Electronic version of an article published as *International Journal of Foundations of Computer Science* **18**(5) (2007) 975–986. DOI: 10.1142/S012905410700508X.

ON THE EFFECT OF VARIABLE IDENTIFICATION ON THE ESSENTIAL ARITY OF FUNCTIONS ON FINITE SETS

MIGUEL COUCEIRO AND ERKKO LEHTONEN

Abstract. We show that every function of several variables on a finite set of k elements with $n > k$ essential variables has a variable identification minor with at least $n - k$ essential variables. This is a generalization of a theorem of Salomaa on the essential variables of Boolean functions. We also strengthen Salomaa's theorem by characterizing all the Boolean functions f having a variable identification minor that has just one essential variable less than f.

1. Introduction

Theory of essential variables of functions has been developed by several authors [2, 5, 6, 7, 14, 16]. In this paper, we discuss the problem how the number of essential variables is affected by identification of variables (diagonalization). Salomaa [14] proved the following two theorems: one deals with operations on arbitrary finite sets, while the other deals specifically with Boolean functions. We denote the number of essential variables of f by $\operatorname{ess} f$.

Theorem 1.1. *Let A be a finite set with k elements. For every $n \leq k$, there exists an n-ary operation f on A such that $\operatorname{ess} f = n$ and every identification of variables produces a constant function.*

Thus, in general, essential variables can be preserved when variables are identified only in the case that $n > k$.

Theorem 1.2. *For every Boolean function f with $\operatorname{ess} f \geq 2$, there is a function g obtained from f by identification of variables such that $\operatorname{ess} g \geq \operatorname{ess} f - 2$.*

Identification of variables together with permutation of variables and cylindrification induces a quasi-order on operations whose relevance has been made apparent by several authors [3, 8, 9, 10, 12, 15, 18]. In the case of Boolean functions, this quasi-order was studied in [4] where Theorem 1.2 was fundamental in deriving certain bounds on the essential arity of functions.

In this paper, we will generalize Theorem 1.2 to operations on arbitrary finite sets in Theorem 3.1. We will also strengthen Theorem 1.2 on Boolean functions in Theorem 4.1 by determining the Boolean functions f for which there exists a function g obtained from f by identification of variables such that $\operatorname{ess} g = \operatorname{ess} f - 1$.

Key words and phrases. Functions on finite sets; Boolean functions; essential variables; variable identification; arity gap; minors of functions.

2. Variable identification minors

Let A and B be arbitrary nonempty sets. A *B-valued function of several variables on A* is a mapping $f : A^n \to B$ for some positive integer n, called the *arity* of f. A-valued functions on A are called *operations on A.* Operations on $\{0,1\}$ are called *Boolean functions.*

We say that the i-th variable is *essential* in f, or f depends on x_i, if there are elements $a_1, \dots, a_n, b \in A$ such that

$$f(a_1, \dots, a_i, \dots, a_n) \neq f(a_1, \dots, a_{i-1}, b, a_{i+1}, \dots, a_n). \tag{1}$$

The number of essential variables in f is called the *essential arity* of f, and it is denoted by $\operatorname{ess} f$. Thus the only functions with essential arity zero are the constant functions.

For an n-ary function f, we say that an m-ary function g is obtained from f by *simple variable substitution* if there is a mapping $\sigma : \{1, \dots, n\} \to \{1, \dots, m\}$ such that

$$g(x_1, \dots, x_m) = f(x_{\sigma(1)}, \dots, x_{\sigma(n)}). \tag{2}$$

In the particular case that $n = m$ and σ is a permutation of $\{1, \dots, n\}$, we say that g is obtained from f by *permutation of variables.* For indices $i, j \in \{1, \dots, n\}$, $i \neq j$, if x_i and x_j are essential in f, then the function $f_{i\leftarrow j}$ obtained from f by the simple variable substitution

$$f_{i\leftarrow j}(x_1, \dots, x_n) = f(x_1, \dots, x_{i-1}, x_j, x_{i+1}, \dots, x_n) \tag{3}$$

is called a *variable identification minor* of f, obtained by identifying x_i with x_j. Note that $\operatorname{ess} f_{i\leftarrow j} < \operatorname{ess} f$, because x_i is not essential in $f_{i\leftarrow j}$ even though it is essential in f.

We define a quasiorder on the set of all B-valued functions of several variables on A as follows: $f \leq g$ if and only if f is obtained from g by simple variable substitution. If $f \leq g$ and $g \leq f$, we denote $f \equiv g$. If $f \leq g$ but $g \not\leq f$, we denote $f < g$. It can be easily observed that if $f \leq g$ then $\operatorname{ess} f \leq \operatorname{ess} g$, with equality if and only if $f \equiv g$.

For a B-valued function f of several variables on A, we denote the maximum essential arity of a variable identification minor of f by

$$\operatorname{ess}^< f = \max_{g<f} \operatorname{ess} g, \tag{4}$$

and we define the *arity gap* of f by $\operatorname{gap} f = \operatorname{ess} f - \operatorname{ess}^< f$.

3. Generalization of Theorem 1.2

Theorem 3.1. *Let A be a finite set of $k \geq 2$ elements, and let B be a set with at least two elements. Every B-valued function of several variables on A with $n > k$ essential variables has a variable identification minor with at least $n - k$ essential variables.*

In the proof of Theorem 3.1, we will make use of the following theorem due to Salomaa (Theorem 1 in [14]), which is a strengthening of Yablonski's [16] "fundamental lemma".

Theorem 3.2. *Let the function $f : M_1 \times \cdots \times M_n \to N$ depend essentially on all of its n variables, $n \geq 2$. Then there is an index j and an element $c \in M_j$ such that the function*

$$f(x_1, \dots, x_{j-1}, c, x_{j+1}, \dots, x_n) \tag{5}$$

depends essentially on all of its $n-1$ variables.

We also need the following auxiliary lemma.

