The formula size of PARITY

Troy Lee *
troyjlee@gmail.com

Abstract

We exactly determine the formula size of the parity function. If $n = 2^\ell + k$, where $0 \leq k < 2^\ell$, then the formula size of parity on n bits is $2^\ell(2^\ell + 3k) = n^2 + k2^\ell - k^2$. Khraphchenko 1971 had previously shown a n^2 lower bound on the formula size of parity—our result shows that when n is not a power of two parity requires larger formulas, and in fact that $\lim_{n\to\infty} \sup$ of the formula size of parity is $(9/8)n^2$.

To obtain this result, we introduce a new technique for proving formula size lower bounds based on matrix rank. This result cannot be proven by any of the lower bound techniques of Khrapchenko, Nečiporuk, Koutsoupias, or the quantum adversary method, which are all limited by n^2.

1 Introduction

One of the most important open problems in complexity theory is to prove superlinear lower bounds on the circuit size of an explicit Boolean function. While this seems quite difficult, a modest amount of success has been achieved in the weaker model of formula size, a formula being a circuit where every gate has fan-out exactly one. The current best lower bound on the formula size of an explicit function is $n^{3-o(1)}$ [Hås98].

The difference between circuit and formula size is well illustrated by the parity function. Parity can be computed by a linear size circuit, but when creating a formula for parity in the obvious way, the restriction that each gate has fan-out one results in a quadratic blowup in size. Khrapchenko was the first to show that this blowup is necessary, showing a n^2 lower bound for parity [Khr71].

Let us see this in more detail. To compute the parity of a single bit, we can make do with a formula of size one, where we measure the size of a formula by the number of literals: $\phi(x_1) = x_1$. To compute the parity of two bits, that is an XOR gate, we need a formula of size four: $\phi(x_1, x_2) = (x_1 \wedge \neg x_2) \vee (\neg x_1 \wedge x_2)$. We can then inductively construct a formula for parity on n bits in a straightforward way

$$\phi(x_1, \ldots, x_n) = (\phi(x_1, \ldots, x_{\lfloor n/2 \rfloor}) \wedge \neg\phi(x_{\lfloor n/2 \rfloor + 1}, \ldots, x_n)) \vee (\neg\phi(x_1, \ldots, x_{\lfloor n/2 \rfloor}) \wedge \phi(x_{\lfloor n/2 \rfloor + 1}, \ldots, x_n)).$$

Letting $\oplus_n$ to be the parity function on n bits, and $L(f)$ the formula size of a function f, we see that $L(\oplus_n) \leq 2(L(\oplus_{\lfloor n/2 \rfloor}) + L(\oplus_{\lceil n/2 \rceil}))$.

Thus an upper bound on the formula size of parity is given by the solution to the recurrence relation $a(1) = 1, a(2) = 4, a(n) = 2(a(\lfloor n/2 \rfloor) + a(\lceil n/2 \rceil))$. This turns out to have solution $a(n) = 2^\ell(2^\ell + 3k) = n^2 + k2^\ell - k^2$, for $n = 2^\ell + k$, where $0 \leq k < 2^\ell$. Our main result gives a matching lower bound:

*Department of Computer Science, Rutgers University. Work conducted while at LRI, Université Paris-Sud and CWI, Amsterdam. This work was supported by a Rubicon grant from the Netherlands Organisation for Scientific Research (NWO). An earlier version of this paper appeared in STACS 2007 under the title "A new rank technique for formula size lower bounds"

Theorem 1 *If $n = 2^\ell + k$ where $0 \leq k < 2^\ell$, then*

$$L(\oplus_n) = 2^\ell(2^\ell + 3k) = n^2 + k2^\ell - k^2.$$

While Khrapchenko's bound of n^2 is tight when n is a power of two, this result shows that parity sometimes requires larger formulas and, in fact, $\lim_{n\to\infty} \sup L(\oplus_n) = (9/8)n^2$.

After we obtained our result it was pointed out to us that Rychkov [Ryc94] shows a lower bound of $n^2 + 3$ for odd $n \geq 5$, and $n^2 + 2$ for even $n \geq 6$ which are not powers of 2.

To prove this theorem we introduce a new formula size lower bound technique based on matrix rank. We use the setting of Karchmer and Wigderson [KW88] who characterize formula size as a communication complexity game, specifically as the communication complexity of a relation. Although matrix rank is one of the best tools available for proving lower bounds on the communication complexity of *functions* it has proved difficult to adapt to the relational case. We do this by means of a "selection function" which restricts a relation into a (non-Boolean) function, for which rank can be applied in a similar way to the usual case of Boolean functions. Razborov [Raz90] has previously used matrix rank in a different way to show superpolynomial lower bounds on *monotone* formula size, but also shows [Raz92] that his method is limited to $O(n)$ bounds for general formulas.

Most of the known generic techniques for proving formula size lower bounds cannot prove lower bounds larger than n^2. The technique of Nečiporuk [Neč66] is limited to bounds of size $n^2/\log n$—we should mention, however, that this technique works in the more general setting where any set of binary gates is allowed; the methods of Khrapchenko [Khr71], its generalization by Koutsoupias [Kou93], and further generalization by the quantum adversary method [LLS06], all cannot prove lower bounds larger than n^2; Karchmer, Kushilevitz, and Nisan [KKN95] introduce a promising technique based on linear programming but at the same stroke show that it cannot prove lower bounds larger than $4n^2$.

