Lower Bounds on Formula Size of Boolean Functions using Hypergraph-Entropy

Ilan Newman
Avi Wigderson *
Hebrew University

February 13, 1995

Abstract

Körner [7] defined the notion of graph-entropy. He used it in [8] to simplify the proof of the Fredman-Komlos lower bound for the family size of perfect hash functions.

We use this information theoretic notion to obtain a general method for formula size lower bounds. This method can be applied to low-complexity functions for which the other known general methods ([11, 12, 3] and see also [17]) do not apply. Specifically the results are:

1. A new general lower bound on the formula size of quadratic Boolean functions.
2. As a corollary we get an $\Omega(n^2 logn)$ lower bound for the function that decides whether a graph of n vertices has a cycle of length four, and to the function that decides whether a graph has a vertex of degree at least two.
3. A simple proof of a result of Krichevskii, [10] , stating that the formula size for the threshold-2 Boolean function with n variables is at least $n \log n$.
4. A simple proof of a lower bound first proved by Snir, [16], stating that a $\bigvee \bigwedge \bigvee$ formula for n- variable threshold-k function, where all $\wedge$ gates have fan in k, has the size of
$$\Omega(\ n \ \frac{\log n - log(k-1)}{\log k - log(k-1)} \ ) = \Omega(nk \log \frac{n}{k} \ )$$

Notation:

1. Let X be a finite set, interpreted as Boolean variables. A **formula** is a rooted tree whose leaves are labeled with members of X or their negations, and whose internal

*Partially supported by American - Israeli binational science foundation, grant no. 87-00082.

nodes are labeled with the Boolean operations AND, OR. The root of the tree computes a Boolean function $f; \{0,1\}^X \mapsto \{0,1\}$ in the natural way. If no negations appear, we say that the formula and the function computed are both monotone. The size of a formula is the number of leaves in the tree.

Let f be any Boolean function. The **formula size** of f is the minimum size of a formula that computes f and it is denoted by $L(f)$. For a monotone function f, the minimum size of a monotone formula for f is denoted by $L_M(f)$.

2. The threshold-k Boolean function, denoted by T_k^n is a Boolean function on n variables, that gets the value 1 if and only if the input has at least k variables assigned 1.

3. The set $\{1..n\}$ is denoted as [n].

4. For a Boolean function f on n variables we will assume that the variables are numbered from 1 to n, and by writing $f(T) = 0 \quad (f(T) = 1), \quad T \subseteq$[n], we mean that $f(x) = 0$ $(f(x) = 1)$ for the characteristic vector x of T.

5. A hypergraph $G = (V, E)$ is a set of 'vertices' V, and a set E of subsets of V (also called the set of 'edges'). If all edges $e \in E$ have a constant size k, the hypergraph is called k-uniform. A 2-uniform hypergraph is simply a graph.

6. The s-uniform hypergraph on n vertices that contains all subsets of size s, is called the complete s-uniform hypergraph and is denoted by K_n^s. The complete graph on n vertices is denoted by K_n.

7. An independent set I is a subset of V that contains no edges of E.

8. For a probability distribution Q_{XY} on a cross product $A \times B$, Q_X (Q_Y) denotes the marginal distribution of Q_{XY} on A (B).

9. All logarithms are to the base 2.

1 Definition and basic properties of Entropy and Hypergraph Entropy:

1. Let X, Y be random variables in some probability space. The entropy of X is defined as:
$$H(X) = \sum_x p(x) log \frac{1}{p(x)}$$
The mutual information between X, Y is defined as $I(X, Y) = H(X) + H(Y) - H((X, Y))$ and it may be written also as:
$$I(X, Y) = H(X) \ - \ \sum_{x,y} p(x, y) log \frac{1}{p(x|y)}$$

Further information on information theory may be found in [1].

2. Körner, [7, 8] defined the notion of hypergraph entropy as follows: Let $G(V,E)$ be a hypergraph and P a probability distribution on V. Let $A(G)$ be the collection of all maximal independent sets of G.

