DEVIATION THEOREMS FOR SOLUTIONS OF DIFFERENTIAL EQUATIONS AND APPLICATIONS TO LOWER BOUNDS ON PARALLEL COMPLEXITY OF SIGMOIDS

(Extended abstract)

D. Yu. Grigoriev

Departments of Computer Science and Mathematics
Pennsylvania State University
University Park, PA 16802 USA
e-mail: dima@cs.psu.edu

Under a sigmoid with a depth d we understand a circuit with d layers where each real function computed at $(i+1)$-th layer is obtained as $G(q)$ where q is a rational expression in the functions computed at i-th layer and G is a gate operator from some admitted family. Two types of the families of gate operators are considered: first, we admit to substitute $g(q)$ where g is a solution of a linear ordinary differential equation with the polynomial coefficients and second, as $G(q)$ we take a solution of nonlinear first-order differential equation. The sigmoids of the first type compute any composition of the functions like exp, log, sin (thus, it includes, in particular, standard sigmoids corresponding to the gate $g = (1 + \exp(-x))^{-1}$), the sigmoids of the second type compute Pfaffian functions. The main result states that if two different functions f_1, f_2 are computed by means of the sigmoids with the parallel complexity d, then the difference $|f_1 - f_2|$ grows not slower than $(\exp^{(d)}(p))^{-1}$ (and not faster than $\exp^{(d)}(p)$) where $\exp^{(d)}$ is d times iteration of the exponential function and p is a certain polynomial, thus one can not rather good approximate f_1 with a precise parallel complexity d by means of a function f_2 with a less parallel complexity. Also we estimate the number of zeroes in the intervals of a function computed by a sigmoid of the first type. All the obtained bounds are sharp.

1. Deviation theorems for the functions computed by sigmoids.

Denote the ring $K_0 = \mathbb{R}[X]$, $F_0 = \mathbb{R}(X)$, $D = d/dX$. By Γ denote the set of real functions $u : \mathbb{R} \to \mathbb{R}$ being solutions of linear ordinary differential equations of the kind

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

where the coefficients $a_j \in F_0$ are defined everywhere on $\mathbb{R}$, in other words, their denominators have no real roots. The elements of Γ will play the role of gate functions in the sigmoids. As the operator L has no real singularities ([H]), the function u is analytic on $\mathbb{R}$ (actually, one could get rid of this requirement and consider gate functions with real singularities, but we shall not dwell on it for the sake of simplifying the exposition). Now we define K_i and F_i by induction on i, namely K_{i+1} for $i \ge 0$ is a differential ring [K] generated by the functions of the form $u(q)$ where $u \in \Gamma$ and $q \in F_i$. Define F_{i+1} as a (differential) field of quotients of K_{i+1}.

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

$$w_{i+1}^{(j)} = u((g_1/g_2)(w_i^{(1)}, w_i^{(2)}, \dots, X)) \qquad (2)$$

for a certain gate function $u \in \Gamma$ and $g_1, g_2 \in \mathbb{R}[W_i^{(1)}, W_i^{(2)}, \dots, X]$ 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.

Let $u = u_1, \dots, u_n$ where $u_\ell : \mathbb{R} \to \mathbb{R}$, $1 \le \ell \le n$ be a basis (over $\mathbb{R}$) of the space of solutions of the equation (1) [H]. Extend the sigmoid without changing its depth adding also the instructions $u_\ell((g_1/g_2)(w_i^{(1)}, w_i^{(2)}, \dots, X))$ at $(i+1)$-th layer. One can show by induction on i that each function $w_{i+1}^{(j)}$ belongs to K_{i+1} (and conversely, any element from K_{i+1} can be obtained as a polynomial in the functions computed at $(i+1)$-th layer of a suitable sigmoid). Usually, u is taken from a certain subset of Γ, for example, in the case of the standard sigmoid one takes $u = \exp(-X)$ (see [MSS]).

Henceforth, we fix a sigmoid and by $\mathcal{D}_i \subset K_i$ denote a differential ring generated over $\mathbb{R}(X)$ by $w_i^{(1)}, w_i^{(2)}, \dots$; so as an algebraic ring $\mathcal{D}_i$ is generated by all the derivatives $w_i^{(1)}, Dw_i^{(1)}, \dots, w_i^{(2)}, Dw_i^{2)} \dots$. Denote by $\exp^{(d)} \in K_d$ the iteration of the exponential function d times. Now we are able to formulate the first main result of the paper (deviation theorem for the functions computed by sigmoids).

THEOREM 1. *Let a function $0 \not\equiv f$ be computed by a sigmoid with a depth d. For any ρ_1 there exist ρ_0, ρ_2 where $\rho_0, \rho_1, \rho_2 \in K_0$ are univariate nonconstant polynomials, being positive everywhere on $\mathbb{R}$ such that for any $x \in \mathbb{R}$ the measure of the points y from an interval $I = (x - (\rho_1(x))^{-1}, x)$ at which $|f(y)| \ge \exp^{(d)}(\rho_0(x))$ or $|f(y)| \le (\exp^{(d)} \rho_0(x))^{-1}$ is less than $\frac{(\rho_1(x))^{-1}}{\exp(\rho_2(x))} = \frac{|I|}{\exp(\rho_2(x))}$.*

Corollary 1. *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 point y with $|y| \geq x_0$ for any $x_0 \geq 0$ is less than $(\exp \rho_4(x_0))^{-1}$ for suitable nonconstant polynomials $\rho_3, \rho_4 \in K_0$ being positive everywhere on $\mathbb{R}$.*

Remark 1. The polynomials $\rho_0, \rho_2, \rho_3, \rho_4$ could be calculated explicitly in terms of the size of the sigmoid and in the coefficients a_j of the differential operators L (see (1)) to which satisfy the gate functions u occuring in the sigmoid.

