CSCI 1590
Intro to Computational Complexity
Formula Size

John E. Savage

Brown University

March 5, 2008

Summary

1. Review
2. Application of Neciporuk's Lower Bound
3. Krapchenko's Formula Size Lower Bound

Review of Results Concerning Circuit Size

Lemma

For $n \geq 3$, $Q_{2,3}^{(n)}$ *contains* $f_{\text{mod } 3,c}^{(n)}$ *where*

$$f_{\text{mod } 3,c}^{(n)}(x_1, x_2, \ldots, x_n) = ((y + c) \text{ mod } 3) \text{ mod } 2)$$

for $c \in \{0, 1, 2\}$ *where* $y = x_1 + x_2 + \ldots + x_n$.

Theorem

Over the basis of all Boolean functions on 2 inputs, $f \in Q_{2,3}^{(n)}$ *for* $n \geq 3$ *has* $C(f) \geq 2n - 3$.

Review of Results Concerning Formula Size

Definition

Given $f : \mathcal{B}^n \mapsto \mathcal{B}$, partition its n variables $\mathbf{X}$ into p disjoint sets $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_p$.

Let $r_j(f)$ be the number of different subfunctions of f over $\mathbf{X}_j$ when the variables in $\mathbf{X} - \mathbf{X}_j$ range over all values.

Theorem

Let Ω be a complete basis of fan-in d and let $c_\Omega = 1/(d+2)$. Then, for every $f : \mathcal{B}^n \mapsto \mathcal{B}$ its formula size satisfies

$$L_\Omega(f) \geq c_\Omega \sum_{i=1}^{p} \log_2 r_i(f)$$

Neciporuk's Lower Bound to Formula Size

Indirect Storage Access Function

$f_{ISA}(k,l) : \mathcal{B}^n \mapsto \mathcal{B}$ has $n = k + l2^k + 2^l$, $\mathbf{a} \in B^k$, $\mathbf{x}_j \in B^l$, $\mathbf{y} \in B^{2^l}$.

$$f_{ISA}^{(k,l)}(\mathbf{a}, \mathbf{x}_{2^k-1}, \ldots, \mathbf{x}_0, \mathbf{y}) = y_{|\mathbf{x}_{|\mathbf{a}|}|}$$

$$L_\Omega(f_{ISA}^{(k,l)}) = O(\frac{n^2}{\log n})$$

where $n = k + l2^k + 2^l$. (See Problem 9.24 of Savage's book.)

Lemma

$$L_\Omega(f_{ISA}^{(k,l)}) = \Omega(\frac{n^2}{\log n})$$

Lower Bounds to Formula Size

Proof.

Let $K = 2^k$, $L = 2^l$. Create K sets $\mathbf{X}_1, \ldots, \mathbf{X}_K$. Let $\mathbf{X}_j$ be the variables in $\mathbf{x}_j$ and possibly other variables that are fixed. (This cannot increase $r_j(f)$.)

$r_j(f) \geq 2^{2^l}$ because when $|\mathbf{a}| = j$ each of the 2^{2^l} assignments of $\mathbf{y}$ defines a different function of the variables in $\mathbf{X}_j$. It follows that $L_\Omega(f_{ISA}^{(k,l)}) \geq c_\Omega \sum_{i=1}^{p} \log_2 r_i(f) \geq c_\Omega KL$.

Let $k = \lceil \log_2 \frac{L}{l} \rceil$. Then, $L/l \leq K = 2^k < 2L/l$. Since, $l = \log_2 L$, $n = \log_2 K + K \log_2 L + L$ satisfies the following bounds.

$$2L \leq n \leq \lceil \log_2 \frac{L}{l} \rceil + 3L \leq 4L$$

Thus, $n = \Theta(L)$ and $L_\Omega(f_{ISA}^{(k,l)}) \geq c_\Omega L^2 / \log_2 L = \Omega(n^2 / \log_2 n)$, the desired result. ☐

Krapchenko's Formula Size Lower Bound

- Krapchenko's lower bound to formula size applies to the bases:
  - a) standard, b) $\{\text{AND}, \text{NOT}\}$, and c) $\{\text{OR}, \text{NOT}\}$.
- It provides slightly stronger results than Neciporuk's.
- We use Krapchenko's method to show that parity on n inputs satisfies $L(f_{\oplus}^{(n)}) \geq \Omega(n^2)$.

Krapchenko's Formula Size Lower Bound

We now give an upper bound on formula size for $f_{\oplus}^{(n)}$. $L(f_{\oplus}^{(2)}) \leq 4$ because the following works on two inputs:

$$f_{\oplus}^{(2)}(x_1, x_2) = (x_1 \wedge \bar{x}_2) \vee (\bar{x}_1 \wedge x_2)$$

We compute $f_{\oplus}^{(n)}$ on n variables by separately computing it on each half of the variables and then combining the results with the above formula.