   Define $\mathcal{Q}(G,P)$ to be the set of all probability distributions Q_{XY} on $V \times A(G)$ such that for every $v \in V$,
   (a) $Q_{XY}(v,I) = 0$ if $v \notin I$.
   (b) $Q_X(v) = P(v)$. (where $Q_X(.)$ is the marginal distribution of Q_{XY} on V).

   The hypergraph entropy $H(G,P)$ is defined:

   $$H(G,P) = min\{I(X,Y)|Q_{XY} \in \mathcal{Q}(G,P)\}$$

   Where $I(X,Y)$ is the mutual information between two random variables X and Y that are distributed according to the marginal distributions Q_X and Q_Y.

3. Here after we consider only uniform distributions on V, so we refer to the hypergraph entropy of G as H(G).

4. We shall need the following basic properties of H(G) proved by Körner and Marton, [8, 7, 9].

   (a) For two hypergraphs on the same vertex set $G(V,E_G), F(V,E_F)$ let $K = G \cup F$ be the hypergraph on V with $E = E_G \cup E_F$, then $H(K) \leq H(G) + H(F)$.
   (b) The hypergraph entropy is monotone, that is deleting an edge can only decrease the entropy.
   (c) The entropy of the complete k-uniform hypergraph is $H(K_n^k) = \log n - log(k-1)$.
   (d) The entropy of a bipartite graph on m (out of n) vertices does not exceed m/n.
   (e) The generalization for hypergraphs: Let $G = (V,E)$ be a k-uniform hypergraph. We call G a k-partite hypergraph if there is a partition of V into k parts $V_1 \dots V_k$, such that for every edge $e \in E$ and every V_i, $i \in$[k], $|e \cap V_i| = 1$.

       We have that the entropy of k-partite hypergraph on m (out of n) vertices is no more than $\frac{m}{n}(\log k - \log(k-1))$.

2 A lower bound for formula size of Boolean functions.

In this section we develop a general technique for formula lower bounds. A natural approach is to associate a nonnegative cost function μ to each Boolean function with the property that if $f = g \diamond h$ then $\mu(f) \leq \mu(g) + \mu(h)$ (where $\diamond$ is either $\bigwedge$ or $\bigvee$). Such a cost function is called 'abstract complexity meassure' [17, 15]. It directly gives a lower bound on the formula size of a Boolean function in terms of its cost. We find that for monotone formulae graph entropy is a natural choice for such a measure. It leads to nontrivial lower bounds for monotone formulae for quadtratic functions (sect 2.1). Then we extend it to the nonmonotone case using a lemma of Krichevskii.

2.1 A lower bound for monotone formula size

We assign each monotone function a cost function and prove that the cost of a function computed by $\bigvee$,$\bigwedge$ gates is no more than the sum of costs of the inputs. (thus the cost function is 'abstract complexity meassure' for monotone formulae). The cost of a single variable will be $1/n$. Hence, (by induction on the formula) for a function of cost μ one gets a lower bound of $n\mu$.

Definion of the cost function:

Let $g : \{0,1\}^n \mapsto \{0,1\}$ be Boolean function g on n variables. We will identify the variable set with the set $[n]$. Define:

1. $(g)_k$ is the set of 'minterms' of g of size k, formally; $(g)_k = \{S |\ S \subseteq [\mathrm{n}],\ |S| = k,\ g(S) = 1,\ \forall T \subset S\ g(T) = 0\}$.
2. We will be interested only in $(g)_1, (g)_2$. Observe that $(g)_1$ is a subset of [n] and $(g)_2$ is a set of unordered pairs on [n]. (g_2) will be identified with the graph $G(g) = (V, E),\ V = [\mathrm{n}],\ E = (g)_2$.
3. The cost μ of a function g will be defined as:

$$\mu(g) = H(G(g)) + \frac{|(g)_1|}{n}$$

Theorem 1 Let g be a monotone Boolean function. Let $L_M(g)$ be the monotone formula size of g. Then, $L_M(g) \geq n\mu(g)$.

