ON THE FREQUENCY OF THE MOST FREQUENTLY
OCcurring VARIABLE IN DUAL MONOTONE DNFs

V. Gurvich† and L. Khachiyan

LCSR-TR-252

Department of Computer Science
Hill Center for the Mathematical Sciences
Rutgers University
New Brunswick, NJ 08903.

* Supported in part by ONR grant N00014-92-J-1375.
† RUTCOR, Rutgers University, New Brunswick, NJ 08903. On leave from the
International Institute of Earthquake Prediction Theory and Mathematical
Geophysics, Moscow.
ON THE FREQUENCY OF THE MOST FREQUENTLY
OCcurring VARIABLE IN DUAL MONOTONE DNFs

Vladimir Gurvich and Leonid Khachiyan

Abstract. Let $f(x_1, \ldots, x_n) = \bigvee_{i \in E} \bigwedge_{i \in I} x_i$ and $g(x_1, \ldots, x_n) = \bigvee_{i \in G} \bigwedge_{i \in I} x_i$ be a pair of dual monotone irredundant disjunctive normal forms, where $E$ and $G$ are the sets of the prime implicants of $f$ and $g$, respectively. For a variable $x_i$, $i = 1, \ldots, n$, let $\mu_i = \#\{I \in E | i \in I\}/|E|$ and $\nu_i = \#\{I \in G | i \in I\}/|G|$ be the frequencies with which $x_i$ occurs in $f$ and $g$. It is easily seen that $\max\{\mu_1, \nu_1, \ldots, \mu_n, \nu_n\} \geq 1/\log(|E| + |G|)$. We give examples of arbitrarily large $E$ and $G$ for which the above bound is tight up to a factor of 2.

Key words: monotone Boolean function, disjunctive normal form, prime implicant, duality, short implicat, frequent variable, transversal hypergraph, clutter, blocker, quasi-polynomial time.

*Supported in part by ONR grant N00014-92-J-1375.
† RUTCOR, Rutgers University, New Brunswick, NJ 08903. On leave from the International Institute of Earthquake Prediction Theory and Mathematical Geophysics, Moscow.
‡ Department of Computer Science, Rutgers University, New Brunswick, NJ 08903.
1. Introduction

Let \( f = f(x_1, \ldots, x_n) \) and \( g = g(x_1, \ldots, x_n) \) be a pair of monotone Boolean functions given by their irredundant disjunctive normal forms:

\[
f = \bigvee_{I \in F} \bigwedge_{i \in I} x_i, \quad g = \bigvee_{I \in G} \bigwedge_{i \in I} x_i,
\]

where \( F \) and \( G \) are the sets of the prime (\( \equiv \) minimal) implicants \( I \subseteq \{1, \ldots, n\} \) of \( f \) and \( g \), respectively. It is well known that if \( f \) and \( g \) are mutually dual, i.e., \( f(x_1, \ldots, x_n) \equiv -g(-x_1, \ldots, -x_n) \) for all \( (x_1, \ldots, x_n) \in \{0,1\}^n \), then \( F \) or \( G \) contains an implicant of only logarithmic size:

\[
\min \{|I| : I \in F \cup G\} \leq \log(|F| + |G|) \tag{1}
\]

(see e.g. Seymour (1974), p. 310; see also Beck (1978)). It is also well known that any dual monotone DNFs \( f \) and \( g \) satisfy the conditions

\[ I \cap J \neq \emptyset \text{ for any } I \in F \text{ and } J \in G, \]

for otherwise there exist disjoint implicants \( I \in F \) and \( J \in G \), and then \( f(x_1, \ldots, x_n) = g(-x_1, \ldots, -x_n) = 1 \) for the characteristic vector \( x \) of \( I \).

Assume that \( f \) and \( g \) are not constant, i.e., \(|F| \geq 1\), and let

\[
\mu_i = \frac{\#\{I \in F \mid i \in I\}}{|F|} \quad \text{and} \quad \nu_i = \frac{\#\{I \in G \mid i \in I\}}{|G|}
\]

be the frequencies with which variable \( x_i, \ i \in \{1, \ldots, n\} \), occurs in \( f \) and \( g \), respectively. From (1) and (*) it follows that any pair of dual DNFs \( f \) and \( g \) contains a variable of logarithmically high frequency:

\[
\max\{\mu_1, \nu_1, \ldots, \mu_n, \nu_n\} \geq \frac{1}{\log(|F| + |G|)} \tag{2}
\]

Equivalently, if \( F \) and \( G \) is a pair of transversal hypergraphs (\( \equiv \) mutually blocking clutters), then either \( F \) contains a vertex of degree \( \geq |F|/\log(|F| + |G|) \), or \( G \) has a vertex of degree \( \geq |G|/\log(|F| + |G|) \).