As our rank method can surpass this n^2 limitation, we hope that it can make progress on some of the many open questions remaining about the formula size of basic functions. One of the most dramatic such questions is the gap in our knowledge about the formula size of the majority function: the best lower bound is $\lceil n/2 \rceil^2$, provable by Khrapchenko's method, while the best upper bound is $O(n^{4.57})$ [PPZ92]. Even in the monotone case, where a formula consists of only AND and OR gates, the best lower bound is $\lfloor n/2 \rfloor n$ [Rad97], while the best upper bound is $O(n^{5.3})$ by Valiant's beautiful construction [Val84].

2 Preliminaries

We will make use of Jensen's inequality. We will use the following form:

Lemma 2 (Jensen's Inequality) *Let $\phi : \mathbb{R} \to \mathbb{R}$ be a convex function and a_i a set of positive real numbers for $i = 1, \ldots, n$. Then*

$$\phi\left(\frac{\sum_{i=1}^n a_i x_i}{\sum_{i=1}^n a_i}\right) \leq \frac{\sum_{i=1}^n a_i \phi(x_i)}{\sum_{i=1}^n a_i}.$$

2.1 Linear algebra

We will use some basic concepts from linear algebra. For a matrix A, let A^* be the transpose conjugate of A, that is $A^*[i,j] = \overline{A[j,i]}$. A matrix is Hermitian if $A = A^*$. We will use $\leq$ to refer to entrywise comparision

of matrices: that is $A \leq B$ if $A[i,j] \leq B[i,j]$ for all (i,j). The shorthand $A \geq 0$ means that all entries of A are nonnegative. The rank of A, denoted by $\mathrm{rk}(A)$, is the number of linearly independent columns of A. The trace of A, written $\mathrm{Tr}(A)$, is the sum of the diagonal entries of A. For a Hermitian n-by-n matrix A, let $\lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_n(A)$ be the eigenvalues of A. Let $\sigma_i(A) = \sqrt{\lambda_i(A^*A)}$ be the i^{th} singular value of A.

We will make use of three matrix norms. The Frobenius norm is the ℓ_2 norm of a matrix thought of as a long vector—that is

$$\|A\|_F = \sqrt{\sum_{i,j} A[i,j]^2}.$$

Notice also that $\|A\|_F^2 = \mathrm{Tr}(A^*A) = \sum_i \sigma_i^2(A)$. We will also use the trace norm, $\|A\|_{tr} = \sum_i \sigma_i(A)$. Finally, the spectral norm $\|A\| = \sigma_1(A)$. A very useful relationship between Frobenius norm, trace norm, and rank is the following:

Lemma 3 *Let A be a n-by-m matrix with $n \leq m$.*

$$\left\lceil \frac{\|A\|_{tr}^2}{\|A\|_F^2} \right\rceil \leq \mathrm{rk}(A).$$

Proof. The rank of A equals the number of nonzero singular values of A. Thus by the Cauchy–Schwarz inequality,

$$\left(\sum_{i=1}^{n} \sigma_i \right)^2 \leq \mathrm{rk}(A) \cdot \sum_{i=1}^{n} \sigma_i^2.$$

As rank is an integer, we obtain

$$\left\lceil \frac{\|A\|_{tr}^2}{\|A\|_F^2} \right\rceil \leq \mathrm{rk}(A).$$

□

A useful tool to lower bound the trace norm is the following:

Lemma 4

$$\|A\|_{tr} = \max_B \frac{|\mathrm{Tr}(A^*B)|}{\|B\|}.$$

For Theorem 1 we in fact need only the following simple bound on the trace norm: if there are k distinct rows $x_1, \ldots, x_k$ and k distinct columns $y_1, \ldots, y_k$ such that $A[x_i, y_i] = 1$ for all $1 \leq i \leq k$, then $\|A\|_{tr} \geq k$.

2.2 Formula size and communication complexity

A formula is a binary tree with nodes labeled by AND and OR gates, and leaves labeled by literals, that is either a variable or its negation. The size of a formula is its number of leaves. The formula size of a Boolean function f, written $L(f)$, is the size of a smallest formula which computes f.

Karchmer and Wigderson [KW88] characterize formula size in terms of a communication game. Since this characterization, nearly all formula size lower bounds have been phrased in the language of communication complexity.

Let X, Y, Z be finite sets and $R \subseteq X \times Y \times Z$ a relation. In the communication problem for R, Alice is given some $x \in X$, Bob some $y \in Y$, and they wish to output some $z \in Z$ such that $(x, y, z) \in R$. A communication protocol is a binary tree with each internal node v labeled either by a function $a_v : X \to \{0, 1\}$ if Alice speaks at this node, or by a function $b_v : Y \to \{0, 1\}$ if Bob speaks. Each leaf is labeled by an element $z \in Z$. We say that a protocol P computes a relation R if for every $(x, y) \in X \times Y$, walking down the tree according to the functions a_v, b_v leads to a leaf labeled with z such that $(x, y, z) \in R$. We let $C^P(R)$ denote the number of leaves in a smallest protocol which computes R.

For a Boolean function $f : \{0, 1\}^n \to \{0, 1\}$, let $X = f^{-1}(0)$ and $Y = f^{-1}(1)$. We associate with f a relation $R_f \subseteq X \times Y \times [n]$, where $R_f = \{(x, y, i) : x \in X, y \in Y, x_i \neq y_i\}$.

