Serdica

Mathematical Journal

Сердика

Математическо списание

Serdica Math. J. **42** (2016), 277–286

Serdica
Mathematical Journal

Bulgarian Academy of Sciences
Institute of Mathematics and Informatics

DELETION-CORRECTING CODES AND DOMINANT VECTORS

Emil Kolev

Communicated by L. Storme

Abstract. In this paper we describe all pairs of binary vectors $(\mathbf{u}, \mathbf{v})$ such that the set of vectors obtained by t deletions in $\mathbf{v}$ is a subset of the set of vectors obtained by t deletions in $\mathbf{u}$ for $t = 1, 2$. Such pairs play an important role for finding the value of $L_2(n, t)$, the maximum cardinality of binary t-deletion-correcting codes of length n

1. Introduction. When a binary message is transmitted through a noisy channel some of its symbols may change. The receiver needs reliable tools for recovering the message. This is done by adding some extra symbols (called check symbols) to the original message and the result is a longer message. The set of all such messages is called an error-correcting code. One of the main goals of coding theory is finding codes with good error-correcting capabilities.

Another possible distortion of the message is the loss of some of its symbols or insertion of some extra symbols. In this case the receiver gets shorter or

2010 *Mathematics Subject Classification:* 94B05.
Key words: insertion/deletion codes, Varshamov-Tenengolts codes, multiple insertion/deletion codes.

longer message and he does not know which of the symbols were lost or inserted. Deletion-correcting codes and insertion-correcting codes are designed to correct such deletions or insertions. Levenshtein has shown [5] that deletion-correcting codes and insertion-correcting codes are essentially the same objects. In this paper we consider only deletion-correcting codes. A code is called t-deletion-correcting if it corrects any t deletions. For more information and useful results the reader is referred to [2, 3, 5, 6, 7, 9, 10, 11, 12].

Example 1. Consider the binary code $\mathcal{C} = \{00000, 11111, 00011, 11000, 10101, 01110\}$. For a given codeword we may delete any of its five symbols. As a result we obtain a set of vectors of length 4. Direct verification shows that all six sets obtained from the six codewords are disjoint. Therefore $\mathcal{C}$ is 1-deletion-correcting code.

Definition 1. *The Levenshtein distance $d_L(\mathbf{x}, \mathbf{y})$ of two binary vectors is defined as the minimum number of deletions and insertions needed to transform $\mathbf{x}$ into $\mathbf{y}$.*

For example, $d_L(0100, 110101) = 4$. Note that in the above definition the vectors $\mathbf{x}$ and $\mathbf{y}$ do not need to be of the same length.

Definition 2. Deletion distance *$dd(\mathbf{u}, \mathbf{v})$ between two vectors $\mathbf{u}$ and $\mathbf{v}$ of equal length is defined as one-half of the smallest number of deletions and insertions needed to change $\mathbf{u}$ to $\mathbf{v}$,* [10].

For example, $\mathrm{dd}(00000, 11111) = 5$ whereas $\mathrm{dd}(00011, 10101) = 2$. It is clear that for vectors $\mathbf{u}$ and $\mathbf{v}$ of equal length we have

$$\mathrm{dd}(\mathbf{u}, \mathbf{v}) = \frac{1}{2} d_L(\mathbf{u}, \mathbf{v}).$$

For a given code $\mathcal{C}$ the deletion distance $\mathrm{dd}(\mathcal{C})$ is defined as

$$\mathrm{dd}(\mathcal{C}) = \min\{\mathrm{dd}(\mathbf{u}, \mathbf{v}) \mid \mathbf{u}, \mathbf{v} \in \mathcal{C}, \mathbf{u} \neq \mathbf{v}\}.$$

For any two distinct codewords $\mathbf{u}$ and $\mathbf{v}$ from t-deletion-correcting code $\mathcal{C}$ of length n we have $\mathrm{dd}(\mathbf{u}, \mathbf{v}) > t$ (or, equivalently $d_L(\mathbf{u}, \mathbf{v}) > 2t$).

Denote by $L_2(n, t)$ the maximum cardinality of a binary t-deletion-correcting code $\mathcal{C}$ of length n. A binary t-deletion-correcting code $\mathcal{C}$ of length n and cardinality $L_2(n, t)$ is called optimal.

For a binary vector $\mathbf{u}$ of length n denote by $D_t(\mathbf{u})$ the set of all vectors of length $n - t$ obtained from $\mathbf{u}$ by deleting t entries in $\mathbf{u}$. In other words, $D_t(\mathbf{u})$ contains all subsequences of $\mathbf{u}$ of length $n - t$.

The size of $D_t(\mathbf{u})$ depends on $\mathbf{u}$. The minimal size of $D_t(\mathbf{u})$ equals 1 and is achieved only for $\mathbf{u} = p^n$ where $p \in \{0, 1\}$. The problem of finding the maximal size of $D_t(\mathbf{u})$ is discussed in [1, 8].