Fredman and Khachiyan (1994) used (2) to show that the duality of any monotone DNFs \( f \) and \( g \) can be tested in quasi-polynomial time \((|F| + |G|)^{\text{polylog(|F|+|G|)}} \). Here we give examples of arbitrarily large dual monotone DNFs for which both bounds (1) and (2) are tight up to a factor of 2:

**Proposition.** There exist dual monotone irredundant DNFs \( f \) and \( g \) with arbitrarily large \(|F|\) and \(|G|\) such that

\[
\min \{|I| : I \in F \cup G\} \geq \frac{\log(|F| + |G|)}{2} \tag{1'}
\]

and

\[
\max\{\mu_1, \nu_1, \ldots, \mu_n, \nu_n\} \leq \frac{2}{\log(|F| + |G|)} \tag{2'}
\]

In the remainder of this note we prove the proposition by using monotone Boolean formulae that correspond to binary \( \land, \lor \)-alternating trees. It should be noted that \( \land, \lor \)-alternating trees can also be used to derive an \((|F| + |G|)^{\text{polylog(|F|+|G|)}} \) lower bound on the running time of the first of the two duality testing algorithm suggested by Fredman and Khachiyan (1994). The second of these algorithm runs in time \((|F| + |G|)^{\text{polylog(|F|+|G|)}} \) for any \( f \) and \( g \).
2. Alternating trees

Let \( f_k \) be the monotone Boolean function of \( n(k) = 2^{2k-1} \) variables defined by the recurrence:

\[
\begin{align*}
    f_1 &= x_1 \lor x_2, \\
    f_2 &= (x_1 \lor x_2)(x_3 \lor x_4) \lor (x_5 \lor x_6)(x_7 \lor x_8), \\
    & \quad \vdots \\
    f_{k+1} &= f_k(x_1, \ldots, x_{n(k)})f_k(x_{n(k)+1}, \ldots, x_{2n(k)}) \lor f_k(x_{2n(k)+1}, \ldots, x_{3n(k)})f_k(x_{3n(k)+1}, \ldots, x_{4n(k)}).
\end{align*}
\]

Denote by \( F_k \) the set of the prime implicants of \( f_k \). Clearly, \( |F_1| = 2, \ |F_2| = 2 \cdot 2^2, \ldots, |F_{k+1}| = 2|F_k|^2 \), which implies

\[
|F_k| = 2^{2^k-1}, \quad k = 1, 2, \ldots
\]

Next, let \( \mu(k) = \mu_1 = \ldots = \mu_{n(k)} \) be the frequency with which each variable \( x_i \) occurs in the irredundant disjunctive normal form of \( f_k \). Since the size of any prime implicant of \( f_k \) is \( 2^{k-1} \), it follows that \( 2^{k-1} = \mu(k)n(k) \), and consequently \( \mu(k) = 2^{2^{k-1}} \).

The dual functions \( g_k \) are defined by the dual recurrence:

\[
\begin{align*}
    g_1 &= x_1x_2, \\
    g_2 &= (x_1x_2 \lor x_3x_4)(x_5x_6 \lor x_7x_8), \\
    & \quad \vdots \\
    g_{k+1} &= (g_k(x_1, \ldots, x_{n(k)}) \lor g_k(x_{n(k)+1}, \ldots, x_{2n(k)}))(g_k(x_{2n(k)+1}, \ldots, x_{3n(k)}) \lor g_k(x_{3n(k)+1}, \ldots, x_{4n(k)}))
\end{align*}
\]

Denoting by \( G_k \) the set of the prime implicants of \( g_k \), we obtain \( |G_1| = 1, \ |G_2| = 2^2, \ldots, |G_{k+1}| = (2|G_k|)^2 \). This gives

\[
|G_k| = \frac{1}{2}|F_k| = 2^{2^k-2}, \quad k = 1, 2, \ldots
\]

We also have \( |I| = 2^k \) for any prime implicant \( I \in G_k \). Therefore each variable \( x_i \) occurs in \( g_k \) with frequency \( \nu(k) = \nu_1 = \ldots = \nu_{n(k)} = 2^{-k+1} \). Hence

\[
\max\{\mu(k), \nu(k)\} = \max\{\mu_1, \nu_1, \ldots, \mu_{n(k)}, \nu_{n(k)}\} = 1/\min\{|I| : I \in F_k \cup G_k\} = 2^{-k+1}
\]

\[
\leq \frac{2^k + \log(3/4)}{\log(|F_k| + |G_k|)}.
\]

References

