D. Yu. Grigoriev, Deviation theorems for pfaffian sigmoids, *Алгебра и анализ*, 1994, том 6, выпуск 1, 127–131

Алгебра и анализ
Том 6 (1994), вып. 1

DEVIATION THEOREMS FOR PFAFFIAN SIGMOIDS

D. Yu. Grigoriev

Abstract. By a Pfaffian sigmoid of depth d we mean a circuit with d layers in which rational operations are admitted at each layer, and to jump to the next layer one solves an ordinary differential equation of the type $v' = p(v)$ where p is a polynomial whose coefficients are functions computed at the previous layers of the sigmoid. Thus, a Pfaffian sigmoid computes Pfaffian functions (in the sense of A. Khovanskii). A deviation theorem is proved which states that for a real function f, $f \not\equiv 0$, computed by a Pfaffian sigmoid of depth (or parallel complexity) d there exists an integer n such that for a certain x_0 the inequalities $(\exp(\cdots(\exp(|x|^n))\cdots))^{-1} \leqslant |f(x)| \leqslant \exp(\cdots(\exp(|x|^n))\cdots)$ hold for all $|x| \geqslant x_0$, where the iteration of the exponential function is taken d times. One can treat the deviation theorem as an analogue of the Liouville theorem (on algebraic numbers) for Pfaffian functions.

Introduction

By a Pfaffian sigmoid (cf. [MSS],[G]) of depth d we understand a computational circuit having d layers such that at each layer the rational operations are admitted, and to jump to the $(i+1)$-th layer, it is allowed to compute a function $w_{i+1} : \mathbb{R} \to \mathbb{R}$ satisfying a first-order differential equation $w'_{i+1} = p(w_{i+1})$ (cf. [Kh]), where p is a polynomial with coefficients computed at the previous layers of the sigmoid (see (1) and Section 1 for the precise definitions). In particular, a jump can be made by taking $\exp$ or $\log$ of a function computed at previous layers. Pfaffian sigmoids of this kind are called elementary and were considered in [G] along with more general sigmoids in which to jump to the next layer one can substitute a function from a previous layer into a solution of a linear ordinary differential equation with polynomial coefficients, in particular, into $\exp$ or $\log$. Another particular case of elementary sigmoids is presented by "standard" sigmoids (see [MSS]), where the jump is made by an application of the function $(1 + \exp(-x))^{-1}$. So, a function computed at the $(i + 1)$-th layer of a sigmoid introduced in [G], satisfies a linear ordinary differential equation with coefficients from the previous layers $1, \dots, i$. In the present paper the corresponding differential equation can be nonlinear, but it is of special form (see (1) below), ensuring that the computed functions are Pfaffian (see [Kh]).

The main result (see the theorem in Section 1) states that two different functions $f_1, f_2 : \mathbb{R} \to \mathbb{R}$ computed by Pfaffian sigmoids of depth (or parallel complexity) d cannot be too close to each other. Namely, for a suitable polynomial $p \in \mathbb{R}[x]$ and some $x_0 \in \mathbb{R}$,

Key words and phrases. Pfaffian sigmoid, deviation theorems, parallel complexity.

we have $(\exp(\cdots(\exp p(x)\cdots)))^{-1} \leqslant |(f_1 - f_2)(x)| \leqslant \exp(\cdots(\exp p(x)))$ whenever $x \geqslant x_0$; here the number of iterations of the exponential function equals d. In [G] results of this type were proved for sigmoids introduced there and called deviation theorems. The deviation theorem can also be treated as an analogue of the Liouville theorem (concerning the bound for the difference of two distinct algebraic numbers) for the functions computed by Pfaffian sigmoids. In particular, it gives a lower bound for the approximation of a function computed by a Pfaffian sigmoid by means of a rational function (see the corollary in Section 1). On the other hand, one can interpret the theorem and the corollary as certain lower bounds for the depth (parallel complexity) of a Pfaffian sigmoid, provided that it can compute a function admitting a rather good approximation by a "simple" (for example, rational) function.

The proof of the theorem is conducted by induction on the depth of the sigmoid. In Section 2 we deal with an upper bound for a function computed by a Pfaffian sigmoid (see the lemma), and in Section 3 we obtain a lower bound and thus complete the inductive step.