Remark 2. The bounds in the theorem 1 and in the corollary 1 are sharp. As an example consider a function $f = \sin \cdot (\exp^{(d)})^{-1}$ with the parallel sigmoidal complexity equal to d: the set of the points y at which $|f(y)| \leq (\exp^{(d)} \rho_5(y))^{-1}$ consists of a union of intervals where n-th interval $(n = 0, 1, \dots)$ has a length $(\exp^{(d)} \rho_6(n))^{-1}$ and contains the point πn (for appropriate polynomials $\rho_5, \rho_6 \in K_0$).

One can treat the theorem 1 and the corollary 1 as the impossibility of "rather good" approximation of a function with the parallel sigmoidal complexity d by means of a function with less parallel complexity (in particular, by a rational function), thus if such an approximation does exist, it gives a lower bound on the parallel sigmoidal complexity.

The corollary 1 could be easily extended to the sigmoids with branching instructions as the resulting function would be piecewise and one could apply the corollary to each piece. In particular, when we consider only rational computations, it gives a lower bound (the similar as in the corollary 1) on the approximation by means of Blum-Shub-Smale computation ([BSS]).

Finally, we estimate the number of zeroes of a function computed by a sigmoid.

In the next proposition let us adopt a convention that $\exp^{(-1)} \equiv \text{const}$.

Proposition. *Let a function f be computed by a sigmoid with a depth $d \geq 1$. There exists a set $J \subset \mathbb{R}$ with a finite measure such that for any $x \in \mathbb{R}$ the number of zeroes of f in the set $[0, x] \setminus J$ does not exceed $\exp^{(d-1)} \rho_7(x)$ for a suitable polynomial $\rho_7 \in K_0$, moreover the intersection $[0, x] \cap J$ is a union of at most $\exp^{(d-2)} \rho_7(x)$ intervals.*

2. Upper bounds on the functions computed by sigmoids.

¿From now on $p_1, p_2, \dots$ will denote polynomials from K_0 each having a form $p_j = \overline{p_j}(X^2)$ where a polynomial $\overline{p_j}$ monotonically increases on $\mathbb{R}^+$ and $\overline{p_j}(0) \geq 1$. The proof of the following lemma is based on the Gronwall's inequality [H]. Let u satisfy (1).

Lemma 1. *For each $j \geq 0$ there exists a polynomial $p_j^{(0)}$ such that $|D^j u| \leq \exp(p_j^{(0)})$.*

The proof of the theorem 1 is conducted by induction on d. The next lemma serves to get upper bounds in the inductive step, its proof relies on (2) and lemma 1.

Lemma 2. *Let $0 \leq i < d$. For a family of differential polynomials $G_1, \dots, G_k \in \mathcal{D}_{i+1}$ one can produce a family of differential polynomials $0 \not\equiv H_0, \dots, H_m \in \mathcal{D}_i$ such that for* *every p_1, p_2 there exists p_3 satisfying the following property: for arbitrary $x \in \mathbb{R}$ if the inequalities $|H_0| \geq (\exp^{(i)}(p_2))^{-1}$, $|H_j| \leq \exp^{(i)}(p_2)$, $0 \leq j \leq m$ hold everywhere on an interval $I_i = (x - (\exp^{(i)} p_1(x))^{-1}, x)$ then $|G_\ell| \leq \exp^{(i+1)}(p_3)$, $1 \leq \ell \leq k$ everywhere on I_i.*

3. Upper and lower bounds on Wronskians of the functions computed by sigmoids.

Denote by W_u the Wronskian of (1) (see [H])

$$W_u = \det \begin{pmatrix} u_1 & \dots & u_n \\ Du_1 & \dots & Du_n \\ \vdots & & \vdots \\ D^{n-1}u_1 & \dots & D^{n-1}u_n \end{pmatrix}$$

As $W_u(x) = W_u(x_0) \exp \int_{x_0}^{x} (-a_{n-1})$ ([H]) we get the following lemma.

LEMMA 3. *For a suitable p_4 $(\exp\ p_4)^{-1} \leq |W_u| \leq \exp\ p_4$.*

A function $w = w_{i+1}^{(j)}$ (see (2)) computed by the sigmoid, satisfies a linear ordinary differential equation $0 = \left(\sum_{0 \leq \ell \leq n} b_\ell D^\ell \right) w$ with the coefficients $b_\ell \in \mathcal{D}_i$. Without loss of generality we can assume that $w \not\equiv$ const, then $u_1((g_1/g_2)(w_i^{(1)}, w_i^{(2)}, \dots, X)), \dots, u_n((g_1/g_2)$ $(w_i^{(1)}, w_i^{(2)}, \dots, X)) \in \mathcal{D}_{i+1}$ (see (2)) constitute a basis (over $\mathbb{R}$) of the space of solutions of this equation. Denote by W_w the Wronskian of this equation. One can prove the following lemma using lemma 3 and the identity

$$W_w = W_u((g_1/g_2)(w_i^{(1)}, w_i^{(2)}, \dots, X)) \cdot (D((g_1/g_2)(w_i^{(1)}, w_i^{(2)}, \dots, X)))^{n(n-1)/2}$$

