There is no reverse derivation mode for discrete derivatives

Éric Schost, STIX, École polytechnique
91128 Palaiseau, France
schost@stix.polytechnique.fr

November 5, 2004

Abstract

In the straight-line program model, it is known that computing all partial derivatives of a single polynomial induces only a constant increase in complexity, using the reverse derivation mode. We show that no such result holds for shifts, divided differences, q-shifts or q-divided differences.

Keywords lower bounds; discrete derivatives; degree bound.

1 Introduction, main result

Ore operators are useful generalizations of the notion of partial derivatives. To give their definition, let us denote by A_n the ring $k[X_1, \ldots, X_n]$, where k is a field and n is a positive integer. The partial derivatives $\partial_1, \ldots, \partial_n$ are defined by $\partial_i : P \in A_n \mapsto \partial P/\partial X_i$; they satisfy the relation $\partial_i(PQ) = \partial_i(P)Q + P\partial_i(Q)$. Ore operators are defined by allowing functional equations more general than the above, of the form $\partial(PQ) = \delta(P)Q + \sigma(P)\partial(Q)$, where σ is a ring homomorphism, and δ is a σ-derivation [9].

Some standard examples of such operators will be considered here. We will use the *shift* operators $S_i : P \mapsto P(X_1, \ldots, X_i + 1, \ldots, X_n)$ and the *q-shift* operators $Q_i : P \mapsto P(X_1, \ldots, qX_i, \ldots, X_n)$, where q is in k, together with the associated (q-)*difference operators* (or *discrete derivatives*) $\Delta_i : P \mapsto S_i(P) - P$ and $\Lambda_i : P \mapsto Q_i(P) - P$.

From the algorithmic point of view, some tools are common to a large class of such operators. This is for instance the case for elimination techniques, based either on suitable versions of the Euclidean algorithm, or on more involved non-commutative variants of Buchberger's algorithm: one can see applications of such techniques for multivariate identities proving in [4], which follows notably [6]. For such questions, partial derivatives and more general operators are treated on an equal footing; the algorithms are common to a whole class of Ore structures.

It seems interesting to pursue these investigations, and study what algorithmic and complexity properties pass from partial derivatives to more general operators. This is our goal in this note: we show that the operators defined above strongly differ from the partial derivatives with respect to some basic complexity questions.

We will work in the straight-line model of computation, and measure complexity using the total number of operations (see [3, Ch. 4] for definitions); for any polynomials $P_1, \ldots, P_s \in A_n$, we write $L(P_1, \ldots, P_s)$ for the minimal size of a straight-line program that computes $P_1, \ldots, P_s$. Thus, we count at unit cost all operations (actually, similar results would also hold for the multiplicative complexity measure).

One easily sees that computing the gradient of a polynomial $P \in A_n$ can be done for about n times the cost of computing P, by propagating forward its n partial derivatives, that is, $L(\partial_1 P, \ldots, \partial_n P) \in O(nL(P))$. However, better can be done: using the so-called reverse mode of derivation, it is known that $L(\partial_1 P, \ldots, \partial_n P) \leq 4L(P)$. This idea goes back at least to [8], and is presented in the straight-line model in [1] (see for instance [5] for a much more comprehensive presentation). Apart from its algorithmic uses, notably for optimization algorithms, this result is also the basis of lower bound estimates, see [1, 3].

Our main result is that no such inequality holds, either for shifts, divided differences, q-shifts or q-divided differences, and that an overhead of about n is unavoidable in the worst case.

Theorem 1 *For any $L \geq 1$, $n \geq 1$ and $\varepsilon > 0$, there exist P_S and P_Δ in A_n such that:*

$$\frac{L(S_1(P_S), \ldots, S_n(P_S))}{L(P_S)} \geq n(1-\varepsilon) \quad \text{and} \quad L(P_S) \geq L,$$

$$\frac{L(\Delta_1(P_\Delta), \ldots, \Delta_n(P_\Delta))}{L(P_\Delta)} \geq n(1-\varepsilon) - 1 \quad \text{and} \quad L(P_\Delta) \geq L.$$

Furthermore, for any $q \neq 1$, there exist R_Q and R_Λ in A_n such that:

$$\frac{L(Q_1(R_Q), \ldots, Q_n(R_Q))}{L(R_Q)} \geq (n-1)(1-\varepsilon) \quad \text{and} \quad L(R_Q) \geq L,$$

$$\frac{L(\Lambda_1(R_\Lambda), \ldots, \Lambda_n(R_\Lambda))}{L(R_\Lambda)} \geq (n-1)(1-\varepsilon) - 1 \quad \text{and} \quad L(R_\Lambda) \geq L.$$

Given a straight-line program that computes a polynomial P, it is immediate to deduce a straight-line program that computes $S_i(P)$, increasing the complexity by at most 1 (which accounts for the cost of computing $X_i + 1$). Thus, $L(S_i(P)) \leq L(P) + 1$, from which we deduce $L(S_1 P, \ldots, S_n P) \leq n(L(P) + 1)$. Similar estimates hold for the other operators considered here, so our lower bound are sharp.