Lemma 3.3. *Let f be an n-ary function with $\operatorname{ess} f = n > k$. Then there are indices $1 \leq i < j \leq k+1$ such that at least one of the variables $x_1, \dots, x_{k+1}$ is essential in $f_{i \leftarrow j}$.*

Proof. Since x_1 is essential in f, there are elements $a_1, \dots, a_n, b \in A$ such that

$$f(a_1, a_2, \dots, a_n) \neq f(b, a_2, \dots, a_n). \tag{6}$$

Thus there are indices $1 \leq i < j \leq k+1$ such that $a_i = a_j$. If $i \neq 1$, then it is clear that x_1 is essential in $f_{i \leftarrow j}$. If there are no such i and j with $i \neq 1$, then $i = 1 < j$ and we have that $b = a_l$ for some $1 < l \leq k+1$, $l \neq j$. For $m = 1, \dots, n$, let $c_m = a_m$ if $m \notin \{1, j, l\}$ and let $c_m = a_1$ if $m \in \{1, j, l\}$. Then $f(c_1, c_2, \dots, c_n)$ is distinct from at least one of $f(a_1, a_2, \dots, a_n)$ and $f(b, a_2, \dots, a_n)$. If $f(c_1, c_2, \dots, c_n) \neq f(a_1, a_2, \dots, a_n)$, then x_l is essential in $f_{1 \leftarrow j}$. If $f(c_1, c_2, \dots, c_n) \neq f(b, a_2, \dots, a_n)$, then x_l is essential in $f_{1 \leftarrow l}$. □

Proof of Theorem 3.1. By Theorem 3.2, there exist $k+1$ constants $c_1, \dots, c_{k+1} \in A$ such that, after a suitable permutation of variables, the function

$$f(c_1, \dots, c_{k+1}, x_{k+2}, \dots, x_n) \tag{7}$$

depends on all of its $n-k-1$ variables. There are indices $1 \leq i < j \leq k+1$ such that $c_i = c_j$, and by Lemma 3.3 there are indices $1 \leq l < m \leq k+1$ such that at least one of the variables $x_1, \dots, x_{k+1}$ is essential in $f_{l \leftarrow m}$. With a suitable permutation of variables, we may assume that $i = 1$, $j = 2$, $1 \leq l \leq 3$, $m = l+1$.

If one of the variables $x_1, \dots, x_{k+1}$ is essential in $f_{1 \leftarrow 2}$, then we are done. Otherwise we have that for all $a_{k+2}, \dots, a_n \in A$,

$$\begin{multline} f(c_1, c_1, c_3, c_4, \dots, c_{k+1}, a_{k+2}, \dots, a_n) = \\ f(c_3, c_3, c_3, c_4, \dots, c_{k+1}, a_{k+2}, \dots, a_n). \end{multline} \tag{8}$$

Thus the variables $x_{k+2}, \dots, x_n$ are essential in $f_{2 \leftarrow 3}$. If one of the variables $x_1, \dots, x_{k+1}$ is essential in $f_{2 \leftarrow 3}$, then we are done. Otherwise we have that for all $a_{k+2}, \dots, a_n \in A$,

$$\begin{multline} f(c_3, c_3, c_3, c_4, \dots, c_{k+1}, a_{k+2}, \dots, a_n) = \\ f(c_3, c_4, c_4, c_4, \dots, c_{k+1}, a_{k+2}, \dots, a_n), \end{multline} \tag{9}$$

and so the variables $x_{k+2}, \dots, x_n$ are essential in $f_{3 \leftarrow 4}$ and also at least one of $x_1, \dots, x_{k+1}$ is essential in $f_{3 \leftarrow 4}$. This completes the proof of Theorem 3.1. □

We would like to remark that our proof is considerably simpler than Salomaa's original proof of Theorem 1.2.

4. Strengthening of Theorem 1.2

It is well-known that every Boolean function is represented by a unique multilinear polynomial over the two-element field. Such a representation is called the *Zhegalkin polynomial* (or the *Reed–Muller polynomial*) of f [11, 13, 17]. It is clear that a variable is essential in f if and only if it occurs in the Zhegalkin polynomial of f. We denote by $\deg \mathfrak{p}$ the degree of polynomial $\mathfrak{p}$. If $\mathfrak{p}$ is the Zhegalkin polynomial of f, then we denote the Zhegalkin polynomial of $f_{i \leftarrow j}$ by $\mathfrak{p}_{i \leftarrow j}$. Note that the only polynomials of degree zero are the constant polynomials.

Theorem 4.1. *Let f be a Boolean function with at least two essential variables. Then the arity gap of f is two if and only if the Zhegalkin polynomial of f is of one of the following special forms:*

- $x_{i_1} + x_{i_2} + \dots + x_{i_n} + c$,
- $x_i x_j + x_i + c$,
- $x_i x_j + x_i x_k + x_j x_k + c$,
- $x_i x_j + x_i x_k + x_j x_k + x_i + x_j + c$,

where $c \in \{0, 1\}$. Otherwise the arity gap of f is one.

We prove first an auxiliary lemma that takes care of the functions of essential arity at least four whose Zhegalkin polynomial has degree two.

Lemma 4.2. *If f is a Boolean function with at least four essential variables and the Zhegalkin polynomial of f has degree two, then the arity gap of f is one.*

Proof. Denote the Zhegalkin polynomial of f by $\mathfrak{p}$. We need to consider several cases and subcases.

Case 1. Assume first that $\mathfrak{p}$ is of the form

$$\mathfrak{p} = x_i x_j + x_i x_k + x_j x_k + x_i \mathfrak{a}_i + x_j \mathfrak{a}_j + x_k \mathfrak{a}_k + \mathfrak{a}, \tag{10}$$

where $\mathfrak{a}_i$, $\mathfrak{a}_j$, $\mathfrak{a}_k$ are polynomials of degree at most 1 and $\mathfrak{a}$ is a polynomial of degree at most 2 such that there are no occurrences of variables x_i, x_j, x_k in $\mathfrak{a}_i$, $\mathfrak{a}_j$, $\mathfrak{a}_k$, $\mathfrak{a}$.

Subcase 1.1. Assume that $\deg \mathfrak{a}_i = \deg \mathfrak{a}_j = \deg \mathfrak{a}_k = 0$. Then $\mathfrak{a}$ contains a variable x_l distinct from x_i, x_j, x_k, and we can write $\mathfrak{a} = x_l \mathfrak{a}' + \mathfrak{a}''$, where $\mathfrak{a}'$ and $\mathfrak{a}''$ do not contain x_l. Then $f_{l \leftarrow i}$ is represented by the polynomial

$$\mathfrak{p}_{l \leftarrow i} = x_i x_j + x_i x_k + x_j x_k + x_i \mathfrak{a}' + \mathfrak{a}'', \tag{11}$$

where all essential variables of f except for x_l occur, because no terms cancel, and hence $\operatorname{gap} f = 1$.

Subcase 1.2. Assume that at least one of $\mathfrak{a}_i$, $\mathfrak{a}_j$, $\mathfrak{a}_k$ has degree 1, say $\deg \mathfrak{a}_i = 1$. Then $\mathfrak{a}_i$ contains a variable x_l distinct from x_i, x_j, x_k, and so $\mathfrak{a}_i = x_l + \mathfrak{a}'_i$, where $\mathfrak{a}'_i$ has degree at most 1 and does not contain x_l. Consider

$$\mathfrak{p}_{j \leftarrow k} = x_k (1 + \mathfrak{a}_j + \mathfrak{a}_k) + x_i \mathfrak{a}_i + \mathfrak{a}. \tag{12}$$

If all essential variables of f except for x_j occur in $\mathfrak{p}_{j \leftarrow k}$, then $\operatorname{gap} f = 1$ and we are done. Otherwise we need to analyze three different subcases.