Theorem 5 (Karchmer–Wigderson) $L(f) = C^P(R_f)$

An important notion in communication complexity is that of a combinatorial rectangle. A combinatorial rectangle of $X \times Y$ is a set which can be expressed as $X' \times Y'$ for some $X' \subseteq X$ and $Y' \subseteq Y$. A set $S \subseteq X \times Y$ is called monochromatic for the relation R if there is some $z \in Z$ such that $(x, y, z) \in R$ for all $(x, y) \in S$. Let $C^D(R)$ be the number of rectangles in a smallest partition of $X \times Y$ into combinatorial rectangles monochromatic for R. We will often refer to this informally as the rectangle bound. A basic fact, which can be found in [KN97], is that $C^D(R) \leq C^P(R)$. The rectangle bound is also somewhat tight—Karchmer, Kushilevitz, and Nisan [KKN95] show that $C^P(R) \leq C^D(R)^{\log C^D(R)}$.

3 Rank technique

One of the best techniques for showing lower bounds on the communication complexity of a function $f : X \times Y \to \{0, 1\}$ is matrix rank, originally used by Melhorn and Schmidt [MS82]. If M_f is a matrix with rows labeled from X, columns labeled from Y and where $M_f[x, y] = f(x, y)$, then $\mathrm{rk}(M_f)$ lower bounds the number of leaves in a communication protocol for f. This follows as rank is subadditive and a communication protocol partitions the communication matrix into monochromatic rectangles, which are rank one matrices.

Let X, Y, Z be finite sets and $R \subseteq X \times Y \times Z$ a relation. In order to apply the rank bound, we first restrict the relation to a (non-Boolean) function by means of what we call a selection function. A selection function $S : X \times Y \to Z$ for the relation R takes input (x, y) and outputs some z such that $(x, y, z) \in R$. That is, it simply selects one of the possible valid outputs of the relation on input (x, y). We let $R|_S = \{(x, y, z) : S(x, y) = z\}$.

Theorem 6 $C^P(R) = \min_S C^P(R|_S)$.

Proof. For any selection function S, we have $C^P(R) \leq C^P(R|_S)$, as a protocol for $R|_S$ is in particular a protocol for R.

To see $C^P(R) \geq \min_S C^P(R_S)$, let P be an optimal protocol for R. We define a selection function based on this protocol, that is, let $S(x, y) = z$ if and only if (x, y) lead to a leaf labeled z by P. Now the protocol P also solves $R|_S$ and the claim follows. □

With the help of selection functions, we can now use rank as in the functional case.

Theorem 7 *Let $R \subseteq X \times Y \times Z$ be a relation. To a selection function S, we associate a set of matrices $\{S_z\}$ over $X \times Y$ where $S_z[x,y] = 1$ if $S(x,y) = z$ and $S_z[x,y] = 0$ otherwise. Then*

$$C^D(R) \geq \min_S \sum_{z \in Z} \mathrm{rk}(S_z).$$

Proof. Let $\mathcal{R}$ be an optimal rectangle partition of R satisfying $|\mathcal{R}| = \mathcal{C}^{\mathcal{D}}(\mathcal{R})$. We let $\mathcal{R}$ define a selection function in the natural way, setting $S(x,y) = z$ where z is the lexicographically least color of the rectangle in $\mathcal{R}$ which contains (x,y).

We now show for this particular choice

$$C^D(R) \geq \sum_{z \in Z} \mathrm{rk}(S_z),$$

which gives the theorem. Clearly $C^D(R)$ is equal to the sum over all z of the number of rectangles labeled z by the partition $\mathcal{R}$. Thus it suffices to show that $\mathrm{rk}(S_z)$ lower bounds the number of rectangles labeled by z. Consider some z and say that there are k monochromatic rectangles $B_1, \ldots, B_k$ labeled z. As each B_i is a combinatorial rectangle we can write it as $B_i = V_i \times W_i$ for $V_i \subseteq X$ and $W_i \subseteq Y$. Let v_i be the characteristic vector of V_i, that is $v_i[x] = 1$ if $x \in V_i$ and $v_i[x] = 0$ otherwise, and similarly for w_i with W_i. Then we can express S_z as $S_z = \sum_{i=1}^k v_i w_i^*$ and so $\mathrm{rk}(S_z) \leq k$. □

In general, this bound seems quite difficult to apply because of the minimization over all selection functions. We will now look at a simplified form of this method where we get around this difficulty by using Lemma 3 to lower bound the rank.

Corollary 8 *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function, and let $X = f^{-1}(0), Y = f^{-1}(1)$. Let c_i be the number of pairs $(x,y) \in X \times Y$ which differ only in position i, and let $s_1, \ldots, s_n$ be n nonnegative integers which sum to $|X||Y|$. Then*

$$C^D(R_f) \geq \min_{\substack{s_i \\ \sum_i s_i = |X||Y|}} \sum_i \left\lceil \frac{c_i^2}{s_i} \right\rceil.$$

Proof. By Theorem 7 and Lemma 3

$$C^D(R_f) \geq \min_S \sum_i \mathrm{rk}(S_i) \geq \min_S \sum_i \left\lceil \frac{\|S_i\|_{tr}^2}{\|S_i\|_F^2} \right\rceil. \tag{1}$$