1. Pfaffian sigmoids and differential fields. Let the field P_0 be equal to $\mathbb{R}(X)$; by induction on i, the field P_{i+1} is generated over P_i by all the functions $w_{i+1}^{(j)} : \mathbb{R} \to \mathbb{R}$ (possibly with a finite number of singularities) satisfying a first-order nonlinear differential equation of the form

$$\left(w_{i+1}^{(j)}\right)' = q\left(w_{i+1}^{(j)}\right), \tag{1}$$

where $q(Z)$ is a polynomial belonging to $P_i[Z]$.

According to [Kh], since any function $f \in P_i$ is Pfaffian, it has a finite number of singularities and roots (for $i = 1$ see also [B]). 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 $x_0 \in \mathbb{R}$; these two cases will be denoted $f_1 \succ f_2$ and $f_1 \prec f_2$ respectively. The symbols $p_1, p_2, \dots \in \mathbb{R}[X]$ will stand for polynomials with positive leading coefficient. By $\exp^{(i)} = \exp(\cdots(\exp)\cdots)$ we denote the i-fold iteration of the exponential function. Obviously, $\exp^{(i)}(p_1)$, $(\exp^{(i)}(p_1))^{-1} \in P_i$. Now we are able to formulate the main result of the paper (cf. [G]).

Theorem. *For any function $f \in P_i$, $f \not\equiv 0$, there exists a polynomial p_1 such that*

$$(\exp^{(i)} p_1)^{-1} \prec |f| \prec \exp^{(i)} p_1.$$

The theorem will be proved in the next two sections, now we indicate its relationships with the sigmoids (cf. [MSS]). By a Pfaffian sigmoid of depth d we understand a computation consisting of d layers such that any function $w_{i+1}^{(j)}$ at the $(i+1)$-th layer satisfies an equation of the form (1), where the coefficients q_l of the polynomial $q(Z) = \sum_{0 \leqslant l \leqslant n} q_l Z^l$ are rational functions in the functions of the form $w_i^{(j_1)}$ computed at the previous layers of the sigmoid (cf. [G]). Thus, at each layer of the Pfaffian sigmoid the rational operations are admitted, and the jump from the previous layer to the next one is done by solving an equation of the form (1). Inducion on i shows that any function $w_{i+1}^{(j)}$ computed at the $(i+1)$-th layer of a Pfaffian sigmoid belongs to P_{i+1}.

In the elementary sigmoids considered in [G], the function $w_{i+1}^{(j)}$ is either $\exp(f_1/f_2)$ or $\log(f_1/f_2)$, where f_1, f_2 are polynomials in the functions of the form $w_i^{(j_1)}$ computed at the previous layers of the elementary sigmoid. Clearly, an elementary sigmoid is a particular case of Pfaffian sigmoids. In its turn, the so-called "standard" sigmoid, where the jump from the previous layers to the next one is made by applying the gate function $(1+\exp(-x))^{-1}$ ([MSS]), is a particular case of elementary sigmoids.

The following important question arises: how well can one approximate a function computed by a Pfaffian sigmoid by means of a rational function (cf. [G])?

Corollary. *If a function $w_d \in P_d$ is computed by a Pfaffian sigmoid of depth d, then for any rational function $r \in P_0$ we have $|w_d - r| \succ (\exp^{(d)}(p_2))^{-1}$ for a suitable polynomial p_2, unless $w_d = r$.*

2. Upper bound for a Pfaffian function. We start proving the theorem by induction on i. The base of induction for P_0 being obvious, we proceed to the inductive step. In the present section we prove the required upper bound for the function $w_{i+1}^{(j)}$ (see (1)).

We assume by the inductive hypothesis that the coefficients $q_l \in P_i$, $q_n \neq 0$, of the polynomial q satisfy the inequalities $(\exp^{(i)}(p_2))^{-1} \prec |q_l| \prec \exp^{(i)}(p_2)$, $0 \leqslant l \leqslant n$, for an appropriate polynomial p_2.