Subcase 1.2.1. Assume that variable x_k occurs in $\mathfrak{p}_{j \leftarrow k}$ but there is a variable x_l that occurs in $\mathfrak{a}_{\mathrm{j}}$ and $\mathfrak{a}_k$ but not in $\mathfrak{a}_i$ nor in $\mathfrak{a}$ such that x_l does not occur in $\mathfrak{p}_{j \leftarrow k}$ (due to some cancelling terms in $\mathfrak{a}_j$ and $\mathfrak{a}_k$). Write $\mathfrak{a}_j = x_l + \mathfrak{a}'_j$, $\mathfrak{a}_k = x_l + \mathfrak{a}'_k$, and consider

$$\begin{aligned} \mathfrak{p}_{j \leftarrow l} &= x_i x_l + x_i x_k + x_l x_k + x_i \mathfrak{a}_i + x_l + x_l \mathfrak{a}'_j + x_k x_l + x_k \mathfrak{a}'_k + \mathfrak{a} \\ &= x_i x_l + x_i x_k + x_i \mathfrak{a}_i + x_l + x_l \mathfrak{a}'_j + x_k \mathfrak{a}'_k + \mathfrak{a}. \end{aligned} \tag{13}$$

Every essential variable of f except for x_j occurs in $\mathfrak{p}_{j\leftarrow l}$, and hence $\operatorname{gap} f = 1$.

Subcase 1.2.2. Assume that x_k does not occur in $\mathfrak{p}_{j\leftarrow k}$. In this case $\mathfrak{a}_j = \mathfrak{a}_k + 1$. Consider

$$\mathfrak{p}_{j\leftarrow i} = x_i(1 + \mathfrak{a}_i + \mathfrak{a}_j) + x_k\mathfrak{a}_k + \mathfrak{a}. \tag{14}$$

If any term of $\mathfrak{a}_j$ is cancelled by a term of $\mathfrak{a}_i$, it still remains as a term of $\mathfrak{a}_k$, and hence all variables occurring in $\mathfrak{a}_i$, $\mathfrak{a}_j$, $\mathfrak{a}_k$ occur in $\mathfrak{p}_{j\leftarrow i}$. If both x_i and x_k also occur in $\mathfrak{p}_{j\leftarrow i}$, then all essential variables of f except for x_j occur in $\mathfrak{p}_{j\leftarrow i}$, and so $\operatorname{gap} f = 1$.

If x_k does not occur in $\mathfrak{p}_{j\leftarrow i}$, then $\mathfrak{a}_k = 0$ and so $\mathfrak{a}_j = 1$. Then

$$\mathfrak{p}_{l\leftarrow i} = x_i x_j + x_i x_k + x_j x_k + x_i + x_i\mathfrak{a}'_i + x_j + \mathfrak{a}, \tag{15}$$

and every essential variable of f except for x_l occurs in $\mathfrak{p}_{l\leftarrow i}$. Thus $\operatorname{gap} f = 1$.

If x_i does not occur in $\mathfrak{p}_{j\leftarrow i}$, then $\mathfrak{a}_j = \mathfrak{a}_i + 1$, and hence $\mathfrak{a}_i = \mathfrak{a}_k$. Consider then

$$\mathfrak{p}_{i\leftarrow k} = x_k(1 + \mathfrak{a}_i + \mathfrak{a}_k) + x_j\mathfrak{a}_j + \mathfrak{a} = x_k + x_j\mathfrak{a}_j + \mathfrak{a}. \tag{16}$$

Again all essential variables of f except for x_i occur in $\mathfrak{p}_{i\leftarrow k}$, and so $\operatorname{gap} f = 1$.

Subcase 1.2.3. Assume that both x_i and x_k occur in $\mathfrak{p}_{j\leftarrow k}$ but there is a variable x_l occurring in $\mathfrak{a}_i$ and in $\mathfrak{a}_j$ but neither in $\mathfrak{a}_k$ nor in $\mathfrak{a}$ such that x_l does not occur in $\mathfrak{p}_{j\leftarrow k}$ (due to some cancelling terms in $\mathfrak{a}_i$ and $\mathfrak{a}_j$). Write $\mathfrak{a}_i = x_l + \mathfrak{a}'_i$, $\mathfrak{a}_j = x_l + \mathfrak{a}'_j$, and consider

$$\begin{aligned} \mathfrak{p}_{j\leftarrow l} &= x_i x_l + x_i x_k + x_l x_k + x_i x_l + x_i\mathfrak{a}'_i + x_l + x_l\mathfrak{a}'_j + x_k\mathfrak{a}_k + \mathfrak{a} \\ &= x_i x_k + x_l x_k + x_i\mathfrak{a}'_i + x_l + x_l\mathfrak{a}'_j + x_k\mathfrak{a}_k + \mathfrak{a}. \end{aligned} \tag{17}$$

Every essential variable of f except for x_j occurs in $\mathfrak{p}_{j\leftarrow l}$, and so $\operatorname{gap} f = 1$.

Case 2. Assume then that $\mathfrak{p}$ is of the form

$$\mathfrak{p} = x_i x_j + x_i x_k\mathfrak{a}_{ik} + x_i\mathfrak{a}_i + x_j\mathfrak{a}_j + x_k\mathfrak{a}_k + \mathfrak{a}, \tag{18}$$

where $\mathfrak{a}_{ik}$ is a polynomial of degree 0; $\mathfrak{a}_i$, $\mathfrak{a}_j$, $\mathfrak{a}_k$ are polynomials of degree at most 1; and $\mathfrak{a}$ is a polynomial of degree at most 2 such that variables x_i, x_j, x_k do not occur in $\mathfrak{a}_{ik}$, $\mathfrak{a}_i$, $\mathfrak{a}_j$, $\mathfrak{a}_k$, $\mathfrak{a}$. Note that $\mathfrak{a}_{ik}$ and $\mathfrak{a}_k$ cannot both be 0, for otherwise x_k would not occur in $\mathfrak{p}$. Consider

$$\mathfrak{p}_{j\leftarrow i} = x_i(1 + \mathfrak{a}_i + \mathfrak{a}_j) + x_i x_k\mathfrak{a}_{ik} + x_k\mathfrak{a}_k + \mathfrak{a}. \tag{19}$$

By the above observation that $\mathfrak{a}_{ik}$ and $\mathfrak{a}_k$ are not both 0, x_k occurs in $\mathfrak{p}_{j\leftarrow i}$. If all essential variables of f except for x_j occur in $\mathfrak{p}_{j\leftarrow i}$, then $\operatorname{gap} f = 1$ and we are done. Otherwise we distinguish between two cases.

