Theoretical Computer Science 11 (1980) 49–56

ON ANOTHER BOOLEAN MATRIX

Nicholas PIPPENGER

Mathematical Sciences Department, IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598, U.S.A.

Communicated by A. Meyer
Received September 1978

Abstract. In his paper "On a Boolean matrix", Nechiporuk gave an explicit example of a set of n homogeneous monotone Boolean functions of the first degree in n variables that require $\Omega(n^{3/2})$ two-input gates in any monotone Boolean network computing them. In this note we show how this can be extended to $\Omega(n^{5/3})$ two-input gates.

1. Introduction

In his paper "On a Boolean matrix", Nechiporuk [5] gave an explicit construction for an $n \times n$ Boolean matrix that requires $\Omega(n^{3/2})$ diodes in any network realizing it. (The notion of a diode network realizing a Boolean matrix is due to Lupanov [4], who showed by a counting argument that 'almost all' $m \times n$ Boolean matrices require at least $nm/\log_2 nm$ diodes when $m = 2^{o(n)}$ and $n = 2^{o(m)}$. An upper bound asymptotic to $nm/\log_2 nm$ has been given by Pippenger [6].)

With any $m \times n$ Boolean matrix M we can associate a set of m Boolean functions

$$f_1 = (M_{11} \wedge x_1) \vee \cdots \vee (M_{1n} \wedge x_n)$$
$$\vdots$$
$$f_m = (M_{m1} \wedge x_1) \vee \cdots \vee (M_{mn} \wedge x_n)$$

of the n Boolean variables $x_1, \ldots, x_n$. These functions are homogeneous monotone functions of degree 1, that is, each function is a disjunction over a subset of the variables. Nechiporuk's $n \times n$ Boolean matrix translates in this way to an explicit construction for n homogeneous monotone Boolean functions of degree 1 that require $\Omega(n^{3/2})$ gates in any monotone network computing them. (Without loss of generality, we shall assume every gate in a monotone network is either an OR-gate (two-argument disjunction) or an AND-gate (two-argument conjunction). The results of Lupanov and Pippenger cited above apply to monotone networks computing Boolean functions of degree 1 as well as to diode networks realizing Boolean matrices.)

Our goal in this note is to show how $\Omega(n^{3/2})$ in the results cited above can be improved to $\Omega(n^{5/3})$. This will be done as follows. With any $m \times n$ Boolean matrix we can associate a bipartite graph (with m inputs and n outputs) in an obvious way. The conditions we desire for Boolean matrices can then be translated into conditions on bipartite graphs; specifically, we seek bipartite graphs that have many edges but no large complete bipartite subgraphs. Nechiporuk's results then follow from a construction (due to Kővári, Sós and Turán [3]) of a bipartite graph with $\Omega(n^{3/2})$ edges but no copy of $K_{2,2}$ (that is, no complete bipartite subgraph with 2 inputs and 2 outputs). Our results follow from a similar construction (due to Brown [1]) with $\Omega(n^{5/3})$ edges but no copy of $K_{3,3}$. (If we had further constructions, with $\Omega(n^{2-1/r})$ edges but no copy of $K_{t,t}$, we could further improve our results, perhaps as far as $\Omega(n^2/(\log n)^2)$. This would still fall short of the bounds obtained by counting arguments.)

A key point in Nechiporuk's proof is that the functions associated with his matrix have 'nothing in common'; as a consequence of this, nothing can be gained by using AND-gates or by having overlap among the subnetworks computing the various functions. As a generalization of this, the functions associated with our matrix will have 'little in common', and thus little can be gained by using AND-gates or overlap. Our results will follow from precise quantitative versions of these ideas.

Wegener [7] has recently given an explicit construction for n homogeneous monotone Boolean functions of n Boolean functions that require $\Omega(n^2/(\log n)^2)$ gates in any monotone network computing them. (He has also given an upper bound of $O(n^2/\log n)$ for these functions.) This result does not supercede the results described above, however, since these functions have degree $\lfloor \log_2 n \rfloor$ rather than degree 1.

