 1). For the question of approximation the most important complexity parameter of a sigmoid appears to be its parallel complexity or depth. By the depth we understand the number of layers of the sigmoid, at each layer the applications of gate functions to the functions computed at the previous layers

Subject classification. Primary 68C25
Key words and phrases: Sigmoid, parallel complexity, deviation theorems, bounds on Wronskian

are admitted and also rational operations are admitted. In the case of the threshold sigmoids (see [MSS] ) the functions at the next layer are taken as $u(\sum w_i h_i)$ where $h_i$ are the functions computed at the previous layers, $w_i \in \mathbb{R}$ are some weights and $u$ is a gate function; so linear functions are admitted. Observe also that the "standard" sigmoid can be simulated by the sigmoids considered in the present paper as $(1+\exp(-x))^{-1}$ is a composition of a rational function and $\exp(-x)$ being a solution of a linear ordinary differential equation with const coefficients. Denote by $\exp^{(i)} = \exp(\cdots(\exp)\cdots)$ the iteration of the exponential function $i$ times. This is an example of a function computed by a sigmoid with the depth $i$. We prove (see corollary 2 in the section 5) that any function computed by a sigmoid with a depth at most $d$ cannot be approximated roughly speaking closer than $(\exp^{(d)})^{-1}$ by a (different) function also computed by a sigmoid with the depth at most $d$ (this bound is sharp). The results of this kind we call deviation theorems, as we prove that two (different) functions computed by sigmoids of small parallel complexities deviate from each other. Also we prove that a function computed by a sigmoid, cannot grow too fast. One can also treat the deviation theorem as an extension of Liouvillean type theorems on algebraic numbers to the functions computed by sigmoids (cf. also [CC] ). A similar lower bound on an approximation of a function computed by a sigmoid by means of a Blum-Shub-Smale computations [BSS] (which admits rational operations and branching) is obtained in corollary 3 (see the section 5).

In the main theorem (see the section 5) we prove a stronger statement that the measure of all the points from a given interval in which the mentioned beforehand either deviation or the growth bound fails, is small. Whence we get (see the corollary) that the measure of all such points on a real line is finite. The theorem is proved by the induction on the depth and in the inductive step in order to get a lower bound on the function $f$ computed at the next $d$-th layer of the sigmoid, we estimate by inductive hypothesis the coefficients $a^{(0)}, \dots, a^{(m)}$ (being the functions computed by a sigmoid with the depth $d-1$) of a certain linear ordinary differential equation $Lf = \sum_{0 \le i \le m} a^{(i)} f^{(i)} = 0$ to which satisfies $f$.

The bound on the coefficients $a^{(0)}, \ldots, a^{(m)}$ implies an upper bound on $|f|$ (see lemma 4 in the section 3), then we obtain a lower bound on the Wronskian (see [H], [K] ) of the equation $Lf = 0$ (see lemmas 5, 6 in the section 4) and finally, we get a lower bound on $|f|$ (see lemma 7 in the section 5). In the section 2 we obtain some auxiliary bounds on the gate functions $u$, in other words on the solutions of the linear ordinary differential equations with polynomial coefficients, namely an upper bound (see lemma 1), a bound on the Wronskian (see lemma 3) and also a polynomial upper bound (being sharp) on the growth of the number of zeroes (in intervals) of $u$ (see lemma 2), the latter was known as a consequence of Sturm theory for second-order equations (see [H] ). Moreover, we give an upper bound of the form $\exp^{(d-1)}(x^{0(1)})$ (see the proposition in the section 5) on the number of zeroes in the interval $[0, x]$ for arbitrary $x$, of a function computed by a sigmoid with the depth $d$ (provided that we exclude a certain subset of $\mathbb{R}$ with a finite measure such that its intersection with the interval $[0, x]$ consist of at most $\exp^{(d-2)}(x^{0(1)})$ intervals). In the section 1 we give necessary definitions and notations.

In the last section 6 we consider "elementary" sigmoids namely the ones with the gates exp, log and algebraic functions (in particular, containing "standard" sigmoids). As log has a singularity (at zero) one cannot directly apply the main theorem and the corollary, but it is still possible to extend the theorem and the corollary imposing a certain restriction on an elementary sigmoid (see the corollary 4).

## 1. Differential fields and sigmoids

Denote the ring $K_0 = \mathbb{R}[X]$, the field $F_0 = \mathbb{R}(X)$, differentiation $D = d/dX$, and by $\Gamma$ denote the set of real functions $u : \mathbb{R} \to \mathbb{R}$ (perhaps, with a finite number of singularities) being the solutions of linear ordinary differential equations of the kind

$$Lu = \left( \sum_{0 \le j \le n} a_j D^j \right) u = 0 \qquad (1)$$

where $a_j \in K_0$. Moreover, we impose a requirement that $u$ could be obtained as the restriction on $\mathbb{R}$ of some branch of a certain complex analytic function.

The elements of $\Gamma$ will play the role of gate functions in the sigmoids. The operator $L$ may have singularities only in the roots $\Delta = \{\delta_1, \dots, \delta_\theta\}$ of the leading coefficient $a_n$ ([H] ). Out of the points $\Delta \cap \mathbb{R}$ the function $u$ is analytic [H]. Now we define $K_{i+1}$ and $F_{i+1}$ by induction on $i$. Namely, $K_{i+1}$ for $i \geq 0$ is a differential ring ([K] ) generated by the functions of the form $u(g)$, where $u \in \Gamma,\ g \in F_i$. Let $F_{i+1}$ be a quotient field of $K_{i+1}$ (so, $F_{i+1}$ is a differential field). Thus, any element of $F_{i+1}$ is a real function and could be obtained as the restriction on $\mathbb{R}$ of some branch of a certain complex analytic function.

Under a sigmoid we'll understand a circuit with a certain depth $d$ in which any function $w_{i+1}^{(j)}$ at $(i+1)$-th layer $(0 \leq i < d)$ is computed as

$$w_{i+1}^{(j)} = u\left(\frac{f}{g}(w_i^{(1)}, w_i^{(2)}, \dots, X)\right) \qquad (2)$$

for some gate function $u \in \Gamma$ and $f, g \in \mathbb{R}[W_i^{(1)}, W_i^{(2)}, \dots, X]$ being polynomials in the functions $w_i^{(1)}, w_i^{(2)}, \dots$ computed at the previous layers of the sigmoid, and in the variable $X$. One can show by induction on $i$ that each function $w_{i+1}^{(j)}$ is real and belongs to $K_{i+1}$, conversely, any element from $K_{i+1}$ can be obtained as a rational function in functions computed at $(i+1)$-th layer of a suitable sigmoid. Usually, $u$ is taken from a certain subclass of $\Gamma$, for example, in the case of the "standard" sigmoid one takes $u = \exp(-X)$ (see [MSS] ).

Henceforth, we fix a sigmoid and by $\mathcal{D}_i \supset K_i$ denote a differential ring generated over $\mathbb{R}$ by $w_i^{(1)}, w_i^{(2)}, \dots$; so as an algebraic ring $\mathcal{D}_i$ is generated by all the derivatives $w_i^{(1)}, Dw_i^{(1)}, \dots, w_i^{(2)}, Dw_i^{(2)} \dots$.

Consider a basis (over $\mathbb{C}$) $\overline{u_1}, \dots, \overline{u_n}$ of the space of solution of the equation (1), where each of $\overline{u_1}, \dots, \overline{u_n}$ could be considered as a branch of a complex analytic function [H], let $u_1 = u$ be the restriction of $\overline{u}_1$ on $\mathbb{R}$, denote by $u_2, \dots, u_n$ the restrictions of $\overline{u}_2, \dots, \overline{u}_n$ on $\mathbb{R}$, respectively. Denote by $\overline{\mathcal{D}}_i \supset \mathcal{D}_i$ a differential ring generated over $\mathbb{C}$ by the functions of the form $u_\ell(\frac{f}{g}(w_i^{(1)}, w_i^{(2)}, \dots, X)),\ 1 \leq \ell \leq n$ for all the functions $w_{i+1}^{(j)}$ computed in the sigmoid at the steps of the kind (2) at $(i+1)$-th layer. Again as above every element from $\overline{\mathcal{D}}_i$ could be considered as the restriction on $\mathbb{R}$ of some branch of a certain complex analytic function.

## 2. Bounds on the solutions and on the Wronskian of a linear ordinary differential equations with polynomial coefficients.

First, obtain upper bounds on the absolute values of a complex analytic function $\overline{u}$ (and its derivatives $D\overline{u}, D^2\overline{u}, \dots$) satisfying (1).

¿From now on $p_0, p_1, \dots$ will denote polynomials from $K_0$ each having a form $p_i = \overline{p_i}(X^2)$ where a polynomial $\overline{p}_i \in K_0$ monotonically increases on $\mathbb{R}^+$ and $\overline{p_i}(0) \geq 1$. The class of all such polynomials denote by $\mathcal{P}$.

For an operator $L$ (see (1)) and $x \in \mathbb{C}$ denote $\|x\|_L = \max\{\, |x|,\ |\,x - \delta_1|^{-1}, \dots, |x - \delta_\theta|^{-1}\}$.

LEMMA 1. *Let $\overline{u}$ satisfy (1). One can find a suitable $p_0 \in \mathcal{P}$ such that for any $x \in \mathbb{C}$ $|\overline{u}(x)|$, $|D\overline{u}(x)|, \dots, |D^n\overline{u}(x)| \leq \exp(p_0(\,\|x\|_L))$.*

*Proof.* Rewrite (1) as a system of linear first-order equations: $DY = AY$ where the vector $Y = (\overline{u}, D\overline{u}, \dots, D^{n-1}\overline{u})$ and the matrix

$$A = \begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & \bigcirc & \\ \bigcirc & & \ddots & 0 & 1 \\ -\frac{a_0}{a_n} & -\frac{a_1}{a_n} & \dots & & -\frac{a_{n-1}}{a_n} \end{pmatrix} \tag{3}$$

Then $Y(x) = Y(x_0) + \int\limits_{x_0}^{x} A(\xi)Y(\xi)d\xi$ ([H] ), provided that the path along which we integrate with the endpoints $x_0, x$, does not pass through the singularities $\delta_1, \dots, \delta_\theta$, hence $\|Y(x)\| \leq \|Y(x_0)\| + \int\limits_{x_0}^{x} \|A(\xi)\|\ \|Y(\xi)\|\, d\xi$ where $\|\quad\|$ denotes the euclidean norm for vectors and matrices. Then Gronwall's inequality [H] implies that $\|Y(x)\| \leq \|Y(x_0)\| \exp \int\limits_{x_0}^{x} \|A(\xi)\|\, d\xi \leq \exp(p_0(\|x\|_L))$ for a suitably chosen path.