Subcase 2.1. Assume that x_i does not occur in $\mathfrak{p}_{j\leftarrow i}$. In this case $\mathfrak{a}_j = \mathfrak{a}_i + 1$, $\mathfrak{a}_{ik} = 0$, and $\mathfrak{a}_k \neq 0$. Consider

$$\begin{aligned} \mathfrak{p}_{i\leftarrow k} &= x_j x_k + x_k\mathfrak{a}_{ik} + x_k\mathfrak{a}_i + x_j\mathfrak{a}_j + x_k\mathfrak{a}_k + \mathfrak{a} \\ &= x_j x_k + x_k(\mathfrak{a}_i + \mathfrak{a}_k) + x_j + x_j\mathfrak{a}_i + \mathfrak{a}. \end{aligned} \tag{20}$$

Both x_j and x_k occur in $\mathfrak{p}_{i\leftarrow k}$, because the term $x_j x_k$ cannot be cancelled. If any term of $\mathfrak{a}_i$ is cancelled by a term of $\mathfrak{a}_k$, it still remains in $x_j\mathfrak{a}_i$. Thus, all essential variables of f except for x_i occur in $\mathfrak{p}_{i\leftarrow k}$, and hence $\operatorname{gap} f = 1$.

Subcase 2.2. Assume that x_i occurs in $\mathfrak{p}_{j\leftarrow i}$ but there is a variable x_l occurring in $\mathfrak{a}_i$ and $\mathfrak{a}_j$ but not in $\mathfrak{a}_{ik}$, $\mathfrak{a}_k$, nor in $\mathfrak{a}$ such that x_l does not occur in $\mathfrak{p}_{j\leftarrow i}$ (due to some cancelling terms in $\mathfrak{a}_i$ and $\mathfrak{a}_j$). Consider

$$\mathfrak{p}_{k\leftarrow l} = x_i x_j + x_i x_l\mathfrak{a}_{ik} + x_i\mathfrak{a}_i + x_j\mathfrak{a}_j + x_l\mathfrak{a}_k + \mathfrak{a}. \tag{21}$$

If $\mathfrak{a}_{ik} = 1$, then the terms $x_i x_l$ in $x_i \mathfrak{a}_i$ and in $x_i x_l \mathfrak{a}_{ik}$ cancel each other. These are the only terms that may be cancelled out. Nevertheless, x_l occurs also in $\mathfrak{a}_j$, and so all essential variables of f except for x_k occur in $\mathfrak{p}_{k \leftarrow l}$. Therefore $\operatorname{gap} f = 1$ also in this case. □

Proof of Theorem 4.1. Denote the Zhegalkin polynomial of f by $\mathfrak{p}$. It is straightforward to verify that if $\mathfrak{p}$ has one of the special forms listed in the statement of the theorem, then f does not have a variable identification minor of essential arity $\operatorname{ess} f - 1$ but it has one of essential arity $\operatorname{ess} f - 2$. For the converse implication, we will prove by induction on $\operatorname{ess} f$ that if $\mathfrak{p}$ is not of any of the special forms, then there is a variable identification minor g of f such that $\operatorname{ess} g = \operatorname{ess} f - 1$, i.e., f has arity gap 1.

If $\operatorname{ess} f = 2$ and $\mathfrak{p}$ is not of any of the special forms, then $\mathfrak{p} = x_i x_j + c$ or $\mathfrak{p} = x_i x_j + x_i + x_j + c$ where $c \in \{0, 1\}$, and in both cases $\mathfrak{p}_{j \leftarrow i} = x_i + c$. In this case $\operatorname{gap} f = 1$.

If $\operatorname{ess} f = 3$, then $\mathfrak{p}$ has one of the following forms:

- $x_i x_j x_k + x_i x_j + x_i x_k + x_j x_k + a_i x_i + a_j x_j + a_k x_k + c$,
- $x_i x_j x_k + x_i x_k + x_j x_k + a_i x_i + a_j x_j + a_k x_k + c$,
- $x_i x_j x_k + x_i x_j + a_i x_i + a_j x_j + a_k x_k + c$,
- $x_i x_j x_k + a_i x_i + a_j x_j + a_k x_k + c$,
- $x_i x_j + x_i x_k + x_j x_k + x_k + c$,
- $x_i x_j + x_i x_k + x_j x_k + x_i + x_j + x_k + c$,
- $x_i x_j + x_i x_k + a_i x_i + a_j x_j + a_k x_k + c$,
- $x_i x_k + a_i x_i + a_j x_j + a_k x_k + c$,

where $a_i, a_j, a_k, c \in \{0, 1\}$. It is easy to verify that in each case $\mathfrak{p}_{j \leftarrow i}$ contains the term $x_i x_k$, and hence both x_i and x_k are essential in $f_{j \leftarrow i}$, and so $\operatorname{gap} f = 1$.

For the sake of induction, assume then that the claim holds for $2 \leq \operatorname{ess} f < n$, $n \geq 4$. Consider the case that $\operatorname{ess} f = n$. Since the case where $\deg \mathfrak{p} = 1$ is ruled out by the assumption that $\mathfrak{p}$ does not have any of the special forms and the case where $\deg \mathfrak{p} = 2$ is settled by Lemma 4.2, we can assume that $\deg \mathfrak{p} \geq 3$. Choose a variable x_m from a term of the highest possible degree in $\mathfrak{p}$, and write

$$\mathfrak{p} = x_m \mathfrak{q} + \mathfrak{r}, \tag{22}$$

where the polynomials $\mathfrak{q}$ and $\mathfrak{r}$ do not contain x_m. We clearly have that $\deg \mathfrak{q} = \deg \mathfrak{p} - 1$, and $\mathfrak{q}$ and $\mathfrak{r}$ represent functions with less than n essential variables. Of course, every essential variable of f except for x_m occurs in $\mathfrak{q}$ or $\mathfrak{r}$. We have three different cases to consider, depending on the comparability under inclusion of the sets of variables occurring in $\mathfrak{q}$ and $\mathfrak{r}$.

Case 1. Assume that there is a variable x_i that occurs in $\mathfrak{q}$ but does not occur in $\mathfrak{r}$, and there is a variable x_j that occurs in $\mathfrak{r}$ but does not occur in $\mathfrak{q}$. Write

$$\mathfrak{q} = x_i \mathfrak{q}' + \mathfrak{q}'', \qquad \mathfrak{r} = x_j \mathfrak{r}' + \mathfrak{r}'', \tag{23}$$

where $\mathfrak{q}'$, $\mathfrak{q}''$, $\mathfrak{r}'$, $\mathfrak{r}''$ do not contain x_i, x_j. Then

$$\mathfrak{p} = x_m x_i \mathfrak{q}' + x_m \mathfrak{q}'' + x_j \mathfrak{r}' + \mathfrak{r}'', \tag{24}$$

and we have that

$$\mathfrak{p}_{j \leftarrow i} = x_m x_i \mathfrak{q}' + x_m \mathfrak{q}'' + x_i \mathfrak{r}' + \mathfrak{r}'', \tag{25}$$

where no terms can cancel. Hence all essential variables of f except for x_j are essential in $f_{j \leftarrow i}$ and so $\operatorname{gap} f = 1$.