In addition to using $O(\cdot)$ and $o(\cdot)$ to denote error terms in the usual way, we shall use $U(\cdot)$ to denote a factor of the form $\exp O(\cdot)$ and $u(\cdot)$ to denote a factor of the form $\exp o(\cdot)$; these are equivalent to factors of the form $1+O(\cdot)$ and $1+o(\cdot)$, respectively, when the quantity denoted by the ellipsis tends to 0. Thus $U(1)$ denotes a factor bounded between positive constants, and $u(1)$, a factor tending to unity.

Our results will require upper bounds on the length of gaps between successive primes. These will take the following form: if p is the smallest prime not less than ξ, then $p = \xi U(\xi^{\theta})$. It is an open problem to determine the largest admissible value of θ; the best result known is that of [2], which shows that any θ in the range $0 < \theta < \frac{9}{20}$ is admissible. In what follows, θ will denote some fixed admissible value.

2. Matrices and diode networks

If F is a bipartite graph, $\#(F)$ will denote the number of edges in F, and $L(F)$ will denote the minimum possible number of diodes in a network realizing the matrix associated with F.

Lemma 1. *If F contains no copy of $K_{t+1,t+1}$, then*

$$L(F) \geq \#(F)/t^2.$$

Proof. Let G be a network that realizes the matrix associated with F using $L(F)$ diodes. The network G can be regarded as a directed graph (not necessarily bipartite) with

(1) the same inputs as F,

(2) the same outputs as F,

(3) the same paths as F (that is, a directed path from an input to an output if and only if there is a corresponding edge in F), and

(4) $\#(G) = L(F)$ edges.

We must show that

$$\#(G) \geq \#(F)/t^2.$$

We shall establish an accounting scheme whereby edges in F are 'charged against' edges in G. The scheme will be such that

(I) at most t^2 edges in F are charged against each edge in G, and

(II) each edge in F is charged against some edge in G.

The lemma will then follow immediately.

Let (a, b) be an edge in F. Since G has the same paths as F, there must be at least one path from a to b in G. Let

$$a = v_0, v_1, \ldots, v_k = b$$

be the vertices of such a path and let

$$(v_0, v_1), (v_1, v_2), \ldots, (v_{k-1}, v_k)$$

be its edges.

We shall say that an edge (v_{j-1}, v_j) in G is 'eligible' if

(i) there are paths from at most t inputs of G to v_{j-1}, and

(ii) there are paths from v_j to at most t outputs of G.

Our accounting scheme is as follows: an edge (a, b) in F will be charged against all eligible edges that lie on paths from a to b in G. We must now show that conditions (I) and (II) are satisfied.

Let (v_{j-1}, v_j) be an edge in G. If (v_{j-1}, v_j) is not eligible, then no edges in F are charged against it. If (v_{j-1}, v_j) is eligible, then the set A of inputs from which there are paths to v_{j-1} contains at most t inputs; similarly, the set B of outputs to which there are paths from v_j contains at most t outputs. If the edge (a, b) in F is charged against (v_{j-1}, v_j), then (v_{j-1}, v_j) lies on a path from a to b in G. Thus a appears in A, b appears in B, and (a, b) appears in $A \times B$, which contains at most t^2 edges. This verifies condition (I); it remains to show that condition (II) is satisfied.

Let (a, b) be an edge in F and consider a path from a to b in G. The set of all edges (v_{j-1}, v_j) on this path that satisfy condition (i) forms an initial segment of the path, for

if there are paths from more than t inputs to v_{j-1}, there will be paths from these inputs to v_j as well. Similarly, the set of all edges on the path that satisfy condition (ii) forms a final segment of the path. If these two segments have an edge in common, this edge will be eligible and the edge (a, b) in F will be charged against it; this will verify condition (II). If the two segments do not have an edge in common, there must be a vertex v_i that separates them, that is, a vertex v_i such that

(1) $j \leqslant i$ for all edges (v_{j-1}, v_j) that satisfy (i), and

(2) $i \leqslant j-1$ for all edges (v_{j-1}, v_j) that satisfy (ii).