COROLLARY 1. *For every $j \geq 0$ there exists $p_0^{(j)} \in \mathcal{P}$ such that $|D^j\overline{u}(x)| \leq \exp(p_0^{(j)}(\|x\|_L))$ for all $x \in \mathbb{R}$.*

*One can prove the corollary by induction on $j$ taking into account that by applying $D^j$ to (1) one gets $(a_n D^{n+j} + \sum_{0\le\ell<n+j} a_\ell^{(j)} D^\ell)\overline{u} = 0$ for some $a_\ell^{(j)} \in K_0$.*

In the next lemma we bound from above the number $N(x)$ of the roots of the real function $u$ satisfying (1), in the interval $[0, x]$, provided that $\Delta \cap \mathbb{R} = \phi$, so when $u$ has no (real) singularities. For second-order equations ($n = 2$) one can find a polynomial bound on $N(x)$ in [H], being a consequence of Sturm theory.

LEMMA 2. *$N(x) \le p_1^{(0)}(x)$ for a certain $p_1^{(0)} \in \mathcal{P}$, provided that $\Delta \cap \mathbb{R} = \phi$.*

*Proof.* Let $|a_i(y)| < p_1(x)$ for each $y \in [0, x]$ for an appropriate polynomial $p_1 \in \mathcal{P}$. We show that each interval $I = (y, y + (2p_1(x))^{-1}) \subset [0, x]$ contains at most $(n-1)$ roots of $u$. Suppose the contrary, then (by Rolle's theorem) each derivative $Du, \dots D^{n-1}u$ has at least one root in the interval $I$. Denote by $M^{(j)} = \max_{z\in I} |D^j u(z)|$. Then $M^{(j+1)} \ge M^{(j)} 2p_1(x)$ by the theorem of the mean of the derivative, $0 \le j < n$. Let $D^n u$ reach $M^{(n)}$ at a point $x_0 \in I$. Then $M^{(n)} = |D^n u(x_0)| = |\sum_{0\le j\le n-1} a_j(x_0) D^j u(x_0)| \le p_1(x)$ $\left(\frac{1}{2p_1(x)} + \left(\frac{1}{2p_1(x)}\right)^2 + \cdots + \left(\frac{1}{2p_1(x)}\right)^n\right) M^{(n)} < p_1(x) \cdot \frac{2}{(2p_1(x))} M^{(n)}$. Obtained contradiction shows that $I$ contains less that $n$ roots of $u$, therefore $N(x) \le 2xp_1(x)(n-1)$, lemma is proved.

*Remark.* The bound in lemma 2 is sharp, take, for example, as $u = \sin(X^n)$.

Consider Wronskian [H]

$$W = \det \begin{pmatrix} \overline{u}_1 & \dots & \overline{u}_n \\ D\overline{u}_1 & \dots & D\overline{u}_n \\ \vdots & & \vdots \\ D^{n-1}\overline{u}_1 & \dots & D^{n-1}\overline{u}_n \end{pmatrix}$$

As $DW = -\frac{a_{n-1}}{a_n} W$, one gets $W(x) = W(x_0) \exp \int_{x_0}^{x} -\frac{a_{n-1}}{a_n}$, provided that the path along which we integrate, does not pass through the singularities $\delta_1, \dots, \delta_\theta$. This formula entails the following lemma.

LEMMA 3. *There exists a polynomial* $p_2 \in \mathcal{P}$ *such that* $(\exp(p_2(\|x\|_L)))^{-1} \leq |W(x)| \leq \exp(p_2(\|x\|_L))$ *hold for all* $x \in \mathbb{C}\backslash\Delta$.

*Remark.* As below we will be interested in the functions on $\mathbb{R}$, we will apply lemmas 1, 3 to the functions $u_1, \ldots, u_n$.

## 3. Upper bounds for the differential polynomials in the functions computed by a sigmoid

Denote by $\exp^{(i)} = \exp(\cdots(\exp)\cdots) \in K_i$ the iteration of the exponential function $i$ times.

By induction on $i$ we'll estimate (from above in the present section and from below in the section 5) differential polynomials $G_1, \ldots, G_K \in \overline{\mathcal{D}_{i+1}}$ (see section 1). For this purpose we'll produce a family of differential polynomials $H^{(0)}, \ldots, H^{(m)} \in \mathcal{D}_i$ such that a lower bound on $|H^{(0)}|$ and upper bounds on $|H^{(0)}|, \ldots, |H^{(m)}|$ would entail upper bounds on $|G_1|, \ldots, |G_k|$ (a similar statement for lower bounds one can find in the section 5). This will be used in the proof of the inductive step in the main theorem.

LEMMA 4. *For a family of differential polynomials* $G_1, \ldots, G_k \in \overline{\mathcal{D}_{i+1}}$ *one can construct a family of differential polynomials* $0 \not\equiv H^{(0)}, \ldots, H^{(m)} \in \mathcal{D}_i$ *such that for every polynomials* $p_7, \tilde{p} \in \mathcal{P}$ *there exists a polynomial* $\overline{p} \in \mathcal{P}$ *satisfying the following property: for arbitrary* $x \in \mathbb{R}$ *if the inequalities* $|H^{(0)}| \geq (\exp^{(i)} \tilde{p})^{-1}$, $|H^{(j)}| \leq \exp^{(i)} \tilde{p}$, $0 \leq j \leq m$ *hold everywhere on an interval* $I_i = (x - (\exp^{(i)} p_7(x))^{-1}, x)$ *then* $|G_\ell| \leq \exp^{(i+1)} \overline{p}$, $1 \leq \ell \leq k$ *hold everywhere on* $I_i$.

*Proof.* For each function $w = w_{i+1}^{(j)}$ computed by the sigmoid (see (2)) consider the following elements from $\mathcal{D}_i$:

$$H_w^{(0)} = g(w_1^{(i)}, \ldots, X) \prod_{1 \leq \ell \leq \theta, \delta_\ell \in \mathbb{R}} (f(w_i^{(1)}, \ldots, X) - \delta_\ell g(w_i^{(1)}, \ldots, X))$$

where the product is taken over the real $\delta_\ell \in \Delta \cap \mathbb{R}$, $H_w^{(1)} = g(w_1^{(i)}, \ldots, X)$, $H_w^{(2)} = f(w_i^{(1)}, \ldots, X)$. Observe that if $|H_w^{(0)}| \geq (\exp^{(i)} \tilde{p})^{-1}$; $|H_w^{(1)}|$, $|H_w^{(2)}| \leq \exp^{(i)} \tilde{p}$ everywhere on $I_i$ then $|g(w_1^{(i)},\dots,X)|$, $|f(w_i^{(1)},\dots,X)-\delta_\ell g(w_i^{(1)},\dots,X)|\geq(\exp^{(i)}p_3)^{-1}$ everywhere on $I_i$ for an appropriate polynomial $p_3\in\mathcal{P}$. Hence $\|\frac{f}{g}(w_i^{(1)},\dots,X)\|_L\leq\exp^{(i)}p_4$ everywhere on $I_i$ for some $p_4\in\mathcal{P}$.

Let in the differential polynomials $G_1,\dots,G_k$ occur the derivatives $D^s u_{æ}((f/g)(w_i^{(1)}, w_i^{(2)},\dots,X))$ (see (2) and the section 1) up to orders $s\leq s_0$. We have $D^s u_{æ}((f/g)(w_i^{(1)}, w_i^{(2)},\dots,X))=\frac{1}{g^{s+1}}(u_{æ}(\frac{f}{g}(w_i^{(1)},\dots))h_0^{(s)}+\frac{\partial u_{æ}}{\partial z}(\frac{f}{g}(w_i^{(1)},\dots))h_1^{(s)}+\dots+\frac{\partial^s u_{æ}}{\partial z^s}(\frac{f}{g}(w_i^{(1)},\dots))$ $h_s^{(s)})$ for suitable differential polynomials $h_0^{(s)},\cdots,h_s^{(s)}\in\mathcal{D}_i$. As $H^{(0)}$ take the product of $H_w^{(0)}$ for all the functions $w$ computed by the sigmoid, as $H^{(1)},\dots,H^{(m)}$ take the union of $H_w^{(1)},H_w^{(2)}$ for all $w$ and in addition $h_0^{(s)},\dots,h_s^{(s)}$ for all $s\leq s_0$, and all $w$. Then by the assumption of the lemma and by the proved above $\|\frac{f}{g}(w_i^{(1)},\dots,X)\|_L\leq\exp^{(i)}p_5$ everywhere on $I_i$ for a certain $p_5\in\mathcal{P}$, therefore $\left|\frac{\partial^t u_{æ}}{\partial z^t}\Big(\frac{f}{g}(w_i^{(1)},\dots)\Big)\right|\leq\exp^{(i+1)}p_6, 0\leq t\leq s_0$ everywhere on $I_i$ for an appropriate $p_6\in\mathcal{P}$ because of lemma 1 and corollary 1. Hence $|D^s u_{æ}((f/g)(w_i^{(1)},\dots,X))|\leq\exp^{(i+1)}\overline{p}_6$ everywhere on $I_i$ for a certain $\overline{p}_6\in\mathcal{P}$ by virtue of the construction of $H^{(1)},\dots,H^{(m)}$.