Proof: We note that

1. For a variable x_i (a leaf of a formula), $(x_i)_1 = \{i\},\ \ G(x_i) = \phi$, (the empty graph), and so $\mu(x_i) = \frac{1}{n}$.
2. The cost function is monotone with respect to inclusion; if $(g)_1 \subseteq (h)_1$ and $(g)_2 \subseteq (h)_2$ then $\mu(g) \leq \mu(h)$.

Sub-Additivity for $\bigvee$ gate: Let $g = h \bigvee f$. We have, $(g)_1 = (h)_1 \cup (f)_1$, and $(g)_2 \subseteq G(h) \cup G(f)$, thus:

$$\mu(g) \le \frac{|(h)_1 \cup (f)_1|}{n} + H(G(h) \cup G(f))$$

$$\le \frac{|(h)_1|}{n} + \frac{|(f)_1|}{n} + H(G(h)) + H(G(f)) \ = \mu(h) + \mu(f)$$

The first inequality is by the monotonicity of μ. The second is by 4a in section 1.

Sub-additivity for $\bigwedge$ gate: Let $g = h \bigwedge f$. Denote $A = (h)_1,\ B = (f)_1$.

We get that $(g)_1 = A \cap B$, and $(g)_2 \subseteq G(h) \cup G(f) \cup G((A-B),(B-A))$, where $G(L, M)$ denotes the complete bipartite graph G with parts L and M. Thus

$$\mu(g) \le H(\ G(h)\ \cup G(f)\ \cup\ G((A-B),(B-A))\ ) + \frac{|A \cap B|}{n}$$

$$\le H(G(h)) + H(G(f)) + \frac{|A-B| + |B-A|}{n} + \frac{|A \cap B|}{n}$$

$$\le H(G(h)) + H(G(f)) + \frac{|A| + |B|}{n} = \mu(h) + \mu(f)$$

The first inequality is by the monotonicity of μ. The second is by 4a and 4d in section 1.

The theorem now follows since $\mu(t) = \frac{1}{n}$ for any leaf t of the formula and the cost of the output function does not exceed the sum of costs of all the leaves. $\Box$

Remark We note here that the best this method can give (by direct application) are lower bounds of at most $n \log n$.

2.2 The general lower bound

We use here a lemma (Krichevskii [10]) to extend our monotone lower bound method to nonmonotone formulae. We get:

Theorem 2 Let f be a Boolean function with $f(S) = 0$ for every $S,\ |S| = 1$. Then $L(f) \ge n\mu(f)$.

Proof:
Lemma: [10] Let $f(x_1, .., x_n)$ be any Boolean function for which $f(S) = 0$ for every $S,\ |S| = 1$. Then, there is a monotone function ψ_f such that

1. $\psi_f(S) = 0$ for any $S,\ |S| = 1$.
2. $\psi_f(S) \ge f(S)$ for every $S,\ |S| = 2$.
3. $L_M(\psi_f) \le L(f)$.

Proof (lemma): The proof is by induction on the formula size $L(f)$. For $L(f) = 2$ the claim is true. Let F be an optimal formula for f. If $F = G \bigvee H$, where G (H) is optimal formula for g (h), then by induction there are ψ_g and ψ_h for which (a) (b) and (c) are satisfied. It is easy to see that $\psi_f = \psi_g \bigvee \psi_h$ satisfies (a) (b) and (c) for f.