It follows that the set A of inputs from which there are paths to v_i contains at least $t+1$ inputs and, similarly, that the set B of outputs to which there are paths from v_i contains at least $t+1$ outputs. Since G contains paths from all the inputs that appear in A to all the outputs that appear in B, F contains all the edges that appear in $A \times B$. Thus F contains a copy of $K_{t+1,t+1}$, a contradiction. This completes the verification of condition (II) and the proof of the lemma.

Construction 1. Given m and n, the graph $H_2(m, n)$ is constructed as follows. Let $w = \max\{m, n\}$. Define ξ by $\xi^2 + \xi + 1 = w$ so that

$$\xi = w^{1/2}\mathrm{U}(w^{-1/2}).$$

Let p be the smallest odd prime not less than ξ, so that

$$p = \xi\mathrm{U}(\xi^{-\theta}) = w^{1/2}\mathrm{U}(w^{-\theta/2}).$$

Let H be the bipartite graph corresponding to the projective plane of order p: the inputs correspond to the points of the plane, the outputs correspond to the lines, and there is an edge from an input to an output if and only if the corresponding point and lines are incident. This graph has

$$p^2 + p + 1 \begin{cases} \geqslant w \\ = w\mathrm{U}(w^{-\theta/2}) \end{cases}$$

inputs and outputs, each of which is incident with

$$p + 1 = w^{1/2}\mathrm{U}(w^{-\theta/2})$$

edges. Since any two points are incident with exactly one line in a projective plane, H contains no copy of $K_{2,2}$. (The foregoing is a slight modification of a construction due to Kővári, Sós and Turán [3].)

Let $H_2(m, n)$ be a graph obtained from H by first deleting $p^2 + p + 1 - m$ inputs and all the edges incident with them, then deleting $p^2 + p + 1 - n$ of the outputs that are incident with the fewest remaining edges, and all these edges incident with them. The resulting graph contains m inputs, n outputs, at least

$$\frac{mn(p+1)}{p^2+p+1} = \frac{mn}{w^{1/2}}\mathrm{U}(w^{-\theta/2}) = \frac{mn}{(\max\{m, n\})^{1/2}}\mathrm{U}((\max\{m, n\})^{-\theta/2})$$

edges, and no copy of $K_{2,2}$.

Combining this construction with Lemma 1 we obtain

Theorem 1.

$$L(H_2(m, n)) \geqslant \frac{mn}{(\max\{m, n\})^{1/2}} u(1)$$

and, in particular

$$L(H_2(n, n)) = \Omega(n^{3/2}).$$

This is the result of Nechiporuk [5].

The following lemma shows that Lemma 1 is very nearly the best possible.

Lemma 2. *For every positive integer t, and every $\varepsilon > 0$, there are graphs F and G as in the proof of Lemma 1 for which*

$$\#(G) \leqslant (1+\varepsilon)\#(F)/t^2.$$

Proof. Let G be the graph obtained by identifying each input of $H_2(n, n)$ with the output of a copy of $K_{t,1}$, and each output of the resulting graph with the input of a copy of $K_{1,t}$. Let F be the bipartite graph with the same paths.

If there were a set A of $t+1$ inputs and a set B of $t+1$ outputs such that A and B induce a copy of $K_{t+1,t+1}$ in F, then there must be paths in G from each input of A to each output of B. These paths must pass through at least 2 distinct inputs of $H_2(n, n)$ and also through at least 2 distinct outputs of $H_2(n, n)$. Thus $H_2(n, n)$ must contain a copy of $K_{2,2}$, which is impossible. It follows that F does not contain a copy of $K_{t+1,t+1}$, and therefore that F and G are as in the proof of Lemma 1.

On the other hand,

$$\#(G) = 2tn + n^{3/2}U(n^{-\theta/2}),$$

so the conclusion of the lemma holds for all sufficiently large n.