Finally, each of $G_1,\dots,G_k$ can be considered as an (algebraic) polynomial in $D^s u_{æ}$ $((f/g)(w_i^{(1)},\dots,X))$, this completes the proof of the lemma.

*Remark.* Actually we did not utilize the particular form of an interval $I_i$, thus the lemma is valid for any interval, this form of $I_i$ will be used later in the section 5 when we'll deal with the lower bounds. The same remark concerns also lemmas 5, 6 below in the section 4.

## 4. <u>Upper and lower bounds on Wronskians of the functions computed by sigmoids</u>

Any function $w=w_{i+1}^{(j)}$ computed by the sigmoid (see (2)) satisfies a linear ordinary differential equation with the coefficients from $\mathcal{D}_i$. One can represent (cf. the proof of lemma 4) $w,Dw,\dots,D^n w$ as linear combinations of $u((f/g)(w_i^{(1)},\dots))$, $\frac{\partial u}{\partial z}((f/g)(w_i^{(1)},\dots))$ $,\dots,\frac{\partial^{n-1}u}{\partial z^{n-1}}((f/g)(w_i^{(1)},\dots))$ with the coefficients being differential polynomials in $(f/g)$ $(w_i^{(1)},\dots)$. Therefore, $w$ satisfies a certain $n$-order equation $0=\mathcal{L}w=(\sum_{0\leq\ell\leq n}b_\ell D^\ell)w$

where $b_0, \dots, b_n$ are differential polynomials in $(f/g)(w_i^{(1)}, \dots)$. By the same token $u_2((f/g)(w_i^{(1)}, \dots)), \dots, u_n((f/g)(w_i^{(1)}, \dots))$ also satisfy the same equation $\mathcal{L}w = 0$ (see the section 1). Let us prove that $w = u_1((f/g)(w_i^{(1)}, \dots)), \dots, u_n((f/g)(w_i^{(1)}, \dots))$ are linearly independent over $\mathbb{C}$ and thereby these functions constitute a basis of the space of solutions of the equation $\mathcal{L}w = 0$, provided that $(f/g)(w_i^{(1)}, \dots) \not\equiv \text{const}$ (if the latter is not the case then $w \equiv \text{const}$ and we can omit computational step (2) in the sigmoid). Indeed, consider the Wronskian

$$W_w = \det \begin{pmatrix} u_1(f/g)(w_i^{(1)}, \dots) & \dots\dots & u_n(f/g)(w_i^{(1)}, \dots) \\ D(u_1(f/g)(w_i^{(1)}, \dots)) & \dots\dots & D(u_n(f/g)(w_i^{(1)}, \dots)) \\ \vdots & & \vdots \\ D^{n-1}(u_1(f/g)(w_i^{(1)}, \dots)) & \dots\dots & D^{n-1}(u_n(f/g)(w_i^{(1)}, \dots)) \end{pmatrix}$$

$$= W((f/g)(w_i^{(1)}, \dots)) \cdot (D((f/g)(w_i^{(1)}, \dots)))^{\frac{n(n-1)}{2}} \not\equiv 0$$

where $W = \det \begin{pmatrix} u_1 & \dots & u_n \\ \vdots & & \vdots \\ D^{n-1}u_1 & \dots & D^{n-1}u_n \end{pmatrix}$ is the Wronskian of the equation $Lu = 0$. One can express $D((f/g)(w_i^{(1)}, \dots,)) = \frac{h_1(w_i^{(1)}, \dots,)}{g^2(w_i^{(1)}, \dots,)}$ where $h_1(w_i^{(1)}, \dots) \in \mathcal{D}_i$ is a certain differential polynomial. Consider (cf. the proof of lemma 4) the following elements from $\mathcal{D}_i$ : $\overline{H}_w^{(0)} = h_1(w_i^{(1)}, \dots) g(w_i^{(1)}, \dots) \prod_{1 \le \ell \le \theta} (f(w_i^{(1)}, \dots) - \delta_\ell g(w_i^{(1)}, \dots))$, $\overline{H}_w^{(1)} = g(w_i^{(1)}, \dots)$, $\overline{H}_w^{(2)} = f(w_i^{(1)}, \dots)$, $\overline{H}_w^{(3)} = h_1(w_i^{(1)}, \dots)$. Suppose that for a certain polynomial $p_8 \in \mathcal{P}$ for arbitrary $x \in \mathbb{R}$ the following inequalities $|\overline{H}_w^{(0)}| \ge (\exp^{(i)} p_8)^{-1}$, $|\overline{H}_w^{(j)}| \le \exp^{(i)} p_8$, $0 \le j \le 3$ hold everywhere on an interval $I_i = (x - (\exp^{(i)} p_7(x))^{-1}, x)$. Then $|\overline{H}_w^{(1)}|$, $|\overline{H}_w^{(3)}| \ge (\exp^{(i)} p_8^2)^{-1}$ hold everywhere on $I_i$. Therefore, $(\exp^{(i)} p_8^4)^{-1} \le |D((f/g)(w_i^{(1)}, \dots))| \le \exp^{(i)} p_8^5$ is true everywhere on $I_i$. Since $\|(f/g)(w_i^{(1)}, \dots)\|_L \le \exp^{(i)} p_9$ is valid everywhere on $I_i$ for an appropriate $p_9 \in \mathcal{P}$, lemma 3 implies inequalities $(\exp^{(i+1)} \overline{p})^{-1} \le |W((f/g)(w_i^{(1)}, \dots))| \le \exp^{(i+1)} \overline{p}$ everywhere on $I_i$ a suitable $\overline{p} \in \mathcal{P}$, this gives similar bounds on $|W_w|$.

Thus, the following lemma is proved.

Lemma 5. *For each function $w = w_{i+1}^{(j)}$ (recall that we assume that $w \not\equiv$ const without loss of generality) computed by a sigmoid (see (2)) one can construct differential polynomials*

$\overline{H}_w^{(j)} \in \mathcal{D}_i,\ 0 \le j \le 3,\ \overline{H}_w^{(0)} \not\equiv 0$ *such that for any polynomials* $p_7, \tilde{p}_0 \in \mathcal{P}$ *there exists a polynomial* $\overline{p}_0 \in \mathcal{P}$ *satisfying the following property: for arbitrary* $x \in \mathbb{R}$ *if the inequalities* $|\overline{H}_w^{(0)}| \ge (\exp^{(i)} \tilde{p}_0)^{-1},\ |\overline{H}_w^{(j)}| \le \exp^{(i)} \tilde{p}_0,\ 0 \le j \le 3$ *are valid everywhere on an interval* $I_i = (x - (\exp^{(i)} p_7(x))^{-1}, x)$ *(cf. lemma 4 and the remark just after it) then the Wronskian* $W_w = W(w_{i+1}^{(j)} = u_1((f/g)(w_i^{(1)}, \dots)), \dots, u_n((f/g)(w_i^{(1)}, \dots)))$ *satisfies inequalities* $(\exp^{(i+1)} \overline{p}_0)^{-1} \le |W_w| \le \exp^{(i+1)} \overline{p}_0$ *everywhere on* $I_i$.

*Thus, we've bounded Wronskian of the functions* $w = w_{i+1}^{(j)}$ *computed by the sigmoid (see (2)). Now we proceed to estimating Wronskian of the differential polynomials in the functions* $w$*, i.e. the elements from* $\mathcal{D}_{i+1}$*, more precisely, by the latter we mean Wronskian of some linear ordinary differential equation with the coefficients from* $\mathcal{D}_i$ *to which satisfies this differential polynomial under consideration. Arguing by induction on the construction of the differential polynomial, we estimate Wronskian in three cases: for the derivative (see 4.1 below), for the sum of the functions (see 4.2 below) and for the product (see 4.3 below). Lemma 5 plays the role of the base of induction.*

*So, we assume that two function* $w, v \in \mathcal{D}_{i+1}$ *satisfy linear ordinary differential equations* $0 = Q_1 w = Q_2 v$*, where* $Q_1 = \sum_{0 \le \ell \le k} \alpha^{(\ell)} D^\ell, Q_2 = \sum_{0 \le \ell \le m} \beta^{(\ell)} D^\ell$ *and the coefficients* $\alpha^{(\ell)}, \beta^{(\ell)} \in \mathcal{D}_i$*. We assume also that by induction some differential polynomials* $0 \not\equiv H_w^{(0)}, \dots, H_w^{(æ_w)}, 0 \not\equiv H_v^{(0)}, \dots, H_v^{(æ_v)} \in \mathcal{D}_i$ *are produced such that for any polynomials* $p_7, p_{10} \in \mathcal{P}$ *there exists* $p_{11} \in \mathcal{P}$ *satisfying the following property: if*

$$|H_w^{(0)}|,\ |H_v^{(0)}| \ge (\exp^{(i)}(p_{10}))^{-1};\ |H_w^{(j)}|,\ |H_v^{(j)}| \le \exp^{(i)}(p_{10}) \tag{4}$$

*for all* $j \ge 0$ *everywhere on an interval* $I_i = (x - (\exp^{(i)} p_7(x))^{-1}, x)$*, then*

$$(\exp^{(i+1)}(p_{11}))^{-1} \le |W_w|,\ |W_v| \le \exp^{(i+1)}(p_{11}) \tag{5}$$

*everywhere on* $I_i$*; here* $W_w, W_v$ *denote the Wronskians of the operators* $Q_1, Q_2$*, resp. Recall (see the section 1) that* $u = u_1, \dots, u_n$ *is a basis of the equation (1) and* $w_{i+1}^{(j)} = u((f/g)(w_i^{(1)}, \dots))$ *(see (2)). Finally, we assume that the equation* $Q_1 w = 0 (Q_2 v = 0$ *resp.)*

*has a basis (over $\mathbb{C}$) $w = w^{(1)}, \dots, w^{(k)}$($v = v^{(1)}, \dots, v^{(m)}$ resp.) where $w^{(1)}, \dots, w^{(k)}$, $v^{(1)}, \dots, v^{(m)} \in \overline{\mathcal{D}_{i+1}}$.*