LEMMA 4. *For each function const $\not\equiv w = w_{i+1}^{(j)}$ computed by the sigmoid (see (2)) one can produce differential polynomials $0 \not\equiv H_0^{(w)}, \dots, H_s^{(w)} \in \mathcal{D}_i$ such that for every p_1, p_5 there exists p_6 satisfying the following property: for arbitary $x \in \mathbb{R}$ if the inequalities $|H_0^{(w)}| \geq (\exp^{(i)} p_5)^{-1}$, $|H_j^{(w)}| \leq \exp^{(i)}(p_5)$, $0 \leq j \leq s$ hold everywhere on an interval $I_i = (x - (\exp^{(i)} p_1(x))^{-1}, x)$ then the Wronskian W_w satisfies inequalities*

$$(\exp^{(i+1)}(p_6))^{-1} \leq |W_w| \leq \exp^{(i+1)}(p_6) \text{ everywhere on } I_i.$$

Let a differential polynomial $G \in \mathcal{D}_{i+1}$. Then G satisfies a certain linear ordinary differential equation $0 = \mathcal{L}G = \left(\sum_{0 \leq j \leq m} h_j D^j \right) G$ with the coefficients $h_j \in \mathcal{D}_i$ and with a basis (over $\mathbb{R}$) of the space of solutions from $\mathcal{D}_{i+1}$ (one could produce the operator $\mathcal{L}$ by induction on the construction of G, see [S]). Denote by W_G the Wronskian of the operator $\mathcal{L}$. The main purpose of this section is to establish the bounds on W_G. An upper bound is provided by applying lemma 2 to W_G and getting a family $\widetilde{H}_0, \ldots, \widetilde{H}_{m_1} \in \mathcal{D}_i$, a lower bound is proved by induction on the construction of G (so, on the number of operations of differentiating, adding and multiplying), lemma 4 gives the base of this induction.

To prove the inductive step we assume that two functions $v_1, v_2 \in \mathcal{D}_{i+1}$ satisfy linear ordinary differential equations $0 = Q_1 v_1 = Q_2 v_2$, where $Q_1 = \sum_{0 \le \ell \le k_1} \alpha^{(\ell)} D^\ell$, $Q_2 = \sum_{0 \le \ell \le k_1} \beta^{(\ell)} D^\ell$ and the coefficients $\alpha^{(\ell)}, \beta^{(\ell)} \in \mathcal{D}_i$. We assume also that by induction some differential polynomials $0 \not\equiv H_0^{(v_1)}, \ldots, H_{æ_1}^{(v_1)}, 0 \not\equiv H_0^{(v_2)}, \ldots, H_{æ_2}^{(v_2)} \in \mathcal{D}_i$ are produced such that for every p_7 there exists p_8 satisfying the following property for any $x \in \mathbb{R}$: if

$$|H_0^{(v_1)}|,\ |H_0^{(v_2)}| \ge (\exp^{(i)} p_7)^{-1};\ |H_j^{(v_1)}|,\ |H_j^{(v_2)}| \le \exp^{(i)} p_7 \tag{3}$$

for all $j \ge 0$ everywhere on an interval $I_i = (x - (\exp^{(i)} p_1(x))^{-1}, x)$ then

$$(\exp^{(i+1)} p_8)^{-1} \le |W_{v_1}|,\ |W_{v_2}| \le \exp^{(i+1)} p_8 \tag{4}$$

everywhere on I_i, here W_{v_1}, W_{v_2} denote the Wronskians of the operators Q_1, Q_2, resp.

One can produce (cf. [S]) the linear ordinary differential operators $Q_D, Q_+, Q_\times$ of the minimal orders with the coefficients from $\mathcal{D}_i$, namely, being differential polynomials in $\alpha^{(\ell)}, \beta^{(\ell)}$ and with basis of the spaces of solutions from $\mathcal{D}_{i+1}$ such that $0 = Q_D(D\overline{v}_1) = Q_+(\overline{v}_1 + \overline{v}_2) = Q_\times(\overline{v}_1 \overline{v}_2)$ for all the solutions of the equations $0 = Q_1 \overline{v}_1 = Q_2 \overline{v}_2$. The main task is to estimate their Wronskians W_{Dv_1}, $W_{v_1+v_2}$, $W_{v_1 v_2}$, that would prove bounds (4) for the inductive step. As estimating W_{Dv_1} is comparatively easy and on the other hand considering $W_{v_1+v_2}$ and $W_{v_1 v_2}$ are similar, let us dwell on estimating $W_{v_1+v_2}$.