Construction 2. Given m and n, the graph $H_3(m, n)$ is constructed as follows. Let $w = \max\{m, n\}$. Define $\xi = w^{1/3}$. Let p be the smallest prime not less than ξ, so that

$$p = \xi U(\xi^{-\theta}) = w^{1/3}U(w^{-\theta/3}).$$

Let d be a non-zero element of GF(p) (the field of integers modulo p) such that d is a quadratic non-residue modulo p if $p \equiv 1 \pmod 4$, and a quadratic residue modulo p if $p \equiv 3 \pmod 4$. Let H be the bipartite graph with inputs corresponding to the points of the 3-dimensional affine space over GF(p), with outputs corresponding to the points of the same space, and with an edge from the input $a = (a_1, a_2, a_3)$ to the output $b = (b_1, b_2, b_3)$ if and only if

$$(a_1 - b_1)^2 + (a_2 - b_2)^2 + (a_3 - b_3)^2 \equiv d \pmod p.$$

Brown [1] has shown that this graph has

$$p^3\begin{cases}\geqslant w,\\ = wU(w^{-\theta/3})\end{cases}$$

inputs and outputs, each incident with

$$p^2-p=w^{2/3}U(w^{-\theta/3})$$

edges, and that it contains no copy of $K_{3,3}$.

Let $H_3(m, n)$ be a graph obtained from H by first deleting p^3-m inputs and all the edges incident with them, then deleting p^3-n of the outputs that are incident with the fewest remaining edges, and all these edges incident with them. The resulting graph contains m inputs, n outputs, at least

$$\frac{mn(p^2-p)}{p^3}=\frac{mn}{w^{1/3}}U(w^{-\theta/3})=\frac{mn}{(\max\{m, n\})^{1/3}}U((\max\{m, n\})^{-\theta/3})$$

edges, and no copy of $K_{3,3}$.

Combining this construction with Lemma 1 we obtain

Theorem 2.

$$L(H_3(m, n))\geqslant\frac{mn}{4(\max\{m, n\})^{1/3}}u(1)$$

and, in particular

$$L(H_3(n, n))=\Omega(n^{5/3}).$$

This is the result we have sought.

A comment is in order concerning the extent to which the foregoing constructions may be regarded as explicit. It will certainly be admitted that they are more explicit than Lupanov's lower bound of $\Omega(n^2/\log n)$. Nevertheless, three objections may be advanced. First, they assume the ability to find a prime not much larger than a prescribed real number. In fact, we know of no way to find such a prime without some exhaustive searching which can hardly be considered explicit. Second, the lower bound of $\Omega(n^{5/3})$ assumes the ability to find a quadratic non-residue modulo a prescribed prime. This can be avoided (at the cost of some deterioration in the error factors) by confining attention to primes congruent to 3 (mod) 4, for which only a quadratic residue (such as 1) is needed. Finally, these constructions assume the ability to delete from a graph a prescribed number of the vertices having the lowest degrees. This is not a completely explicit specification of the vertices to be deleted, but this too can be avoided (again at the cost of some deterioration in the error factors): since the original graphs are regular, it does not matter much which vertices are deleted.

3. Functions and monotone networks

First we shall consider monotone networks containing only OR-gates. If F is a bipartite graph, $L_1(F)$ will denote the minimum possible number of gates in such a monotone network computing the functions associated with the matrix associated with F.

Lemma 3. $L_1(F) \geq \frac{1}{2}L(F)$.

Proof. Consider a monotone network, containing only OR-gates, that computes $f_1, \ldots, f_m$ from $x_1, \ldots, x_n$. From this monotone network we can construct a directed graph which has a vertex for each of the variables $x_1, \ldots, x_n$ (these n vertices will be the outputs of the graph), a vertex for each OR-gate (the vertices corresponding to the gates computing the functions $f_1, \ldots, f_m$ will be the inputs of the graph), and two edges for each OR-gate (from the vertex corresponding to the gate to the vertices corresponding to its arguments). This directed graph can be regarded as a diode network, and it is easy to see that it realizes the matrix associated with F. Since the number of diodes is just twice the number of gates, the lemma follows.

Let us now consider monotone networks containing AND-gates as well as OR-gates. If F is a bipartite graph, $L_2(F)$ will denote the minimum possible number of gates in such a monotone network computing the functions associated with the matrix associated with F.

Lemma 4. *If F contains no copy of $K_{t+1,t+1}$, then*

$$L_2(F) \geq L_1(F)/\max\{t-1, 1\}.$$

Proof. Consider a monotone network that computes $f_1, \ldots, f_m$ from $x_1, \ldots, x_n$. We shall describe a surgical procedure that eliminates an AND-gate, introduces $t-1$ or fewer OR-gates, and preserves the property of computing $f_1, \ldots, f_m$ from $x_1, \ldots, x_n$. The lemma will then follow immediately by induction on the number of AND-gates in the network.