Case 2. Assume that every variable occurring in $\mathfrak{r}$ occurs in $\mathfrak{q}$. In this case $\mathfrak{q}$ represents a function q of essential arity $\operatorname{ess} f - 1$, containing all essential variables of f except for x_m. We also have that $\deg \mathfrak{q} = \deg \mathfrak{p} - 1 \geq 2$.

Subcase 2.1. If $\operatorname{ess} f \geq 5$, then $\operatorname{ess} q \geq 4$, and we can apply the inductive hypothesis, which tells us that there are variables x_i and x_j such that $\operatorname{ess} q_{i \leftarrow j} = \operatorname{ess} q - 1$. Hence $f_{i \leftarrow j}$ is represented by the polynomial $\mathfrak{p}_{i \leftarrow j} = x_m \mathfrak{q}_{i \leftarrow j} + \mathfrak{r}_{i \leftarrow j}$, and all essential variables of f except for x_i occur in $\mathfrak{p}_{i \leftarrow j}$, since no terms can cancel between $x_m \mathfrak{q}_{i \leftarrow j}$ and $\mathfrak{r}_{i \leftarrow j}$. Thus $\operatorname{gap} f = 1$.

Subcase 2.2. If $\operatorname{ess} f = 4$, then $\operatorname{ess} q = 3$, and we can apply the inductive hypothesis as above unless $\mathfrak{q} = x_i x_j + x_i x_k + x_j x_k + c$ or $\mathfrak{q} = x_i x_j + x_i x_k + x_j x_k + x_i + x_j + c$. If this is the case, consider first the case where $\mathfrak{q}$ contains a variable $x_l \in \{x_i, x_j, x_k\}$ that does not occur in $\mathfrak{r}$. Consider then

$$\mathfrak{p}_{m \leftarrow l} = x_l \mathfrak{q} + \mathfrak{r}. \tag{26}$$

Then $x_l \mathfrak{q}$ contains the term $x_i x_j x_k$, which cannot be cancelled. Namely, all other terms of $x_l \mathfrak{q}$ have degree at most 2, and since there are at most two variables occurring in $\mathfrak{r}$, the terms of $\mathfrak{r}$ also have degree at most 2. Thus, all variables of f except for x_m occur in $\mathfrak{p}_{m \leftarrow l}$, and so the arity gap of f is 1.

Consider then the case that $\mathfrak{q}$ and $\mathfrak{r}$ contain the same variables, i.e., x_i, x_j, x_k. If $\deg \mathfrak{r} \leq 2$, then it is easily seen that $\mathfrak{p}_{m \leftarrow i}$ contains the term $x_i x_j x_k$, and all essential variables of f except for x_m are essential in $f_{m \leftarrow i}$. Otherwise, we can apply the inductive hypothesis on the function r represented by $\mathfrak{r}$ and we obtain variables x_α and x_β such that $\operatorname{ess} r_{\alpha \leftarrow \beta} = \operatorname{ess} r - 1$. It can be easily verified that no identification of variables brings $\mathfrak{q}$ into the zero polynomial, so x_m and two other variables will occur in $\mathfrak{p}_{\alpha \leftarrow \beta} = x_m \mathfrak{q}_{\alpha \leftarrow \beta} + \mathfrak{r}_{\alpha \leftarrow \beta}$. We have that $\operatorname{gap} f = 1$ also in this case.

Case 3. Assume that every variable occurring in $\mathfrak{q}$ occurs in $\mathfrak{r}$ but there is a variable x_l that occurs in $\mathfrak{r}$ but does not occur in $\mathfrak{q}$. If $\deg \mathfrak{r} = 1$, then $\mathfrak{r} = x_l + \mathfrak{r}'$ where $\mathfrak{r}'$ does not contain x_l. Then $\mathfrak{p}_{m \leftarrow l} = x_l \mathfrak{q} + x_l + \mathfrak{r}'$, where the only term that may cancel out is x_l, and this happens if $\mathfrak{q}$ has a constant term 1. Nevertheless, x_l occurs in $\mathfrak{r}_{m \leftarrow l}$ because $\deg \mathfrak{q} \geq 2$. Of course, all other essential variables of f except for x_m also occur in $\mathfrak{p}_{m \leftarrow l}$, so $\operatorname{gap} f = 1$. We may thus assume that $\deg \mathfrak{r} \geq 2$.

Subcase 3.1. Assume first that $\operatorname{ess} f = 4$ (in which case $\mathfrak{r}$ contains three variables and $\mathfrak{q}$ contains at most two variables) and $\mathfrak{r} = x_i x_j + x_i x_k + x_j x_k + c$ or $\mathfrak{r} = x_i x_j + x_i x_k + x_j x_k + x_i + x_j + c$. Since we assume that $\deg \mathfrak{p} \geq 3$, we have that $\deg \mathfrak{q} \geq 2$ and hence $\mathfrak{q}$ contains at least two variables. Thus exactly two variables occur in $\mathfrak{q}$ and so also $\deg \mathfrak{q} = 2$. Then $\mathfrak{q} = x_\alpha x_\beta + b_1 x_\alpha + b_2 x_\beta + d$ where $\alpha, \beta \in \{i, j, k\}$ and $b_1, b_2, d \in \{0, 1\}$. Let $\gamma \in \{i, j, k\} \setminus \{\alpha, \beta\}$. Then $\mathfrak{p}_{m \leftarrow \gamma}$ contains the term $x_i x_j x_k$, and hence all essential variables of f except for x_m occur in $\mathfrak{p}_{m \leftarrow \gamma}$, and so $\operatorname{gap} f = 1$.

Subcase 3.2. Assume then that $\operatorname{ess} f > 4$ or $\operatorname{ess} f = 4$ but $\mathfrak{r}$ does not have any of the special forms. In this case we can apply the inductive hypothesis on the function r represented by $\mathfrak{r}$. Let x_i and x_j be such that $\operatorname{ess} r_{j \leftarrow i} = \operatorname{ess} r - 1$. If $\mathfrak{q}_{j \leftarrow i} \neq 0$, then x_m and all other essential variables of f except for x_j occur in $\mathfrak{p}_{j \leftarrow i}$, and we are done—the arity gap of f is 1. We may thus assume that $\mathfrak{q}_{j \leftarrow i} = 0$. Write $\mathfrak{q}$ and $\mathfrak{r}$ in the form

$$\mathfrak{q} = x_i x_j \mathfrak{a}_1 + x_i \mathfrak{a}_2 + x_j \mathfrak{a}_3 + \mathfrak{a}_4, \tag{27}$$

$$\mathfrak{r} = x_i x_j \mathfrak{b}_1 + x_i \mathfrak{b}_2 + x_j \mathfrak{b}_3 + \mathfrak{b}_4, \tag{28}$$

where the polynomials $\mathfrak{a}_1$, $\mathfrak{a}_2$, $\mathfrak{a}_3$, $\mathfrak{a}_4$, $\mathfrak{b}_1$, $\mathfrak{b}_2$, $\mathfrak{b}_3$, $\mathfrak{b}_4$ do not contain x_i, x_j. Define the polynomials $\mathfrak{q}_1, \ldots, \mathfrak{q}_7$ as follows (cf. the proof of Theorem 4 in Salomaa [14]):

$\mathfrak{q}_1$ consists of the terms common to $\mathfrak{a}_1$, $\mathfrak{a}_2$, and $\mathfrak{a}_3$.

