Some Algebraic Problems with Connections to Circuit Complexity of Dynamic Data Structures

William Hesse
Clarkson University
Potsdam, New York USA
whesse@clarkson.edu

January 10, 2007

Abstract

1 Introduction

While researching dynamic data structures of polynomial size that are updated by extremely simple circuits, we have come across many interesting algebraic problems. Some of these simple questions about small sums and products in an algebra would give lower bounds on the complexity of dynamic data structures.

2 Small Sums

We consider problems in two algebras: The vector space over $\mathbf{Z}_2$ of all linear functions of n bits, and the algebra over $\mathbf{Z}_2$ of all Boolean functions of n bits. The questions we ask are simpler to answer over the vector space, but have stronger consequences if they can be shown for arbitrary Boolean functions.

Because we are trying to prove limits on the power of extremely small circuits, containing $O(1)$ or $O(\log \log n)$ gates, we consider sums of only a few elements. We consider, for a small integer r, such as 2 or 3, all sums of r elements from a set A of elements. Define

$$\mathbf{Span}_k(A) = \{\sum_{a \in S} a \mid S \subseteq A, \ |S| \leq r\}$$

Instead of the subspace spanned by A, containing all linear combinations of elements of A, we consider only the set of small linear combinations of elements of A. Note that $|\mathbf{Span}_k(A)| \leq |A|^k$.

We are interested in finding a set B that cannot be covered by the $\mathbf{Span}_k$ of any set A smaller than a given bound. We consider $\mathbf{Z}_2^n$, the n-dimensional vector space over Z_2. For sets $B \subseteq \mathbf{Z}_2^n$ with $|B| \leq n$, the question is trivial, since all the elements of B can be linearly independent, and thus spanned by no smaller set. We are instead interested in sets B that are hard to cover, of size $n^{O(1)} > n$ or $n^{O(\log \log n)}$. If $|A|^k \leq |B|$, then, trivially, B cannot be contained in $\mathbf{Span}_k(A)$. So the interesting problem is when $n < |A| < |B| < |A|^k$.

Question 2.1 *Let r be a small integer, and let m and s be exponents satisfying $1 < s < m < rs$. Find an explicit construction of a set $B \subseteq \mathbf{Z}_2^n$ such that:*

$$|B| = O(n^m)$$

$$\textit{For all } A \subseteq \mathbf{Z}_2^n, \quad |A| = O(n^s) \Rightarrow B \not\subseteq \mathbf{Span}_r(A)$$

We have found a partial solution to this problem, when $m = rs - r/2$, described later in this note. A simple counting argument shows that non-coverable sets B should exist whenever $m > s$, so there is much room to reduce m further. The counting argument simply observes that there are many more sets B of size n^m than there are sets A of size n^s, and for each A, $\mathbf{Span}_r(A)$ contains a relatively small number of subsets of size n^m. Thus, some sets B are not covered by any set A.

The most important way to generalize this question, for our purposes, is to extend it to a non-constant r, such as $r = O(\log n)$, and a non-constant m, such as $O(\log \log n)$.

2.1 Arbitrary Boolean Functions

The sets B constructed in Question 2.1 can be considered as sets of (linear) queries on n bits that cannot all be answered using r XOR gates, even if we are allowed $O(n^s)$ bits of precomputed auxiliary (linear) data.

We are more interested in what cannot be computed using a small number of AND, OR, and NOT gates from $O(n^s)$ bits of precomputed auxiliary data. So we consider the analogous problem in the algebra of arbitrary Boolean functions of n inputs. If we consider this as an algebra over $\mathbf{Z}_2$, then addition is XOR, (pointwise) multiplication is AND, and OR and NOT are computable as low-degree polynomials. (We may also want to consider the algebra over $\mathbf{Z}_3$ or R, with 0 and 1 represented by 1 and -1, to use the alternate basis of Fourier basis functions, which is a group algebra over the group $\mathbf{Z}_2^n$.)

We restrict the computation performed by a small circuit by limiting the number of input bits accessed by the circuit, not by explicitly limiting the number of gates. We allow an arbitrary function (a truth table lookup) of a constant number of input bits. In the algebra of Boolean functions, this is expressed as a polynomial in r variables. There are 2^{2^r} different polynomials on r variables, but if $r = O(1)$, this is constant.

