Boolean Complexity and Ramsey Theorems

Pavel Pudlák

1. General Remarks

The aim of this paper is to bring attention to some connections between these two fields. In the complexity theory there are difficult open problems most of which are essentially of combinatorial character. It is generally believed that some interaction between complexity theory and combinatorics may help to solve these problems.

An n-dimensional *Boolean function* is any mapping $f : \{0,1\}^n \to \{0,1\}$. Thus a Boolean function can be also viewed as a partition of the n-cube $\{0,1\}^n$. A Boolean function is called symmetric if $f(a_1, \ldots, a_n)$ depends only on the number of 1's among $a_1, \ldots, a_n$. We call the set of all vectors with exactly k ones the k-th *level* of the n-cube. Hence a symmetric function is a function which is constant on every single level. Given a complete basis of connectives, we define *the formula size complexity* $L(f)$ of a function f to be the size of the smallest formula realizing f, where the size of a formula is conveniently defined to be the number of all the occurrences of variables in it. (E.g. $x_1 \wedge (\neg x_1 \vee x_2)$ has size 3).

Theorem A. *For every basis there exists $\epsilon > 0$ such that if f is n-dimensional and*

$$L(f) \leq \epsilon \cdot n(\log\log n - \log r),$$

then there exists an interval $I = (\mathbf{0}, \mathbf{a})$ of length r in the n-cube such that

(1) $f \mid I$ *is symmetric;*

(2) in $f \mid I$ all the even levels, with a possible exeption of the 0-th level, are of the same color, and all the odd levels are of the same color. (It is not excluded that f is constant on I.)

Theorem B. *For the basis of all at most binary connectives there exists $\epsilon > 0$ such that it f is n-dimensional and*

$$L(f) \leq \epsilon \cdot n \cdot (\log n - \log r),$$

J. Nešetil et al. (eds.), *Mathematics of Ramsey Theory*
© Springer-Verlag Berlin Heidelberg 1990

then there exists an interval $I = (\mathbf{a}, \mathbf{b})$ of length r in the middle of the n-cube such that

(1) $f|I$ is symmetric;

(2) in $f|I$ all the even levels are of the same color, and all the odd levels are of the same color. (The exact meaning of "the middle" is that the number of 1's in $\mathbf{a}$ equals to the number of 0's in $\mathbf{b}$ possibly -1.)

Theorem A is a reformulation of the Hodes-Specker theorem (Hodes, Specker 1968) with the bound proved in (Pudlák 1984), the second theorem is a reformulation of the main theorem of Fischer, Meyer, Paterson (1982). The bounds are known to be of the best growth rate. I see at least three connections of these theorems to Ramsey theory.

1. The general form of the statement is:"If an object is of small complexity, then it is locally very simple". If we consider e.g. the number of colors as the complexity of a coloration (say of a complete graph), then the Ramsey theorem is of this form.
2. Using Ramsey theorem one can prove e.g. that there exists a function $r(n)$, with $\lim_{n\to\infty} r(n) = \infty$, such that for every n-dimensional Boolean function f there exists an interval $I = (\mathbf{0}, \mathbf{a})$ of length $r(n)$ such that $f \mid I$ is symmetric. In case f is of small complexity Theorem A extends the information about $f|I$ in two ways: gives us a larger interval I and the condition (2).
3. The original proof of the Hodes-Specker theorem and the proof of Fischer-Meyer-Paterson theorem use the standard heuristic "divide and take the largest one" used also for Ramsey theorems. Ramsey theorem was also used in the proof of a generalization of Hodes-Specker theorem by Vilfan (1976). Ramsey theorem is the corner-stone of the proof of the bound of Theorem A in Pudlák (1984). Roughly speaking the proof goes as follows. Given a Boolean formula $\alpha(x_1, \ldots, x_n)$, where $x_1, \ldots, x_n$ are the propositional variables of the formula, we define *the induced formula* αX for every $X \subseteq \{x_1, \ldots, x_n\}$ in a suitable way. The formula is called *homogeneous* if for every $X, Y, |X| = |Y| = 2$, αX is isomorphic to αY, (which means that if we substitute the first variable of X for the first variable of Y and the second variable of X for the second variable of Y in αY, then we obtain αX). Given r, if the complexity of $\alpha(x_1, \ldots, x_n)$ is small (i.e. $\leq \epsilon \cdot n \cdot (\log\log n - \log r)$), then using the Ramsey theorem one can find a subset of variables H of cardinality r such that αH is homogeneous. Then it is shown (and this is the difficult part of the proof) that every homogeneous formula determines a Boolean function which satisfies (1), (2) of Theorem A.

The theorem of Ajtai (1983) and Furst, Saxe and Sipser (1981) can be stated also in a form resembling the Ramsey theorem. A theorem of Hodes-Specker type for branching programs was announced in Pudlák (1984).

During the preparation of this book several new lower bounds to the complexity of Boolean functions have been obtained. A large part of these results uses some version of the Ramsey theorem. The lower bound of this paper