If $|w_{i+1}^{(j)}| \prec 4(\exp^{(i)}(p_2))^2$, then the required upper bound is proved. Supposing the contrary, we have

$$\frac{1}{2}(\exp^{(i)}(p_2))^{-1} \prec \left| q_n + \frac{q_{n-1}}{w_{i+1}^{(j)}} + \cdots + \frac{q_0}{(w_{i+1}^{(j)})^n} \right| = \left| \frac{(w_{i+1}^{(j)})'}{(w_{i+1}^{(j)})^n} \right| \prec exp^{(i)}(p_2) + 1$$

If $n = 0$, then $|w_{i+1}^{(j)}(x)| \leqslant C_0 + \int_{x_0}^{x} (\exp^{(i)}(p_2) + 1) \prec \exp^{(i)}(p_3(x))$ for sufficiently large x, say $x > x_0$ (p_3 is a suitable polynomial, $C_0 \in \mathbb{R}$). If $n \geqslant 2$, then $|(w_{i+1}^{(j)}(x))^{-n+1}| \geqslant (n-1)\int_{x}^{\infty} \frac{1}{2}(\exp^{(i)}(p_2))^{-1} \succ (\exp^{(i)}(p_4(x)))^{-1}$ for all sufficiently large x and a suitable polynomial p_4, therefore $|w_{i+1}^{(j)}| \prec (\exp^{(i)}(p_4))$. Finally, if $n = 1$, then $\log|w_{i+1}^{(j)}(x)| \leqslant C_1 + \int_{x_0}^{x} (\exp^{(i)}(p_2) + 1) \prec \exp^{(i)}(p_5(x))$ for all sufficiently large x, $x > x_0$, and a suitable polynomial p_5 ($C_1 \in \mathbb{R}$), therefore $|w_{i+1}^{(j)}| \prec \exp^{(i+1)}(p_6)$ for a suitable polynomial p_6.

We summarize all that in the following lemma (for $i = 0$ one can find its proof in [B]).

Lemma. *Assume that the statement of the theorem is proved for P_i, and that $w_{i+1}^{(j)}$ satisfies* (1), *where* $\deg(q) = n$. *Then*

a) *if $n = 0$ or $n \geqslant 2$, then $|w_{i+1}^{(j)}| \prec \exp^{(i)}(p_7)$;*

b) *if $n = 1$, then $|w_{i+1}^{(j)}| \prec \exp^{(i+1)}(p_7)$*

for an appropriate polynomial p_7.

Remark. For any polynomial $h \in P_i[Y_1, \dots, Y_m]$ we have a similar estimate: $|h(w_{i+1}^{(j_1)}, \dots, w_{i+1}^{(j_m)})| \prec \exp^{(i+1)}(p_8)$. Here the functions $w_{i+1}^{(j_1)}, \dots, w_{i+1}^{(j_m)}$ satisfy certain equations similar to (1), namely $(w_{i+1}^{(j_s)})' = q^{(s)}(w_{i+1}^{(j_s)})$, where $q^{(s)} \in P_i[Z]$, $1 \leqslant s \leqslant m$. This is the required upper bound for the function $h(w_{i+1}^{(j_1)}, \dots, w_{i+1}^{(j_m)}) \in P_{i+1}$.

3. Lower bound for a Pfaffian function. Now we proceed to the proof of the lower estimate for $|h(w_{i+1}^{(j_1)}, \dots, w_{i+1}^{(j_m)})|$ required in the theorem. First, we consider the case where $w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)}$ are algebraically independent over P_i. Assuming that the required lower estimate fails, we have $|h(w_{i+1}^{(j_1)}, \dots, w_{i+1}^{(j_m)})| \prec (\exp^{(i+1)}(p_l))^{-1}$ for all polynomials p_l. We say in this case that the function $h(w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)})$ is small. Also we assume that m is the smallest possible integer with this property. Finally, without loss of generality one can assume that the polynomial h is irreducible over P_i.

Since the derivative $(h(w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)}))' \in P_{i+1}$ is a Pfaffian function [Kh], it is also small. One can write $(h(w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)}))' = \sum_{1 \leqslant s \leqslant m} \frac{\partial h}{\partial w_{i+1}^{(j_s)}} \cdot q^{(s)}(w_{i+1}^{(j_s)}) = g(w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)})$ for a certain polynomial $g \in P_i[Y_1, \cdots, Y_m]$.

