Toward Showing Equality in Lemma 25 from Bshouty's "Learning with Errors in Answers to Membership Queries"

Livia Overand

Aug 1, 2005

1 Introduction

In "Learning with Errors in Answers to Membership Queries" it is shown that for any two boolean functions $f : \{0,1\}^{n_1} \to \{0,1\}$ and $g : \{0,1\}^{n_2} \to \{0,1\}$ and two sets of disjoint variables $x = (x_1, ...x_{n_1})$ and $y = (y_1, ...y_{n_2})$ we have,

$$size_{DCD}(f(x) \oplus g(y)) \leq size_{DCD}(f(x)) \cdot size_{DCD}(g(y)).$$

I would like to extend this to prove equality. I have made progress in doing so, but lack the proof for one crucial step. This paper documents the progress I have made and describes the problems that I have encountered in attempting to complete this proof.

2 Definitions

1. The number of *conflicts* between two terms is the number of variables occurring unnegated in one term and negated in the other.
2. A DNF is *disjoint* if any two of its terms have at least one conflict.
3. Any two terms that have at least two conflicts can not be covered by a single term of fewer variables.
4. The *minimal disjoint CDNF representation for the always false function* is $(0,\ 1)$, where $size_{DCD}((0,1)) = 1$.
5. The *minimal disjoint DNF representation for the always false function* is 0, where $size_{DDNF}(0) = 0$.
6. The *minimal disjoint CDNF representation for the always true function* is $(1,\ 0)$, where $size_{DCD}((1,\ 0)) = 1$.
7. The *minimal disjoint DNF representation for the always true functon* is 1, where $size_{DDNF}(1) = 1$.
8. $size_{DCD}(f) = size_{DDNF}(f) + size_{DDNF}(\bar{f})$
9. $(f \oplus g) \equiv (f \wedge \bar{g}) \vee (\bar{f} \wedge g) \equiv \overline{(f \wedge g) \vee (\bar{f} \wedge \bar{g})}$

3 Completed Progress

Remark 1 *By definitions 5 and 6 we have,*

$$size_{DCD}(f \oplus g) = size_{DDNF}((f \wedge \bar{g}) \vee (\bar{f} \wedge g)) + size_{DDNF}((f \wedge g) \vee (\bar{f} \wedge \bar{g})).$$

Lemma 1 *For any two boolean functions $f : \{0,1\}^{n_1} \rightarrow \{0,1\}$ and $g : \{0,1\}^{n_2} \rightarrow \{0,1\}$ and two sets of disjoint variables $x = (x_1, ...x_{n_1})$ and $y = (y_1, ...y_{n_2})$,*

$$size_{DDNF}((f \wedge \bar{g}) \vee (\bar{f} \wedge g)) = size_{DDNF}(f \wedge \bar{g}) + size_{DDNF}(\bar{f} \wedge g).$$

Proof : First notice that for any two functions on disjoint variables we have,

$$f \wedge \bar{g} \equiv (f \wedge \bar{g}) \wedge (f \vee \bar{g}) \equiv (f \wedge \bar{g}) \wedge \overline{(\bar{f} \wedge g)}$$

and

$$\bar{f} \wedge g \equiv (\bar{f} \wedge g) \wedge (\bar{f} \vee g) \equiv (\bar{f} \wedge g) \wedge \overline{(f \wedge \bar{g})}.$$

So,

$$(f \wedge \bar{g}) \vee (\bar{f} \wedge g) \equiv \left[(f \wedge \bar{g}) \wedge \overline{(\bar{f} \wedge g)}\right] \vee \left[(\bar{f} \wedge g) \wedge \overline{(f \wedge \bar{g})}\right] \equiv (f \wedge \bar{g}) \oplus (\bar{f} \wedge g).$$

I first show

$$size_{DDNF}((f \wedge \bar{g}) \vee (\bar{f} \wedge g)) \leq size_{DDNF}(f \wedge \bar{g}) + size_{DDNF}(\bar{f} \wedge g).$$

Let P and Q be a minimal disjoint DNF for $(f \wedge \bar{g})$ and $(\bar{f} \wedge g)$ of size s_1 and s_2 respectively. Then $P \vee Q$ is a disjoint DNF for $((f \wedge \bar{g}) \vee (\bar{f} \wedge g))$ of size $s_1 + s_2$.