*Thus, we've formulated an inductive hypothesis of the induction on the construction of $w \in \mathcal{D}_{i+1}$ (the base was proved in lemma 5).*

### 4.1. Estimating Wronskian for the derivative Dw.

Consider 2 cases. In the first one $\alpha^{(0)} \neq 0$. Then $Dw$, $D^2 w, \dots, D^{k+1} w$ can be expressed as the linear combinations of $w, Dw, \dots, D^{k-1} w$ with the coefficients being the differential polynomials in $\alpha^{(0)}, \dots, \alpha^{(k)}$ and thereby, being the elements from $\mathcal{D}_i$. Therefore, we get a required equation for $Dw$, namely, $0 = \overline{Q}_1 Dw = (\sum_{0 \leq \ell \leq k} \overline{\alpha}^{(\ell)} D^\ell) Dw$. The functions $Dw^{(1)}, \dots, Dw^{(k)}$ constitute a basis of the space of its solutions since

$$W_{Dw} = \det \begin{pmatrix} Dw^{(1)} & \dots & Dw^{(k)} \\ \vdots & & \\ D^k w^{(1)} & \dots & D^k w^{(k)} \end{pmatrix} = (-1)^{k+1} \frac{\alpha^{(0)}}{\alpha^{(k)}} W_w \not\equiv 0$$

Set $H_{Dw}^{(0)} = H_w^{(0)} \cdot \alpha^{(0)} \alpha^{(k)}, H_{Dw}^{(1)} = H_w^{(1)}, \dots, H_{Dw}^{(æ_w)} = H_w^{(æ_w)}, H_{Dw}^{(æ_w+1)} = H_w^{(0)}, H_{Dw}^{(æ_w+2)} = \alpha^{(0)}, H_{Dw}^{(æ_w+3)} = \alpha^{(k)}$ (see (4)), this gives the desired bounds on the Wronskian $W_{Dw}$ (see (5)).

In the second case $\alpha^{(0)} = 0$. Then $0 = \overline{Q}_1 (Dw) = (\sum_{0 \leq \ell \leq k-1} \alpha^{(\ell+1)} D^\ell) Dw$ provides the required equation for $Dw$. Since $0 = Q_1\, 1$ we can take as a basis of the space of solutions of the equation $0 = Q_1 w$ the functions $1, w, \overline{w}^{(2)}, \dots, \overline{w}^{(k-1)}$ for some $\overline{w}^{(2)}, \dots, \overline{w}^{(k-1)} \in \overline{\mathcal{D}}_{i+1}$. Then the Wronskian $W_w$ of the equation $Q_1 w = 0$

$$\det \begin{pmatrix} 1 & w & \overline{w}^{(2)} & \dots & \overline{w}^{(k+1)} \\ 0 & Dw & D\overline{w}^{(2)} & \dots & D\overline{w}^{(k+1)} \\ \vdots & & & & \\ 0 & D^{k-1} w & D^{k-1} \overline{w}^{(2)} & \dots & D^{k-1} \overline{w}^{(k+1)} \end{pmatrix} \text{ equals to}$$

the Wronskian of the basis $Dw, D\overline{w}^{(2)}, \dots, D\overline{w}^{(k-1)}$ of the space of solutions of the equation $0 = \overline{Q}_1 (Dw)$. This proves the inductive hypothesis for $W_{Dw}$ (taking $H_{Dw}^{(0)} = H_w^{(0)}, \dots, H_{Dw}^{(æ_w)} = H_w^{(æ_w)}$).

### 4.2. Estimating Wronskian for the sum $\mathbf{w}+\mathbf{v}$.

All the functions $\overline{w}+\overline{v}$ where $\overline{w},\overline{v}$ are solutions of the equations $0=Q_1\overline{w}=Q_2\overline{v}$, satisfy a linear ordinary differential equation $(\sum_{0\le p\le s}\gamma^{(\ell)}D^{\ell})(\overline{w}+\overline{v})=Q_+(\overline{w}+\overline{v})=0$ of the minimal order $s\le m+k$ because one can express $\overline{w}+\overline{v},D(\overline{w}+\overline{v}),\cdots,D^{k+m}(\overline{w}+\overline{v})$ as the linear combinations of $\overline{w},D\overline{w},\ldots,D^{k-1}\overline{w},\overline{v},\ D\overline{v},\ldots,D^{m-1}\overline{v}$ with the coefficients being the quotients of differential polynomials in $\alpha^{(0)},\ldots,\alpha^{(k)},\beta^{(0)},\ldots,\beta^{(m)}$. Hence $\gamma^{(\ell)}$ can be taken as differential polynomials in $\alpha^{(0)},\ldots,\alpha^{(k)},\beta^{(0)},\ldots,\beta^{(m)}$ and thereby one can assume $\gamma^{(\ell)}\in\mathcal{D}_i$. As a basis of the space of solutions of the equation $Q_+z=0$ one can take a subset of $w^{(1)},\ldots,w^{(k)},v^{(1)},\ldots,v^{(m)}$ and $s$ equals to the maximal number of linearly independent (over $\mathbb{C}$) functions among $w^{(1)},\ldots,w^{(k)},v^{(1)},\ldots,v^{(m)}$ (see [S81] ). Denote by $W_{w+v}$ the Wronskian of the equation $Q_+z=0$ (with respect to the chosen basis). Therefore, relying on lemma 4 one can construct $\widetilde{H}^{(0)}_{w+v},\ldots,\widetilde{H}^{(\tilde{æ}_{w+v})}_{w+v}\in\mathcal{D}_i$ such that if $|\widetilde{H}^{(0)}_{w+v}|\ge(\exp^{(i)}(p_{12}))^{-1},\ |\widetilde{H}^{(j)}_{w+v}|\le\exp^{(i)}(p_{13})$ everywhere on $I_i$ then $|W_{w+v}|\le\exp^{(i+1)}(p_{14})$ everywhere on $I_i$.

Proceed to a lower bound on $|W_{w+v}|$. Rewrite the equations $0=Q_1w=Q_2v$ as first-order linear systems $DY_1=B_1Y_1,\ DY_2=B_2Y_2$ where (cf. (3))

$$B_1=\begin{pmatrix} 0 & 1 & & \bigcirc \\ & 0 & & \\ \bigcirc & & \ddots & 1 \\ -\frac{\alpha^{(0)}}{\alpha^{(k)}} & -\frac{\alpha^{(1)}}{\alpha^{(k)}} & \cdots & -\frac{\alpha^{(k-1)}}{\alpha^{(k)}} \end{pmatrix},\quad B_2=\begin{pmatrix} 0 & 1 & & \bigcirc \\ & 0 & & \\ \bigcirc & & \ddots & 1 \\ -\frac{\beta^{(0)}}{\beta^{(m)}} & -\frac{\beta^{(1)}}{\beta^{(m)}} & \cdots & -\frac{\beta^{(m-1)}}{\beta^{(m)}} \end{pmatrix}$$

Consider a direct sum of the latter 2 systems: $DU=(B_1\oplus B_2)U$ where $B_1\oplus B_2=\left(\begin{smallmatrix}B_1&0\\0&B_2\end{smallmatrix}\right)$. Denote by $\mathcal{T}_1$ (resp $\mathcal{T}_2$) $k$-dimensional (resp. $m$-dimensional) space of solutions of the system $DY_1=B_1Y_1$ (resp $DY_2=B_2Y_2$), a certain basis of $\mathcal{T}_1$ (resp. $\mathcal{T}_2$) consists of the vectors $\{(w^{(j)},Dw^{(j)},\ldots,D^{k-1}w^{(j)})\}_{1\le j\le k}$ (resp. $\{(v^{(j)},Dv^{(j)},\ldots,D^{m-1}v^{(j)})\}_{1\le j\le m}$). The space of solutions of the system $DU=(B_1\oplus B_2)U$ is the direct sum $\mathcal{T}_1\oplus\mathcal{T}_2$. Therefore, the Wronskian $W_{B_1\oplus B_2}$ of the system $DU=(B_1\oplus B_2)U$ equals to $W_{B_1\oplus B_2}=W_{B_1}\cdot W_{B_2}=W_w\cdot W_v$.

Rewrite also the equation $Q_+(w+v)=0$ as a first-order linear differential system:

$DT = CT$ where $C = \begin{pmatrix} 0 & 1 & & \bigcirc \\ & 0 & & \\ \bigcirc & & \ddots & 1 \\ \frac{-\gamma^{(0)}}{\gamma^{(s)}} & & \cdots & \frac{-\gamma^{(s-1)}}{\gamma^{(s)}} \end{pmatrix}$. Denote by $\mathcal{T}$ ($s$-dimensional) space of solutions of the system $DT = C\,T$. Consider a mapping $\mathcal{L} : \mathcal{T}_1 \oplus \mathcal{T}_2 \to \mathcal{T}$ which maps

$$(w^{(j_1)}, Dw^{(j_1)}, \dots, D^{k-1}w^{(j_1)}) \oplus (v^{(j_2)}, Dv^{(j_2)}, \dots, D^{m-1}v^{(j_2)}) \to$$
$$(w^{(j_1)} + v^{(j_2)},\ D(w^{(j_1)} + v^{(j_2)}), \dots, D^{s-1}(w^{(j_1)} + v^{(j_2)}))$$

Obviously, $\mathcal{L}$ is surjective.

