On random orderings of variables for parity OBDDs

Petr Savický *

Abstract

Ordered binary decision diagrams (OBDDs) are a model for representing Boolean functions. There is also a more powerful variant called parity OBDDs. The size of the representation of a given function depends in both these models on an ordering of the variables.

It is known that there are functions such that almost all orderings of its variables yield an OBDD of polynomial size, but there are also some exceptional orderings, for which the size is exponential. We prove that for parity OBDDs, the size for a random ordering and the size for the worst ordering are polynomially related.

More exactly, for every $\varepsilon > 0$ there is a number $c > 0$ such that the following holds. If a Boolean function f is such that a random ordering of the variables yields a parity OBDD for f of size at most s with probability at least ε, then every ordering of the variables yields a parity OBDD for f of size at most s^c.

1 Introduction

Parity OBDDs were introduced by Gergov and Meinel [4] and simplified by Waack [11]. The structure of a parity OBDD depends on an ordering of the variables represented by a permutation π of $\{1, 2, \ldots, n\}$. For a given permutation π, a parity π-OBDD over the variables $x_1, x_2, \ldots, x_n$ means a directed acyclic graph with at most one source and at most one sink satisfying the following. Every non-sink node is labeled by a variable x_i for $i \in \{1, 2, \ldots, n\}$ and every edge is labeled by 0 or 1 or both. Moreover, it is required that if an edge leads from a node labeled by x_i to a node labeled by x_j, then $\pi(i) < \pi(j)$. If the ordering of the variables is not specified, the structure is called just a parity OBDD.

Let an assignment $a = (a_1, a_2, \ldots, a_n)$ of the variables be given. An edge starting in a node labeled by x_i is called consistent with the assignment, if the set of its labels contains a_i. A path from the source to the sink is called consistent with the assignment, if all its edges are consistent with it. The assignment is accepted, if the number of paths from the source to the sink consistent with the assignment is odd. In particular, if the graph is empty then no assignment is accepted and if the source coincides with the sink then all assignments are accepted. A parity OBDD represents a Boolean function f, if it accepts an assignment a if and only if $f(a) = 1$. Although the number of paths may be exponential, there is a simple algorithm that decides if an assignment is accepted or not, which works in time linear in the number of edges of the graph.

For a given ordering π, any parity π-OBDD with the minimal number of nodes among all parity π-OBDDs for a given function is called reduced. By a result of Waack [11], there is a polynomial time algorithm transforming a parity π-OBDD into a reduced one for the given

*Institute of Computer Science, Academy of Sciences of Czech Republic, Pod vodárenskou věží 2, 182 07 Praha 8, Czech Republic, e-mail: savicky@uivt.cas.cz. The research was partially supported by GA of the Czech Republic, Grant No. 201/98/0717, EU Grant No. INCO-COOP 96-0195 "ALTEC-KIT" and Ministry of Education grant OK-304.

ordering π. If the ordering is not fixed, no efficient algorithm to minimize the size of a parity OBDD is known.

The ordering of variables influences the size of a parity OBDD remarkably. For any two 0,1 vectors x and y of length n let $g(x,y)$ be 1 if and only if $x=y$. The function g can be represented by a parity OBDD of size polynomial in n for the ordering $x_1, y_1, x_2, y_2, \ldots, x_n, y_n$, but requires exponential size for the ordering $x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n$. For more information on parity OBDDs see [4], [11], [12].

Parity OBDDs were obtained as a generalization of OBDDs. The difference is that an OBDD contains two sinks, accepting and rejecting, and for every non-sink node there is at most one edge labeled by 0 and at most one edge labeled by 1 starting in the node. Hence, for every input, there is exactly one path from the source to one of the sinks. The path may be obtained by a deterministic step by step computation starting in the source and the input is accepted or rejected according to the sink, where the path leads to. If an OBDD is understood as a parity OBDD, it computes the same function, since every input has at most one accepting path.

OBDDs are used in applications as a data structure for representing the Boolean functions, since there are efficient algorithms for several required operations with the functions, if they are represented by OBDDs, see e.g. [2], [12]. Analogous algorithms exist also for parity OBDDs, see [4] and [11], however, some of them inlcude Gaussian elimination and hence, these are less efficient that the corresponding algorithms for OBDDs, although they also work in a polynomial time, see also [5].

From the point of view of the size of the representation, parity OBDDs are strictly more powerful than OBDDs. There are functions for which the difference in the required size is exponential, see [4], [11].

