Approximation and complexity: Liouvillean type theorems for linear differential equations on an interval

Dima Grigoriev*

Abstract

Let u, v be solutions on an interval I of linear differential equations (LDE) $P = 0, Q = 0$, respectively. We obtain a lower bound on the approximation of v by u in terms of bounds on the coefficients of LDE $S_i = 0$ (for several i) satisfied by the i-th derivative of v and by the i-th derivatives of a basis of the LDE $P = 0$.

One could view this result as a differential analog of the Liouville's theorem which states that two different algebraic numbers are well separated if they satisfy algebraic equations with small enough integer coefficients. Unlike the algebraic situation, in the differential setting, in order to bound from below the difference $|u - v|$ we need to involve not only the coefficients of P, Q themselves, but also those of S_i.

Introduction

The well-known Liouville's theorem states that if $f(a) = g(b) = 0$ where

$$f = \sum_{0 \le i \le n} f_i X^i, g = \sum_{0 \le i \le m} g_i X^i \in \mathbf{Z}[X]$$

and $a \neq b$ then one can bound from below the difference $|a - b|$. For the sake of simplicity assume that f, g have no common roots. One possible approach to its proof is to consider the resultant

*IRMAR, Université de Rennes, Campus de Beaulieu, 35042 Rennes, cedex France

$$R = f_n^m g_m^n \prod (a_i - b_j) \in \mathbf{Z}$$

where the product is taken over the roots a_i of f and b_j of g, respectively. Say, for definiteness $a_1 = a, b_1 = b$. Then

$$1 \leq |R| = |f_n^m g_m^n (a_1 - b_1) \prod_{(l,j) \neq (1,1)} (a_l - b_j)|$$

and from the upper bounds on the roots $|a_l| \leq \max\{1, |f_i/f_n|\}, |b_j| \leq \max\{1, |g_i/g_m|\}$ one obtains a lower bound on $|a_1 - b_1|$. Note that this argument provides a lower bound on $|a_l - b_j|$ for any pair of the roots of f and g. We recall also that $R = g_m^n \prod f(b_j)$ and alternatively, one could obtain similar to above a lower bound on $|f(b_1)|$. The proof of the Liouville's theorem uses two basic ingredients: a lower bound on $|R|$ and an upper bound on the roots $|a_i|, |b_j|$.

If one would try to transfer this argument to the solutions $P(u) = Q(v) = 0$ of linear ordinary differential operators

$$P = \sum_{0 \leq i \leq n} p_i \frac{d^i}{dX^i}, Q = \sum_{0 \leq i \leq m} q_i \frac{d^i}{dX^i}$$

one needs a replacement of the resultant.

Informally speaking, the approach can be viewed as follows. We have

$$\sum_{0 \leq i \leq n} p_i \frac{d^i}{dX^i}(u - v) = P(u - v) = -P(v).$$

In fact, $P(v)$ plays a role of the resultant: it could be represented as the determinant of an appropriate $(n+1) \times (n+1)$ matrix with the last column formed by the derivatives $\frac{d^i}{dX^i}(u - v)$, $0 \leq i \leq n$ and other entries being the derivatives of the elements of a basis $u_1, \ldots, u_n$ of the space of solutions of LDE $P = 0$. Assume that a certain lower bound on $||P(v)||$ is given where $||, ||$ denotes some norm, this replaces the first of the mentioned ingredients. We want to derive a contradiction from the supposition that $||u - v||$ is small. For this purpose we have to guarantee that under the supposition the norms of the derivatives $||\frac{d^i}{dX^i}(u - v)||, 1 \leq i \leq n$, should be small as well (it is an extra effort in comparison with the algebraic situation). Together with the upper bounds on the entries of the matrix which occur in the rate of approximation (this replaces the second of the mentioned ingredients), that leads to a contradiction by means of expanding $P(v)$ with respect to its last column.