Replace the equations $0 = Q_1 \overline{v}_1 = Q_2 \overline{v}_2 = Q_+(\overline{v}_1 + \overline{v}_2)$ by the corresponding first-order linear systems $DV_1 = A_1 V_1$, $DV_2 = A_2 V_2$, $DV_+ = A_+ V_+$, resp. where the matrix

$$A_1 = \begin{pmatrix} 0 & 1 & & \bigcirc \\ & 0 & \ddots & \\ & & & \\ & \bigcirc & & 1 \\ -\frac{\alpha^{(0)}}{\alpha^{(k_1)}} & -\frac{\alpha^{(1)}}{\alpha^{(k_1)}} & \cdots & -\frac{\alpha^{(k_1-1)}}{\alpha^{(k_1)}} \end{pmatrix}$$

is in the Frobenius form (see [H]), the similar for A_2, A_+. Denote by $\mathcal{V}_1, \mathcal{V}_2, \mathcal{V}_+$ the corresponding spaces (over $\mathbb{R}$) of solutions of the linear systems. Consider a natural epimorphism $\sigma : \mathcal{V}_1 \oplus \mathcal{V}_2 \to \mathcal{V}_+$ mapping $\sigma((v_1^{(j_1)}, Dv_1^{(j_1)}, \ldots, D^{k_1-1} v_1^{(j_1)}) \oplus (v_2^{(j_2)}, Dv_2^{(j_2)}, \ldots, D^{k_2-1} v_2^{(j_2)})) \to (v_1^{(j_1)} + v_2^{(j_2)}, D(v_1^{(j_1)} + v_2^{(j_2)}), D^2(v_1^{(j_1)} + v_2^{(j_2)}), \ldots)$. The direct sum $\mathcal{V}_1 \oplus \mathcal{V}_2$ is the space of solutions of the system $DV = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} V$. The subspace $\mathrm{Ker}(\sigma) \subset \mathcal{V}_1 \oplus \mathcal{V}_2$

is invariant under the differential Galois group of the latter system (see [K], also [BBH], [G90b]). Therefore, any nonsingular linear transformation of the space $\mathcal{V}_1 \oplus \mathcal{V}_2$ being of the form $\overline{\sigma} = \begin{pmatrix} * \\ \sigma \end{pmatrix}$ where $*$ is a matrix with the entries from $\mathbb{R}$ reduces (see e.g. [BBH]) the system $DV = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} V$ to the block-triangular form $D\overline{V} = \begin{pmatrix} C_1 & 0 \\ C_2 & C_3 \end{pmatrix} \overline{V}$, where $\overline{V} = \overline{\sigma} V$. The space of solutions of the system $DV_3 = C_3 V_3$ coincides with $\mathrm{Ker}(\sigma)$ and the space of solutions of the system $D\overline{V}_1 = C_1 \overline{V}_1$ equals to $\mathcal{V}_+$ (in [G90a], [G90b] one can find the complexity bounds on reducing a system to the block-triangular form).

Using the formula for the Wronskian $W_{C_1} = \exp \int \mathrm{tr} C_1$ ([H]), we obtain equalities

$$W_{C_1} W_{C_3} = W_{\begin{pmatrix} C_1 & 0 \\ C_2 & C_3 \end{pmatrix}} = (\det \overline{\sigma}) \cdot W_{\begin{pmatrix} B_1 & 0 \\ 0 & B_2 \end{pmatrix}} = (\det \overline{\sigma}) W_{v_1} W_{v_2}.$$

As the coefficients of the vectors from $\mathrm{Ker}(\sigma) \subset \mathcal{V}_1 \oplus \mathcal{V}_2$ belong to $\mathcal{D}_{i+1}$ one can apply to them lemma 2 and get a family $H_0^{(\sigma)}, \dots, H_{m_2}^{(\sigma)} \in \mathcal{D}_i$. Also $\det(\overline{\sigma}) = \overline{H}_0 / \overline{H}_1$ for suitable $\overline{H}_0, \overline{H}_1 \in \mathcal{D}_i$. Finally, using (3) we take $H_0^{(v_1+v_2)} = H_0^{(v_1)} \cdot H_0^{(v_2)} \cdot \widetilde{H}_0 \cdot H_0^{(\sigma)} \cdot \overline{H}_0$ and as $H_1^{(v_1+v_2)}, \dots$ we take the union of $H_j^{(v_1)}, H_j^{(v_2)}, \widetilde{H}_j, H_j^{(\sigma)}, \overline{H}_0, \overline{H}_1$ for all $j \geq 0$ and using (4) prove the inductive step for $v_1 + v_2$. Namely, for every p_9 there exists p_{10} satisfying the following property: for any $x \in \mathbb{R}$ (cf. (3), (4)) if $|H_0^{(v_1+v_2)}| \geq (\exp^{(i)}(p_9))^{-1}$, $|H_j^{(v_1+v_2)}| \leq \exp^{(i)}(p_9)$ for all $j \geq 0$ everywhere on an interval I_i then $(\exp^{(i+1)}(p_{10}))^{-1} \leq |W_{v_1+v_2}| \leq \exp^{(i+1)}(p_{10})$ everywhere on I_i.

This completes the consideration of the inductive step for the sum $v_1 + v_2$. The bound on $W_{v_1 v_2}$ is proved in a similar way, the role of the direct sum $\mathcal{V}_1 \oplus \mathcal{V}_2$ is being replaced by the tensor product $\mathcal{V}_1 \otimes_{\mathbb{R}} \mathcal{V}_2$ and the role of the matrix $\begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}$ is played by $A_1 \otimes E_{k_2} + E_{k_1} \otimes A_2$ where E_{k_1} denotes the unit $k_1 \times k_1$ matrix.Thus, by induction on the construction of the differential polynomial $G \in \mathcal{D}_{i+1}$ we get the following lemma.