If $h \nmid g$ in the ring $P_i[Y_1, \cdots, Y_m]$, then there exist polynomials $h_1, g_1 \in P_i[Y_1, \cdots, Y_m]$ such that $0 \not\equiv h h_1 + g g_1 \in P_i[Y_1, \cdots, Y_{m-1}]$ (because h is irreducible). Applying the remark at the end of Section 2 to the polynomials h_1, g_1, we conclude that the function $(hh_1 + gg_1)(w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_{m-1})})$ is small, and this contradicts the minimality of m.

Now we assume that $g = hg_0$, where $g_0 \in P_i[Y_1, \cdots, Y_m]$. Consider any s, $1 \leqslant s \leqslant m$, for which $\deg_Z(q^{(s)}) \leqslant 1$. We have

$$\deg_{w_{i+1}^{(j_s)}} \left(\frac{\partial h}{\partial w_{i+1}^{(j_s)}} q^{(s)}(w_{i+1}^{(j_s)}) \right) \leqslant \deg_{w_{i+1}^{(j_s)}} (h(w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)}))$$

and

$$\deg_{w_{i+1}^{(j_s)}} \left(\frac{\partial h}{\partial w_{i+1}^{(j_l)}} q^{(l)}(w_{i+1}^{(j_l)}) \right) \leqslant \deg_{w_{i+1}^{(j_s)}} (h(w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)}))$$

for every $l \neq s$, whence $w_{i+1}^{(j_s)}$ does not occur in the polynomial $g_0(w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)})$. If $\deg_Z(q^{(s)}) \geqslant 2$ for some s, $1 \leqslant s \leqslant m$, then the lemma implies that $|w_{i+1}^{(j_s)}| \prec \exp^{(i)}(p_9)$ for an appropriate polynomial p_9. Therefore $|g_0(w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)})| \prec \exp^{(i)}(p_{10})$ for a certain p_{10}. Thus,

$$\left| \frac{(h(w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)}))'}{h(w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)})} \right| \prec \exp^{(i)}(p_{10}),$$

whence

$$\left|\log\left|h(w_{i+1}^{(j_1)}), \cdots, w_{i+1}^{(j_m)}\right|\right| \prec \exp^{(i)}(p_{11}) \quad \text{and} \quad \left|h(w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)})\right| \succ (\exp^{(i+1)}(p_{11}))^{-1}.$$

This contradicts the assumption that $h(w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)})$ is small and proves the required lower bound in the case where $w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)}$ are algebraically independent over P_i.

In the general case we choose among $w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)}$ a transcendental over P_i basis (without loss of generality, let it be $w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_s)}$). Then there exists a polynomial $t(Z) = \sum_{0 \leqslant l \leqslant K} t^{(l)} Z^l \in P_i[w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_s)}][Z]$ with coefficients $t^{(l)}$ belonging to $P_i[w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_s)}]$, $0 \leqslant l \leqslant K$, $t^{(0)} \not\equiv 0$, such that $t(h(w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)})) \equiv 0$. Since we have proved that $|t^{(0)}| \succ (\exp^{(i+1)}(p_{12}))^{-1}$ and, by the lemma and the remark after it, $|t^{(l)}| \prec (\exp^{(i+1)}(p_{12}))$, $0 \leqslant l \leqslant K$, for a certain p_{12}, we see that $|h(w_{i+1}^{(j_1)}, \ldots, w_{i+1}^{(j_m)})| \succ \frac{1}{2}(\exp^{(i+1)}(p_{12}))^{-2}$.

This completes the inductive step in the proof of the theorem (see the beginning of Section 2) because any element of P_{i+1} can be represented as a quotient $h^{(1)}(w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)})/h^{(2)}(w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)})$ for some elements $w_{i+1}^{(j_1)}, \cdots, w_{i+1}^{(j_m)} \in P_{i+1}$, each satisfying an equation of the kind (1), namely, $(w_{i+1}^{(j_s)})' = q^{(s)}(w_{i+1}^{(j_s)}), 1 \leqslant s \leqslant m$, and for some polynomials $h^{(1)}, h^{(2)} \in P_i[Y_1, \cdots, Y_m]$. The theorem is proved.

References

[B] Bellman R., *Stability theory of differential equations.*, McGraw-Hill Book Co., 1953.

[G] Grigoriev D., *Deviation theorems for solutions of linear ordinary differential equations and applications to lower bounds on parallel complexity of sigmoids*, this issue.

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

[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.

Поступило 13 апреля 1993 г.

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