2 Proof of the statements

For $m \in \mathbb{N}$, we denote by $\mathbb{A}^m(\overline{k})$ the m-dimensional affine space over an algebraic closure of k. If V is an r-equidimensional algebraic variety in $\mathbb{A}^m(\overline{k})$, its degree $\deg(V)$ is the generic, and maximal, number of intersection points with a linear subspace of codimension r, when this intersection is finite. We will use Strassen's degree bound [10]: let $P_1, \ldots, P_s$ be in

$k[X_1, \ldots, X_m]$, and let $V \subset \mathbb{A}^{m+s}(\overline{k})$ be the graph of $P_1, \ldots, P_s$. Then V is equidimensional, and the inequality $L(P_1, \ldots, P_s) \geq \log(\deg(V))$ holds. Here, and in all that follows, all logarithms are taken in base 2. Finally, we denote by $\mathrm{char}(k)$ the characteristic of k. In all this section, n is a fixed positive integer.

2.1 Shift operators

For $M \geq 0$, we define $P_{n,M} = (X_1 \cdots X_n)^M$. Since $\deg(P_{n,M}) = nM$, we get the following lower bound:

Lemma 1 *The inequality $L(P_{n,M}) \geq \log(nM)$ holds.*

On the other hand, by first computing the product $X_1 \cdots X_n$ and then raising it to Mth power by binary powering, we obtain the inequality $L(P_{n,M}) \leq n + 2\log(M)$. However, a better asymptotic estimate holds. Let us indeed denote by $\ell(M)$ the minimal length of an addition chain that computes M. It is known [2] (see also [7] for more bibliography) that $\ell(M)$ is asymptotically equivalent to $\log(M)$. We deduce the following improved bound for $L(P_{n,M})$.

Lemma 2 *Let $\varepsilon > 0$. The inequality $L(P_{n,M}) \leq n+(1+\varepsilon)\log(M)$ holds for M large enough.*

We now give a lower bound on the complexity of $L(S_1(P_{n,M}), \ldots, S_n(P_{n,M}))$.

Lemma 3 *If M and $\mathrm{char}(k)$ are coprime, the inequality $L(S_1(P_{n,M}), \ldots, S_n(P_{n,M})) \geq n\log(M)$ holds.*

Proof. We can suppose $n \geq 2$ (the case $n = 1$ is immediate). Let $V \subset \mathbb{A}^{2n}(\overline{k})$ be the graph of the map

$$\begin{array}{cccc} \varphi : & \mathbb{A}^n(\overline{k}) & \to & \mathbb{A}^n(\overline{k}) \\ & \mathbf{x} = (x_1, \ldots, x_n) & \mapsto & (S_1(P_{n,M})(\mathbf{x}), \ldots, S_n(P_{n,M})(\mathbf{x})). \end{array}$$

By the degree bound, it suffices to prove that $\deg(V) \geq M^n$. Let $v \subset \mathbb{A}^{2n}(\overline{k})$ be the fiber $\varphi^{-1}(1, \ldots, 1)$. We will now prove the following fact: *v is finite and has cardinality at least M^n*. Note that v is obtained by cutting V through n hyperplanes; thus, this claim implies that $\deg(V) \geq M^n$, which will prove the lemma.

The fiber v is isomorphic to the zero-set $v' \subset \mathbb{A}^n(\overline{k})$ of the system

$$\left\{ \begin{array}{rcl} S_1(P_{n,M})(X_1, \ldots, X_n) & = & 1, \\ & \vdots & \\ S_n(P_{n,M})(X_1, \ldots, X_n) & = & 1, \end{array} \right.$$

which can be rewritten as

$$\left\{ \begin{array}{rcl} ((X_1+1)X_2 \cdots X_n)^M & = & 1, \\ & \vdots & \\ (X_1 X_2 \cdots (X_n+1))^M & = & 1. \end{array} \right.$$

Let us denote by $\omega_1, \ldots, \omega_M$ the Mth roots of 1 in $\overline{k}$; our assumption on M and $\mathrm{char}(k)$ implies that $\omega_1, \ldots, \omega_M$ are pairwise distinct. To any map $\lambda : \{1, \ldots, n\} \to \{1, \ldots, M\}$, we associate the system $\mathfrak{S}_\lambda$ (with coefficients in $\overline{k}$):

$$\mathfrak{S}_\lambda \left\{ \begin{array}{rcl} ((X_1+1)X_2 \cdots X_n) & = & \omega_{\lambda(1)}, \\ & \vdots & \\ (X_1 X_2 \cdots (X_n+1)) & = & \omega_{\lambda(n)}. \end{array} \right.$$

For any such λ, let $v_\lambda \subset \mathbb{A}^n(\overline{k})$ be the zero-set of $\mathfrak{S}_\lambda$. Then, v' is the disjoint union of all v_λ. There are M^n choices for λ, so to prove our claim, it suffices to prove that all v_λ are finite and non-empty.