If $F = G \bigwedge H$ with the functions g, h respectively; Define $G_1 = \{x_i | g(\{x_i\}) = 1, \ and\ x_i\ appears\ in\ G\}$. Define H_1 similarly. By the assumption on f, it follows that $G_1 \cap H_1 = \phi$. Assume (w.l.o.g) that $G_1 = \{x_1, .., x_k\}$ and $H_1 = \{x_{k+1}, .., x_{k+l}\}$. Let

$$F^* = (x_1 \vee .. \vee x_k \vee G(0, ..0, x_{k+1}, ..x_n)) \bigwedge (x_{k+1} \vee .. \vee x_{k+l} \vee H(x_1, .., x_k, 0, .., 0, x_{k+l+1}, .., x_n))$$

We have that F^* is a formula for some function f^*. It is easy to verify that for any S with $|S| = 1,\ f^*(S) = 0$, and for any S with $|S| = 2,\ f^*(S) \geq f(S)$. In addition $G(0, ..0, x_{k+1}, ..x_n)$ and $H(x_1, .., x_k, 0, .., 0, x_{k+l+1}, .., x_n)$ are formulae of some functions g^*, h^* that meet the requirements of the lemma, so by induction there are monotone functions $\psi_{g^*},\ \psi_{h^*}$ with monotone formulae G^*, H^* as required. Observe that by plugging G^*, H^* into F^* we get a monotone formula for ψ_f that satisfies (a) (b) and (c).

We proceed now with the proof of the theorem. By the previous lemma there is a monotone function ψ_f for which $L(f) \geq L_M(\psi_f)$, from theorem 1 we get $L_M(\psi_f) \geq n\mu(\psi_f)$. Since $\psi_f(S) \geq f(S)$ for $|S| \leq 2$, the monotonicity of the cost function μ implies the result. □

2.3 Application to specific functions

Let $C4(n)$ be the Boolean function that decides '1' on a graph of n vertices if the graph contains a cycle of length 4.

Let $D2(n)$ be the Boolean function that decides '1' on a graph of n vertices if the graph contains a vertex of degree at least 2.

Note, $C4(n)$ and $D2(n)$ are Boolean function on $N = \binom{n}{2}$ variables.

Corollary 1: Any fromula for $D2(n)$ has size of $\Omega(n^2 \log n)$.

Proof: By theorem 2 it is enough to show that $\mu(D2(n)) = \Omega(\log n)$. Observe that $(D2(n))_1 = \phi$ and $(D2(n))_2 = L(K_n)$, the line graph of K_n (the graph whose vertices are the edges of K_n and two edges are connected if they have a common vertex in K_n).

We show that $\mu(D2(n)) = \Omega(\log n)$ by explicitly specifying the optimal distributions according to the definition of graph entropy. We do that by showing an upper bound of $\log(n-1)$ and $\log n/2$ on the graph entropies of $L(K_n)$ and its complement, respectively. (Note that the sum of these two numbers is $\log \binom{n}{2}$). However, by 4a and 4c in sec 1, the sum of the two graph entropies must be at least $\log \binom{n}{2}$, thus the upper bounds are in fact tight.

The independent sets of $L(K_n)$ are matchings (in K_n). The cliques in $L(K_n)$ are stars and triangles (in K_n). (A star is a set of edges all adjacent to a vertex). Let $\mathcal{M}$ denote the

set of perfect matchings, $\mathcal{S}$ denote the set of maximal stars and E denote the edge set of K_n. Define the probability Q_1 on $E \times \mathcal{M}$; $Q_1(e|M) = \frac{2}{n}$ for every matching $M \in \mathcal{M}$ and $e \in M$, and such that the induced probability on $\mathcal{M}$ is uniform. Let (X_1, Y_1) be a random variable on $E \times \mathcal{M}$ distributed according to Q_1. Define a probability distribution Q_2 on $E \times \mathcal{S}$; $Q_2(e|S) = \frac{1}{n-1}$ for every star in $\mathcal{S}$, $e \in \mathcal{S}$ and such that the induced probability on $\mathcal{S}$ is uniform. Let (X_2, Y_2) be a random variable on $E \times \mathcal{S}$ distributed according to Q_2.
It is easy to check, using the definitions in section 1.1, that: $I_1(X_1, Y_1) = \log(n-1)$, $I_2(X_2, Y_2) = \log n/2$ But,

$$\log \binom{n}{2} \leq H(L(G)) + H(L(G)^C) \leq I_1 + I_2 = \log \binom{n}{2}$$

(The first inequality is by 4a in section 1, and by the fact that $L(G) \cup L(G)^C$ is the complete graph on $\binom{n}{2}$ vertices. The second inequality is by the definition of graph entropy, section 1.2). Thus we get equality all the way and $H(L(G)) = I_1 = \Omega(\log n)$. $\square$

Corollary 2: Any formula for $C4(n)$ has size of $\Omega(n \log n)$.