I now show

$$size_{DDNF}((f \wedge \bar{g}) \vee (\bar{f} \wedge g)) \geq size_{DDNF}(f \wedge \bar{g}) + size_{DDNF}(\bar{f} \wedge g).$$

Suppose $size_{DDNF}((f \wedge \bar{g}) \vee (\bar{f} \wedge g)) < size_{DDNF}(f \wedge \bar{g}) + size_{DDNF}(\bar{f} \wedge g)$. Then, there exists some term in the disjoint DNF for $(f \wedge \bar{g}) \vee (\bar{f} \wedge g)$ that covers a portion of $f \wedge \bar{g}$ and a portion of $\bar{f} \wedge g$. Clearly, any such term must have less than $n_1 + n_2$ literals, since any term with $n_1 + n_2$ literals must either be in $f \wedge \bar{g}$ or in $\bar{f} \wedge g$, but not both. So consider a term that covers a portion of both $f \wedge \bar{g}$ and $\bar{f} \wedge g$ that has less than $n_1 + n_2$ literals.

Case 1: All absent variables are from the domain of f. Then this term covers a portion of f and of $\bar{f}$. However, if no variables are removed from the domain of g, then this term still only covers a portion of g or $\bar{g}$, but not both.

Case 2: All absent variables are from the domain of g. Then this term covers a portion of g and of $\bar{g}$. However, if no variables are removed from the domain of f, then this term still only covers a portion of f or $\bar{f}$, but not both.

Case 3: Some variables are removed from the domain of f and from the domain of g. Then some x_i has been removed such that when the value of that variable changes, the value of f changes. Also some y_i has been removed such that when the value of that variable changes, the value of g changes. So this term covers assignments that satisfy $f \wedge \bar{g}$ and $\bar{f} \wedge g$. However, it also covers assignments that satisfy $f \wedge g$ and $\bar{f} \wedge \bar{g}$. This is a contradiction, because $(f \wedge \bar{g}) \vee (\bar{f} \wedge g)$ is zero when either $f \wedge g$ is satisfied or when $\bar{f} \wedge \bar{g}$ is satisfied. Therefore, $size_{DDNF}((f \wedge \bar{g}) \vee (\bar{f} \wedge g)) \geq size_{DDNF}(f \wedge \bar{g}) + size_{DDNF}(\bar{f} \wedge g)$. $\square$

Lemma 2 *For any minimal disjoint DNF T of size s, the expression obtained by deleting any term from T is a minimal disjoint DNF of size $s-1$.*

Proof : Let $T = t_1 \vee t_2 \vee \ldots \vee t_s$ and T' be the expression obtained by deleting some t_i from T. Clearly, $T' = t_1 \vee t_2 \vee \ldots \vee t_{i-1} \vee t_{i+1} \vee \ldots \vee t_s$ is a disjoint DNF of size at least $s-1$. Suppose $size_{DDNF}(T') < s-1$. Then there is some covering of all but one of the terms in T of size less than $s-1$. This, however, is a contradiction to the minimality of T. $\square$

Fact 1 *For any two boolean functions $f : \{0,1\}^{n_1} \rightarrow \{0,1\}$ and $g : \{0,1\}^{n_2} \rightarrow \{0,1\}$ and two sets of disjoint variables $x = (x_1, ...x_{n_1})$ and $y = (y_1, ...y_{n_2})$, there are*

$$(2^{2^{n_1}} - 1) \cdot (2^{2^{n_2}} - 1) + 1$$

different boolean functions $h : \{0,1\}^{n_1+n_2} \rightarrow \{0,1\}$ where h is of the form $f \wedge g$.