LEMMA 5. *For every differential polynomials $G_0 = G, G_1, \dots, G_{æ} \in \mathcal{D}_{i+1}$ one can produce differential polynomials $0 \not\equiv H_0, \dots, H_\eta \in \mathcal{D}_i$ such that for any polynomials p_1, p_{11} there exists a polynomial p_{12} satisfying the following property: for arbitrary $x \in \mathbb{R}$ if $|H_0| \geq (\exp^{(i)} p_{11})^{-1}$, $|H_\ell| \leq \exp^{(i)} p_{11}$, $0 \leq \ell \leq \eta$ hold everywhere on an interval $I_i = (x - (\exp^{(i)} p_1(x))^{-1}, x)$ (cf. lemmas 2,4) then $|G_\ell| \leq \exp^{(i+1)}(p_{12})$, $0 \leq \ell \leq æ$ and $|W_G| \geq (\exp^{(i+1)} p_{12})^{-1}$ everywhere on I_i, where W_G denotes the Wronskian of a certain linear ordinary differential equation $0 = \left(\sum_{0 \leq j \leq m} \gamma^{(j)} D^j \right) G$ with the coefficients $\gamma^{(j)} \in \mathcal{D}_i$ and with a basis (over $\mathbb{R}$) of the space of solutions $G_{0,0,\dots,}G_{0,m-1} \in \mathcal{D}_{i+1}$.*

4. Lower bounds on functions computed by a sigmoid.

Relying on a lower bound on the Wronskian W_G (see lemma 5) one can obtain a lower bound on a differential polynomial $0 \not\equiv G \in \mathcal{D}_{i+1}$

Lemma 6. *For every differential polynomials* $0 \not\equiv G_0 = G, G_1, \dots, G_{æ} \in \mathcal{D}_{i+1}$ *one can produce differential polynomials* $0 \not\equiv \mathcal{H}_0, \dots, \mathcal{H}_\nu \in \mathcal{D}_i$ *such that for any polynomials* p_1, p_{13} *there exist polynomials* p_{14}, p_{15}, p_{16} *satisfying the following property: for arbitrary* $x \in \mathbb{R}$ *if* $|\mathcal{H}_0| \geq (\exp^{(i)} p_{13})^{-1}$, $|\mathcal{H}_\ell| \leq \exp^{(i)} p_{13}$, $0 \leq \ell \leq \nu$ *hold everywhere on an interval* $I_i = (x - (\exp^{(i)} p_1(x))^{-1}, x)$, *then* $|G_\ell| \leq \exp^{(i+1)} p_{14}$ *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* $|I_{i+1}^{(\alpha)}| = (\exp^{(i+1)} p_{15}(x))^{-1}$, *moreover the lower bound* $|G| \geq (\exp^{(i+1)} p_{14}(x))^{-1}$ *holds everywhere on* $I_{i+1}^{(\alpha)}$ *for each* α *and finally* $\sum_\alpha |I_{i+1}^{(\alpha)}| \geq |I_i|(1 - (\exp^{(i+1)} p_{16}(x))^{-1})$. *In addition, the complement* $I_i \setminus \cup_\alpha I_{i+1}^{(\alpha)}$ *consists of at most* $\exp^{(i)} p_{15}(x)$ *intervals and* G *has at most* $\exp^{(i)} p_{15}(x)$ *zeroes in* I_i.

The latter inequality informally means that the desired lower bound on $|G|$ *holds "almost everywhere" on* I_i.

To prove lemma 6 first apply lemma 5 to $G_0, \dots, G_{æ}$ *and produce* $H_0, \dots, H_\eta$. *Then taking an equation* $0 = (\sum_{0 \leq j \leq m} \gamma^{(j)} D^j)G$ *from lemma 5 produce an equation* $0 = (\sum_{0 \leq j \leq m} \sigma^{(j)} D^j) DG$ *with the coefficients* $\sigma^{(j)} \in \mathcal{D}_i$, $\gamma^{(m)} = \sigma^{(m)} \not\equiv 0$. *Apply lemma 2 to a family* $\{D^j G_{0,\ell}\}_{0 \leq j, \ell \leq m-1}$ *(see lemma 5) and get a family* $\overline{H}_0, \dots, \overline{H}_\mu \in \mathcal{D}_i$. *As the required in the lemma differential polynomials take* $\mathcal{H}_0 = H_0 \overline{H}_0 \sigma^{(m)}$ *and as* $\mathcal{H}_1, \dots, \mathcal{H}_\nu$ *take the union of* $H_0, \dots, H_\eta, \overline{H}_0, \dots, \overline{H}_\mu, \sigma^{(0)}, \dots, \sigma^{(m)}$. *Hence lemmas 2, 5 imply the existence of polynomials* $p_{17}, p_{18}, p_{19}, p_{20}$ *such that inequalities* $|\sigma^{(m)}| \geq (\exp^{(i)} p_{17})^{-1}$, $|W_G| \geq (\exp^{(i+1)} p_{18})^{-1}$, $|G_\ell| \leq \exp^{(i+1)} p_{19}$, $|D^j G_{0,\ell}| \leq \exp^{(i+1)} p_{20}$, $|\beta^{(\ell)}| \leq \exp^{(i)} p_{17}$ *hold everywhere on* I_i. *When* $G \equiv const$, *these inequalities give the lemma, so assume that* $G \not\equiv const$.