Proof: Let f be the restriction of $C4(n+1)$ obtained by: Take a special vertex z and set all edges adjacent to it to '1'. Clearly if there is a vertex of degree at least two in the remaining graph (the graph induced by unset edges), then there is a cycle of length 4 in the original graph. We have $(f)_1 = \phi$, $(f)_2 \supseteq (D2)_2$ so by the monotonicity of the cost we get that $\mu(D2) \leq \mu(f)$ and the result follows by corollary 2.

Corollary 3:[10] Let T be any formula that computes T_2^n, then the size of T is at least $nlogn$. (We note here that this is best possible).

Proof: $(T_2^n)_1 = \phi$, $G(T_2^n) = K_n$, (the complete graph on n vertices), thus by 4c in section 1. $\mu(T_2^n) = log\ n$. $\square$

Remark: A proof of the monotone formula lower bound for T_2^n was given also by Hansel [5]. (See also [13]).

3 A lower bound on the size of $\bigvee\bigwedge\bigvee$ formula for threshold-k function, where $\wedge$ gates fan in is k.

A $\bigvee\bigwedge\bigvee$ formula, where $\wedge$ gates have fan in k, is a formula of the form $\bigvee_{i=1}^{p} \bigwedge_{j=1}^{k} \bigvee_{q \in S_{ij}} t_q$, where $t_q \in \{x_q, \neg x_q\}$ for every q.

Theorem: [16] The size of a $\bigvee\bigwedge\bigvee$ formula for T_k^n, where $\wedge$ gates have fan in k, is at least

$$s \geq \frac{n\ log \frac{n}{k-1}}{log \frac{k}{k-1}}$$

Remarks:

1. This result was significantly improved recently by J. Radhakrishnan [14] using graph entropy methods. He proved a near optimal lower bound for any $\bigvee\bigwedge\bigvee$ formulae of $e^{\delta(k)} n \log n$ where $\delta(k) = \Omega(\frac{\sqrt{k}}{log^2 k})$ and $k < \log n$.

2. For constant k there are (optimal) construction of $O(n \log n)$ $\bigvee\bigwedge\bigvee$ formulae for T_k^n [6, 4].

The original proof was based on some ad-hoc combinatorial considerations. We will go along the lines of the proof of the previous section.

Proof: Consider a minimum size $\bigvee\bigwedge\bigvee$ formula for T_k^n. That is, of the form $\bigvee_{i=1}^{p} \bigwedge_{j=1}^{k} \bigvee_{q\in S_{ij}} t_q$. Let $\{g_i,\ i = 1, ..., p\}$ be the functions computed at the $\bigwedge$ gates.
Clearly, the formula must be monotone (that is, no negations), and for every fixed i, $1 \leq i \leq p$, the sets S_{ij}, $1 \leq j \leq k$ are pairwise disjoint.

Let g be any Boolean function define, as in the previous section, $(g)_k = \{S \subseteq [\text{n}] : |S| = k,\ g(S) = 1,\ \forall T \subset S,\ g(T) = 0\}$ Define the hypergraph G_i whose edge set is $(g_i)_k$. We get that G_i is a k-partite hypergraph on vertex set $\bigcup_j S_{ij}$. (the 'parts' are S_{ij}, $1 \leq j \leq k$). Similarly, define T to be the hypergraph whose edge set is $(T_k^n)_k$. T is the complete k-regular hypergraph K_n^k. Since $T_k^n = \bigvee_i g_i$ we get that $(T_k^n)_k = \cup_i (g_i)_k$. That is, a formula of this kind for T_k^n defines a way to decompose the complete k-regular hypergraph to a union of k-partite hypergraphs.