Proof : The number of functions on n_1 variables is $2^{2^{n_1}}$. Likewise, the number of functions on n_2 variables is $2^{2^{n_2}}$. Since f and g are on disjoint variables $f(x_1, ...x_{n_1}) \wedge g(y_1, ...y_{n_2}) = h(x_1, ...x_{n_1}, y_1, ...y_{n_2})$. By the product rule the number of functions $h : \{0,1\}^{n_1+n_2} \rightarrow \{0,1\}$ where h is of the form $f \wedge g$ is $2^{2^{n_1}} \cdot 2^{2^{n_2}}$. However, one of the $2^{2^{n_1}}$ functions is the always false function. Likewise, one of the $2^{2^{n_2}}$ functions is the always false function. Since $0 \wedge g = 0$ and $f \wedge 0 = 0$, $2^{2^{n_1}} + 2^{2^{n_2}} - 1$ functions, h, will be the always false function. Therefore, the number of different functions $h : \{0,1\}^{n_1+n_2} \rightarrow \{0,1\}$ where h is of the form $f \wedge g$ is

$$2^{2^{n_1}} \cdot 2^{2^{n_2}} - (2^{2^{n_1}} + 2^{2^{n_2}} - 1) + 1 = (2^{2^{n_1}} - 1) \cdot (2^{2^{n_2}} - 1) + 1$$

$\square$

Fact 2 *For any two boolean functions $f : \{0,1\}^{n_1} \rightarrow \{0,1\}$ and $g : \{0,1\}^{n_2} \rightarrow \{0,1\}$ and two sets of disjoint variables $x = (x_1, ...x_{n_1})$ and $y = (y_1, ...y_{n_2})$, if P is a minimal disjoint DNF for $f(x)$ and Q is a minimal disjoint DNF for $g(y)$, then no two terms in $P \wedge Q$ can be covered by a single term of fewer variables.*

Proof : Since any two terms in P have at least one conflict, any two terms in Q have at least one conflict, and P and Q are on disjoint variables, any two terms in $P \wedge Q$ have at least two conflicts. Any two terms that have two conflcits can not be covered by a single term of fewer variables. $\square$

4 Future Work

In order to finish proving

$$size_{DCD}(f(x) \oplus g(y)) = size_{DCD}(f(x)) \cdot size_{DCD}(g(y))$$

it is necessary to show that for any two boolean functions $f : \{0,1\}^{n_1} \rightarrow \{0,1\}$ and $g : \{0,1\}^{n_2} \rightarrow \{0,1\}$ and two sets of disjoint variables $x = (x_1, ...x_{n_1})$ and $y = (y_1, ...y_{n_2})$,

$$size_{DDNF}(f \wedge g) \geq size_{DDNF}(f) \cdot size_{DDNF}(g). \tag{1}$$

If this fact can be proven then it would imply that

$$size_{DDNF}((f \wedge \bar{g}) \vee (\bar{f} \wedge g)) \geq size_{DDNF}(f) \cdot size_{DDNF}(\bar{g}) + size_{DDNF}(\bar{f}) \cdot size_{DDNF}(g),$$

which would then imply that

$$\begin{aligned} size_{DCD}(f \oplus g) &\geq size_{DDNF}(f) \cdot size_{DDNF}(\bar{g}) + size_{DDNF}(\bar{f}) \cdot size_{DDNF}(g) \\ &+ size_{DDNF}(f) \cdot size_{DDNF}(g) + size_{DDNF}(\bar{f}) \cdot size_{DDNF}(\bar{g}) \\ &= (size_{DDNF}(f) + size_{DDNF}(\bar{f})) \cdot (size_{DDNF}(g) + size_{DDNF}(\bar{g})) \\ &= size_{DCD}(f) \cdot size_{DCD}(g) \end{aligned}$$