A code $\mathcal{C}$ is t-deletion-correcting code if the sets $D_t(\mathbf{u})$ for $u \in \mathcal{C}$ are disjoint. Further, if the sets $D_t(\mathbf{u})$ for $u \in \mathcal{C}$ partition the set F_q^{n-t} then the code is called perfect.

As in the case of error-correcting codes the two main research problems for deletion-correcting codes are:

1. For given n and t find $L_2(n, t)$, the maximum cardinality of a binary t-deletion-correcting code of length n.
2. When $L_2(n, t)$ is known, find all distinct (in some sense) optimal codes.

In general, finding the value of $L_2(n, t)$ is an open problem in coding theory. The efforts are concentrated on specific values of n and t. Tables with known values of $L_2(n, t)$ for different n and t can be found in [3] and [4].

2. Preliminaries.

Any permutation of coordinates of a given code $\mathcal{C}$ does not alter its error-correcting capabilities. On the contrary, for deletion-correcting codes a permutation of coordinates, in general, does not result in a code with the same deletion-correcting properties. Nevertheless, there are two simple observations that describe when two deletion-correcting codes are essentially the same and allow to adopt different notion for equivalence. First, we may read the codewords backwards and second, we may change 0 and 1. This leads to the following

Definition 3. *Two deletion-correcting codes $\mathcal{C}_1$ and $\mathcal{C}_2$ are equivalent if one of the following is true:*

1. $(u_1, u_2, \ldots, u_n) \in \mathcal{C}_1$ *if and only if* $(\overline{u_1}, \overline{u_2}, \ldots, \overline{u_n}) \in \mathcal{C}_2$*;*
2. $(u_1, u_2, \ldots, u_n) \in \mathcal{C}_1$ *if and only if* $(u_n, u_{n-1}, \ldots, u_1) \in \mathcal{C}_2$*;*
3. $(u_1, u_2, \ldots, u_n) \in \mathcal{C}_1$ *if and only if* $(\overline{u_n}, \overline{u_{n-1}}, \ldots, \overline{u_1}) \in \mathcal{C}_2$*.*

Here, for $x \in \{0, 1\}$ the element $\overline{x} \in \{0, 1\}$ is such that $\{x, \overline{x}\} = \{0, 1\}$.

In finding the exact value of $L_2(n, t)$ usually at some stage an exhaustive computer search is performed. As in any computer search a good pruning technique is required. It turns out that when choosing the codewords of optimal deletion-correcting code some of the vectors may be left out.

Definition 4. *We say that a vector $\mathbf{u}$ is t-dominant if there exists a vector $\mathbf{v}$ such that $\mathbf{u} \neq \mathbf{v}$ and $D_t(\mathbf{v}) \subseteq D_t(\mathbf{u})$. Alternatively, $\mathbf{v}$ is t-subordinate of $\mathbf{u}$ when $\mathbf{u}$ is t-dominant over $\mathbf{v}$.*

It is clear that if $\mathbf{u}$ is t-dominant over $\mathbf{v}$ then for any $s > t$ the vector $\mathbf{u}$ is s-dominant over the vector $\mathbf{v}$. If a codeword $\mathbf{u}$ is t-dominant over the vector $\mathbf{v}$ then

$$(\mathcal{C} \setminus \{\mathbf{u}\}) \cup \{\mathbf{v}\}$$

is also t-deletion-correcting code. In other words a dominant codeword may be replaced by its subordinate vector. Hence, in computer search we may exclude all dominant vectors from consideration. Therefore it is important to know all pairs of vectors $(\mathbf{u}, \mathbf{v})$ such that $D_t(\mathbf{v}) \subseteq D_t(\mathbf{u})$.

Furthermore, we may assume that an optimal code $\mathcal{C}$ includes the vectors 0^n and 1^n as codewords. Indeed, for $p \in \{0, 1\}$:

- if $p^{n-t} \in D_t(\mathbf{u})$ for a codeword $\mathbf{u}$ then, as above, replace $\mathbf{u}$ by p^n and
- if $p^{n-t} \notin D_t(\mathbf{u})$ for any codeword $\mathbf{u}$ then $\mathcal{C} \cup \{p^n\}$ is t-deletion-correcting code, i.e. $\mathcal{C}$ is not optimal.

A code $\mathcal{C}$ is called basic if it does not contain dominant vectors. In the lights of the last two definitions the main problems for deletion-correcting codes become:

1. For certain n and t find $L_2(n, t)$;
2. Find all inequivalent basic optimal codes.

3. Results. As explained in the previous section knowing the pairs of t dominant vectors plays an important role in finding $L_2(n, t)$. In what follows we describe all pairs of binary vectors $(\mathbf{u}, \mathbf{v})$ such that $\mathbf{u}$ is t-dominant over $\mathbf{v}$ for $t = 1$ and $t = 2$.

