One Property of Cross-Intersecting Families

Alexander A. Razborov*
Steklov Mathematical Institute
Gubkina 8, 117966, GSP–1
Moscow, RUSSIA

Nikolai K. Vereshchagin†
Dept. of Mathematical Logic and Theory of Algorithms,
Moscow State University,
Moscow 119899, RUSSIA
E-mail: ver@mech.math.msu.su

Theorem 1. *Assume that $\mathcal{A}, \mathcal{B}$ are finite families of sets such that every set in $\mathcal{A}$ has at most m elements, every set in $\mathcal{B}$ has at most n elements, and every set in $\mathcal{A}$ intersects with every set in $\mathcal{B}$. Then there exists an element c such that*

$$\frac{|\{A \in \mathcal{A} \mid c \in A\}|}{|\mathcal{A}|} \geq \frac{1}{2n}, \qquad \frac{|\{B \in \mathcal{B} \mid c \in B\}|}{|\mathcal{B}|} \geq \frac{1}{2m}.$$

Proof. Assume the contrary and let $\mathbf{A}, \mathbf{B}$ be independent random variables that are uniformly distributed in $\mathcal{A}, \mathcal{B}$ respectively. Then the probability of the event $\exists c\ (c \in \mathbf{A} \cap \mathbf{B})$ is equal to 1. Hence

$$\sum_c \mathrm{Prob}[c \in \mathbf{A} \cap \mathbf{B}] \geq 1.$$

Let C_0 consist of those c for which $\frac{|\{A \in \mathcal{A} | c \in A\}|}{|\mathcal{A}|} = \mathrm{Prob}[c \in \mathbf{A}] < \frac{1}{2n}$, and C_1 of the remaining c's. Note that by our assumption for any $c \in C_1$, $\mathrm{Prob}[c \in \mathbf{B}] = \frac{|\{B \in \mathcal{B} | c \in B\}|}{|\mathcal{B}|} < \frac{1}{2m}$ holds. We have therefore

$$\sum_{c \in C_1} \mathrm{Prob}[c \in \mathbf{A} \cap \mathbf{B}] = \sum_{c \in C_1} (\mathrm{Prob}[c \in \mathbf{A}] \cdot \mathrm{Prob}[c \in \mathbf{B}])$$
$$< \frac{1}{2m} \cdot \sum_{c \in C_1} \mathrm{Prob}[c \in \mathbf{A}] \leq \frac{1}{2m} \cdot \sum_c \mathrm{Prob}[c \in \mathbf{A}] = \frac{1}{2m} \cdot \mathrm{E}[|\mathbf{A}|] \leq \frac{1}{2}.$$

In a similar way we obtain

$$\sum_{c \in C_0} \mathrm{Prob}[c \in \mathbf{A} \cap \mathbf{B}] < \frac{1}{2},$$

*Supported by the Russian Basic Research Foundation
†Supported by the grant MQT300 from the International Science Foundation

a contradiction. ☐

This theorem can be used to obtain a new statement of the following type:

> if DNFs F_0, F_1 are small and the formula $F_0 \wedge F_1$ is not satisfiable, then given an assignment, we can either certify that F_0 is false on that assignment or certify that F_1 is false on that assignment by probing only a small number of variables.

In the known result of this kind [1, 2, 3] the number of probed variables is at most mn, where m, n are maximum fanins of ANDs in F_0, F_1, respectively. Theorem 1 yields a sometimes better bound of $2m \ln N + 2n \ln M$, where M, N are the *numbers* of ANDs in F_0, F_1 respectively.

Theorem 2. *Assume that F_0 is a DNF that is an OR of M ANDs of fanin at most m, and F_1 is a DNF that is an OR of N ANDs of fanin at most n. Assume that the formula $F_0 \wedge F_1$ is not satisfiable. Then given an assignment a, we can either certify that F_0 is false on a or certify that F_1 is false on a by probing at most $2m \ln N + 2n \ln M + 2$ variables.*

Proof. Let C be an AND from F_0. We interpret C as the set of all its literals, and let $\mathcal{A} = \{C \mid C \text{ is an AND from } F_0\}$. Let $\mathcal{B}$ be obtained in the same way using F_1 instead of F_0, but this time we flip all literals, i.e., $x \in C \in F_1$ gets replaced by $\bar{x}$, and $\bar{x}$ gets replaced by x. Then the unsatisfiability of $F_0 \wedge F_1$ means that each set in $\mathcal{A}$ intersects with each set in $\mathcal{B}$. Applying Theorem 1 we find a literal that belongs to many sets from both $\mathcal{A}, \mathcal{B}$ and we probe its underlying variable x. Then either we learn that at least $\frac{M}{2n}$ ANDs from F_0 are false on a or we learn that at least $\frac{N}{2m}$ ANDs from F_1 are false on a.

Update F_0, F_1 by deleting false ANDs. Again $F_0 \wedge F_1$ is unsatisfiable thus we can form new $\mathcal{A}, \mathcal{B}$ and apply Theorem 1. Repeat this until one of F_0, F_1 has no ANDs. If this is F_0 then it is false on the given assignment. Otherwise F_1 is false.

Let us estimate the number of evaluated variables. Let t_0 be the number of times when at least a fraction $1/2n$ of ANDs from the current F_0 was deleted, and t_1 be the number of times when at least a fraction $1/2m$ of ANDs from F_1 was deleted. We have then

$$M\left(1 - \frac{1}{2n}\right)^{t_0 - 1} \geq 1, \qquad N\left(1 - \frac{1}{2m}\right)^{t_1 - 1} \geq 1.$$

Therefore

$$t_0 - 1 \leq -\frac{1}{\ln(1 - \frac{1}{2n})} \ln M \leq 2n \ln M.$$

Analogously, $t_1 \leq 2m \ln N + 1$. ☐

The bound in Theorem 1 is tight up to a multiplicative factor of 2, as the following example shows:

$$\mathcal{A} = \{A_i \mid i = 1, 2, \dots, n\}, \text{ where } A_i = \{\langle i, j\rangle \mid j = 1, 2, \dots, m\},$$
$$\mathcal{B} = \{B_j \mid j = 1, 2, \dots, m\}, \text{ where } B_j = \{\langle i, j\rangle \mid i = 1, 2, \dots, n\}.$$

For any c we have

$$\frac{|\{A \in \mathcal{A} \mid c \in A\}|}{|\mathcal{A}|} = \frac{1}{n}, \qquad \frac{|\{B \in \mathcal{B} \mid c \in B\}|}{|\mathcal{B}|} = \frac{1}{m}.$$

References

[1] M. Blum and R. Impagliazzo. Generic oracles and oracle classes. In *Proc. of the 28th IEEE Ann. Symp. on Foundations of Computer Science (FOCS)*, pages 118–126, 1987.

[2] R. Impagliazzo and M. Naor. Decision trees and downward closures. In *Proc. of the 3rd ACM Ann. Conf. on Structure in Complexity Theory*, pages 29–38, 1988.

[3] G. Tardos. Query complexity, or why it is difficult to separate $NP^A \cap coNP^A$ from P^A by a random oracle. *Combinatorica*, 9:385–392, 1989.