We claim that G *takes every value* ϵ *in the interval* I_i *at most* $\max\{m+1, 2(m+1)|I_i| \exp^{(i)} p_{17}^2(x)\}$ *times. Suppose the contrary. Then there exists a subinterval* $I \subset I_i$ *of a length* $\lambda = \min\{|I_i|, (2\exp^{(i)} p_{17}^2(x))^{-1}\}$ *in which* G *takes value* ϵ *at least* $m+1$ *times. Therefore, each derivative* $DG, \dots, D^m G$ *has a least one root in the interval* I. *Denote* $M^{(j)} = \max_{y \in I} |D^j G(y)|$. *Then* $\lambda M^{(j+1)} \geq M^{(j)}$, $1 \leq j \leq m$. *Let* $D^{m+1}G$ *reach* $M^{(m+1)}$ *at a point* $y_0 \in I$. *Then* $M^{(m+1)} = \left|(\sigma^{(m)}(y_0))^{-1} \left(\sum_{0 \leq j \leq m-1} \sigma^{(j)} D^{j+1} G(y_0)\right)\right| \leq (\exp^{(i)} p_{17}(x))^2 M^{(m+1)} (\lambda + \lambda^2 + \cdots + \lambda^m) < M^{(m+1)}$, *the contradiction proves the claim.*

Construct a sequence of polynomials $r_0, \dots, r_{m-1} \in K_0$ by (inverse) induction: set $r_{m-1} = p_{18} + (m-1)p_{20} + m^2$, and $r_j = 2r_{j+1} + p_{20} + 2$, $0 \leq j < m-1$, then $r_0 \geq r_1 \geq \cdots \geq r_{m-1}$ everywhere on $\mathbb{R}$. First, assume that at some point $x_1 \in I_i$ inequalities $|D^j G(x_1)| \leq (\exp^{(i+1)} r_j(x))^{-1} \leq (\exp^{(i+1)} r_{m-1}(x))^{-1}$ hold for all $0 \leq j \leq m-1$. Then expanding Wronskian W_G with respect to the column consisting of $G, DG, \dots, D^{m-1}G$ (as $G \not\equiv 0$ we can take G as one of the elements of a basis of the space of solutions of $0 = (\sum_{0 \leq j \leq m} \gamma^{(j)} D^j)G$ we obtain inequalities $(\exp^{(i+1)} p_{18}(x))^{-1} \leq$

$|W_G(x_1)| \leq (\exp^{(i+1)} r_{m-1}(x))^{-1} m! (\exp^{(i+1)} p_{20}(x))^{m-1}$, that contradicts to the choice of r_{m-1}.

Consider a subinterval $I^{(0)} \subset I_i$ on which $|G| \leq (\exp^{(i+1)} r_0(x))^{-1}$ everywhere. Take the minimal $1 \leq j_0 \leq m-1$ such that there exists a point $x_0 \in I^{(0)}$ for which $|D^{j_0} G(x_0)| \geq (\exp^{(i+1)} r_{j_0}(x))^{-1}$. Since $|D^{j_0+1} G| \leq \exp^{(i+1)} p_{20}(x)$ everywhere on I_i we get for arbitrary $x_2 \in I_i$ inequalities $|D^{j_0} G(x_2)| \geq |D^{j_0} G(x_0)| - |x_2 - x_0| \exp^{(i+1)} p_{20}(x) \geq (\exp^{(i+1)} r_{j_0}(x))^{-1} - |x_2 - x_0| \exp^{(i+1)} p_{20}(x)$. Assume that at least one of two points $x_3 = x_0 \pm (\exp^{(i+1)}(r_{j_0} + p_{20})(x))^{-1}$ belong to $I^{(0)}$, then $|D^{j_0-1} G(x_0) - D^{j_0-1} G(x_3)| = |\int_{x_0}^{x_3} D^{j_0} G| \geq (2 \exp^{(i+1)}(r_{j_0} + p_{20})(x))^{-1} (\exp^{(i+1)} r_{j_0}(x))^{-1} \geq 2(\exp^{(i+1)} r_{j_0-1}(x))^{-1}$, that leads to a contradiction with the minimality of j_0. Thus, neither of two considered points belong to $I^{(0)}$, therefore, $|I^{(0)}| \leq 2(\exp^{(i+1)}(r_{j_0} + p_{20})(x))^{-1}$.

Partition the interval I_i on the subintervals with the endpoints in which G takes the values $\pm(\exp^{(i+1)} r_0(x))^{-1}$. By the proved above there are at most $2 \max\{m+1, 2(m+1)|I_i| \exp^{(i)} p_{17}^2(x)\}$ such subintervals. Also we have proved that the length of any subinterval on which $|G| \leq (\exp^{(i+1)} r_0(x))^{-1}$ everywhere, is less than $2(\exp^{(i+1)}(r_{j_0} + p_{20})(x))^{-1}$ (this is used in the proof of the proposition from the section 1), partitioning all the other subintervals into disjoint subintervals $I_{i+1}^{(\alpha)}$ completes the proof of lemma 6.