$\mathfrak{q}_i$, $i = 2, 3$, consists of those terms common to $\mathfrak{a}_1$ and $\mathfrak{a}_i$ which are not in $\mathfrak{q}_1$.

$\mathfrak{q}_4$ consists of those terms common to $\mathfrak{a}_2$ and $\mathfrak{a}_3$ which are not in $\mathfrak{q}_1$.

$\mathfrak{q}_{4+i}$, $i = 1, 2, 3$, consists of the remaining terms in $\mathfrak{a}_i$.

Define the polynomials and $\mathfrak{r}_1, \ldots, \mathfrak{r}_7$ similarly in terms of the $\mathfrak{b}_i$'s. Note that for any $i \neq j$, $\mathfrak{q}_i$ and $\mathfrak{q}_j$ do not have any terms in common, and similarly $\mathfrak{r}_i$ and $\mathfrak{r}_j$ do not have any terms in common. Hence,

$$\begin{aligned}\mathfrak{q} &= x_i x_j(\mathfrak{q}_1 + \mathfrak{q}_2 + \mathfrak{q}_3 + \mathfrak{q}_5) + \\ &\quad x_i(\mathfrak{q}_1 + \mathfrak{q}_2 + \mathfrak{q}_4 + \mathfrak{q}_6) + \\ &\quad x_j(\mathfrak{q}_1 + \mathfrak{q}_3 + \mathfrak{q}_4 + \mathfrak{q}_7) + \mathfrak{a}_4, \end{aligned}\tag{29}$$

$$\begin{aligned}\mathfrak{r} &= x_i x_j(\mathfrak{r}_1 + \mathfrak{r}_2 + \mathfrak{r}_3 + \mathfrak{r}_5) + \\ &\quad x_i(\mathfrak{r}_1 + \mathfrak{r}_2 + \mathfrak{r}_4 + \mathfrak{r}_6) + \\ &\quad x_j(\mathfrak{r}_1 + \mathfrak{r}_3 + \mathfrak{r}_4 + \mathfrak{r}_7) + \mathfrak{b}_4. \end{aligned}\tag{30}$$

Identification of x_i with x_j yields

$$\mathfrak{q}_{j\leftarrow i} = x_i(\mathfrak{q}_1 + \mathfrak{q}_5 + \mathfrak{q}_6 + \mathfrak{q}_7) + \mathfrak{a}_4, \tag{31}$$

$$\mathfrak{r}_{j\leftarrow i} = x_i(\mathfrak{r}_1 + \mathfrak{r}_5 + \mathfrak{r}_6 + \mathfrak{r}_7) + \mathfrak{b}_4. \tag{32}$$

Since we are assuming that $\mathfrak{q}_{j\leftarrow i} = 0$, we have that $\mathfrak{q}_1 = \mathfrak{q}_5 = \mathfrak{q}_6 = \mathfrak{q}_7 = \mathfrak{a}_4 = 0$. On the other hand, $\mathfrak{q} \neq 0$, so $\mathfrak{q}_2$, $\mathfrak{q}_3$, $\mathfrak{q}_4$ are not all zero. Thus

$$\mathfrak{q} = x_i x_j(\mathfrak{q}_2 + \mathfrak{q}_3) + x_i(\mathfrak{q}_2 + \mathfrak{q}_4) + x_j(\mathfrak{q}_3 + \mathfrak{q}_4). \tag{33}$$

All essential variables of f except for x_j are contained in $\mathfrak{r}_{j\leftarrow i}$.

Subcase 3.2.1. Assume that there is a variable x_t occurring in $\mathfrak{b}_4$ that does not occur in $\mathfrak{r}_1$, $\mathfrak{r}_5$, $\mathfrak{r}_6$, $\mathfrak{r}_7$. Consider

$$\mathfrak{p}_{m\leftarrow t} = x_t\mathfrak{q} + \mathfrak{r} = x_l\mathfrak{q} + x_i x_j\mathfrak{b}_1 + x_i\mathfrak{b}_2 + x_j\mathfrak{b}_3 + \mathfrak{b}_4. \tag{34}$$

Cancelling may only happen between a term of $x_t\mathfrak{q}$ and a term of $\mathfrak{r}$. No term of $\mathfrak{b}_4$ can be cancelled, because every term of $x_t\mathfrak{q}$ contains x_i or x_j but the terms of $\mathfrak{b}_4$ do not contain either. The variables that do not occur in $\mathfrak{b}_4$ occur in some terms of $\mathfrak{b}_1$, $\mathfrak{b}_2$, $\mathfrak{b}_3$ that do not contain x_t. Thus, all essential variables of f except for x_m occur in $\mathfrak{p}_{m\leftarrow t}$, and so in this case f has arity gap 1.

Subcase 3.2.2. Assume that all variables of $\mathfrak{r}$ except for x_i, x_j occur already in $\mathfrak{r}_1 + \mathfrak{r}_5 + \mathfrak{r}_6 + \mathfrak{r}_7$. Consider

$$\begin{aligned}\mathfrak{p}_{m\leftarrow i} &= x_i x_j(\mathfrak{q}_2 + \mathfrak{q}_4 + \mathfrak{r}_1 + \mathfrak{r}_2 + \mathfrak{r}_3 + \mathfrak{r}_5) + \\ &\quad x_i(\mathfrak{q}_2 + \mathfrak{q}_4 + \mathfrak{r}_1 + \mathfrak{r}_2 + \mathfrak{r}_4 + \mathfrak{r}_6) + \\ &\quad x_j(\mathfrak{r}_1 + \mathfrak{r}_3 + \mathfrak{r}_4 + \mathfrak{r}_7) + \mathfrak{b}_4. \end{aligned}\tag{35}$$