Let us thus fix a map $\lambda : \{1, \ldots, n\} \to \{1, \ldots, M\}$. Since $n \geq 2$, all coordinates of all solutions of $\mathfrak{S}_\lambda$ are non-zero. Letting $Y_i = 1/X_i$, the system $\mathfrak{S}_\lambda$ can then be rewritten in the form

$$\left\{ \begin{array}{rcl} 1 + Y_1 & = & \omega_{\lambda(1)} Y_1 \cdots Y_n, \\ & \vdots & \\ 1 + Y_n & = & \omega_{\lambda(n)} Y_1 \cdots Y_n, \end{array} \right.$$

which yields the equivalent set of equations

$$\left\{ \begin{array}{rcl} Y_1 & = & \omega_{\lambda(1)} Y_1 \cdots Y_n - 1, \\ & \vdots & \\ Y_n & = & \omega_{\lambda(n)} Y_1 \cdots Y_n - 1. \end{array} \right. \tag{1}$$

Let $\Delta \in \overline{k}[T]$ be the polynomial $\prod_{1 \leq i \leq n} (\omega_{\lambda(i)} T - 1)$. Let next $\delta \subset \overline{k}$ be the set of the roots of the polynomial $\Delta - T$; since $n \geq 2$, $\Delta - T$ is non-zero and non-constant, so δ is finite and non-empty. We conclude by showing that v_λ itself is finite and non-empty:

- Taking the product of Equations (1), we see that for all $(y_1, \ldots, y_n)$ in v_λ, the product $y_1 \cdots y_n$ belongs to δ; thus, the function $Y_1 \cdots Y_n$ takes a finite number of values on v_λ. Furthermore, Equations (1) show that the value of the product $y_1 \cdots y_n$ uniquely determines $y_1, \ldots, y_n$. Thus, v_λ is finite.

- Conversely, let us consider p in δ, and define $y_i = \omega_{\lambda(1)} p - 1$, for $i = 1, \ldots, n$. Taking the product of these equalities, we deduce that $y_1 \cdots y_n = \Delta(p)$. By definition, $\Delta(p) = p$, so $y_1 \cdots y_n = p$. Thus, the point $(y_1, \ldots, y_n)$ is a solution of Equations (1), and so v_λ is non-empty. $\square$

We can now conclude the proof of the first two assertions in Theorem 1. Let thus $\varepsilon > 0$ and $L \geq 1$, and let $\varepsilon' > 0$ be such that $\frac{1-\varepsilon'}{1+\varepsilon'} \geq 1 - \varepsilon$. Let next M be coprime with the characteristic of k and large enough to satisfy the inequalities

$$\frac{n \log(M)}{n + (1+\varepsilon') \log(M)} \geq n \, \frac{1-\varepsilon'}{1+\varepsilon'},$$

$$L(P_{n,M}) \leq n + (1+\varepsilon') \log(M)$$

and

$$\log(nM) \geq L.$$

We deduce from the above lemmas that

$$\frac{L(S_1(P_{n,M}), \ldots, S_n(P_{n,M}))}{L(P_{n,M})} \geq n\frac{1-\varepsilon'}{1+\varepsilon'} \geq n(1-\varepsilon) \quad \text{and} \quad L(P_{n,M}) \geq L.$$

This proves the first assertion in the theorem.

To prove the second assertion, note that for any polynomial P and any $1 \leq i \leq n$, we have the inequality $L(S_i(P)) \leq L(\Delta_i(P)) + L(P) + 1$, since $S_i(P)$ is obtained as $\Delta_i(P) + P$. Taking all i into account, this rewrites as

$$L(\Delta_1(P), \ldots, \Delta_n(P)) \geq L(S_1(P), \ldots, S_n(P)) - n - L(P),$$

so that

$$\frac{L(\Delta_1(P), \ldots, \Delta_n(P))}{L(P)} \geq \frac{L(S_1(P), \ldots, S_n(P))}{L(P)} - \frac{n}{L(P)} - 1.$$

Then, the previous result easily yields the second point in the theorem.