Finally, one can prove theorem 1 (and as well the proposition). First, apply lemma 6 to $i = d-1$ and a family of differential polynomials consisting of a single element $f \in \mathcal{D}_d$, then to the obtained family from $\mathcal{D}_{d-1}$ and $i = d-2$ again apply lemma 6 and so on decreasing i until we get a family of the rational functions from $\mathcal{D}_0 = \mathbb{R}(X)$. Then we ascertain the necessary bounds by induction on (increasing i) again using lemma 6 for the inductive step.

5. Deviation theorems for the functions computed by Pfaffian sigmoids.

Consider another class of sigmoids which are called Pfaffian and which also contain, in particular, "standard" sigmoids. Denote the field $P_0 = \mathbb{R}(X)$, then by induction on i the field P_{i+1} is generated over P_i by all the functions $v_{i+1}^{(j)} : \mathbb{R} \to \mathbb{R}$ (possibly having a finite number of singularities) satisfying first-order nonlinear differential equations of the form

$$D\, v_{i+1}^{(j)} = q(v_{i+1}^{(j)}) \tag{5}$$

where a polynomial $q(Z) \in P_i[Z]$. Obviously $\exp^{(i)} \in P_i$.

According to [Kh] any function $f \in P_i$, being Pfaffian, has a finite number of singularities and roots. Hence for every two functions $f_1, f_2 \in P_i$, $f_1 \not\equiv f_2$, the difference $(f_1 - f_2)(x)$ is either positive or negative everywhere on an interval $x \in [x_0, \infty)$ for a certain $x_0 \in \mathbb{R}$, we write $f_1 \succ f_2$ or $f_1 \prec f_2$, respectively. Now we can formulate the deviation theorem for Pfaffian sigmoids.

THEOREM 2. *For any function $0 \not\equiv f \in P_i$ there exists a polynomial p_{21} such that*

$$(\exp^{(i)} p_{21})^{-1} \prec |f| \prec \exp^{(i)} p_{21}$$

The bounds are obviously sharp. For Pfaffian sigmoids the necessary bounds are valid starting with some point x_0 unlike corollary 1 where the bounds were valid out of a finite-measure set. Analogues of the remark 1 and the remark concerning Blum-Shub-Smale model are true also for Pfaffian sigmoids.

The theorem is proved by induction on i and firstly we prove an upper bound (for $i = 0$ one can find its proof in [B]).

Lemma 7. *Assume that the statement of the theorem 2 is proved for P_i and $v_{i+1}^{(j)}$ satisfies (5) where $\deg(q) = n$. Then for an appropriate polynomial p_{22}*

a) *if $n = 0$ or $n \geq 2$ then $|v_{i+1}^{(j)}| \prec \exp^{(i)} p_{22}$*

b) *if $n = 1$ then $|v_{i+1}^{(j)}| \prec \exp^{(i+1)} p_{22}$*

Let each of the functions $v_{i+1}^{(j_1)}, \dots, v_{i+1}^{(j_m)} \in P_{i+1}$ satisfy an equation similar to (5), namely $Dv_{i+1}^{(j_\ell)} = q_\ell(v_{i+1}^{(j_\ell)})$. Then for any polynomial $0 \not\equiv h \in P_i[Z_1, \dots, Z_m]$ the bound $|h(v_{i+1}^{(j_1)}, \dots, v_{i+1}^{(j_m)})| \prec \exp^{(i+1)} p_{23}$ holds for a suitable polynomial p_{23} because of lemma 7. Thus, to prove theorem 2 it remains to prove a lower bound on $h(v_{i+1}^{(j_1)}, \dots, v_{i+1}^{(j_m)}) \in P_{i+1}$.

Firstly, we consider the case when $v_{i+1}^{(j_1)}, \dots, v_{i+1}^{(j_m)}$ are algebraically independent over P_i. Suppose that $|h(v_{i+1}^{(j_1)}, \dots, v_{i+1}^{(j_m)})| \prec (\exp^{(i+1)} p)^{-1}$ for all the polynomials p. Then we say that $h(v_{i+1}^{(j_1)}, \dots, v_{i+1}^{(j_m)})$ is small. Also we suppose that m is the least possible with this property. Finally, without loss of generality, one can suppose that the polynomial h is irreducible over P_i.

Since the derivative $D(h(v_{i+1}^{(j_1)}, \dots, v_{i+1}^{(j_m)})) = \sum\limits_{1 \leq \ell \leq m} \frac{\partial h}{\partial v_{i+1}^{(j_\ell)}} q_\ell(v_{i+1}^{(j_\ell)}) = g(v_{i+1}^{(j_1)}, \dots, v_{i+1}^{(j_m)})$ $\in P_{i+1}$ for a certain polynomial $g \in P_i[Z_1, \dots, Z_m]$, the derivative should be also small (as being also a Pfaffian function). If $h \nmid g$ in the ring $P_i[Z_1, \dots, Z_m]$ then there exist polynomials $h_1, g_1 \in P_i[Z_1, \dots, Z_m]$ such that $0 \not\equiv hh_1 + gg_1 \in P_i[Z_1, \dots, Z_{m-1}]$. But then the function $(hh_1 + gg_1)(v_{i+1}^{(j_1)}, \dots, v_{i+1}^{(j_m)})$ is small by virtue of lemma 7, this contradicts to the choice of m.