Let $(x_1, y_1), \ldots, (x_{c_i}, y_{c_i})$ be the c_i many pairs which differ only in position i. For these pairs, any selection function S must choose i. As for every x, the string y differing from x only in position i is unique, this gives us c_i many distinct rows $x_1, \ldots, x_{c_i}$ and c_i many distinct columns $y_1, \ldots, y_{c_i}$ for which $S_i[x_k, y_k] = 1$. Thus by the comment following Lemma 4 we have $\|S_i\|_{tr} \geq c_i$. The Frobenius norm squared of a zero/one matrix is simply the number of ones, thus $\|S_i\|_F^2$ is simply the number of (x,y) pairs for which the selection function S chooses i. As the selection function is total, $\sum_i \|S_i\|_F^2 = |X||Y|$. The claim follows. □

The simplified version of the rank method given in Corollary 8 is already strong enough to imply Khrapchenko's method. Khrapchenko's method works as follows: let f be a Boolean function, and as before let $X = f^{-1}(0), Y = f^{-1}(1)$. Let C be the set of $(x,y) \in X \times Y$ which have Hamming distance one. Khrapchenko's bound is then $|C|^2/|X||Y|$.

Theorem 9 *The bound given in Corollary 8 is at least as large as that of Khrapchenko.*

Proof. Let c_i be the number of $(x, y) \in X \times Y$ which differ only in position i, and let $\{s_i\}$ be such that $\sum_i s_i = |X||Y|$ and which minimize the bound given in Corollary 8. We now apply Jensen's inequality, Lemma 2, with $\phi(x) = 1/x, x_i = s_i/c_i$, and $a_i = c_i$ to obtain

$$\sum_i \frac{c_i^2}{s_i} \geq \frac{(\sum_i c_i)^2}{\sum_i s_i} = \frac{|C|^2}{|X||Y|}.$$

□

4 Application to parity

In this section, we apply the rank technique to the parity function and thereby obtain Theorem 1. We first show the upper bound.

Proposition 10 *Let $n = 2^\ell + k$, where $0 \leq k < 2^\ell$. Then $L(\oplus_n) \leq 2^\ell(2^\ell + 3k)$.*

Proof. We construct our formula inductively. For $n = 1$, we have a formula of size one, $\phi(x_1) = x_1$. For $n = 2$, we have a formula of size four, $\phi(x_1, x_2) = (x_1 \wedge \neg x_2) \vee (\neg x_1 \wedge x_2)$. We can then inductively construct a formula for parity on n bits as

$$\phi(x_1, \ldots, x_n) = (\phi(x_1, \ldots, x_{\lfloor n/2 \rfloor}) \wedge \neg\phi(x_{\lfloor n/2 \rfloor + 1}, \ldots, x_n)) \vee (\neg\phi(x_1, \ldots, x_{\lfloor n/2 \rfloor}) \wedge \phi(x_{\lfloor n/2 \rfloor + 1}, \ldots, x_n)).$$

This construction leads us to consider the recurrence relation $a(1) = 1, a(2) = 4, a(n) = 2(a(\lfloor n/2 \rfloor) + a(\lceil n/2 \rceil))$, which has solution $a(n) = 2^\ell(2^\ell + 3k)$, where $n = 2^\ell + k$. This can be easily verified as it satisfies the initial conditions and

$$\begin{aligned} 2(a(\lfloor n/2 \rfloor) + a(\lceil n/2 \rceil) &= 2\left(2^{\ell-1}(2^{\ell-1} + 3\lfloor k/2 \rfloor) + 2^{\ell-1}(2^{\ell-1} + 3\lceil k/2 \rceil)\right) \\ &= 2^\ell\left(2^\ell + 3(\lfloor k/2 \rfloor + \lceil k/2 \rceil)\right) \\ &= 2^\ell\left(2^\ell + 3k\right) \end{aligned}$$

□

We now turn to the proof of the lower bound.

Proposition 11 *Let $n = 2^\ell + k$, where $0 \leq k < 2^\ell$. Then $L(\oplus_n) \geq 2^\ell(2^\ell + 3k)$.*

Proof. Let S be any selection function. For every i, there are 2^{n-1} entries of the matrix S_i which *must* be one, namely the entries x, y which differ only on position i. If S only assigns these entries to have the label i, then S_i is a permutation matrix and so has rank 2^{n-1}. Thus to reduce the rank of S_i, the selection function S must therefore assign more (x, y) pairs to also have the label i. The catch is that S must do this for all i simultaneously, and we will bound how well it can do this.

Notice that for parity on n-bits, $|X| = |Y| = 2^{n-1}$. For every i there are 2^{n-1} pairs (x, y) which differ only in position i. Thus applying Corollary 8 with $c_i = 2^{n-1}$ for all i, we obtain

$$C^D(R) \geq \min_{s_i : \sum_i s_i = 2^{2n-2}} \sum_{i=1}^{n} \left\lceil \frac{2^{2n-2}}{s_i} \right\rceil . \tag{2}$$

If we were to ignore the ceilings, then we are minimizing over a convex function $\phi(x) = 1/x$ and so Jensen's inequality gives that the minimum is obtained when all s_i are equal. In this case $s_i = 2^{2n-2}/n$ and so $\sum_i 2^{2n-2}/s_i = n^2$.

To get bound larger than n^2 we need to take the ceiling functions into account. If n is not a power of two, then $2^{2n-2}/n$ will not be an integer, whereas each s_i is an integer—this means that it is no longer possible to have all s_i values equal and $\sum_i s_i = 2^{2n-2}$. It is this imbalance that will lead to a larger lower bound.