Recall (see [K], also [Si 90] ) that a subspace (over $\mathbb{C}$) $\mathcal{U}_1 \subset \mathcal{T}_1 \oplus \mathcal{T}_2$ is called invariant (with repsect to the system $DU = (B_1 \oplus B_2)U$) if $\mathcal{U}_1$ is invariant under the action of the (differential) Galois group. Then the kernel $\mathcal{U} = \ker \mathcal{L} \subset \mathcal{T}_1 \oplus \mathcal{T}_2$ is invariant (see e.g. [BBH], [G90b] ), hence any nonsingular linear transformation of the space $\mathcal{T}_1 \oplus \mathcal{T}_2$ being of the form $\overline{\mathcal{L}} = \binom{*}{\mathcal{L}}$, where $*$ is $(k + m - s) \times (k + m)$ matrix over $\mathbb{R}$, reduces (see e.g. [BBH] ) the system $DU = (B_1 \oplus B_2)U$ to the block-triangular form $D\overline{U} = \left(\begin{smallmatrix} C_1 & 0 \\ C_2 & C_3 \end{smallmatrix}\right)\overline{U}$ where $\overline{U} = \overline{\mathcal{L}}U$, $\left(\begin{smallmatrix} C_1 & 0 \\ C_2 & C_3 \end{smallmatrix}\right) = (D\overline{\mathcal{L}}) \cdot (\overline{\mathcal{L}})^{-1} + \overline{\mathcal{L}}\left(\begin{smallmatrix} B_1 & O \\ O & B_2 \end{smallmatrix}\right)(\overline{\mathcal{L}})^{-1}$ and the space of solutions of the system $DU_3 = C_3U_3$ equals to $\mathcal{U}$ (in [G90b], see also [G90a] one can find the complexity bounds on reducing a system to the block-triangular form). The space of solutions of the system $DU_1 = C_1U_1$ equals to $\mathcal{T}$, hence the Wronskian $W_{C_1}$ of this system equals to $W_{w+v}$.

It is known (see [H] ) that the Wronskian $W_{C_1} = \exp \int \, tr\, C_1$, therefore, $W_{C_1} W_{C_3} W_{\left(\begin{smallmatrix} C_1 & 0 \\ C_2 & C_3 \end{smallmatrix}\right)}$ $= (\det \overline{\mathcal{L}}) \cdot W_{\left(\begin{smallmatrix} B_1 & 0 \\ 0 & B_2 \end{smallmatrix}\right)} = (\det \overline{\mathcal{L}}) \cdot W_w \cdot W_v$. As each coefficient of any vector in the space $\mathcal{U} \subset \mathcal{T}_1 \oplus \mathcal{T}_2$ belongs to $\overline{\mathcal{D}_{i+1}}$ one can construct by lemma 4 the differential polynomials $\overline{H}^{(0)}_{\mathcal{U},\dots,}\overline{H}^{(æ_{\mathcal{U}})}_{\mathcal{U}} \in \mathcal{D}_i$ such that if $|\overline{H}^{(0)}_{\mathcal{U}}| \ge (\exp^{(i)}(p_{15}))^{-1}$, $|\overline{H}^{(j)}_{\mathcal{U}}| \le \exp^{(i)}(p_{16})$, $0 \le j \le æ_{\mathcal{U}}$ everywhere on an interval $I_i = (x - (\exp^{(i)}(p_7(x)))^{-1}, x)$ then $|W_{C_3}| \le \exp^{(i+1)}(p_{17})$ everywhere on $I_i$. As each entry of the matrix $\overline{\mathcal{L}}$ belongs to a quotient field of the ring $\mathcal{D}_i$, one can represent $\det(\overline{\mathcal{L}}) = \overline{H}^{(0)} / \overline{H}^{(1)}$ where $\overline{H}^{(0)}, \overline{H}^{(1)} \in \mathcal{D}_i$.

Finally, taking $H^{(0)}_{w+v} = H^{(0)}_w\, H^{(0)}_v\, \widetilde{H}^{(0)}_{w+v}\, \overline{H}^{(0)}_{\mathcal{U}}\, \overline{H}^{(0)}$ (see (4)) and as a family $H^{(1)}_{w+v}, \dots,$ $H^{(æ_{w+v})}_{w+v}$ taking the union of $H^{(j)}_w$, $H^{(j)}_v$, $\widetilde{H}^{(j)}_{w+v}$, $\overline{H}^{(j)}_{\mathcal{U}}$, $\overline{H}^{(0)}, \overline{H}^{(1)}$ for all $j \ge 0$, we obtain (using (5)) that if $|H^{(0)}_{w+v}| \ge (\exp^{(i)}(p_{18}))^{-1}$, $|H^{(j)}_{w+v}| \le \exp^{(i)}(p_{19})$ for all $j \ge 0$ everywhere on $I_i$ then $(\exp^{(i+1)}(p_{20}))^{-1} \leq |W_{C_1}| = |W_{w+v}| \leq \exp^{(i+1)}(p_{14})$ everywhere on $I_i$, which completes the proof of the inductive hypothesis for the sum $w+v \in \mathcal{D}_{i+1}$.

### 4.3. Estimating Wronskian for the product wv.

All the functions $\overline{w}\,\overline{v}$ where $\overline{w}, \overline{v}$ are the solutions of the equations $0 = Q_1\overline{w} = Q_2\overline{v}$, satisfy a linear ordinary differential equation $0 = Q_\times(\overline{w}\,\overline{v}) = (\sum_{0\leq j\leq t} \rho^{(j)} D^j)(\overline{w}\,\overline{v})$ of the (minimal) order $t \leq mk$ because one can express $\overline{w}\,\overline{v}$, $D(\overline{w}\,\overline{v}), \ldots, D^{mk}(\overline{w}\,\overline{v})$ as the linear combinations of $D^{j_1}\overline{w}\,D^{j_2}\overline{v}$ for $0 \leq j_1 \leq k-1,\ 0 \leq j_2 \leq m-1$ with the coefficients being quotients of differential polynomials in $\alpha^{(0)}, \ldots, \alpha^{(k)}$, $\beta^{(0)}, \ldots, \beta^{(m)}$. Hence $\rho^{(j)}$ can be taken as differential polynomials in $\alpha^{(0)}, \ldots, \alpha^{(k)}$, $\beta^{(0)}, \ldots, \beta^{(m)}$ and thereby one can assume that $\rho^{(j)} \in \mathcal{D}_i$. As a basis of the space of solutions of the equation $Q_\times z = 0$ one can take a maximal linearly independent (over $\mathbb{C}$) subset of the elements $w^{(j_1)}v^{(j_2)}$, $1 \leq j_1 \leq k,\ 1 \leq j_2 \leq m$ (see [S81] ). Denote by $W_{wv}$ the Wronskian of the equation $Q_\times z = 0$ (with respect to the chosen basis). Therefore, relying on lemma 4 one can construct differential polynomials $\widetilde{H}_{wv,\ldots,}^{(0)} \widetilde{H}_{wv}^{(\tilde{æ}_{wv})} \in \mathcal{D}_i$ such that if $|\widetilde{H}_{wv}^{(0)}| \geq (\exp^{(i)}(p_{21}))^{-1}$, $|\widetilde{H}_{wv}^{(j)}| \leq \exp^{(i)}(p_{22})$, for all $j \geq 0$ everywhere on an interval $I_i$ (see above the subsection 4.2) then $|W_{wv}| \leq \exp^{(i+1)}(p_{23})$ everywhere on $I_i$.

Thus, similar to the case of the sume $w+v$, it remains to prove the lower bound on $|W_{wv}|$. Consider a tensor product of the systems $DY_1 = B_1Y_1$, $DY_2 = B_2Y_2$ (see above): $DU = (B_1 \otimes E_m + E_k \otimes B_2)U$ where $E_m,\ E_k$ denote the unit matrices of the corresponding sizes. Then the space of solutions of the latter system coincides with the tensor product $\mathcal{T}_1 \otimes_{\mathbb{R}} \mathcal{T}_2$ (cf. e.g. [BBH] ). Hence Wronskian $W_{B_1\otimes E_m + E_k \otimes B_2}$ of this system equals to $W_{B_1}^m\, W_{B_2}^k = W_w^m\, W_v^k$.

Rewrite (cf. (3)) also the equation $Q_\times z = 0$ as a first-order differential system: $DR = \mathcal{J}R$ where $\mathcal{J} = \begin{pmatrix} 0 & 1 & & \bigcirc \\ & 0 & & \\ \bigcirc & & \ddots & 1 \\ -\frac{\rho^{(0)}}{\rho^{(t)}} & & & -\frac{\rho^{(t-1)}}{\rho^{(t)}} \end{pmatrix}$. Denote by $\mathcal{R}$ ($t$-dimensional) space of solutions of the system $DR = \mathcal{J}R$. Consider a mapping $\mathcal{Y} : \mathcal{T}_1 \otimes \mathcal{T}_2 \to \mathcal{R}$ which maps $(\ldots, (D^{\ell_1}w^{(j_1)})(D^{\ell_2}v^{(j_2)}), \ldots)_{0\leq \ell_1 < k,\, 0 \leq \ell_2 < m} \to (\ldots, D^{\ell}(w^{(j_1)}v^{(j_2)}), \ldots)_{0 \leq \ell < t}$. Obviously,

$\mathcal{Y}$ is surjective.

The kernel $\mathcal{K} = \ker \mathcal{Y} \subset \mathcal{T}_1 \otimes \mathcal{T}_2$ is invariant (cf. above and [BBH], [G90b] ), hence any nonsingular linear transformation of the space $\mathcal{T}_1 \otimes \mathcal{T}_2$ being of the form $\overline{\mathcal{Y}} = \begin{pmatrix} * \\ \mathcal{Y} \end{pmatrix}$, where $*$ is $(km - t) \times km$ matrix over $\mathbb{R}$, reduces (cf. above and [BBH] ) the system $DU = (B_1 \otimes E_m + E_k \otimes B_2)U$ to $D\overline{U} = \left(\begin{smallmatrix} \mathcal{J}_1 & 0 \\ \mathcal{J}_2 & \mathcal{J}_3 \end{smallmatrix}\right)\overline{U}$ where $\overline{U} = \overline{\mathcal{Y}}U$, $\left(\begin{smallmatrix} \mathcal{J}_1 & 0 \\ \mathcal{J}_2 & \mathcal{J}_3 \end{smallmatrix}\right) = (D\overline{\mathcal{Y}})(\overline{\mathcal{Y}})^{-1} + \overline{\mathcal{Y}}(B_1 \otimes E_m + E_k \otimes B_2)(\overline{\mathcal{Y}})^{-1}$ and the space of solutions of the system $DU_3 = \mathcal{J}_3 U_3$ coincides with $\mathcal{K}$. The space of solutions of the system $DU_1 = \mathcal{J}_1 U_1$ equals to $\mathcal{R}$, whence $W_{\mathcal{J}_1} = W_{wv}$. Therefore, (see the subsection 4.2 above and [H] ) $W_{\mathcal{J}_1} W_{\mathcal{J}_3} = W_{\left(\begin{smallmatrix} \mathcal{J}_1 & 0 \\ \mathcal{J}_2 & \mathcal{J}_3 \end{smallmatrix}\right)} = (\det \overline{\mathcal{Y}}) \cdot W_{B_1 \otimes E_m + E_k \otimes B_2} = (\det \overline{\mathcal{Y}}) \cdot W_w^m \cdot W_v^k$.