This plan is fulfilled in the theorem below which provides a lower bound on approximations $\max_{x \in I} |v - \sum_{1 \leq j \leq n} \lambda_j u_j(x)|$ of a function v by means of any linear combination of the form $\sum_{1 \leq j \leq n} \lambda_j u_j, \lambda_j \in \mathbf{R}$, of functions $u_1, \ldots, u_n$ defined on a finite closed interval $I \subset \mathbf{R}$. The bound depends on n which could be informally treated as a complexity measure of an approximation, having in mind that a function v is given and we try to minimize the number n of functions $u_1, \ldots, u_n$ taken from a fixed set of "basic" ones, for example, monomials, or trigonometric monomials $\sin(\theta X)$, or exponential monomials $\exp(\theta X)$ etc.

Let us underline that unlike the case of algebraic equations where the bound depends separately on both polynomials f and g, we consider approximations of a fixed solution of one operator by means of any solution of another operator, rather than the difference of any pair of solutions which could be arbitrary small on the interval, and the bound depends on suitable minors composed of the derivatives of v and of $u_1, \ldots, u_n$.

As an application of the theorem we provide a lower bound for approximations by means of linear combinations of functions of the forms $\sin(\theta X), \cos(\theta X)$.

Observe that it is more difficult to prove analogs of Liouvillean type theorems for solutions of linear differential equations on an interval than on the whole real line. Moreover, one can obtain lower bounds on approximations on $\mathbf{R}$ not only for solutions of LDE, but for their compositions [G 92] making use of a more general approach involving the Wronskian.

It is an interesting question whether one can prove Liouvillean type theorems for approximations on an *interval* for two classes of functions studied in [G 92], [G 93], respectively, namely, compositions of solutions of linear differential equations and Pfaffian functions [Kh] (or in other terms, for nested solutions of first-order non-linear differential equations). We mention also that beyond these two classes of functions one could hardly expect any Liouvillean type theorem due to the example (see [B]) of a second-order non-linear differential equation with arbitrarily closeness to zero.

One could also view the result of the paper as a trade-off between approximations and complexity. It would be interesting to understand more on this trade-off. We mention that in this direction a lower bound was proved in [CG] on the complexity of approximating algebraic computation trees.

Another motivation for this trade-off arises from neural networks (see [MSS] and the references there) where one considers sigmoids (circuits with certain transcendental functions as gates), and the problem of approximating a sigmoid by another of small complexity.

It is worthwhile also to mention that in [K] a version of a differential analog of the Liouville's theorem was proposed in terms of bounds on valuations, while we consider approximations in L_∞-norm.

The author would like to take this opportunity to thank Georg Schnitger for stimulating discussions. Also the author is indebted to the anonymous referees whose remarks have conduced to improve the exposition.

Approximations of solutions of linear differential equations

Let $I \subset \mathbf{R}$ be a finite interval of length $|I|$ and $u_1, \dots, u_n, v$ be $2n+2$-differentiable functions on I. We study the question of how well linear combinations of the form $\sum_{1 \le j \le n} \lambda_j u_j$ for $\lambda_j \in \mathbf{R}$ can approximate v, i.e. the problem of bounding from below the norm

$$||v - \sum_{1 \le j \le n} \lambda_j u_j||_I = \max |(v - \sum_{1 \le j \le n} \lambda_j u_j)(x)|$$

where the maximum of absolute values is taken over the points x from the closure of the interval I.

Consider $(2n+3) \times (n+1)$ matrix $\overline{\Delta}$ with the rows (respectively, columns) numbered from 0 to $2n+2$ (respectively, from 1 to $n+1$) defined as follows. For $1 \le k \le n$ its k-th column is formed by the derivatives $u_k, u_k', \dots, u_k^{(2n+2)}$ and $(n+1)$-th column is formed by the derivatives $v, v', \dots, v^{(2n+2)}$. For $0 \le i \le n+1$, $i \le j \le i+n+1$ let $\overline{\Delta}(i, \hat{j})$ denote the $(n+1) \times (n+1)$ subdeterminant of $\overline{\Delta}$ formed by the rows $i, i+1, \dots, j-1, j+1, \dots, i+n, i+n+1$.