We transform Equation (2) in a series of steps. First, notice that

$$\min_{s_i : \sum_i s_i = 2^{2n-2}} \sum_{i=1}^{n} \left\lceil \frac{2^{2n-2}}{s_i} \right\rceil = \min_{s'_i : \sum_i s'_i \leq 2^{2n-2}} \sum_{i=1}^{n} \left\lceil \frac{2^{2n-2}}{s'_i} \right\rceil . \tag{3}$$

The right hand side is clearly less than the left hand side as the minimization is taken over a larger set. The left hand side is less than the right hand side as given a solution $\{s'_i\}$ to the right hand side, we can obtain a solution to the left hand side which is not larger by setting $s_i = s'_i$ for $i = 1, \ldots, n-1$, and $s_n = 2^{2n-2} - \sum_{i=1}^{n-1} s'_i \geq s'_n$.

Now we observe that there is an optimal solution $\{s_i\}$ to Equation (3) where each $2^{2n-2}/s_i$ is an integer, and so each s_i is a power of two. If $2^{2n-2}/s_i$ is not an integer, then we can set s'_i to the largest power of two less than s_i and $\lceil 2^{2n-2}/s_i \rceil = 2^{2n-2}/s'_i$, and the sum of s'_i does not increase.

Thus assume that each s_i is a power of two, say $s_i = 2^{a_i}$. We can now rewrite Equation (3) as

$$\min_{\substack{a_i \\ \sum_i 2^{a_i} \leq 2^{2n-2}}} \sum_i 2^{2n-2-a_i}$$

As the logarithm is a monotone function, the values $\{a_i\}$ which achieve this minimum will maximize

$$\max_{\substack{a_i \\ \sum_i 2^{a_i} \leq 2^{2n-2}}} \sum_i a_i .$$

We now show that there is an optimal solution to this maximization problem where $|a_i - a_j| \leq 1$ for all i, j. If $a_i - a_j > 2$ then we can let $a'_i = a_i - 1$ and $a'_j = a_j + 2$, so that $a'_i + a'_j > a_i + a_j$ and $2^{a'_j} \leq 2^{a_j} + 2^{a_i - 1}$ so $2^{a'_i} + 2^{a'_j} \leq 2^{a_i} + 2^{a_j}$. If $a_i - a_j = 2$ then by setting $a'_i = a_i - 1$ and $a'_j = a_j + 1$ then we still have $a'_i + a'_j = a_i + a_j$, and have saved on weight, $2^{a'_i} + 2^{a'_j} < 2^{a_i} + 2^{a_j}$.

By performing these transformations, we can turn any solution into one where $|a_i - a_j| \leq 1$ and whose value is at least as good. Now if we have $|a_i - a_j| \leq 1$ and $\sum_i 2^{a_i} = 2^{2n-2}$, it follows that $a_i = 2n - \ell - 2$ for $2^\ell - k$ many values of i and $a_i = 2n - \ell - 3$ for $2k$ many values of i. This gives

$$\begin{aligned} \min_{a_i} \sum_{i=1}^{n} 2^{2n-2-a_i} &= (2^\ell - k)2^\ell + 2k2^{\ell+1} \\ &= 2^\ell(2^\ell + 3k). \end{aligned}$$

□

We take this opportunity to make some remarks on the expression of the formula size of parity. The recurrence relation we consider has arisen before in complexity theory [AS03], and is sequence A073121 in Sloane's online encyclopedia of integer sequences [Slo].

The values of n for which parity is the "hardest" to compute are of the form $n = [(4/3)2^k]$, where $[x]$ denotes the nearest integer to x. The formula size of parity on these instances asymptotically approaches $(9/8)n^2$. This sequence is known as the Jacobsthal sequence (Sloane's A001045), and arises in surprisingly many contexts. We record these observations in the next Corollary.

Corollary 12

$$\lim_{n\to\infty} \sup \frac{L(\oplus_n)}{n^2} = \frac{9}{8}$$

Proof. Simple calculus shows that $f(x) = 2^\ell(2^\ell + 3x)/(2^\ell + x)^2$, a function of the real variable x, achieves its maximum in the interval $x \in [0, 2^k - 1]$ when $x = 2^\ell/3$. At this value, $f(x) = 9/8$. As the second derivative of f is negative for $x \in [0, 2^k - 1]$, the largest value of $f(x)$ over integers will be an integer adjacent to $2^\ell/3$, and in fact is $[2^\ell/3]$ the closest integer to $2^\ell/3$. It is straightforward to check that $f([2^\ell/3])$ approaches $9/8$ as $\ell \to \infty$. □

5 A more general technique

In this section, we highlight one way to generalize the simple bound given by Equation (1). While this bound works well for the parity function, it has the shortcoming that it cannot take advantage of the fact that certain inputs to a function might be harder than others. To give a concrete example, the bound given by Equation (1) on the function $f : \{0,1\}^{2n} \to \{0,1\}$ on $2n$ bits which is just the parity of the first n bits, ignoring the second n bits, is worse than the bound for parity on n bits. To remedy this, we let u be a unit vector of length $|X|$ and v be a unit vector of length $|Y|$ and consider the matrix $S_i \circ uv^*$ instead of the matrix S_i. As $\mathrm{rk}(S_i \circ uv^*) \leq \mathrm{rk}(S_i)$, we can again apply Theorem 7 and Lemma 3 to obtain

$$C^D(R_f) \geq \min_S \max_{\substack{u,v \\ \|u\|=\|v\|=1}} \sum_i \frac{\|S_i \circ uv^*\|_{tr}^2}{\|S_i \circ uv^*\|_F^2}. \tag{4}$$

Remark 13 *As* $\mathrm{rk}(AB) \leq \min\{\mathrm{rk}(A), \mathrm{rk}(B)\}$*, one may ask why we do not consider the stronger bound*

$$\max_{X,Y} \frac{\|XAY\|_{tr}^2}{\|XAY\|_F^2} \leq \mathrm{rk}(A).$$

The left hand side, however, is actually equal to the rank of A, thus this approach gives us the full strength of Theorem 7.