2.2 q-shift operators

Let us now consider the q-shift operators $Q_i(P) = P(X_1, \ldots, qX_i, \ldots, X_n)$ for some $q \in k$. For $M \geq 0$, we define $R_{n,M} = (X_1 + \cdots + X_n)^M$. The following lower bound is immediate in view of the degree of $R_{n,M}$:

Lemma 4 *The inequality $L(R_{n,M}) \geq \log(M)$ holds.*

As in the previous subsection, by first computing the sum $X_1 + \cdots + X_n$ and raising it to Mth power, we obtain the following upper bound:

Lemma 5 *Let $\varepsilon > 0$. The inequality $L(R_{n,M}) \leq n + (1+\varepsilon)\log(M)$ holds for M large enough.*

We now give a lower bound on the complexity of $L(Q_1(R_{n,M}), \ldots, Q_n(R_{n,M}))$.

Lemma 6 *If M and $\mathrm{char}(k)$ are coprime and $(q-1)(q+n-1) \neq 0$ in k, the inequality $L(Q_1(R_{n,M}), \ldots, Q_n(R_{n,M})) \geq n\log(M)$ holds.*

Proof. The proof is similar to that of Lemma 3. Using the degree bound, it is enough to prove the following fact: *the zero-set of the system*

$$\begin{cases} (qX_1 + X_2 + \cdots + X_n)^M &= 1, \\ & \vdots \\ (X_1 + X_2 + \cdots + qX_n)^M &= 1 \end{cases}$$

is finite, of cardinality M^n. Let us denote by $\omega_1, \ldots, \omega_M$ the M pairwise distinct Mth roots of 1 in $\overline{k}$. As above, to any map $\lambda : \{1, \ldots, n\} \to \{1, \ldots, M\}$, we associate the following system (with coefficients in $\overline{k}$):

$$\left\{ \begin{array}{rcl} qX_1 + X_2 + \cdots + X_n & = & \omega_{\lambda(1)}, \\ & \vdots & \\ X_1 + X_2 + \cdots + qX_n & = & \omega_{\lambda(n)}. \end{array} \right.$$

This system is linear, of determinant $(q-1)^{n-1}(q+n-1)$, which is non-zero by assumption. Thus, it has exactly one solution. Since there are M^n such systems, and their zero-sets are disjoint, our claim follows. $\square$

We deduce the following corollary, which lifts the assumption $q-n+1 \neq 0$ of the previous Lemma.

Lemma 7 *If* M *and* $\mathrm{char}(k)$ *are coprime, and* $q \neq 1$*, then the inequality* $L(Q_1(R_{n,M}), \ldots, Q_n(R_{n,M})) \geq (n-1)\log(M)$ *holds.*

Proof. If $q+n-1 \neq 0$, then the above lemma concludes (and actually gives a slightly better bound). Else, suppose that $q+n-1 = 0$. Any straight-line program that computes $Q_1(R_{n,M}), \ldots, Q_n(R_{n,M})$ in $k[X_1, \ldots, X_n]$ yields, by specializing X_n at 0, a straight-line program that computes $Q_1(R_{n-1,M}), \ldots, Q_{n-1}(R_{n-1,M})$ in $k[X_1, \ldots, X_{n-1}]$, without cost increase.

Now, we have $q+(n-1)-1 \neq 0$, so we can apply the previous lemma, which implies that $L(Q_1(R_{n-1,M}), \ldots, Q_{n-1}(R_{n-1,M})) \geq (n-1)\log(M)$ if M and $\mathrm{char}(k)$ are coprime. The remark in the preceding paragraph finishes the proof. $\square$

The proof of the last two statements of Theorem 1 follows as in the previous subsection.

References

[1] W. Baur and V. Strassen. The complexity of partial derivatives. *Theoret. Comput. Sci.*, 22(3):317–330, 1983.

[2] A. Brauer. On addition chains. *Bull. Amer. Math. Soc.*, 45:736–739, 1939.

[3] P. Bürgisser, M. Clausen, and M. A. Shokrollahi. *Algebraic complexity theory*, volume 315 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, 1997.

[4] F. Chyzak and B. Salvy. Non-commutative elimination in Ore algebras proves multivariate identities. *J. Symbolic Comput.*, 26(2):187–227, 1998.

[5] A. Griewank. *Evaluating derivatives*, volume 19 of *Frontiers in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), 2000.

[6] A. Kandri-Rody and V. Weispfenning. Noncommutative Gröbner bases in algebras of solvable type. *J. Symbolic Comput.*, 9(1):1–26, 1990.

[7] D. E. Knuth. *The Art of Computer Programming, vol. 2, Seminumerical Algorithms.* Addison-Wesley, Reading MA, 3rd edition, 1998.

[8] S. Linnainmaa. Taylor expansion of the accumulated rounding error. *Nordisk Tidskr. Informationsbehandling (BIT)*, 16(2):146–160, 1976.

[9] O. Ore. Theory of non-commutative polynomials. *Ann. of Math. (2)*, 34(3):480–508, 1933.

[10] V. Strassen. Die Berechnungskomplexität von elementarsymmetrischen Funktionen und von Interpolationskoeffizienten. *Numer. Math.*, 20:238–251, 1972/73.