We assume the following bounds for any point $x \in I$:

$$|\overline{\Delta}(i, \hat{j})(x)| \le M, |\overline{\Delta}(i, i+\widehat{n}+1)(x)| \ge \delta > 0 \tag{1}$$

for all $0 \leq i \leq n+1, i \leq j \leq i+n+1$ and for certain fixed M, δ. Then the space of solutions of the LDE $S_i = 0$ where

$$S_i = \sum_{0 \leq j \leq n+1} (-1)^j \overline{\Delta}(i, \widehat{j+i}) \frac{d^j}{dX^j}, 0 \leq i \leq n+1$$

has a basis $u_1^{(i)}, \ldots, u_n^{(i)}, v^{(i)}$ due to the condition on δ in (1).

For $0 \leq j \leq n$ let $\Delta(\hat{j})$ be the $n \times n$ subdeterminant of the first n columns of $\overline{\Delta}$ and the rows $0, 1, \ldots, j-1, j+1, \ldots, n$. Then $u_1, \ldots, u_n$ is a basis of the space of solutions of the LDE $S = 0$ where

$$S = \sum_{0 \leq j \leq n} (-1)^j \Delta(\hat{j}) \frac{d^j}{dX^j}.$$

Assume that for any point $x \in I$ we have

$$|\Delta(\hat{j})(x)| \leq M_0, 0 \leq j \leq n \tag{2}$$

Theorem

$$||v - \sum_{1 \leq j \leq n} \lambda_j u_j||_I \geq \frac{\delta}{(n+1)M_0} \left(\frac{\min\{|I|, \delta/2M\}}{(n+1)^3} \right)^n \tag{3}$$

Remark. There exists a closed subinterval $I_n \subset I$ (which depends on $\lambda_1, \ldots, \lambda_n$) of length

$$|I_n| \geq \frac{\min\{|I|, \delta/2M\}}{(n+1)^3}$$

such that $|(v - \sum_{1 \leq j \leq n} \lambda_j u_j)(x)|$ is greater than the right-hand side of (3) for any point $x \in I_n$.

The following lemma was proved as lemma 2 [G 92].

Lemma 1 *For each $0 \leq i \leq n+1$ the number of roots on I of the derivative $w^{(i)} = v^{(i)} - \sum_{1 \leq j \leq n} \lambda_j u_j^{(i)}$, does not exceed $\lceil 2M|I|/\delta \rceil n$; moreover any subinterval of I of length less or equal to $\delta/2M$ contains at most n roots of $w^{(i)}$.*

Proof. Suppose that a certain closed subinterval $I' \subset I$ of length $|I'| \leq \delta/2M$ contains more than n roots of $w^{(i)}$. Then each of the derivatives

$w^{(i)}, w^{(i+1)}, \dots, w^{(i+n)}$ has a root in I'. Let $M^{(j)} = ||w^{(i+j)}||_{I'}, 0 \le j \le n+1$. Then

$$M^{(j+1)} \ge M^{(j)}/|I'| \ge 2MM^{(j)}/\delta, 0 \le j \le n$$

by the Mean Value Theorem.

On the other hand according to the Cramer's rule we have the identity

$$w^{(i+n+1)}\overline{\Delta}(i, i+\widehat{n}+1) = \sum_{i \le j \le i+n} (-1)^{j-i-n} w^{(j)} \overline{\Delta}(i, \hat{j})$$

taking into the account that the minors $\overline{\Delta}(i, \hat{j})$ do not change if to replace the $(n+1)$-th column of the matrix $\overline{\Delta}$ by the derivatives $w, w^{(1)}, \dots, w^{(2n+2)}$, that corresponds to a linear elementary transformation of the columns of $\overline{\Delta}$. Substituting in this identity a point $x \in I'$ at which the derivative $|w^{(i+n+1)}|$ attains its maximum $M^{(n+1)}$, we bound from below the absolute value of the left-hand side by $M^{(n+1)}\delta$, and on the other hand, bound from above the absolute value of the right-hand side by $M(M^{(0)} + \dots + M^{(n)})$ due to (1), i.e. $M^{(n+1)} \le M(M^{(n)} + \dots + M^{(0)})/\delta$. Hence