We define

$$\mathbf{Span}_r(A) = \left\{ p(a_1, a_2, \ldots, a_r) \;\middle|\; a_i \in A,\ p \text{ a polynomial} \right\}$$

Question 2.2 *Consider the 2^n dimensional algebra F of Boolean functions of n inputs.*

For a small integer constant r, is there an easily computable hard set of queries $B \subseteq F$, with $|B| = O(n^m)$, so that no set $A \subseteq F$ of precomputed functions with $|A| = O(n^k)$ contains B in its r-span? As above, we write

$$\textit{For all } A \subseteq F, \quad |A| = O(n^s) \Rightarrow B \not\subseteq \mathbf{Span}_r(A)$$

For this question, we know of no non-trivial answers. Simple non-constructive counting arguments, as before, show that such a set B should exist for any $m > k$, but the only proof for an explicit set B that we know is for the trivial cases of $m > kr$ when $|B| > |\mathbf{Span}_r(A)|$ asymptotically.

To show that no polynomial size dynamic data structure with update computations based on circuits of size $O(1)$ cannot solve a certain problem based on these queries, we need to go beyond a constant number r of inputs, and show that there is a set B of size $n^{O(\log \log n)}$ such that

$$B \not\subseteq Span_{O(\log n)}(A)$$

for any A of size $n^{O(1)}$.

3 Construction of a difficult set of linear queries

For any prime p, we can construct a set B of p^m elements of the vector space $\mathbf{Z}_2^{p^2}$ so that no $p/(m-1)$ elements of B are linearly dependent. We will then show that this set B is not contained in $\mathbf{Span}_r(A)$ for any $A \subseteq Z_2^{p^2}$ with $|A| < p^{m/r+1-1/r}$. This exponent is almost 1 bigger than the trivial bound of m/r.

Label the p^2 basis elements as $e_{i,j}$, with $0 \leq i, j < p$. We use the polynomials in the field Z_p of degree $m-1$ or less to create the elements of B. Let Q be such a polynomial. The corresponding element of B will be

$$\sum_{i=0}^{p-1} e_{i,Q(i)}$$

Since there are p^m polynomials of degree $m-1$ or less, including 0 and other constants, and no distinct polynomials agree on more than $m-1$ values (the roots of the difference polynomial), the size of B is p^m.

We now show that no set of $p/(m-1)$ or fewer elements of B sums to 0. Consider a set S of elements of B that sums to 0. Take one element, b, of S. This element is the sum of p basis elements $e_{i,j}$ of $\mathbf{Z}_2^{p^2}$. For a sum of elements to add to 0, each basis element must be added in an even number of times. Therefore, all p basis elements hit by b must also be hit by another element of S. But any two elements of B have at most $m-1$ basis elements in common, so there must be at least $p/(m-1)$ additional elements in S to hit all the basis elements hit by b.

Now we show that if $|A| < p^{m/r+1-1/r}$ and $B \subseteq \mathbf{Span}_r(A)$, there are subsets $A' \subset A$ and $B' \subset B$ such that

$$B' \subseteq \mathbf{Span}_r(A'), \qquad |A'| < p/(m-1) - 1, \qquad \text{and } |A'| < |B'|.$$

Once we have shown these three things, the proof concludes as follows: Since B' is contained in the $\mathbf{Span}_r(A')$, then B' is contained in the actual (vector space) span of A', which has dimension at most $|A'|$. Since the size of B' is greater than the dimension of a subspace containing it, there must be a set of at most $|A'|+1$ elements of B' that add to 0 (since we are working over $\mathbf{Z}_2$. But $|A'|+1 \leq p/(m-1)$, and by construction, no set of $p/(m-1)$ elements of B adds to zero, so we have a contradiction.

We show that a subset A' of A exists with the above properties by computing the expected number of elements of B covered by the span of a random subset of A of that size. Obviously, a subset A' of A exists that meets or exceeds that value.

Let $g = p/(m-1) - 1$, the desired size of A'. Let $a = |A|$. We will consider subsets of A with g elements, called g-sets, and subsets with r (or fewer) elements, r-sets. Consider an element of B. It is the sum of the elements of an r-set, S. There are at least $\binom{a-r}{g-r}$ g-sets containing S. Thus there are $\binom{a-r}{g-r}|B|$ pairs of a g-set and a covered element of B. But there are only $\binom{a}{g}$ g-sets, so the expected number of elements of $|B|$ covered by the r-span of a random g-set is

$$\frac{\binom{a-r}{g-r}}{\binom{a}{g}}|B| \approx \left(\frac{g}{a}\right)^r |B|$$

If we set $a = |A| < p^{m/r+1-1/r}$, and $g = |A'| = (p/m-1) - 1$, then we find that the expected number is greater than g, so these elements of B must be linearly dependent.