As each coefficient of any vector in the space $\mathcal{K} \subset \mathcal{T}_1 \otimes \mathcal{T}_2$ belongs to $\overline{\mathcal{D}_{i+1}}$, one can construct by lemma 4 the differential polynomials $\overline{H}_{\mathcal{K},\dots,}^{(0)} \overline{H}_{\mathcal{K}}^{(æ_{\mathcal{K}})} \in \mathcal{D}_i$ such that if $|\overline{H}_{\mathcal{K}}^{(0)}| \geq (\exp^{(i)}(p_{24}))^{-1}$, $|\overline{H}_{\mathcal{K}}^{(j)}| \leq \exp^{(i)}(p_{25})$ for all $j \geq 0$ everywhere on an interval $I_i$ (see above), then $|W_{\mathcal{J}_3}| \leq \exp^{(i+1)}(p_{26})$ everywhere on $I_i$. As each entry of the matrix $\overline{\mathcal{Y}}$ belongs to a quotient field of the ring $\mathcal{D}_i$, one can represent $\det(\overline{\mathcal{Y}}) = \widetilde{H}^{(0)} / \widetilde{H}^{(1)}$ where $\widetilde{H}^{(0)}$, $\widetilde{H}^{(1)} \in \mathcal{D}_i$. Finally, taking $H_{wv}^{(0)} = H_w^{(0)} H_v^{(0)} \widetilde{H}_{wv}^{(0)} \overline{H}_{\mathcal{K}}^{(0)} \widetilde{H}^{(0)}$ (see (4)) and as a family $H_{wv}^{(1)}, \dots, H_{wv}^{(æ_{wv})}$ taking the union of $H_w^{(j)}$, $H_v^{(j)}$, $\widetilde{H}_{wv}^{(j)}$, $\overline{H}_{\mathcal{K}}^{(j)}$, $\widetilde{H}^{(0)}$, $\widetilde{H}^{(1)}$ for all $j \geq 0$, we obtain (using (5)) that if $|H_{wv}^{(0)}| \geq (\exp^{(i)}(p_{27}))^{-1}$, $|H_{wv}^{(j)}| \leq \exp^{(i)}(p_{28})$ for all $j \geq 0$ everywhere on $I_i$ then

$$(\exp^{(i+1)}(p_{29}))^{-1} \leq |W_{\mathcal{J}_1}| = |W_{wv}| \leq \exp^{(i+1)}(p_{23})$$

everywhere on $I_i$, which completes the proof of the inductive hypothesis for the product $wv$.

Let us summarize what was proved above in the present section in the following lemma.

Lemma 6. *For any differential polynomials $G_0 \in \mathcal{D}_{i+1}, G_1, \dots, G_æ \in \overline{\mathcal{D}_{i+1}}$ one can construct differential polynomials $0 \not\equiv H_0, H_1, \dots, H_\eta \in \mathcal{D}_i$ such that for any polynomials $p_7, \tilde{p}_i \in \mathcal{P}$ there exists a polynomial $\tilde{p}_{i+1} \in \mathcal{P}$ satisfying the following property: for any $x \in \mathbb{R}$ if $|H_0| \geq (\exp^{(i)}(\tilde{p}_i))^{-1}$, $|H_\ell| \leq \exp^{(i)}(\tilde{p}_i)$, $0 \leq \ell \leq \eta$ hold everywhere* *on an interval* $I_i = (x - (\exp^{(i)}(p_7(x)))^{-1}, x)$ *(cf. lemma 4 and the remark after it) then* $|G_\ell| \leq \exp^{(i+1)}(\tilde{p}_{i+1}),\ 0 \leq \ell \leq æ$ *and* $|W_{G_0}| \geq (\exp^{(i+1)}(\tilde{p}_{i+1}))^{-1}$ *everywhere on* $I_i$ *where* $W_{G_0}$ *denotes the Wronskian of a certain linear ordinary differential equation* $(\sum_{0 \leq j \leq m} \overline{\beta}^{(j)} D^j) G_0 = 0$ *with the coefficients* $\overline{\beta}^{(j)} \in \mathcal{D}_i$ *and with the space of solutions containing* $G_0$ *and having a basis consisting of the elements from* $\overline{\mathcal{D}_{i+1}}$.

The proof is conducted by induction on the construction of $G_0$ (using lemma 5 for the base of induction), so obtain a family $H_0^{(0)}, H_1^{(0)}, \dots \in \mathcal{D}_i$, then apply lemma 4 to $G_0, \dots, G_æ$, obtain as a result a family $H_0^{(1)}, H_1^{(1)}, \dots \in \mathcal{D}_i$ and finally combine $H_0 = H_0^{(0)} H_0^{(1)}$ and as a family $H_1, H_2, \dots$ take the union of both families $H_0^{(0)} H_1^{(0)}, \dots$ and $H_0^{(1)}, H_1^{(1)}, \dots$.

## 5. Lower bound on a differential polynomial in the functions computed by a sigmoid

Lemma 7. *Let* $i \geq 0$. *For any differential polynomials* $0 \not\equiv G_0 \in \mathcal{D}_{i+1}, G_1, \dots, G_æ \in \overline{\mathcal{D}_{i+1}}$ *one can construct differential polynomials* $0 \not\equiv \mathcal{H}_0, \dots, \mathcal{H}_\nu \in \mathcal{D}_i$ *such that for any polynomials* $p_7, \tilde{p}_i \in \mathcal{P}$ *there exist polynomials* $\overline{p_{i+1}},\ p_{i+1}^{(0)},\ p_{i+1}^{(1)} \in \mathcal{P}$ *satisfying the following property: for any* $x \in \mathbb{R}$ *if* $|\mathcal{H}_0| \geq (\exp^{(i)}(\tilde{p}_i))^{-1},\ |\mathcal{H}_e| \leq \exp^{(i)}(\tilde{p}_i),\ 0 \leq e \leq \nu$ *hold everywhere on an interval* $I_i = (x - (\exp^{(i)}(p_7(x)))^{-1}, x)$, *then* $|G_\ell| \leq \exp^{(i+1)}(\overline{p_{i+1}}),\ 0 \leq \ell \leq æ$ *are valid everywhere on* $I_i$ *and there exists a disjoint family of subintervals* $\{I_{i+1}^{(\alpha)}\}_\alpha$ *of the interval* $I_i$ *each with the length* $|I_{i+1}^{(\alpha)}| = (\exp^{(i+1)}(p_{i+1}^{(0)}(x)))^{-1}$, *moreover the lower bound* $|G_0| \geq (\exp^{(i+1)}(\overline{p_{i+1}}))^{-1}$ *holds everywhere on* $I_{i+1}^{(\alpha)}$ *for each* $\alpha$ *and finally* $\sum_\alpha |I_{i+1}^{(\alpha)}| \geq |I_i|(1 - (\exp^{(i+1)}(p_{i+1}^{(1)}(x)))^{-1})$. *Moreover, the complement* $I_i \setminus \cup_\alpha I_{i+1}^{(\alpha)}$ *consists of at most* $\exp^{(i)}(p_{i+1}^{(0)}(x))$ *intervals and* $G_0$ *has at most* $\exp^{(i)}(p_{i+1}^{(0)}(x))$ *zeros in* $I_i$.

*Remark 1.* One can interpret the lemma 7 informally that the desired lower bound on $|G_0|$ holds "almost" everywhere on $I_i$.

*Remark 2.* Observe that the construction of $\mathcal{H}_0, \dots, \mathcal{H}_\nu$ from $G_0, \dots, G_æ$ is independent from the intervals and bounds, the similar remark concerns as well the previous lemmas

4, 5, 6.

*Proof of lemma 7.* Apply lemma 6 to $G_0, \dots, G_{æ}$ and produce $H_0, \dots, H_\eta$. Then consider an equation $0 = \overline{\mathcal{L}} G_0 = (\sum_{0 \le j \le m} \overline{\beta}^{(j)} D^j) G_0$ yielded in lemma 6, where $\overline{\beta}^{(0)}, \dots, \overline{\beta}^{(m)} \in \mathcal{D}_i$, then $DG_0$ satisfies equation $0 = \mathcal{L}(DG_0) = (\sum_{0 \le j \le m} \beta^{(j)} D^j) DG_0$ for suitable differential polynomials $\beta^{(0)}, \dots, \beta^{(m)} \in \mathcal{D}_i$, $\beta^{(m)} = \overline{\beta}^{(m)} \not\equiv 0$. Consider any basis $G_0 = G_{0,1}, \dots, G_{0,m} \in \overline{\mathcal{D}_{i+1}}$ of the space of solutions of the equation $0 = \overline{\mathcal{L}} G_0$ and apply lemma 4 to a family $\{D^j G_{0,\ell}\}_{0 \le j \le m-1,\, 1 \le \ell \le m}$. As a result we get a family $\overline{H}_0, \dots, \overline{H}_\mu \in \mathcal{D}_i$, set $\mathcal{H}_0 = H_0 \overline{H}_0 \beta^{(m)}$ and as a family $\mathcal{H}_1, \dots, \mathcal{H}_\nu$ take the union of $H_0, \dots, H_\eta$, $\overline{H}_0, \dots, \overline{H}_\mu$, $\beta^{(0)}, \dots, \beta^{(m)}$. Thus by lemmas 4, 6 $|\beta^{(m)}| \ge (\exp^{(i)}(p_{30}))^{-1}$, $|W_{G_0}| \ge (\exp^{(i+1)}(p_{31}))^{-1}$, $|G_j| \le \exp^{(i+1)}(p_{32}), |D^j G_{0,\ell}| \le \exp^{(i+1)}(p_{33})$, $|\beta^{(\ell)}| \le \exp^{(i)}(p_{34})$ everywhere on $I_i$. When $G_0 \equiv \text{const}$, these inequalities give the lemma immediately, so furthermore assume that $G_0 \not\equiv \text{const}$.