$$M^{(n+1)} \le \frac{M}{\delta} M^{(n+1)}((\frac{\delta}{2M}) + (\frac{\delta}{2M})^2 + \dots + (\frac{\delta}{2M})^{n+1}) <$$

$$\frac{M}{\delta} M^{(n+1)} \frac{\delta}{2M} \frac{2M}{2M-\delta} \le M^{(n+1)},$$

this contradiction proving the lemma. □

In view of lemma 1, there exists a closed subinterval $I_0 \subset I$ of length

$$|I_0| \ge \frac{\min\{|I|, \delta/2M\}}{(n+1)^2}$$

without roots of any $w^{(i)}, 0 \le i \le n+1$. Under these conditions on I_0 we estimate the norms of the derivatives $||w^{(1)}||_{I_n}, \dots, ||w^{(n)}||_{I_n}$ via the norm $||w||_{I_n}$ for a suitable closed subinterval $I_n \subset I_0$.

Lemma 2 *Assume that the derivatives $w^{(0)}, \dots, w^{(n+1)}$ have no roots in an interval I_0. Then there exists a closed subinterval $I_n \subset I_0$ of length $|I_n| = \frac{|I_0|}{n+1}$ such that*

$$||w^{(j)}||_{I_n} \leq ||w||_{I_0} \Big(\frac{n+1}{|I_0|}\Big)^j, 0 \leq j \leq n.$$

Proof. Assume that one has already produced (by recursion on j) closed subintervals $I_0 \supset I_1 \supset \cdots \supset I_j$ with the lengths $|I_l| = |I_0|\frac{n+1-l}{n+1}$ such that $||w^{(l)}||_{I_l} \leq ||w||_{I_0}\big(\frac{n+1}{|I_0|}\big)^l, 0 \leq l \leq j < n$.

Denote by $a_1 = |w^{(j+1)}(x_1)|, a_2 = |w^{(j+1)}(x_2)|$ the values of the function $|w^{(j+1)}|$ at the endpoints of the interval $I_j = [x_1, x_2]$. If $a_1 < a_2$ then set $x_0 = x_2 - \frac{|I_0|}{n+1}$ and the subinterval $I_{j+1} = [x_1, x_0]$. Otherwise, if $a_1 > a_2$ then $x_0 = x_1 + \frac{|I_0|}{n+1}$ and the subinterval $I_{j+1} = [x_0, x_2]$. Then $||w^{(j+1)}||_{I_{j+1}} = |w^{(j+1)}(x_0)|$ and $|w^{(j+1)}(x)| \geq ||w^{(j+1)}||_{I_{j+1}}$ for any point x from the subinterval $I_j - I_{j+1}$ since $w^{(j+1)}$ is monotone and has no roots in the subinterval $I_j \subset I_0$ (whence $|w^{(j+1)}|$ is also monotone on the same interval). Observe that $a_1 \neq a_2$, indeed, otherwise $w^{(j+2)}$ would vanish identically on the interval I_j. Hence

$$||w^{(j)}||_{I_j - I_{j+1}} \geq ||w^{(j+1)}||_{I_{j+1}}(|I_j| - |I_{j+1}|)$$

because $w^{(j)}$ has no roots in the subinterval $I_j \subset I_0$. Thus,

$$||w^{(j)}||_{I_j} \geq ||w^{(j+1)}||_{I_{j+1}} \frac{|I_0|}{n+1}$$

which proves the recursion hypothesis. Taking $j = n$ and $l = j$, and noting that $I_n \subset I_j$, we get lemma 2. $\square$

To complete the proof of the theorem consider the $(n+1) \times (n+1)$ subdeterminant formed by the first $n+1$ derivatives of the functions $u_1, \ldots, u_n, w$. Since this subdeterminant is equal to $\overline{\Delta}(0, \widehat{n+1})$, we get (using lemma 2) from its expansion with respect to the last column (taking into account (2) and the bound on the length $|I_0|$ following lemma 1) that

$$\delta \leq ||\overline{\Delta}(0, \widehat{n+1})||_{I_n} \leq (n+1)M_0||w||_{I_0}\Big(\frac{n+1}{|I_0|}\Big)^n.$$