It is well-known that there are functions f such that for some orderings π_1, π_2, the sizes of the reduced π_1-OBDD and π_2-OBDD for f differ exponentially. We call an ordering π good for a function f, if it leads to a small size of a reduced π-OBDD for f. The problem of finding a good ordering is hard. In particular, it is NP-complete to test, if the size of a given OBDD may be reduced by changing the variable ordering, see [1]. If the measure of the quality of the ordering π is the size of the resulting π-OBDD, then the best ordering cannot be even approximated, see [8] and [9].

In order to get insight into the structure of the set of good orderings for a given function, random orderings were investigated. Let us mention explicitly the following result [7], see also [12]. There is a sequence of functions f_n of n variables such that the π-OBDD size of f_n for a random ordering π is $O(n^{3/2})$ with probability at least $1 - O(n^{-1/2} \log^{3/2} n)$, but for every n, there is an ordering π_n such that the size of any π_n-OBDD for f_n is exponential in $n^{1/2} \log^{-1/2} n$. We prove that in the case of parity OBDDs, the size for a random ordering and the size for the worst ordering are polynomially related.

Let $\mathrm{size}_\pi(f)$ denote the number of nodes in a reduced parity π-OBDD representing f. Note that the number of edges is at most the square of the size. A parity OBDD is called complete, if every path from the source to the sink in it tests all variables. The width of a complete parity OBDD is the number of nodes on the largest level. Let $\mathrm{width}_\pi(f)$ denote the minimum width of a complete parity π-OBDD representing f.

Every parity OBDD can be made complete by adding new nodes. This transformation increases the size at most by a factor of n and the width of the new OBDD is at most the size of the original one. Clearly, we have $\mathrm{width}_\pi(f) \leq \mathrm{size}_\pi(f) \leq n \cdot \mathrm{width}_\pi(f)$.

We prove that if $\mathrm{width}_\pi(f)$ is polynomial for a random ordering π with probability at least a positive constant, then $\mathrm{width}_\pi(f)$ is polynomial for all orderings, see Theorems 2.7 and 2.10. Because of the inequality above, the same result holds for the size.

2 The result

Let us consider Boolean functions defined on a set of variables $X = \{x_1, x_2, \ldots, x_n\}$. Let $\mathcal{P}(X)$ be the set of all subsets of X. It is convenient to consider $\mathcal{P}(X)$ as a linear space over the two element field, where the addition $\oplus$ means the symmetric difference.

By a subfunction of a Boolean function f on a set $A \subseteq X$, we mean any function obtained from f by any setting of the variables in $X \setminus A$ to constants. If the setting is denoted by c then the subfunction is denoted by $f|_c$.

Definition 2.1 Let $S(f, A)$ be the linear span over the two element field of the set of all subfunctions of f on A. Moreover, let $d(f, A)$ be the dimension of $S(f, A)$.

By results of Waack [11], both $\mathrm{size}_\pi(f)$ and $\mathrm{width}_\pi(f)$ may be expressed in terms of the dimension of linear spaces spanned by appropriately chosen sets of subfunctions of f. In particular, for the width, we have

Lemma 2.2 *For every f and π we have*

$$\mathrm{width}_\pi(f) = \max_{i=1,2,\ldots,n+1} d(f, \{x_{\pi^{-1}(i)}, x_{\pi^{-1}(i+1)}, \ldots, x_{\pi^{-1}(n)}\}).$$

Note that the set of variables used in the lemma is just the set of variables tested in the last $n - i + 1$ non-sink levels of a complete parity π-OBDD.

By a random permutation or a random set we mean an element chosen from the uniform distribution on the corresponding domain. We will need the following corollary of Lemma 2.2.

Corollary 2.3 *Let $\tilde{\pi}$ be a random permutation and let $\tilde{A}$ be a random set of variables. Then, for any f and w we have* $\Pr(\mathrm{width}_{\tilde{\pi}}(f) \le w) \le \Pr(d(f, \tilde{A}) \le w)$.