Similar to the proof of lemma 2 we prove that $G_0$ takes each value $\varepsilon$ on the interval $I_i$ at most $\max\{m+1, 2\exp^{(i)}((p_{30}p_{34})(x))(m+1)|I_i|\}$ times. Suppose the contrary. Then there is a subinterval $I \subset I_i$ of the length $\lambda = \min\{|I_i|, \frac{1}{2}(\exp^{(i)}(p_{30}p_{34}(x)))^{-1}\}$, in which $G_0$ takes value $\varepsilon$ at least $(m+1)$ times. Therefore, each derivative $DG_0, \dots, D^m G_0$ has at least one root in the interval $I$. Denote $M^{(j)} = \max_{y \in I} |D^j G_0(y)|$. Then $M^{(j+1)} \ge \frac{M^{(j)}}{\lambda}, 1 \le j \le m$. Let $|D^{m+1} G_0|$ reach $M^{(m+1)}$ at a point $x_0 \in I$. Then $M^{(m+1)} = \left| \frac{1}{\beta^{(m)}(x_0)} \left( \sum_{0 \le j \le m-1} \beta^{(j)} D^{j+1} G_0(x_0) \right) \right| \le \exp^{(i)}(p_{30}(x_0)) M^{(m+1)} \exp^{(i)}(p_{34}(x_0))(\lambda + \lambda^2 + \dots + \lambda^m) < M^{(m+1)}$. The obtained contradiction proves that $G_0$ takes each value $\varepsilon$ in the interval $I_i$ at most $\max\{m+1, 2\exp^{(i)}((p_{30}p_{34})(x))(m+1)|I_i|\} \le \exp^{(i)}(p_{35})(x)$ times for a certain $p_{35}$.

Construct a sequence of polynomials $r_0, \dots, r_{m-1} \in K_0$ by (inverse) induction: set $r_{m-1} = p_{31} + (m-1)p_{33} + m^2$ and $r_j = 2r_{j+1} + p_{33} + 2$, $0 \le j < m-1$, then $r_0 \ge r_1 \ge \dots \ge r_{m-1}$ everywhere on $\mathbb{R}$.

Firstly, assume that at some point $x_1 \in I_i$ hold inequalities $|D^j G_0(x_1)| \le (\exp^{(i+1)}(r_j \cdot$

$(x)))^{-1} \leq (\exp^{(i+1)}(r_{m-1}(x)))^{-1}$ for all $0 \leq j \leq m-1$. Then expanding $W_{G_0}$ with respect to the column consisting of $G_0, DG_0, \dots, D^{m-1}G_0$ (as $G_0 \not\equiv 0$ we can assume that $G_0$ is one of the elements of a basis of the space of solutions of $\overline{\mathcal{L}}G = 0$) we obtain $(\exp^{(i+1)}(p_{31}(x)))^{-1} \leq |W_{G_0}(x_1)| \leq (\exp^{(i+1)}(r_{m-1}(x)))^{-1} m! (\exp^{(i+1)}(p_{33}(x)))^{m-1}$ that contradicts to the choice of $r_{m-1}$.

Consider a subinterval $I^{(0)} \subset I_i$ on which $|G_0| \leq (\exp^{(i+1)}(r_0(x)))^{-1}$ everywhere. Take a minimal $1 \leq j_0 \leq m-1$ (it does exist, see above) such that there exists a point $x_0 \in I^{(0)}$ for which $|D^{j_0}G_0(x_0)| \geq (\exp^{(i+1)}(r_{j_0}(x)))^{-1}$. As $|D^{j_0+1}G_0| \leq \exp^{(i+1)}(p_{33}(x))$ everywhere on $I_i$ we get for each $x_1 \in I_i$ $|D^{j_0}(x_1)| \geq |D^{j_0}G_0(x_0)| - |x_1 - x_0| \exp^{(i+1)}(p_{33}(x)) \geq (\exp^{(i+1)}(r_{j_0}(x)))^{-1} - |x_1 - x_0| \exp^{(i+1)}(p_{33}(x))$. Assume that at least one of two points $x_2 = x_0 \pm (\exp^{(i+1)}(r_{j_0} + p_{33})(x))^{-1}$ belongs to $I^{(0)}$, then $|D^{j_0-1}G_0(x_0) - D^{j_0-1}G_0(x_2)| = |\int_{x_0}^{x_2} D^{j_0}G_0| \geq 1/2(\exp^{(i+1)}(r_{j_0} + p_{33})(x))^{-1}(\exp^{(i+1)}(r_{j_0}(x)))^{-1} \geq 2(\exp^{(i+1)}(r_{j_0-1}(x)))^{-1}$ that leads to a contradiction with the minimality of $j_0$. Hence, neither of two considered points $x_3$ belong to $I^{(0)}$, therefore, the length $|I^{(0)}| \leq 2(\exp^{(i+1)}(r_{j_0} + p_{33})(x))^{-1} \leq 2(\exp^{(i+1)}(r_{m-1} + p_{33})(x))^{-1} \leq (\exp^{(i+1)}(p_{36}(x)))^{-1}$ for a certain $p_{36}$.

Partition the interval $I_i$ on the subintervals with the endpoints in which $G_0$ takes the values $\pm(\exp^{(i+1)}(r_0(x)))^{-1}$. By the proved above there are at most $2\exp^{(i)}(p_{35}(x))$ such subintervals. Also we've proved that the length of any subinterval on which $|G_0| \leq (\exp^{(i+1)}(r_0(x)))^{-1}$ everywhere, is less than $(\exp^{(i+1)} p_{36}(x))^{-1}$, partitioning all the other subintervals into disjoint subintervals $I_{i+1}^{(\alpha)}$ completes the proof of lemma 7. Now we can formulate the main result of the paper.

THEOREM. *Let a function $0 \not\equiv f$ be computed by a sigmoid of the depth $d$. For any $\rho_1$ there exist $\rho_0, \rho_2$, where $\rho_0, \rho_1, \rho_2$ are univariate nonconstant polynomials, being positive everywhere on $\mathbb{R}$ such that for any $x \in \mathbb{R}$ the measure of the points $y \in (x - (\rho_1(x))^{-1}, x) = I$ at which $|f(y)| \geq \exp^{(d)}(\rho_0(x))$ or $|f(y)| \leq (\exp^{(d)} \rho_0(x))^{-1}$, is less than $\frac{(\rho_1(x))^{-1}}{\exp(\rho_2(x))} = \frac{|I|}{\exp(\rho_2(x))}$.*

Corollary. *The measure of the points* $y \in \mathbb{R}$ *for which* $|f(y)| \geq \exp^{(d)}(\rho_3(y))$ *or* $|f(y)| \leq (\exp^{(d)} \rho_3(y))^{-1}$ *is finite, moreover the measure of such points* $y$ *with* $|y| \geq x_0$ *for any* $x_0 > 0$ *is less than* $\exp(-\rho_4(x_0))$ *for suitable nonconstant polynomials* $\rho_3, \rho_4$ *positive everywhere on* $\mathbb{R}$.

*Remark.* Note that all the polynomials occurring in the theorem and the corollary could be calculated explicitly in terms of the size of the sigmoid and in the parameters of the differential equations defining occurring in the sigmoid functions from $\Gamma$ (see the section 1).

*Proof of the theorem.* Apply lemma 7 (taking also into account remark 2 just after it) to $f$, it will produce a family $\mathcal{H}_0^{(d-1)}, \dots, \mathcal{H}_{æ_{d-1}}^{(d-1)} \in \mathcal{D}_{d-1}$, next apply lemma 7 to $\mathcal{H}_0^{(d-1)}, \dots, \mathcal{H}_{æ_{d-1}}^{(d-1)}$ and obtain as a result a family $\mathcal{H}_0^{(d-2)}, \dots, \mathcal{H}_{æ_{d-2}}^{(d-2)} \in \mathcal{D}_{d-2}$ and so on, until we get a family $\mathcal{H}_0^{(0)}, \dots, \mathcal{H}_{æ_0}^{(0)} \in K_0$ of polynomials.

Let $x_0 > 0$ be larger than the absolute values of all the roots of the polynomial $\mathcal{H}_0^{(0)}$. Take any polynomial $p_{37} \in \mathcal{P}$. We'll prove by induction on $i \geq 1$ that there exist polynomials $\tilde{p}_4^{(i)}$, $\tilde{p}_5^{(i)}$, $\tilde{p}_6^{(i)} \in \mathcal{P}$ such that for any $|x| > x_0 + 1$ the subset of the points $y \in (x - (p_{37}(x))^{-1}, x) = I$ at which $|\mathcal{H}_0^{(i)}(y)| \geq (\exp^{(i)} \tilde{p}_4^{(i)}(x))^{-1}$ and $|\mathcal{H}_j^{(i)}(y)| \leq (\exp^{(i)} \tilde{p}_4^{(i)}(x))$, $0 \leq j \leq æ_i$ contains a certain disjoint union of $Q^{(i)}(x) \subset I$ of the intervals $I_\alpha^{(i)}$ each of the length $I_\alpha^{(i)}| = (\exp^{(i)} \tilde{p}_5^{(i)}(x))^{-1}$ and the measure mes $(Q^{(i)}(x)) \geq$ mes $(I) \cdot (1 - (\exp \tilde{p}_6^{(i)}(x))^{-1})$.