We now show that Equation (4) gives bounds at least as large as the formula size bounds given by the quantum adversary method [LLS06]. Laplante, Lee, and Szegedy have already shown that the quantum adversary method gives lower bounds at least as large as the method of Koutsoupias, which in turn is known to give lower bounds at least as large Khrapchenko's method.

Ambainis [Amb02, Amb03] developed the quantum adversary method to prove lower bounds on bounded-error quantum query complexity. Laplante, Lee, and Szegedy show that the square of the adversary bound is lower bound on formula size. The adversary bound can be phrased as a maximization problem of the spectral norm of a matrix associated with f [BSS03].

Definition 14 (Adversary bound) *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function, and $X = f^{-1}(0)$ and $Y = f^{-1}(1)$. Let Γ be a $|X|$-by-$|Y|$ matrix, and let Γ_i be the matrix such that $\Gamma_i[x,y] = \Gamma[x,y]$ if $x_i \neq y_i$ and $\Gamma_i[x,y] = 0$ otherwise, for $1 \leq i \leq n$. Then*

$$\mathrm{ADV}(f) = \max_{\substack{\Gamma \geq 0 \\ \Gamma \neq 0}} \frac{\|\Gamma\|}{\max_i \|\Gamma_i\|}.$$

Theorem 15 *The bound given by Equation (4) is at least as large as $\mathrm{ADV}(f)^2$.*

Proof. Applying Jensen's Inequality to Equation (4) with $\phi(x) = 1/x$, $x_i = \|S_i \circ uv^*\|_F^2 / \|S_i \circ uv^*\|_{tr}$, and $a_i = \|S_i \circ uv^*\|_{tr}$, we obtain:

$$\min_S \max_{\substack{u,v \\ \|u\|=\|v\|=1}} \sum_i \frac{\|S_i \circ uv^*\|_{tr}^2}{\|S_i \circ uv^*\|_F^2} \geq \min_S \max_{\substack{u,v \\ \|u\|=\|v\|=1}} \frac{\left(\sum_i \|S_i \circ uv^*\|_{tr}\right)^2}{\sum_i \|S_i \circ uv^*\|_F^2}.$$

As the selection function is total we have $\sum_i \|S_i \circ uv^*\|_F^2 = \|uv^*\|_F^2 = 1$.

Now we use Lemma 4 to lower bound $\|S_i\|_{tr}$. One can think of the weight matrix Γ in the adversary bound as the matrix from Lemma 4 which witnesses that the trace norm of the S_i's is large:

$$\begin{aligned}
\min_S \max_{\substack{u,v \\ \|u\|=\|v\|=1}} \left(\sum_i \|S_i \circ uv^*\|_{tr}\right)^2 &\geq \min_S \max_{\substack{\Gamma \geq 0 \\ \Gamma \neq 0}} \max_{\substack{u,v \\ \|u\|=\|v\|=1}} \left(\sum_i \frac{|\mathrm{Tr}((\Gamma \circ S_i)vu^*|}{\|\Gamma \circ S_i\|}\right)^2 \\
&\geq \min_S \max_{\substack{\Gamma \geq 0 \\ \Gamma \neq 0}} \max_{\substack{u,v \\ \|u\|=\|v\|=1}} \left(\sum_i \frac{|\mathrm{Tr}((\Gamma \circ S_i)vu^*|}{\|\Gamma_i\|}\right)^2.
\end{aligned}$$

This step follows as $0 \leq \Gamma \circ S_i \leq \Gamma_i$ and for matrices A, B if $0 \leq A \leq B$ then $\|A\| \leq \|B\|$. We can now continue

$$\begin{aligned}
\min_S \max_{\substack{\Gamma \geq 0 \\ \Gamma \neq 0}} \max_{\substack{u,v \\ \|u\|=\|v\|=1}} \left(\sum_i \frac{|\mathrm{Tr}((\Gamma \circ S_i)vu^*|}{\|\Gamma_i\|}\right)^2 &\geq \min_S \max_{\substack{\Gamma \geq 0 \\ \Gamma \neq 0}} \max_{\substack{u,v \\ \|u\|=\|v\|=1}} \left(\frac{\sum_i \mathrm{Tr}((\Gamma \circ S_i)vu^*)}{\max_i \|\Gamma_i\|}\right)^2 \\
&= \max_{\substack{\Gamma \geq 0 \\ \Gamma \neq 0}} \max_{\substack{u,v \\ \|u\|=\|v\|=1}} \left(\frac{\mathrm{Tr}(\Gamma vu^*)}{\max_i \|\Gamma_i\|}\right)^2 \\
&= \max_{\substack{\Gamma \geq 0 \\ \Gamma \neq 0}} \left(\frac{\|\Gamma\|}{\max_i \|\Gamma_i\|}\right)^2.
\end{aligned}$$

□

6 Hierarchy of techniques

In this section, we give one more result to clarify the hierarchy of available techniques for proving lower bounds on formula size. We show that the linear programming bound of Karchmer, Kushilevitz, and Nisan [KKN95] also gives bounds at least as large as the quantum adversary method.

