Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Sums of read-once formulas: How many summands are necessary? ☆

Meena Mahajan *, Anuj Tawari

The Institute of Mathematical Sciences IMSc, HBNI, Chennai, India

ARTICLE INFO

Article history:
Received 26 October 2016
Received in revised form 1 September 2017
Accepted 22 October 2017
Available online xxxx
Communicated by L.M. Kirousis

Keywords:
Arithmetic formulas
Read-once polynomials
Hardness of representation

ABSTRACT

An arithmetic read-once formula (ROF) is a formula (circuit of fan-out 1) over $+, \times$ where each variable labels at most one leaf. Every multilinear polynomial can be expressed as the sum of (possibly exponentially many) ROFs. In this work, we prove, for certain multilinear polynomials, a tight lower bound on the number of summands in such an expression.

1. Introduction

Read-once formulas (ROF) are formulas (circuits of fan-out 1) in which each variable appears at most once. A formula computing a polynomial that depends on all its variables must read each variable at least once. Therefore, ROFs compute some of the simplest possible functions that depend on all of their variables. The polynomials computed by such formulas are known as read-once polynomials (ROPs). Since every variable is read at most once, ROPs are multilinear. (A polynomial is said to be multilinear if the individual degree of each variable is at most one.) But not every multilinear polynomial is a ROP. For example, $x_1x_2 + x_2x_3 + x_1x_3$.

We investigate the following question: Given an n-variate multilinear polynomial, can it be expressed as a sum of at most k ROPs? It is easy to see that every bivariate multilinear polynomial is a ROP. Any tri-variate multilinear polynomial can be expressed as a sum of 2 ROPs. With a little thought, we can obtain a sum-of-3-ROPs expression for any 4-variate multilinear polynomial. An easy induction on n then shows that any n-variate multilinear polynomial, for $n \geq 4$, can be written as a sum of at most $3 \times 2^{n-4}$ ROPs; see Proposition 5. Also, the sum of two multilinear monomials is a ROP, so any n-variate multilinear polynomial with M monomials can be written as the sum of $\lceil M/2 \rceil$ ROPs. We ask the following question: Does there exist a strict hierarchy among k-sums of ROPs? Formally,

Problem 1. Consider the family of n-variate multilinear polynomials. For $1 < k \leq 3 \times 2^{n-4}$, is $\sum^k \cdot$ROP strictly more powerful than $\sum^{k-1} \cdot$ROP? If so, what explicit polynomials witness the separations?

☆ Preliminary versions of this paper appear in the Proceedings of the 11th Computer Science Symposium in Russia CSR, LNCS Vol. 9691 pp. 266–279 and as Technical Reports on ECCC and CoRR [12].

* Corresponding author.
E-mail address: meena@imsc.res.in (M. Mahajan).

https://doi.org/10.1016/j.tcs.2017.10.019
0304-3975/© 2017 Elsevier B.V. All rights reserved.

We answer this affirmatively for $k \leq \lceil n/2 \rceil$. In particular, for $k = \lceil n/2 \rceil$, there exists an explicit n-variate multilinear polynomial which cannot be written as a sum of less than k ROPs but it admits a sum-of-k-ROPs representation.

Note that n-variate ROPs are computed by linear sized formulas. Thus if an n-variate polynomial p is in $\sum^k \cdot$ROP, then p is computed by a formula of size $O(kn)$ where every intermediate node computes a multilinear polynomial. Since superpolynomial lower bounds are already known for the model of multilinear formulas [14], we know that for those polynomials (including the determinant and the permanent), a $\sum^k \cdot$ROP expression must have k at least quasi-polynomial in n. However the best upper bound on k for these polynomials is only exponential in n, leaving a big gap between the lower and upper bound on k. A lesser but still significant gap also exists in the known exponential lower bound for sums of ROPs; in [13] it is shown that a certain polynomial, explicitly described by Raz and Yehudayoff in [15], requires $2^{\Omega(n^{1/3}/\log n)}$ ROP summands, while 2^n summands is anyway sufficient. On the other hand, our lower bound is provably tight.

A counting argument (see Proposition 7) shows that a random multilinear polynomial requires exponentially many ROPs; there are multilinear polynomials requiring $k = \Omega(2^n/n^2)$. Our general upper bound on k is $O(2^n)$, leaving a gap between the lower and upper bound. One challenge is to close this gap.