The theorem is proved. $\square$

Now we give an application of the theorem in the case of the functions $u_{2j-1} = \sin(\theta_j X), u_{2j} = \cos(\theta_j X)$ for pairwise distinct squares θ_j^2. Since we deal with $2n$ functions $u_1, \ldots, u_{2n}$ the role of n in the bounds from theorem

1 will be played by $2n$. About the function v we assume that the derivatives $v^{(0)}, v^{(2)}, v^{(4)}, \dots, v^{(4n+2)}$ of even orders all have the same sign at each point of I, the same holds for all odd order derivatives $v^{(1)}, v^{(3)}, \dots, v^{(4n+1)}$, furthermore the derivatives are bounded above, and away from zero, with $A \geq |v^{(l)}(x)| \geq a > 0$ for $x \in I, 0 \leq l \leq 4n+2$. In particular, one one could take $v = \exp$.

Denote by σ_i the i-th elementary symmetric function of $\theta_1^2, \dots, \theta_n^2, 0 \leq i \leq n$, in particular, $\sigma_0 = 1$. Denote by $B_0, B_1, \dots$ the rows of the matrix $\overline{\Delta}$, respectively. Then $B_{2n+j} = \sum_{0 \leq i \leq n-1} (-1)^{n-i+1} \sigma_{n-i} B_{2i+j}$ for $j \geq 0$ (in fact, this holds for every $j \geq 0$, but $j \leq 2n+2$ in the matrix $\overline{\Delta}$). Therefore, $|\Delta(0, \hat{j})|$ equals $|\sigma_{n-j/2} \Delta(0, \widehat{2n})|$ when j is even, and equals zero when j is odd. Hence $M_0 \leq \max_{0 \leq l \leq n} \{\sigma_l\}$.

Furthermore, $\delta \geq a \sum_{0 \leq j \leq n} \sigma_j |\Delta(0, \widehat{2n})|$ and

$M \leq A \max_{0 \leq l \leq n} \{\sigma_l\} \sum_{0 \leq l \leq n} |\Delta(0, \widehat{2n})|$.

Let $R = \max\{1, \theta_j^2\}$, then $max_{0 \leq l \leq n} \{\sigma_l\} \leq (2R)^n$. The theorem implies that

$$||v - \sum_{1 \leq j \leq 2n} \lambda_j u_j||_I \geq \frac{\delta}{(n+2)^{n+1} M_0} |I_0|^n = \frac{\delta}{n^n (n+2)^{2n+1} M_0} \min\{|I|^n, (\frac{\delta}{2M})^n\}$$

Thus, we obtain the following corollary.

Corollary.

$$||v - \sum_{1 \leq j \leq n} (\lambda_j \sin(\theta_j X) + \lambda_{j+n} \cos(\theta_j X))||_I \geq \min\{\frac{|I|^n a}{n^{O(n)} A}, \frac{a^{n+1}}{A^n 2^{O(n^2)} R^{n^2}}\}$$

.

References

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

[CG] F.Cucker, D.Grigoriev. Complexity lower bounds for approximation algebraic computation trees, J. Complexity, 1999, 15, 4, p. 499–512.

[G 92] D.Grigoriev. Deviation theorems for solutions of linear ordinary differential equations and applications to parallel complexity of sigmoids, St.Petersburg Math. J., 1995, 6, 1, p. 89–106.

[G 93] D.Grigoriev, Deviation theorems for Pfaffian sigmoids, St.Petersburg Math. J., 1995, 6, 1, p. 107–112.

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

[Kh] A.Khovanskii. Fewnomials. AMS Transl. Math. Monogr., 1991.

[K] E.R.Kolchin. Rational approximations to solutions of algebraic differential equations. Proc. AMS, 1959, 10, p. 238–244.

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