The size of the formula is

$$s \ = \ \sum_{i=1}^{p}\sum_{j=1}^{k} |S_{ij}| \ = \ \sum_{i=1}^{p} |V(G_i)|$$

The hypergraph entropy of K_n^k is $log\ n\ - log(k-1)$. (4c in section 1). For each G_i we have $H(G_i)\ = \leq \frac{|V(G_i)|}{n}(log\ k\ - log(k-1)\ )$ (by 4e in sec. 1). Thus, by the subadditivity of the hypergraph entropy, (4a in section 1):

$$H(K_n^k) \ = \ log\ n\ - log(k-1) \ \leq \sum_{i=1}^{p} H(G_i)$$

$$\leq \ \frac{1}{n}(log\ k\ - log(k-1)\ ) \sum_{i=1}^{p} |V(G_i)| \ = \ \frac{s}{n} log \frac{k}{k-1}$$

And we get the desired lower bound:

$$s \ \geq \ \frac{n\ log \frac{n}{k-1}}{log \frac{k}{k-1}}$$

Acknowledgements
We are grateful to Mauricio Karchmer and Aviad Cohen for helpful discussion.

References

[1] I. Csiszar, J. Körner, Information Theory: Coding Theorems for Discrete Memoryless Systems, Academic Press, New York, 1982.

[2] M. Fredman, J. Komlos, On the size of separating systems and perfect Hash functions, SIAM J. of Algorithms and discrete Meth. 5(1984) 61-68.

[3] Fischer, Meyer, Paterson, $\Omega(n \log n)$ lower bounds on length of Boolean formulas. SIAM Journal on comp. 11 (1982), 416-427.

[4] J. Friedman, Constructing $O(n \log n)$ size monotone formulae for the k-th elementary symmetric polynomial of n Boolean variables, SIAM J. on Computing, 15(1986) 641-654.

[5] G. Hansel, Nombre minimal de contacts de fermature nessecaires pour realiser une function booleenne symetrique de n variables, C. R. Acad. Sci. Paris 258, 1964 pp. 6037-6040.

[6] L.S. Khasin, Complexity bounds for the realization of monotone symmetrical functions by means of formulas in the basis +*-, Soviet Physics, Dokladi 14 (1970) 1149-1151.

[7] J. Körner, Coding of an information source having ambiguous alphabet and the entropy of graphs, Trans. 6-th Prague Conf. on Information Theory, Academia, Prague 1973, pp. 441-425.

[8] J. Körner, Fredman-Komlos bounds and information theory, Siam J. Alg. Disc. Meth. vol. 7, 1986, pp. 560-570

[9] J. Körner, K. Marton, New bounds for perfect hashing via information theory, to appear in European J. combinatorics.

[10] R.E Krichevskii, Complexity of contact circuits realizing a function of logical algebra, Soviet Physics, Dokladi 8(1964) 770-772.

[11] Krapchenko, A Method of obtaining lower bounds for the complexity of π-schemes. Math. Notes Acad. Sci. USSR 11 (1972), 474-479.

[12] Nechiporuk, A Boolean function. Sov. Math. Dokladi 7 (1966) 999-1000.

[13] N. Pippenger, An information-Theoretic Method in Combinatorial Theory, J. Combinatorial theory. Vol. 23 no. 1. July 1977, 99-104.

[14] J. Radhakrishnan, Phd thesis, Rutgers, 1991.

[15] A. A. Razborov, On sub modular complexity measures, to appear in 'Boolean function complexity: selected papers from LMS Symp. Durham, July 1990' Edited by M.S. Paterson, Cambridge U. Press.

[16] M. Snir, The covering problem of complete uniform hypergraphs, a note, Discrete Math. 27(1979) 103-105.

[17] I. Wegener, The Complexity of Boolean functions. Wiley-Teubner Series in Computer Science. 1987.