A natural question to ask is whether stronger lower bounds than the above result can be proven. In particular, to separate $\sum^{k-1} \cdot$ROP from $\sum^k \cdot$ROP, how many variables are needed? Our hierarchy result says that $2k-1$ variables suffice, but there may be simpler polynomials (with fewer variables) witnessing this separation. We demonstrate another technique which improves upon the previous result for $k = 3$, showing that 4 variables suffice. In particular, we show that over the field of reals, there exists an explicit multilinear 4-variate multilinear polynomial which cannot be written as a sum of 2 ROPs. This lower bound is again tight, as there is a sum of 3 ROPs representation for every 4-variate multilinear polynomial.

1.1. Our results and techniques

We now formally state our main results.

The first main result establishes the strict hierarchy among k-sums of ROPs.

Theorem 1. *For each $n \geq 1$, the n-variate degree $n-1$ symmetric polynomial S_n^{n-1} cannot be written as a sum of less than $\lceil n/2 \rceil$ ROPs, but it can be written as a sum of $\lceil n/2 \rceil$ ROPs.*

The idea behind the lower bound is that if $g = S_n^{n-1}$ can be expressed as a sum of less than $\lceil n/2 \rceil$ ROFs, then one of the ROFs can be eliminated by taking partial derivative with respect to one variable and substituting another by a field constant. We then use the inductive hypothesis to arrive at a contradiction. This approach necessitates a stronger hypothesis than the statement of the theorem, and we prove this stronger statement in Lemma 18 as part of Theorem 21.

This result separates $\sum^3 \cdot$ROP from $\sum^2 \cdot$ROP via the polynomials S_5^4 and S_6^5. Our second main result shows that $\sum^3 \cdot$ROP is also separated from $\sum^2 \cdot$ROP by a 4-variate multilinear polynomial.

Theorem 2. *There is an explicit 4-variate multilinear polynomial f which cannot be written as the sum of 2 ROPs over $\mathbb{R}$.*

The proof of this theorem mainly relies on a structural lemma (Lemma 25) for sum of 2 read-once formulas. In particular, we show that if f can be written as a sum of 2 ROPs then one of the following must be true:

1. Some 2-variate restriction is a linear polynomial.
2. There exist variables $x_i, x_j \in \mathrm{Var}(f)$ such that the polynomials x_i, x_j, $\partial_{x_i}(f)$, $\partial_{x_j}(f)$, 1 are linearly dependent.
3. We can represent f as $f = l_1 \cdot l_2 + l_3 \cdot l_4$ where (l_1, l_2) and (l_3, l_4) are variable-disjoint linear forms.

Checking the first two conditions is easy. For the third condition we use the commutator of f, introduced in [16], to find one of the l_i's. The knowledge of one of the l_i's suffices to determine all the linear forms. Finally, we construct a 4-variate polynomial which does not satisfy any of the above mentioned conditions. This construction does not work over algebraically closed fields. We do not yet know how to construct an explicit 4-variate multilinear polynomial not expressible as the sum of 2 ROPs over such fields, or even whether such polynomials exist.

1.2. Related work

Despite their simplicity, ROFs have received a lot of attention both in the arithmetic as well as in the Boolean world [8,5,3,4,16,17]. The most fundamental question that can be asked about polynomials is the polynomial identity testing (PIT) problem: Given an arithmetic circuit $\mathcal{C}$, is the polynomial computed by $\mathcal{C}$ identically zero or not. PIT has a randomized polynomial time algorithm: Evaluate the polynomial at random points. It is not known whether PIT has a deterministic polynomial time algorithm. In 2004, Kabanets and Impagliazzo established a connection between PIT algorithms and proving general circuit lower bounds [10]. Similar results are known for some restricted classes of arithmetic circuits, for instance, constant-depth circuits [6,1]. However, consider the case of multilinear formulas. Even though strong lower bounds are

Please cite this article in press as: M. Mahajan, A. Tawari, Sums of read-once formulas: How many summands are necessary?, Theoret. Comput. Sci. (2017), https://doi.org/10.1016/j.tcs.2017.10.019