We first introduce the linear programming bound. Karchmer, Kushilevitz, and Nisan notice that for a relation $R \subseteq X \times Y \times Z$ the rectangle bound $C^D(R)$ can be written as an integer program. Indeed, let

$\mathcal{R}$ be the set of all rectangles which are monochromatic with respect to the relation R. To represent the relationship between inputs (x, y) and the rectangles of $\mathcal{R}$ we use a $|X| \cdot |Y|$-by-$|\mathcal{R}|$ incidence matrix A, where for $(x, y) \in X \times Y$ and $S \in \mathcal{R}$ we let $A[(x, y), S] = 1$ if $(x, y) \in S$. Now a set of rectangles can be described by a $|\mathcal{R}|$-length vector α, with each entry $\alpha[S] \in \{0, 1\}$. If α represents a partition, then $A\alpha = \vec{1}$, and the number of rectangles in such a partition is simply $\sum_S \alpha[S]$. Karchmer, Kushilevitz, and Nisan relax this integer program to a linear program by replacing the condition $\alpha[S] \in \{0, 1\}$ with $0 \leq \alpha[S] \leq 1$.

Definition 16 (Linear programming bound [KKN95]) *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function, R_f the relation corresponding to f, and α a vector indexed by rectangles monochromatic with respect to R_f. The linear programming bound, denoted $\mathrm{LP}(f)$, is then*

$$\mathrm{LP}(f) = \min_{\substack{\alpha : A\alpha = \vec{1} \\ 0 \leq \alpha[S] \leq 1}} \sum_S \alpha[S].$$

Now we show that the bound given by the linear programming method is also always at least as large as that given by the adversary method.

Theorem 17 $\mathrm{LP}(f) \geq \mathrm{ADV}^2(f)$.

Proof. Let α be a solution to the linear program associated with f. By definition we have $\sum_{S:(x,y)\in S} \alpha[S] = 1$ for every (x, y). Let u, v be unit vectors such that $|u^* \Gamma v| = \|\Gamma\|$. We will need some notation to label submatrices of Γ and portions of u, v. For a combinatorial rectangle $S = U \times V$, let $\Gamma_S[x, y] = A[x, y]$ if $(x, y) \in S$ and $\Gamma[x, y] = 0$ otherwise. Similarly, let $u_S[x] = u[x]$ if $x \in U$ and $u_S[x] = 0$ otherwise, and similarly for v_S. Now

$$\begin{aligned}
\|\Gamma\| &= \sum_{x,y} \Gamma[x, y] u[x] v[y] \\
&= \sum_{x,y} \sum_{S:(x,y)\in S} \alpha[S] \Gamma[x, y] u[x] v[y] \\
&= \sum_S \alpha[S] \sum_{(x,y)\in S} \Gamma[x, y] u[x] v[y] \\
&\leq \sum_S \alpha[S] \|\Gamma_S\| \|u_S\| \|v_S\| \\
&\leq \left(\sum_S \alpha[S] \|\Gamma_S\|^2 \right)^{1/2} \left(\sum_S \alpha[S] \|u_S\|^2 \|v_S\|^2 \right)^{1/2},
\end{aligned}$$

where the first inequality follows from the definition of spectral norm, and the second uses the Cauchy–Schwarz inequality. Notice that

$$\sum_S \alpha[S] \|u_S\|^2 \|v_S\|^2 = \sum_{x,y} \alpha[S] |u[x]|^2 |v[y]|^2 = 1.$$

Thus

$$\|\Gamma\|^2 \leq \sum_S \alpha[S] \|\Gamma_S\|^2 \leq \max_S \|\Gamma_S\|^2 \sum_S \alpha[S],$$

and so

$$\sum_S \alpha[S] \geq \max_\Gamma \frac{\|\Gamma\|^2}{\max_S \|\Gamma_S\|^2} \geq \max_\Gamma \frac{\|\Gamma\|^2}{\max_i \|\Gamma_i\|^2}.$$