If $n = 2^k$, let $P(k) = L(f_{\oplus}^{(n)})$. It follows from above formula that $P(k) = 4 * P(k-1) = 4^k = n^2$. The upper bound will agree with the lower bound for this function.

Krapchenko's Formula Size Lower Bound

Definition

Let A and B be disjoint sets of B^n. The **neighborhood** of A and B, $N(A, B)$, is the set of pairs (x, y) of n-tuples, $x \in A$ and $y \in B$, that differ in exactly one position.

For $f : \mathcal{B}^n \mapsto \mathcal{B}$ let $f^{-1}(0)$ and $f^{-1}(1)$ be the sets of n-tuples that cause f to assume values 0 and 1, respectively.

Theorem

For $f : \mathcal{B}^n \mapsto \mathcal{B}$ and any $A \subseteq f^{-1}(0)$ and any $B \subseteq f^{-1}(1)$ the following holds over the standard basis:

$$L(f) \geq \frac{|N(A, B)|^2}{|A||B|}$$

Krapchenko's Formula Size Lower Bound

Proof

For 2-input bases the number of leaves in the optimal formula is one more than the number of two-input gates. We assume w.l.o.g. that the gates used consist of AND and NOT.

The proof is by induction. Base case: $n = 1$. Here f is either a constant ($|N(A,B)| = 0$ and $L(f) = 0$) or it is x or $\bar{x}$. If $f(x) = x$, $f^{-1}(0) = 0$ and $f^{-1}(1) = 1$, and with $A = \{0\}$, $B = \{1\}$ we have $|N(A,B)| = 1$. A similar argument holds if $f(x) = \bar{x}$. Since $|A| = |B| = 1$, the base case holds.

We assume that $L(f) \geq |N(A,B)|^2/|A||B|$ holds for f such that $L(f) \leq L_0 - 1$ and show that it holds for $L(f) = L_0 \geq 2$.

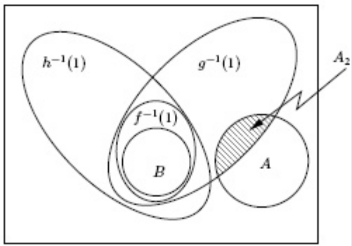
W.l.o.g assume that the output gate is AND and computes $f = g \wedge h$. We assume the formula for f is optimal. This must also be true for g and h.

Krapchenko's Formula Size Lower Bound

Proof (cont.)

Let $A \subseteq f^{-1}(0)$ and $B \subseteq f^{-1}(1)$. That is, $f(x) = 0$ for $x \in A$ and $f(x) = 1$ for $x \in B$.

It follows that $g(x) = h(x) = 1$ when $x \in B$. That is, $f^{-1}(1) \subseteq g^{-1}(1)$ and $f^{-1}(1) \subseteq h^{-1}(1)$. Thus, $B \subseteq g^{-1}(1)$ and $B \subseteq h^{-1}(1)$. Let $B_1 = B_2 = B$. Let $A_1 = A \cap g^{-1}(0)$ and $A_2 = A - A_1$.



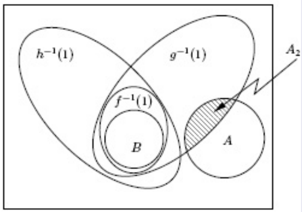
Krapchenko's Formula Size Lower Bound

Proof (cont.)

If $x \in A$, $f(x) = 0$. If x is also in A_2, $g(x) = 1$. This implies $h(x) = 0$ or $A_2 \subseteq h^{-1}(0)$ which implies $N(A_1, B_1)$ and $N(A_2, B_2)$ are disjoint (recall that $B_1 = B_2 = B$) and $N(A, B) = N(A_1, B_1) + N(A_2, B_2)$.
Because g and h are independent, by induction

$$L(g) = |N(A_1, B_1)|^2 / |A_1||B_1|$$

$$L(h) = |N(A_2, B_2)|^2 / |A_2||B_2|$$



Krapchenko's Formula Size Lower Bound

Proof (cont.)

Also, $L(f) = L(g) + L(h)$. In the above we have $|B|$ as common denominator.

The remaining expression is of the form

$$S = (n_1^2/c_1) + (n_2^2/c_2)$$

where $n_1 + n_2 = N(A, B)$, $c_1 = |A_1|$, and $c_2 = |A_2|$. It is straightforward to show that

$$S \geq (n_1 + n_2)^2/(c_1 + c_2)$$

from which the result follows.