Now let $g = h\, g_0$ for some $g_0 \in P_i[Z_1, \dots, Z_m]$. Consider any $1 \leq \ell \leq m$ for which $\deg_Z(q_\ell) \leq 1$, then for each $1 \leq s \leq m$

$$\deg_{v_{i+1}^{(j_\ell)}} \left(\frac{\partial h}{\partial v_{i+1}^{(j_s)}} q_s(v_{i+1}^{(j_s)}) \right) \leq \deg_{v_{i+1}^{(j_\ell)}} (h(v_{i+1}^{(j_1)}, \dots, v_{i+1}^{(j_m)}))$$

and therefore, Z_ℓ does not occur in the polynomial g_0. If for some $1 \leq \ell \leq m$ $\deg_Z(q_\ell) \geq 2$ then lemma 7 entails that $|v_{i+1}^{(j_\ell)}| \prec \exp^{(i)} p_{23}$ for a suitable polynomial p_{23}. Hence $|g_0(v_{i+1}^{(j_1)}, \dots, v_{i+1}^{(j_m)})| \prec \exp^{(i)} p_{24}$ for a certain p_{24}. Thus $|D(h(v_{i+1}^{(j_1)}, \dots, v_{i+1}^{(j_m)}))/ h(v_{i+1}^{(j_1)}, \dots, v_{i+1}^{(j_m)})| \prec \exp^{(i)} p_{24}$, therefore $|\log |h(v_{i+1}^{(j_1)}, \dots, v_{i+1}^{(j_m)})|\,| \prec \exp^{(i)} p_{25}$ for an appropriate p_{25} and finally $|h(v_{i+1}^{(j_1)}, \dots, v_{i+1}^{(j_m)})| \succ (\exp^{(i+1)} p_{25})^{-1}$, this contradicts to the supposition that $h(v_{i+1}^{(j_1)}, \dots, v_{i+1}^{(j_m)})$ is small.

In the general case choose some transcendental over P_i basis (let it be $v_{i+1}^{(j_1)},\dots,v_{i+1}^{(j_s)}$) among $v_{i+1}^{(j_1)},\dots,v_{i+1}^{(j_m)}$. Then there exists a polynomial $t(Y)=\sum_{0\le\ell\le k} t^{(\ell)}Y^{\ell}\in P_i[v_{i+1}^{(j_1)},\dots,v_{i+1}^{(j_s)}][Y]$ with the coefficients $t^{(\ell)}\in P_i[v_{i+1}^{(j_1)},\dots,v_{i+1}^{(j_s)}]$, $0\le\ell\le k$ and $t^{(0)}\not\equiv 0$ such that $t(h(v_{i+1}^{(j_1)},\dots,v_{i+1}^{(j_m)}))\equiv 0$. Since we have proved that $|t^{(0)}|\succ(\exp^{(i+1)}p_{26})^{-1}$ and by lemma 7 $|t^{(\ell)}|\prec\exp^{(i+1)}p_{26}$, $0\le\ell\le k$ for a suitable p_{26}, we obtain that $|h(v_{i+1}^{(j_1)},\dots,v_{i+1}^{(j_m)})|\succ(1/2)(\exp^{(i+1)}p_{26})^{-2}$ and complete the proof of theorem 2 taking into account that any element of the field P_{i+1} can be represented as a quotient $h^{(1)}(v_{i+1}^{(j_1)},\dots,v_{i+1}^{(j_m)})/h^{(2)}(v_{i+1}^{(j_1)},\dots,v_{i+1}^{(j_m)})$ for some elements $v_{i+1}^{(j_1)},\dots,v_{i+1}^{(j_m)}\in P_{i+1}$ each satisfying an equation of the type (1) and polynomials $h^{(1)},h^{(2)}\in P_i[Z_1,\dots,Z_m]$.

6. Deviation theorems for elementary sigmoids.

By an elementary sigmoid with a depth d we mean a sigmoid like in the section 1 where at the computational step (2) the gate function is either $u=\exp$ or $u=\log$, in the latter case we impose a requirement that $(g_1/g_2)(w_i^{(1)},\dots,X)$ is positive everywhere. Then the function $w_{i+1}^{(j)}$ satisfies an equation of the form (5) and therefore $w_{i+1}^{(j)}\in P_{i+1}$. A function computed by an elementary sigmoid is elementary (cf. [S]). A "standard" sigmoid ([MSS]) is a particular case of an elementary one. Theorem 2 implies the following.

Corollary 2. *Let a function $0\not\equiv f$ be computed by an elementary sigmoid with a depth d. Then for a certain polynomial p_{27} we have $(\exp^{(d)}p_{27})^{-1}\prec|f|\prec\exp^{(d)}p_{27}$.*

The remarks similar to remarks 1, 2 and the remark about Blum-Shub-Smale model from the section 1 are true also for the elementary sigmoids.

Acknowledgements. The author would like to thank M. Singer for valuable remarks used in the proofs of lemmas 1, 5 and G. Schnitger for interesting discussions.

REFERENCES

[B] BELLMAN, R., *Stability theory of differential equations*, McGraw-Hill.

[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.

[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.

[Kh] KHOVANSKY, A., *Fewnomials*, Transl. Math. Monogr., AMS, 88 (1991).

[S] SINGER, M., *Liouvillian solutions of nth order linear differential equations*, Amer. J. Math, 103 (1981), pp. 661–682.