The base of induction for $i = 0$ (it is convenient to start with $i = 0$ although the inductive hypothesis is true for $i \geq 1$). In this case for a suitable polynomial $\tilde{p}_4^{(0)} \in \mathcal{P}$ the required inequalities $|\mathcal{H}_0^{(0)}(y)| \geq (\tilde{p}_4^{(0)}(x))^{-1}$, $|\mathcal{H}_j^{(0)}(y)| \leq \tilde{p}_4^{(0)}(x)$ hold everywhere on $I$, so $Q^{(0)} = I$ and mes$Q^{(0)} =$ mes$(I)$.

For the inductive step take any of the intervals $I_\alpha^{(i)}$ and apply to it (and to the family $\mathcal{H}_0^{(i+1)}, \dots, \mathcal{H}_{æ_{i+1}}^{(i+1)}$, see above) lemma 7, this will provide polynomials $\tilde{p}_4^{(i+1)}$, $\tilde{p}_5^{(i+1)} \in \mathcal{P}$ and the necessary intervals which we denote by $I_\beta^{(i+1)}$ (taking the union over all $\alpha$). By lemma 7 and the inductive hypothesis one gets for $i \geq 1$ $\sum_{\beta} |I_{\beta}^{(i+1)}| \geq (1 - (\exp^{(i+1)}(p_{i+1}^{(1)}(x)))^{-1}) \sum_{\alpha}$ $|I_{\alpha}^{(i)}| \geq (1 - (\exp^{(i+1)}(p_{i+1}^{(1)}(x)))^{-1})(1 - (\exp \tilde{p}_6^{(i)}(x))^{-1})|I| \geq (1 - (\exp \tilde{p}_6^{(i+1)}(x))^{-1})|I|$ for a relevant polynomial $\tilde{p}_6^{(i+1)} \in \mathcal{P}$ (when $i = 0$ put $\tilde{p}_6^{(1)} = \tilde{p}_1^{(1)}$), that completes the proof of the theorem.

Following the proofs of lemma 7 (see the end of its proof) and the theorem one can bound the number of zeroes of a function computed by a sigmoid (cf. lemma 2). In the next proposition we agree that $\exp^{(-1)} \equiv \text{const}$.

PROPOSITION. *Let a function $f$ be computed by a sigmoid of the depth $d \geq 1$.*

*a) For any polynomial $\rho_1$ there exist polynomials $\rho_2, \rho_5$, where $\rho_2, \rho_2, \rho_5$ are nonconst positive everywhere on $\mathbb{R}$, such that for arbitrary $x \in \mathbb{R}$ there exists a subset $J \subset I = (x - (\rho_1(x))^{-1}, x)$ being a union of $\exp^{(d-2)} \rho_5(x)$ intervals, with the measure less than $\frac{|I|}{\exp(\rho_2(x))}$ and the number of zeroes of the function $f$ in the set $I \backslash J$ does not exceed $\exp^{(d-1)} \rho_5(x)$;*

*b) There exists a set $J_0 \subset \mathbb{R}$ with a finite measure such that for arbitrary $x \in \mathbb{R}$ the number of zeroes of $f$ in the set $[0, x] \backslash J_0$ is less than $\exp^{(d-1)} \rho_6(x)$ for a suitable polynomial $\rho_6$, moreover the intersection $[0, x] \cap J_0$ is a union of at most $\exp^{(d-2)} \rho_6(x)$ intervals.*

Let us demonstrate the sharpness of the bound in the corollary to the theorem. Namely, consider a function $f = \sin(x) \cdot (\exp^{(i)}(x))^{-1}$, which can be computed by a sigmoid of the depth $i$. Then the set of the points $y$ at which $|f(y)| \leq (\exp^{(i)} p_{39}(y))^{-1}$ consists of a union of intervals where $n$-th interval ($n = 0, 1, \ldots$) has a length $(\exp^{(i)} p_{38}(n))^{-1}$ and contains the point $\pi n$ (for appropriate polynomials $p_{39}$, $p_{38} \in \mathcal{P}$).

Let us give two applications of the theorem and of the corollary to the questions of approximation of the functions computed by sigmoids. One can treat corollary 2 below as an analogue of Liouvillean theorem (on the algebraic numbers) for the functions computed by sigmoids. Here we consider the growth of the difference of two functions as a measure of their deviation, in particular for the solutions of the linear ordinary differential equations with polynomial coefficients. In the latter case another analogue of Liouvillean theorem for another measure of deviation, namely for the order of the power-series expansions, was ascertained in [CC].

Corollary 2. *If different functions $f, g$ are computed by sigmoids with the depths at most $d$ then $\exp^{(d)}(p_{40}(x)) \geq |(f-g)(x)| \geq (\exp^{(d)} p_{40}(x))^{-1}$ for a suitable nonconst polynomial $p_{40}$ positive everywhere on $\mathbb{R}$ for all $x \in \mathbb{R}$, except for a set with a finite measure.*

*Another application concerns Blum-Shub-Smale model* [BSS]. *If one tries to approximate $f$ by BSS-computation, that means the $f$ is approximated by a piecewise rational function. Applying the corollary to the theorem to each piece, we get*

Corollary 3. *If an irrational function $f$ is computed by a sigmoid with a depth $d$ and a function $h$ is computed by some Blum-Shub-Smale computation then $\exp^{(d)}(p_{41}(x)) \geq |(f-h)(x)| \geq (\exp^{(d)} p_{41}(x))^{-1}$ for an appropriate nonconst polynomial $p_{41}$ positive everywhere on $\mathbb{R}$ for all $x \in \mathbb{R}$, except for a set with a finite measure.*

*Actually, both corollaries 2, 3 could be formulated in the stronger ways using the theorem rather than the corollary. One could consider corollary 2 as a lower bound on the parallel complexity (i.e. the depth) of a function, computed by a sigmoid, approximating a given one. Corollary 3 could be treated as a lower bound on the approximation of a function computed by a sigmoid by means of a piecewise rational function. Note also that the bounds in the corollaries 2, 3 are sharp: take $f = (\exp^{(d)})^{-1}$ and $g = h = 0$.*

## 6. Deviation theorems for elementary sigmoids

Recall (see the section 1) that the considered gate (real) functions $u$ from $\Gamma$ were defined everywhere on $\mathbb{R}$ (out of a finite number of singular points), and moreover we required that $u$ was the restriction on $\mathbb{R}$ of some branch of an analytic complex function.

One can weaken this requirement and consider a function $u$ such that for a certain finite union of (possibly infinite) intervals $N \subset \mathbb{R}$ the function $u$ is real on $N$, i.e. $u : N \to \mathbb{R}$ and $u$ is the restriction on $\mathbb{R}$ of some branch of an analytic complex function. We require that at every computational step (2) of the sigmoid the values of the function $(f/g)(w_i^{(1)}, \dots, X)$ should belong to $N$ everywhere on $\mathbb{R}$ (out of singularities), thus $w_{i+1}^{(j)} : \mathbb{R} \to \mathbb{R}$. One can easily (almost literally) extend the proof of the theorem (and the corollary) to this situation.

Under elementary sigmoid we understand a sigmoid in which the role of the gate functions can play exp, log and algebraic functions (in particular, they contain "standard" sigmoids with the gate function $(\exp(-x) + 1)^{-1}$). Each of these gate functions satisfy a certain linear ordinary differential equation with polynomial coefficients. The above requirement when, for example, $u = \log$ means that $(f/g)(w_i^{(1)}, \dots, X)$ is positive everywhere on $\mathbb{R}$. Thus, the corollary implies

COROLLARY 4. *Let a function $f$ be computed by an elementary sigmoid with a depth $d$ with the gate functions* exp, log *and algebraic functions with the requirement that at each computational step of the sigmoid (see (2)) the function $u((f/g)(w_i^{(1)}, \dots, X))$ takes only real values on $\mathbb{R}$. Then there exists a polynomial $p_{42}$ such that the measure of the points $x$ for which one of the inequalities $(\exp^{(d)} p_{42}(x))^{-1} \leq |f(x)| \leq \exp^{(d)} p_{42}(x)$ fails, is finite.*

*The statements similar to the remark just after the corollary and to corollaries 2, 3 are true also for the elementary sigmoids.*

**Acknowledgements** The author would like to thank M. Singer for valuable remarks used in the proofs of lemma 1, 6 and G. Schnitger for interesting discussions.

## REFERENCES


[BBH] Beukers, F., Brownawell, D., Heckman G., *Siegel normality*, Ann. Math., 127 (1988), pp. 279–308.

[BSS] Blum, L., Shub, M., Smale, S., *On a theory of computation and complexity over the real numbers: NP-completeness, recursive functions and universal machines*, Bull. AMS, 21, 1 (1989), pp. 1–46.

[CC] Chudnovsky, D., Chudnovsky, G., *The Wronskian formalism for linear differential equations and Pade approximations*, Adv. Math., 53 (1984), pp. 28–54.

[G90a] Grigoriev, D., *Complexity of factoring and GCD calculating of ordinary linear differential operators*, J. Symb. Comput., 10 (1990), pp. 7–37.

[G90b] Grigoriev, D., *Complexity of irreducibility testing for a system of linear ordinary differential equations*, Proc. Int. Symp. Symb. Alg. Comput., ACM, Tokyo (1990), pp. 225–230.

[H] Hartman, P., Ordinary differential equations, Birkhäuser, 1982.

[K] Kaplansky, I., An introduction to differential algebra, Hermann, 1957.

[MSS] Maass, W., Schnitger, G., Sontag, E., *On the computational power of sigmoid versus boolean threshold circuits*, Proc. 32 FOCS, IEEE (1991), pp. 767–776.

[S81] Singer, M., *Liouvillean solutions of $n$-th order linear differential equations*, Amer. J. Math, 103 (1981), pp. 661–682.

[S90] Singer, M., *Formal solutions of differential equations*, J. Symb. Comput., 10 (1990), pp. 59–94.



Departments of Computer Science and Mathematics
Penn State University
State College, PA 16802 USA

on leave from
Mathematical Institute
Fontanka 27
St. Petersburg 191011 RUSSIA