Proof: Let $\tilde{A}$ be a random set. Consider a random permutation of $\tilde{A}$ and independently a random permutation of $X \setminus \tilde{A}$. Then, construct an ordering $\tilde{\pi}$ by taking the permuted $X \setminus \tilde{A}$ first and then the permuted $\tilde{A}$. For every k, the conditional distribution of the new ordering under the condition $|\tilde{A}| = k$ is uniform. Hence, also the unconditional distribution of the new ordering is uniform. Since $d(f, \tilde{A}) \le \mathrm{width}_{\tilde{\pi}}(f)$, the lemma follows. $\square$

Let $N(f, A)$ denote the number of subfunctions of f on A. In [10], it is proved that $N(f, X \setminus A) \le 2^{N(f,A)}$ and $N(f, A \cap B) \le N(f, A)N(f, B)$. We prove analogous statements for $d(f, A)$. In terminology of [10], the next lemma means that $d(f, A)$ is *operation continuous* with $\alpha(x, y) = xy$ and $\beta(x) = x$.

Lemma 2.4 *For every Boolean function f on the variables X and for every subsets $A, B \subseteq X$, we have*

$$\begin{aligned} d(f, X \setminus A) &= d(f, A), \\ d(f, A \cap B) &\le d(f, A)d(f, B), \\ d(f, A \cup B) &\le d(f, A)d(f, B). \end{aligned}$$

Proof: Let x denote a setting of variables in A and let y denote a setting of variables in $X \setminus A$. Consider the function $f(x, y)$ as a matrix, where x represents the row index and y the column index. Then, the identity $d(f, X \setminus A) = d(f, A)$ follows from the fact that the dimensions of the spaces generated by rows and by columns in this matrix coincide.

Let us prove $d(f, A\cap B) \leq d(f, A)d(f, B)$. Let $\alpha = d(f, A)$ and let $c_1, c_2, \ldots, c_\alpha$ be settings to the variables in $X \setminus A$ such that $f|_{c_i}$ for all $i = 1, 2, \ldots, \alpha$ generate $S(f, A)$.

Let us consider a setting $c = c'c''$ of variables in $X \setminus A$, where c' is a setting of variables in $X \setminus (A \cup B)$ and c'' is a setting of variables in $B \setminus A$. Every subfunction of $f|_c$ on $A \cap B$ may be obtained by restricting an appropriate subfunction of $f|_{c'}$ on B by the setting c''. Moreover, taking the restriction by c'' is a linear map from $S(f|_{c'}, B)$ onto $S(f|_c, A \cap B)$. Hence, $d(f|_c, A \cap B) \leq d(f|_{c'}, B)$.

Clearly, every subfunction of $f|_{c'}$ on B is also a subfunction of f on B. Hence, $d(f|_{c'}, B) \leq d(f, B)$. Altogether, we have $d(f|_c, A \cap B) \leq d(f, B)$.

Let $\beta = d(f, B)$. Consider some i,$1 \leq i \leq \alpha$ and the corresponding setting c_i. By the previous paragraph, there are settings $d_{i,1}, d_{i,2}, \ldots d_{i,\beta}$ to the variables in $A \setminus B$ such that the functions $(f|_{c_i})|_{d_{i,j}}$ generate $S(f|_{c_i}, A \cap B)$.

Every setting of variables in $X \setminus (A\cap B)$ may be written as a combined setting cd consisting from a setting c of variables in $X \setminus A$ and a setting d of the variables in $A \setminus B$. The function $f|_c$ may be expressed as a linear combination of functions $f|_{c_1}, f|_{c_2}, \ldots, f|_{c_\alpha}$. Hence, $f|_{cd}$ may be expressed as a combination of functions $f|_{c_i d}$. Since each of the latter functions may be expressed as a combination of functions $f|_{c_i d_{i,j}}$ for all $i \leq \alpha$ and $j \leq \beta$, the required inequality follows.

In order to prove $d(f, A \cup B) \leq d(f, A)d(f, B)$ note that $d(f, X \setminus (A \cup B)) = d(f, (X \setminus A) \cap (X \setminus B))$ and use the inequalities for the complement and intersection. $\square$

Definition 2.5 A family of sets $\mathcal{A} \subseteq \mathcal{P}(X)$ is called K-complete if every set $A \subseteq X$ may be expressed from sets in $\mathcal{A}$ and their complements by an expression using $\cap$ and $\cup$ of depth at most K.

In order to extend bounds on $d(f, A)$ from sets $A \in \mathcal{A}$ to all subsets A, we use a variant of Lemma 3.5 from [10]. The proof may be done by a straightforward induction on the depth.