I have not, however, been able to prove (1). I attempted to prove this by induction on $n = n_1 + n_2$. The base case is simple. For n=0 we have $(n_1, n_2) = (0, 0)$ The only functions on zero variables are the always true or always false function. If $f \wedge g = 0$ then either $f = 0$ or $g = 0$, and clearly $size_{DDNF}(0) = 0 \geq size_{DDNF}(0) \cdot size_{DDNF}(g) = 0 \cdot size_{DDNF}(g) = 0$. If $f \wedge g = 1$ then $f = g = 1$, and clearly $size_{DDNF}(1) = 1 \geq size_{DDNF}(1) \cdot size_{DDNF}(1) = 1 \cdot 1 = 1$. Then the inductive hypothesis is for a boolean function $f \wedge g : \{0,1\}^k \to \{0,1\}$, where k is an arbitrary number of variables, $size_{DDNF}(f \wedge g) \geq size_{DDNF}(f) \cdot size_{DDNF}(g)$. I have not, however, been able to find a way to use this hypothesis to prove the case for $f \wedge g : \{0,1\}^{k+1} \to \{0,1\}$.

I have also attempted to prove (1) by double induction on (n_1, n_2). Again the base cases are simple, and we get the additional facts that $\forall n_2((0, n_2) \to (0, n_2+1))$ and $\forall n_1((n_1, 0) \to (n_1+1, 0))$. Again the problem is that I have not found a way to use the inductive hypothesis to prove the inductive step.

I believe my most hopeful attempt to prove (1) was by double induction on (s_1, s_2), where $size_{DDNF}(f) = s_1$ and $size_{DDNF}(g) = s_2$. Following is an outline of my progress for this proof.

$\forall s_1 \forall s_2$, if $size_{DDNF}(f) = s_1$ and $size_{DDNF}(g) = s_2$, then $size_{DDNF}(f \wedge g) = s_1 \cdot s_2$.

- **Base case** $\forall s_2$, if $size_{DDNF}(f) = 0$ and $size_{DDNF}(g) = s_2$, then $size_{DDNF}(f \wedge g) = 0 \cdot s_2$.

  - If $size_{DDNF}(f) = 0$, then f is the always false function. For any function g, $0 \wedge g = 0$, so $size_{DDNF}(0 \wedge g) = 0$.

- **Inductive Hypothesis** $\forall s_2$, if $size_{DDNF}(f) = m$ and $size_{DDNF}(g) = s_2$, then $size_{DDNF}(f \wedge g) = m \cdot s_2$.

- **Inductive Step** $\forall s_2$, if $size_{DDNF} = m + 1$ and $size_{DDNF}(g) = s_2$, then $size_{DDNF}(f \wedge g) = (m+1) \cdot s_2$.

  - **Base Case** If $size_{DDNF} = m + 1$ and $size_{DDNF}(g) = 0$, then $size_{DDNF}(f \wedge g) = (m+1) \cdot 0$.

    - If $size_{DDNF}(g) = 0$, then g is the always false function. For any function f, $f \wedge 0 = 0$, so $size_{DDNF}(f \wedge 0) = 0$.

- **Inductive Hypothesis** If $size_{DDNF}(f) = m+1$ and $size_{DDNF}(g) = n$, then $size_{DDNF}(f \wedge g) = (m+1) \cdot n$.

? **Inductive Step** If $size_{DDNF} = m+1$ and $size_{DDNF}(g) = n+1$, then $size_{DDNF}(f \wedge g) = (m+1) \cdot (n+1)$.