□

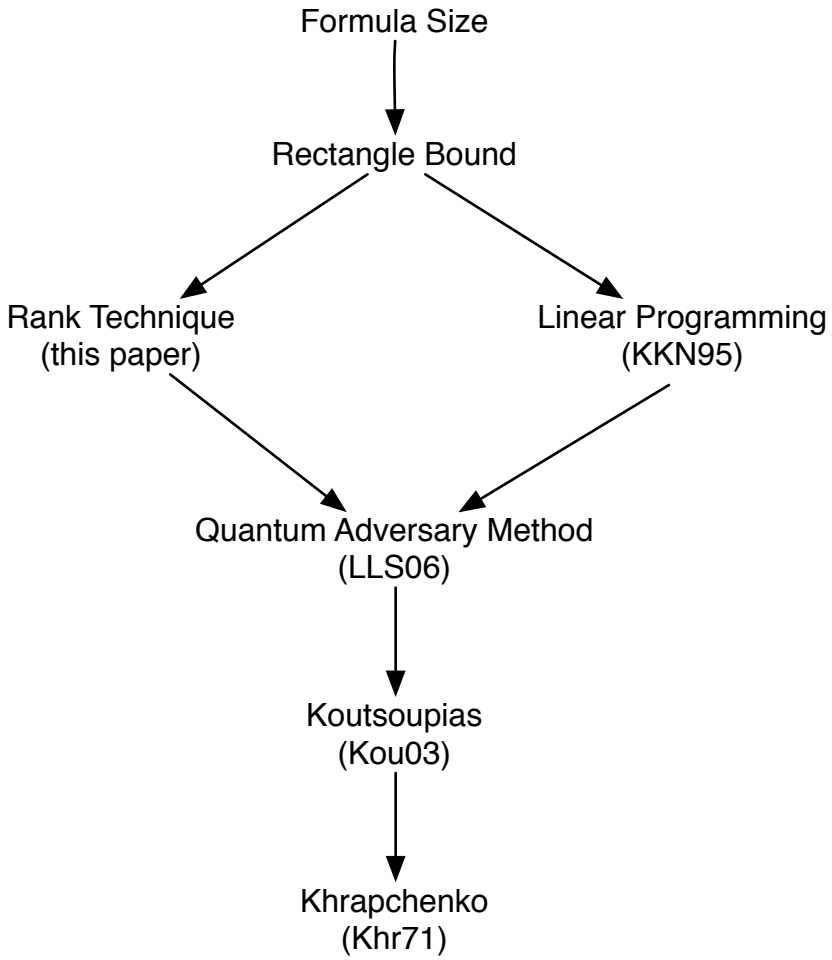

Figure 1: Hierarchy of formula size techniques. Arrows point from larger to smaller.

Acknowledgments

I would like to thank Anna Gál for helpful discussions on the topics of this paper, Dmitry Cherukhin for pointing out the work of Rychkov, Gabor Ivanyos for asking what "one may ask" in Remark 13, and the referees of STACS 2007 for many beneficial comments.

References

[AS03] J.-P. Allouche and J. Shallit. The ring of k-regular sequences, II. *Theoretical Computer Science*, 307:3–29, 2003.

[Amb02] A. Ambainis. Quantum lower bounds by quantum arguments. *Journal of Computer and System Sciences*, 64:750–767, 2002.

[Amb03] A. Ambainis. Polynomial degree vs. quantum query complexity. In *Proceedings of the 44th IEEE Symposium on Foundations of Computer Science*, pages 230–239. IEEE, 2003.

[BSS03] H. Barnum, M. Saks, and M. Szegedy. Quantum decision trees and semidefinite programming. In *Proceedings of the 18th IEEE Conference on Computational Complexity*, pages 179–193, 2003.

[Hås98] J. Håstad. The shrinkage exponent is 2. *SIAM Journal on Computing*, 27:48–64, 1998.

[Khr71] V.M. Khrapchenko. Complexity of the realization of a linear function in the case of Π-circuits. *Math. Notes Acad. Sciences*, 9:21–23, 1971.

[KKN95] M. Karchmer, E. Kushilevitz, and N. Nisan. Fractional covers and communication complexity. *SIAM Journal on Discrete Mathematics*, 8(1):76–92, 1995.

[KN97] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, 1997.

[Kou93] E. Koutsoupias. Improvements on Khrapchenko's theorem. *Theoretical Computer Science*, 116(2):399–403, 1993.

[KW88] M. Karchmer and A. Wigderson. Monotone connectivity circuits require super-logarithmic depth. In *Proceedings of the 20th ACM Symposium on the Theory of Computing*, pages 539–550, 1988.

[LLS06] S. Laplante, T. Lee, and M. Szegedy. The quantum adversary method and classical formula size lower bounds. *Computational Complexity*, 15:163–196, 2006.

[MS82] K. Melhorn and E. Schmidt. Las Vegas is better than determinism in VLSI and distributed computing. In *Proceedings of the 14th ACM Symposium on the Theory of Computing*, pages 330–337. ACM, 1982.

[Neč66] E. I. Nečiporuk. A Boolean function. *Soviet Mathematics–Doklady*, 7:999–1000, 1966.

[PPZ92] M. Paterson, N. Pippenger, and U. Zwick. Optimal carry save networks. In *Boolean function complexity*, pages 174–201. London Mathematical Society Lecture Note Series 169, Cambridge University Press, 1992.

[Rad97] J. Radhakrishnan. Better lower bounds for monotone threshold formulas. *Journal of Computer and System Sciences*, 54(2):221–226, 1997.

[Raz90] A. Razborov. Applications of matrix methods to the theory of lower bounds in computational complexity. *Combinatorica*, 10(1):81–93, 1990.

[Raz92] A. Razborov. On submodular complexity measures. In M. Paterson, editor, *Boolean function complexity*, volume 169 of *London Math. Soc. Lecture Notes Series*, pages 76–83. Cambridge University Press, 1992.

[Ryc94] K. Rychkov. On lower bounds of complexity of series-parallel contact networks which realize linear Boolean functions. *Siberian Journal of Operations Research*, 1(4): 33–52, 1994.

[Slo] N. Sloane. On-line encyclopedia of integer sequences. `http://www.research.att.com/~njas/sequences/`.

[Val84] L.G. Valiant. Short monotone formulae for the majority function. *Journal of Algorithms*, 5:363–366, 1984.