Subcase 3.2.2.1. Assume first that x_i does not occur in $\mathfrak{p}_{m\leftarrow i}$ in (35). Then

$$\mathfrak{q}_2 + \mathfrak{q}_4 + \mathfrak{r}_1 + \mathfrak{r}_2 + \mathfrak{r}_3 + \mathfrak{r}_5 = 0, \tag{36}$$

$$\mathfrak{q}_2 + \mathfrak{q}_4 + \mathfrak{r}_1 + \mathfrak{r}_2 + \mathfrak{r}_4 + \mathfrak{r}_6 = 0, \tag{37}$$

and since the $\mathfrak{r}_i$'s do not have terms in common, we have that

$$\mathfrak{r}_1 + \mathfrak{r}_2 = \mathfrak{q}_2 + \mathfrak{q}_4, \qquad \mathfrak{r}_3 = \mathfrak{r}_4 = \mathfrak{r}_5 = \mathfrak{r}_6 = 0. \tag{38}$$

Then all variables of $\mathfrak{r}$ except for x_i, x_j occur already in $\mathfrak{r}_1 + \mathfrak{r}_7$. Consider

$$\begin{aligned} \mathfrak{p}_{m\leftarrow j} &= x_i x_j(\mathfrak{q}_3 + \mathfrak{q}_4 + \mathfrak{r}_1 + \mathfrak{r}_2 + \mathfrak{r}_3 + \mathfrak{r}_5) + \\ &\quad x_i(\mathfrak{r}_1 + \mathfrak{r}_2 + \mathfrak{r}_4 + \mathfrak{r}_6) + \\ &\quad x_j(\mathfrak{q}_3 + \mathfrak{q}_4 + \mathfrak{r}_1 + \mathfrak{r}_3 + \mathfrak{r}_4 + \mathfrak{r}_7) + \mathfrak{b}_4 \\ &= x_i x_j(\mathfrak{q}_2 + \mathfrak{q}_3) + \\ &\quad x_i(\mathfrak{r}_1 + \mathfrak{r}_2) + \\ &\quad x_j(\mathfrak{q}_2 + \mathfrak{q}_3 + \mathfrak{r}_2 + \mathfrak{r}_7) + \mathfrak{b}_4. \end{aligned} \tag{39}$$

All variables of $\mathfrak{r}_1$ are there on the fifth line of (39). If a term of $\mathfrak{r}_7$ is cancelled by a term of $\mathfrak{q}_2 + \mathfrak{q}_3$ on the sixth line, it still remains on the fourth line, so all variables of $\mathfrak{r}_7$ are also there. We still need to verify that the variables x_i and x_j are not cancelled out from (39). If $\mathfrak{q}_2 + \mathfrak{q}_3 \neq 0$ then we are done. Assume then that $\mathfrak{q}_2 + \mathfrak{q}_3 = 0$, in which case $\mathfrak{q}_4 \neq 0$. Since

$$\mathfrak{r}_1 + \mathfrak{r}_2 + \mathfrak{r}_4 + \mathfrak{r}_6 = \mathfrak{r}_1 + \mathfrak{r}_2 = \mathfrak{q}_2 + \mathfrak{q}_4 = \mathfrak{q}_4 \neq 0, \tag{40}$$

we have x_i in (39). Since

$$\mathfrak{q}_3 + \mathfrak{q}_4 + \mathfrak{r}_1 + \mathfrak{r}_3 + \mathfrak{r}_4 + \mathfrak{r}_7 = \mathfrak{q}_4 + \mathfrak{r}_1 + \mathfrak{r}_7 \tag{41}$$

and $\mathfrak{r}_1 + \mathfrak{r}_7$ contains all variables of $\mathfrak{r}$ except for x_i, x_j, but $\mathfrak{q}_4$ does not, $\mathfrak{q}_4 + \mathfrak{r}_1 + \mathfrak{r}_7 \neq 0$, so we also have x_j in (39). Thus, the arity gap of f equals 1 in this case.

Subcase 3.2.2.2. Assume then that x_i occurs in $\mathfrak{p}_{m\leftarrow i}$ in (35). Nothing cancels out on the third line of (35), and therefore the variables of $\mathfrak{r}_1$ and $\mathfrak{r}_7$ occur in $\mathfrak{p}_{m\leftarrow i}$. Terms of $\mathfrak{r}_5$ may be cancelled out by terms of $\mathfrak{q}_2 + \mathfrak{q}_4$ on the first line of (35) but such terms will remain on the second line. Thus the variables of $\mathfrak{r}_5$ occur in $\mathfrak{p}_{m\leftarrow i}$. A similar argument shows that the variables of $\mathfrak{r}_6$ also occur in $\mathfrak{p}_{m\leftarrow i}$. In order for f to have arity gap 1, we still need to verify that x_j occurs in $\mathfrak{p}_{m\leftarrow i}$. If $\mathfrak{q}_2 + \mathfrak{q}_4 + \mathfrak{r}_1 + \mathfrak{r}_2 + \mathfrak{r}_3 + \mathfrak{r}_5 \neq 0$, then we are done. We may thus assume that

$$\mathfrak{q}_2 + \mathfrak{q}_4 + \mathfrak{r}_1 + \mathfrak{r}_2 + \mathfrak{r}_3 + \mathfrak{r}_5 = 0. \tag{42}$$

By the assumption that x_i occurs in $\mathfrak{p}_{m\leftarrow i}$, the second line of (35) does not vanish, i.e.,

$$0 \neq \mathfrak{q}_2 + \mathfrak{q}_4 + \mathfrak{r}_1 + \mathfrak{r}_2 + \mathfrak{r}_4 + \mathfrak{r}_6 = \mathfrak{r}_3 + \mathfrak{r}_4 + \mathfrak{r}_5 + \mathfrak{r}_6. \tag{43}$$

If the third line of (35) does not vanish either, i.e., $\mathfrak{r}_1 + \mathfrak{r}_3 + \mathfrak{r}_4 + \mathfrak{r}_7 \neq 0$, then we have both x_i and x_j and we are done. We may thus assume that $\mathfrak{r}_1 + \mathfrak{r}_3 + \mathfrak{r}_4 + \mathfrak{r}_7 = 0$, i.e., $\mathfrak{r}_1 = \mathfrak{r}_3 = \mathfrak{r}_4 = \mathfrak{r}_7 = 0$. Then all variables of $\mathfrak{r}$ except for x_i, x_j occur already in $\mathfrak{r}_5 + \mathfrak{r}_6$. Equation (42) implies that $\mathfrak{r}_2 + \mathfrak{r}_5 = \mathfrak{q}_2 + \mathfrak{q}_4$. Consider

$$\begin{aligned} \mathfrak{p}_{m\leftarrow j} &= x_i x_j(\mathfrak{q}_3 + \mathfrak{q}_4 + \mathfrak{r}_1 + \mathfrak{r}_2 + \mathfrak{r}_3 + \mathfrak{r}_5) + \\ &\quad x_i(\mathfrak{r}_1 + \mathfrak{r}_2 + \mathfrak{r}_4 + \mathfrak{r}_6) + \\ &\quad x_j(\mathfrak{q}_3 + \mathfrak{q}_4 + \mathfrak{r}_1 + \mathfrak{r}_3 + \mathfrak{r}_4 + \mathfrak{r}_7) + \mathfrak{b}_4 \\ &= x_i x_j(\mathfrak{q}_2 + \mathfrak{q}_3) + \\ &\quad x_i(\mathfrak{q}_2 + \mathfrak{q}_4 + \mathfrak{r}_5 + \mathfrak{r}_6) + \\ &\quad x_j(\mathfrak{q}_3 + \mathfrak{q}_4) + \mathfrak{b}_4. \end{aligned} \tag{44}$$