Let the network be topologically sorted so that the arguments of a gate precede the gate itself. We shall eliminate the AND-gate that appears last in this order, so that any gate intervening between this AND-gate and a gate computing one of the functions $f_1, \ldots, f_m$ must be an OR-gate. Suppose that this AND-gate computes $g \wedge h$ from g and h.

Let r denote the number of variables among $x_1, \ldots, x_n$ that imply $g \wedge h$ (this is just the number of terms of degree 1 in the disjunctive normal form of $g \wedge h$), and let s denote the number of functions among $f_1, \ldots, f_m$ that are implied by $g \wedge h$. Then we must have $r \leq t$ or $s \leq t$, else F would contain a copy of $K_{t+1,t+1}$.

If $r \leq t$, the disjunction of the terms of degree 1 in $g \wedge h$ can be computed by a subnetwork of $r-1 \leq t-1$ OR-gates. If the AND-gate computing $g \wedge h$ is replaced by this subnetwork, the overall network will still compute $f_1, \ldots, f_m$, else one of these functions would contain a term of degree 2 or greater. Thus the AND-gate can be replaced by $t-1$ or fewer OR-gates.

If on the other hand $s \leq t$, let the AND-gate computing $g \wedge h$ be replaced by the constant 0 and simplify the resulting network. Suppose the network now computes $f'_1, \ldots, f'_m$. We must have $f_i = f'_i$ except possibly for the s values of i for which $g \wedge h$ implies f_i. For the exceptional values of i, either g or h must imply f_i, else the latter would contain a term of degree 2 or greater. Thus for these values of i we must have either $f_i = f'_i \vee g$ or $f_i = f'_i \vee h$. This shows that the AND-gate can be replaced by t OR-gates. In fact it can be replaced by $t-1$ OR-gates, since if no OR-gate was eliminated during the simplification, the AND-gate computed one of the functions $f_1, \ldots, f_m$; in this case there can be at most one exceptional value of i, for which $f'_i = 0$ and thus for which $f_i = g$ or $f_i = h$. This completes the proof.

Combining these lemmas with Theorem 1 yields

Theorem 3. $L_2(H_2(n, n)) = \Omega(n^{3/2})$.

This is the result of Nechiporuk [5].

Combining them with Theorem 2 yields

Theorem 4. $L_2(H_3(n, n)) = \Omega(n^{5/3})$.

This is the result we have sought. In these applications, $t = 1$ or 2, and $\max\{t-1, 1\} = 1$, so nothing can be gained by using AND-gates.

References

[1] W.G. Brown, On graphs that do not contain a Thompson graph, *Canad. Math. Bull.* **9** (1966) 281–285.

[2] D.R. Heath-Brown and H. Iwaniec, On the difference between consecutive primes, *Bull. Amer. Math. Soc.* **1** (1979) 758–760.

[3] T. Kővári, V.T. Sós and P. Turán, On a problem of K. Zarankiewicz, *Colloq. Math.* **3** (1954) 50–57.

[4] O.B. Lupanov, O ventilnykh i kontaktno-ventilnykh skhemakh, *Dokl. Akad. Nauk SSSR* **111**(1) (1963) 50–53.

[5] E.I. Nechiporuk, Ob odnoi bulevskoi matritse, *Problemy Kibernet.* **21** (1969) 237–240; English translation: On a Boolean matrix, *Systems Theory Res.* **21** (1971) 236–239.

[6] N.J. Pippenger, On the evaluation of powers and related problems, in: *Proc. 17th Ann. IEEE Symp. on Found. of Comp. Sci.* (1976) 258–263; see also: The minimum number of edges in graphs with prescribed paths, *Math. Systems Theory* **12** (1979) 325–346.

[7] I. Wegener, Switching functions whose monotone complexity is nearly quadratic, in: *Proc. 10th Ann. ACM Symp. on Theory of Computing* (1978) 143–149; see also: *Theoret. Comput. Sci.* **9** (1979) 83–97.