Theorem 2.6 ([10]) *Let a function $\phi : \mathcal{P}(X) \to \mathrm{R}$ be given such that for every sets $A, B \subseteq X$ we have $\phi(A \cap B) \leq \phi(A)\phi(B)$ and $\phi(X \setminus A) = \phi(A)$. Let $\mathcal{A}$ be a K-complete family such that for all $A \in \mathcal{A}$, we have $\phi(A) \leq w$. Then, for every $A \subseteq X$, we have $\phi(A) \leq w^{2^K}$.*

Theorem 2.7 *Let f be a Boolean function and w an integer such that the random ordering $\tilde{\pi}$ satisfies $\Pr(\mathrm{width}_{\tilde{\pi}}(f) \leq w) > \frac{1}{2}$. Then for every ordering π, we have $\mathrm{width}_\pi(f) \leq w^4$.*

Proof: Let $\mathcal{A} = \{A; d(f, A) \leq w\}$. By Corollary 2.3, a random set $\tilde{A}$ satisfies $\Pr(d(f, \tilde{A}) \leq w) > \frac{1}{2}$ and, hence, $|\mathcal{A}| > 2^{n-1}$. Then, it is easy to see that every $A \in \mathcal{P}(X)$ is expressible as $A = A_1 \oplus A_2$, where $A_1, A_2 \in \mathcal{A}$. Clearly, $A = (A_1 \cap (X \setminus A_2)) \cup ((X \setminus A_1) \cap A_2)$. This implies that $\mathcal{A}$ is 2-complete. Together with Theorem 2.6, this implies $d(f, A) \leq w^4$ for every A. The proof may then be finished using Lemma 2.2. $\square$

For the proof of a stronger bound, we need a bound on the covering radius of binary linear codes defined as follows. For a binary linear code C of length ℓ, let $\mathrm{cov}(C) = \max_{u \in \{0,1\}^\ell} \mathrm{dist}(u, C)$, where $\mathrm{dist}(u, C) = \min_{v \in C} \mathrm{dist}(u, v)$. The following theorem may be found e.g. as Theorem 8.1.21 in [3].

Theorem 2.8 ([3]) *Let C be a binary linear code of length ℓ, minimum distance at least 3 and codimension n. Then, $\mathrm{cov}(C) \leq k$, where k is the maximal integer satisfying $k - \lfloor \log_2 k \rfloor \leq n - \lfloor \log_2 \ell \rfloor$.*

Corollary 2.9 *For every $\varepsilon > 0$ let $k(\varepsilon)$ be the maximal integer k such that $k - \lfloor \log_2 k \rfloor \leq \log_2 \frac{1}{\varepsilon} + 2$. Let $\mathcal{A} \subseteq \mathcal{P}(X)$ be such that $|\mathcal{A}| \geq \varepsilon 2^{|X|}$. Then, for every subset $A \in \mathrm{span}(\mathcal{A})$, there are sets $A_1, A_2, \ldots, A_k \in \mathcal{A}$ such that $A = A_1 \oplus A_2 \oplus \ldots \oplus A_k$ and $k \leq k(\varepsilon)$.*

Proof: W.l.o.g., we may assume that $\mathrm{span}(\mathcal{A}) = \mathcal{P}(X)$, otherwise, we can consider $\mathrm{span}(\mathcal{A})$ as $\mathcal{P}(X')$ for an appropriate X' smaller than X.

Let ℓ be the number of nonempty sets in $\mathcal{A}$. Let M be an n times ℓ matrix, whose columns are the characteristic functions of the nonempty elements of $\mathcal{A}$. Since $\mathrm{span}(\mathcal{A}) = \mathcal{P}(X)$, we have $\mathrm{rank}(M) = n$. Let $C = \{u; Mu = 0\}$. Clearly, C is a linear code of length ℓ, minimum distance at least 3 and codimension n. Since $\ell \geq \varepsilon 2^n - 1$, we have $n - \lfloor \log_2 \ell \rfloor \leq \log_2 \frac{1}{\varepsilon} + 2$. Hence, by Theorem 2.8, we have $\mathrm{cov}(C) \leq k(\varepsilon)$.

Since $A \in \mathrm{span}(\mathcal{A})$, there is a vector x such that $Mx = A$, where we identify set A and its characteristic function. Since $\mathrm{cov}(C) \leq k(\varepsilon)$, there is a vector $u \in C$ such that $\mathrm{dist}(x, u) \leq k(\varepsilon)$. Finally, since $M(x \oplus u) = A$, the nonzero entries of $x \oplus u$ determine the selection of elements $A_1, A_2, \ldots, A_k \in \mathcal{A}$, such that $k \leq k(\varepsilon)$ and $A_1 \oplus A_2 \oplus \ldots A_k = A$. $\Box$

It is easy to verify that $k(\varepsilon) \leq \log \frac{1}{\varepsilon} + \log\log \frac{1}{\varepsilon} + O(1)$.