Assume first that $\mathfrak{q}_2 + \mathfrak{q}_3 = 0$, in which case $\mathfrak{q}_4 \neq 0$. If a term of $\mathfrak{r}_5 + \mathfrak{r}_6$ is cancelled by a term of $\mathfrak{q}_4$ on the fifth line of (44), it will still remain on the sixth line. Therefore we have in $\mathfrak{p}_{m\leftarrow j}$ all variables of $\mathfrak{r}$ except for x_i and x_j. Since $\mathfrak{r}_5 + \mathfrak{r}_6$ contains all variables of $\mathfrak{r}$ except for x_i, x_j but $\mathfrak{q}_2 + \mathfrak{q}_4 = \mathfrak{q}_4$ does not, the fifth line of (44) does not vanish, and so we have x_i. We also have x_j because $\mathfrak{q}_3 + \mathfrak{q}_4 = \mathfrak{q}_4 \neq 0$ on the sixth line. In this case f has arity gap 1.

Assume then that $\mathfrak{q}_2 + \mathfrak{q}_3 \neq 0$. Then the fourth line of (44) does not vanish and both x_i and x_j occur in $\mathfrak{p}_{m \leftarrow j}$. If any term of $\mathfrak{r}_5 + \mathfrak{r}_6$ is cancelled by a term of $\mathfrak{q}_2$ on the fifth line of (44), it still remains on the fourth line, and if it is cancelled by a term of $\mathfrak{q}_4$, it remains on the sixth line. Thus all variables of $\mathfrak{r}$ occur in $\mathfrak{p}_{m \leftarrow j}$, and f has arity gap 1 again. This completes the proof of Theorem 4.1. $\square$

5. Concluding remarks

We do not know whether the upper bound on arity gap given by Theorem 3.1 is sharp. For base sets A with $k \geq 3$ elements, we do not know whether there exists an operation f on A with $\operatorname{ess} f \geq k+1$ and $\operatorname{gap} f \geq 3$. We know that for all $k \geq 2$, there are operations on a k-element set A with arity gap 2. Consider for instance the quasi-linear functions of Burle [1]. A function f is *quasi-linear* if it has the form

$$f = g(h_1(x_1) \oplus h_2(x_2) \oplus \cdots \oplus h_n(x_n)), \tag{45}$$

where $h_1, \ldots, h_n : A \to \{0, 1\}$, $g : \{0, 1\} \to A$ are arbitrary mappings and $\oplus$ denotes addition modulo 2. It is easy to verify that if those h_i's that are nonconstant coincide (and g is not a constant map), then f has arity gap 2.

In general, if there is an operation f on a k-element set A with with $\operatorname{gap} f = m$, then there are operations of arity gap m on all sets B of at least k elements. Namely, it is easy to see that any operation g on B of the form

$$g = \phi(f(\gamma(x_1), \gamma(x_2), \ldots, \gamma(x_n))), \tag{46}$$

where $\gamma : B \to A$ is surjective and $\phi : A \to B$ is injective, satisfies $\operatorname{ess} g = \operatorname{ess} f$ and $\operatorname{gap} g = \operatorname{gap} f$.

References

[1] G. A. Burle, The classes of k-valued logics containing all one-variable functions, *Diskretnyi Analiz* **10** (1967) 3–7 (in Russian).

[2] K. N. Čimev, *Separable Sets of Arguments of Functions,* Studies 180/1986 (Computer and Automation Institute, Hungarian Academy of Sciences, Budapest, 1986).

[3] M. Couceiro, On the lattice of equational classes of Boolean functions and its closed intervals, Technical report A367, University of Tampere, 2006.

[4] M. Couceiro and M. Pouzet, On a quasi-ordering on Boolean functions, arXiv:math.CO/0601218, 2006.

[5] R. O. Davies, Two theorems on essential variables, *J. London Math. Soc.* **41** (1966) 333–335.

[6] K. Denecke and J. Koppitz, Essential variables in hypersubstitutions, *Algebra Universalis* **46** (2001) 443–454.

[7] A. Ehrenfeucht, J. Kahn, R. Maddux and J. Mycielski, On the dependence of functions on their variables, *J. Combin. Theory Ser. A* **33** (1982) 106–108.

[8] O. Ekin, S. Foldes, P. L. Hammer and L. Hellerstein, Equational characterizations of Boolean function classes, *Discrete Math.* **211** (2000) 27–51.

[9] A. Feigelson and L. Hellerstein, The forbidden projections of unate functions, *Discrete Appl. Math.* **77** (1997) 221–236.

[10] E. Lehtonen, Descending chains and antichains of the unary, linear, and monotone subfunction relations, *Order* **23** (2006) 129–142.

[11] D. E. Muller, Application of Boolean algebra to switching circuit design and to error correction, *IRE Trans. Electron. Comput.* **3**(3) (1954) 6–12.

[12] N. Pippenger, Galois theory for minors of finite functions, *Discrete Math.* **254** (2002) 405–419.

[13] I. S. Reed, A class of multiple-error-correcting codes and the decoding scheme, *IRE Trans. Inf. Theory* **4**(4) (1954) 38–49.

[14] A. Salomaa, On essential variables of functions, especially in the algebra of logic, *Ann. Acad. Sci. Fenn. Ser. A I. Math.* **339** (1963) 3–11.

[15] C. Wang, Boolean minors, *Discrete Math.* **141** (1991) 237–258.

[16] S. V. Yablonski, Functional constructions in a k-valued logic, *Tr. Mat. Inst. Steklova* **51** (1958) 5–142 (in Russian).

[17] I. I. Zhegalkin, On the calculation of propositions in symbolic logic, *Mat. Sb.* **34** (1927) 9–28 (in Russian).

[18] I. E. Zverovich, Characterizations of closed classes of Boolean functions in terms of forbidden subfunctions and Post classes, *Discrete Appl. Math.* **149** (2005) 200–218.

(M. Couceiro) Department of Mathematics, Statistics and Philosophy, University of Tampere, FI-33014 Tampereen yliopisto, Finland

E-mail address: `miguel.couceiro@uta.fi`

(E. Lehtonen) Institute of Mathematics, Tampere University of Technology, P.O. Box 553, FI-33101 Tampere, Finland

E-mail address: `erkko.lehtonen@tut.fi`