Intuitively, this last inductive step seems possible to prove. Let $P = p_1 \vee p_2 \vee \ldots \vee p_{m+1}$ be a minimal disjoint DNF for f and $Q = q_1 \vee q_2 \vee \ldots \vee q_{n+1}$ be a minimal disjoint DNF for g. Then, $P \wedge Q = \bigvee_{i=1}^{m+1} \bigvee_{j=1}^{n+1} (p_i \wedge q_j) = \left[\bigvee_{i=1}^{m+1} \bigvee_{j=1}^{n} (p_i \wedge q_j)\right] \vee \left[\bigvee_{i=1}^{m+1} (p_i \wedge q_{n+1})\right]$. By the inductive hypothesis, we know that $\bigvee_{i=1}^{m+1} \bigvee_{j=1}^{n} (p_i \wedge q_j)$ is a minimal disjoint DNF of size $(m+1) \cdot n$ for $f \wedge g$, if $size_{DDNF}(f) = m+1$ and $size_{DDNF}(g) = n$. It is also clear that $\bigvee_{i=1}^{m+1} (p_i \wedge q_{n+1})$ is a minimal disjoint DNF of size $m+1$ for $f \wedge g$, if $size_{DDNF}(f) = m+1$ and $size_{DDNF}(g) = 1$. However, it is unclear how to prove that $\left[\bigvee_{i=1}^{m+1} \bigvee_{j=1}^{n} (p_i \wedge q_j)\right] \vee \left[\bigvee_{i=1}^{m+1} (p_i \wedge q_{n+1})\right]$ is a minimal disjoint DNF for $f \wedge g$, if $size_{DDNF}(f) = m+1$ and $size_{DDNF}(g) = n+1$. In lemma 2, I showed that for any minimal disjoint DNF T of size s, the expression obtained by deleting any term from T is a minimal disjoint DNF of size $s-1$. If something could be said about the opposite direction, that is, if some conditions could be determined about forming a minimal disjoint DNF of size s by adding a term to minimal disjoint DNF of size $s-1$, then I believe the inductive step could be proved.

The only way I have been able to prove (1) for any fixed n is by exhaustively considering all functions on n variables. I have, in fact, done this for $n = 1, 2,$and 3.

I have also attempted to prove that for any two boolean functions $f : \{0,1\}^{n_1} \to \{0,1\}$ and $g : \{0,1\}^{n_2} \to \{0,1\}$ and two sets of disjoint variables $x = (x_1, ...x_{n_1})$ and $y = (y_1, ...y_{n_2})$, if P is a minimal disjoint DNF for $f(x)$ of size s_1 and Q is a minimal disjoint DNF for $g(x)$ of size s_2, then $P \wedge Q$ is minimal disjoint DNF for $f \wedge g$ of size $s_1 \cdot s_2$. Clearly, $P \wedge Q$ is a disjoint DNF for $f \wedge g$ of size $s_1 \cdot s_2$. Showing that $P \wedge Q$ is minimal, however, has proved to be a difficult task. There really is no precise definition for a minimal representation of a function other than its size is smaller than any other representation of the function. A minimal representation is not unique, and there certainly are other minimal disjoint DNF representations other than $P \wedge Q$ for $f \wedge g$.

5 Conclusion

In my attempt to prove that for any two boolean functions $f : \{0,1\}^{n_1} \to \{0,1\}$ and $g : \{0,1\}^{n_2} \to \{0,1\}$ and two sets of disjoint variables $x = (x_1, ...x_{n_1})$ and $y = (y_1, ...y_{n_2})$,

$$size_{DCD}(f(x) \oplus g(y)) = size_{DCD}(f(x)) \cdot size_{DCD}(g(y))$$

I have only managed to show that

$$size_{DCD}(f \oplus g) = size_{DDNF}(f \wedge \bar{g}) + size_{DDNF}(\bar{f} \wedge g) + size_{DDNF}(f \wedge g) + size_{DDNF}(\bar{f} \wedge \bar{g}).$$

It remains to be shown that

$$size_{DDNF}(f \wedge g) \geq size_{DDNF}(f) \cdot size_{DDNF}(g)$$

holds for any two boolean functions on disjoint variables. I am thouroughly convinced that this is true and that it can in fact be proven.