Theorem 2.10 *Let f be a Boolean function and let $w \geq 2$ and ε, $0 < \varepsilon \leq \frac{1}{2}$ be such that the random ordering $\tilde{\pi}$ satisfies $\Pr(\mathrm{width}_{\tilde{\pi}}(f) \leq w) > \varepsilon$. Then, for every ordering π, we have*

$$\mathrm{width}_{\pi}(f) \leq w^{O\left(\log^2 \frac{1}{\varepsilon}\right)}.$$

Proof: Let $\mathcal{A} = \{A; d(f, A) \leq w\}$. By Corollary 2.3, a random set $\tilde{A}$ satisfies $\Pr(d(f, \tilde{A}) \leq w) > \varepsilon$ and, hence, $|\mathcal{A}| > \varepsilon 2^n$. The dimension of the space generated by subfunctions on a one-element set of variables is at most 2. Hence, for every i, we have $\{x_i\} \in \mathcal{A}$. It follows that $\mathrm{span}(\mathcal{A}) = \mathcal{P}(X)$. Let A be any subset of X. By Theorem 2.9, there are sets $A_1, A_2, \ldots, A_{k(\varepsilon)} \in \mathcal{A}$ such that $A = A_1 \oplus A_2 \oplus \ldots \oplus A_{k(\varepsilon)}$.

Represent the expression of A in terms of $A_1, A_2, \ldots, A_{k(\varepsilon)}$ as a binary tree of depth $\lceil \log k(\varepsilon) \rceil$ with inner nodes labeled by $\oplus$. Then, starting from the subtrees of depth 1 replace each subtree with $\oplus$ in its root using the identity $A \oplus B = (A \cap (X \setminus B)) \cup ((X \setminus A) \cap B)$. Finally, propagate the complement operation to the leaves using de Morgan rules. The depth of the obtained expression is at most $2\lceil \log k(\varepsilon) \rceil$. Hence, $\mathcal{A}$ is $2\lceil \log k(\varepsilon) \rceil$-complete. Together with Theorem 2.6, this implies $d(f, A) \leq w^{O(k(\varepsilon)^2)}$ for every set A. Hence, by Lemma 2.2, $\mathrm{width}_{\pi}(f) \leq w^{O(k(\varepsilon)^2)}$. $\Box$

Acknowledgement. The author is very grateful to Pavel Pudlák and Noga Alon for directing his attention to [3] and to Pavel Pudlák and Oleg Verbitsky for helpful discussion on the bounds on the covering radius.

References

[1] B. Bollig, I. Wegener, Improving the variable ordering of OBDDs is NP-complete. *IEEE Trans. on Computers* 45 (1996), pp. 993 – 1002.

[2] R. E. Bryant, Graph based algorithms for Boolean function manipulation. *IEEE Trans. on Computers* 35, pp. 677-691.

[3] G. Cohen, I. Honkala, S. Litsyn, A. Lobstein, Covering Codes. North Holland Mathematical Library, 1997.

[4] J. Gergov, Ch. Meinel, MOD-2-OBDDs - a data structure that generalizes EXOR-sum-of-products and ordered binary decision diagrams. *Formal Methods in System Design* 8, pp. 273–282.

[5] M. Loebbing, D. Sieling, I. Wegener, Parity OBDDs cannot be handled efficiently enough. Information Processing Letters 67 (1998), pp. 163–168.

[6] F. J. MacWilliams, N. J. A. Sloane, The theory of Error-Correcting Codes. Elsevier Science, 1978.

[7] P. Pudlák, personal communication.

[8] D. Sieling, On the existence of polynomial time approximation schemes for OBDD minimization, STACS'98, LNCS 1373, pp. 205–215.

[9] D. Sieling, The Nonapproximability of OBDD Minimization, Revision 1 of ECCC Report TR98-001, 1998, Trier.

[10] M. Szegedy, Functions with Bounded Symmetric Communication Complexity, Programs over Commutative monoids, and ACC, *J. of Computer and System Sciences* 47 (1993), pp. 405–423.

[11] St. Waack, On the descriptive and algorithmic power of parity ordered binary decision diagrams, STACS'97, LNCS 1200, pp. 201–212.

[12] I. Wegener, Branching programs and binary decision diagrams - theory and applications. To appear 1999 in the SIAM monograph series Trends in Discrete Mathematics and Applications (ed. P.L. Hammer).