For the two trivial cases $\mathbf{v} = 0^n$, $\mathbf{v} = 1^n$ and for any t we have:

- if $\mathbf{v} = 0^n$ then $\mathbf{u}$ is t-dominant over $\mathbf{v}$ if and only if $\mathbf{u} \neq \mathbf{v}$ and $\mathrm{wt}(\mathbf{u}) \leq t$;
- if $\mathbf{v} = 1^n$ then $\mathbf{u}$ is t-dominant over $\mathbf{v}$ if and only if $\mathbf{u} \neq \mathbf{v}$ and $\mathrm{wt}(\mathbf{u}) \geq n - t$.

In what follows the vector $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ is t-dominant over $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ and $\{p, q\} = \{0, 1\}$. We begin with a useful observation.

Proposition 1. *Let $n \geq 2$ be positive integer. Consider two vectors $\mathbf{x}$ and $\mathbf{y}$ of lengths n and $n-1$ respectively. If any single deletion changes $\mathbf{x}$ to $\mathbf{y}$ then all entries in $\mathbf{x}$ and $\mathbf{y}$ are equal.*

P r o o f. Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_{n-1})$ and choose a positive integer k such that $1 \leq k \leq n-1$. By deleting x_k we have that $x_{k+1} = y_k$

and by deleting x_{k+1} we infer that $x_k = y_k$. Therefore $x_k = x_{k+1} = y_k$ for any $k = 1, 2, \ldots, n-1$ which implies that $x_1 = x_2 = \cdots = x_n = y_1 = y_2 = \cdots = y_{n-1}$. $\square$

Remark. The above proposition is true also for vectors $\mathbf{x}$ and $\mathbf{y}$ of lengths $n \geq 3$ and $n-2$, respectively, when the result of any two deletions in $\mathbf{x}$ is $\mathbf{y}$. The proof is straightforward.

First, we describe all 1-dominant vectors.

Proposition 2. *Let* $\mathbf{u}$ *be* 1*-dominant over* $\mathbf{v}$ *and* $\mathbf{v} \neq 0^n, 1^n$. *Then* $\mathbf{u} = p^{m-1}qpq^{n-m-1}$ *and* $\mathbf{v} = p^m q^{n-m}$ *for some positive integer* m.

Proof. Let $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_n)$. For $n = 2$ the result is trivial, so let $n \geq 3$. Assume first that $u_1 \neq v_1$ and let $v_1 = p$, $u_1 = q$. Any deletion of v_i for $i \geq 2$ results in a vector $\mathbf{w} \in D_1(\mathbf{u})$ with first coordinate $v_1 \neq u_1$. This is possible only if $\mathbf{w}$ is obtained from $\mathbf{u}$ by deleting its first coordinate and $u_2 = v_1 = p$. Proposition 1 applied for $\mathbf{x} = (v_2, \ldots, v_n)$ and $\mathbf{y} = (u_3, \ldots, u_n)$ implies that $v_2 = v_3 = \cdots = v_n = u_3 = \cdots = u_n$. Since $\mathbf{v} \neq p^n$ we infer that

$$\mathbf{u} = qpq^{n-2} \text{ and } \mathbf{v} = pq^{n-1}. \tag{1}$$

It is easy to check that $\mathbf{u}$ is 1-dominant over $\mathbf{v}$. In this case $m = 1$.

Assume $\mathbf{u} = (p, \ldots, p, u_{k+1}, \ldots, u_n)$ and $\mathbf{v} = (p, \ldots, p, v_{k+1}, \ldots, v_n)$ where $k \geq 1$ and $u_{k+1} \neq v_{k+1}$.

If $v_{k+1} = q$ then $u_{k+1} = p$. By deleting the first coordinate in $\mathbf{v}$ we obtain a vector $\mathbf{w}$ with k-th coordinate equals to q. Note that all vectors from $D_1(\mathbf{u})$ have their first k entries equal to p. Therefore $\mathbf{w} \notin D_1(\mathbf{u})$, a contradiction.

Hence, $v_{k+1} = p$ and $u_{k+1} = q$. Since $\mathbf{v} \neq p^n$ we have that $n \geq k+2$. By deleting v_i for arbitrary $i \geq k+2$ we obtain a vector $\mathbf{w}$ with first $k+1$ entries equal to p. The only way to obtain such a vector by 1 deletion in $\mathbf{u}$ is to have $u_{k+2} = p$ and to delete $u_{k+1} = q$. If $n = k+2$ then $\mathbf{u} = p^{n-2}qp$, $\mathbf{v} = p^{n-1}q$, and this pair is equivalent to the pair described in (1). If $n \geq k+3$ then Proposition 1 applied for $\mathbf{x} = (v_{k+2}, \ldots, v_n)$ and $\mathbf{y} = (u_{k+3}, \ldots, u_n)$ implies that $v_{k+2} = \cdots = v_n = u_{k+3} = \cdots = u_n$. Since $\mathbf{v} \neq p^n$ we conclude that $\mathbf{v} = p^{k+1}q^{n-k-1}$ and $\mathbf{u} = p^k qpq^{n-k-2}$. In this case $m = k+1$. $\square$

All pairs $\mathbf{u}$ and $\mathbf{v}$ such that $\mathbf{u}$ is 1-dominant over $\mathbf{v}$ are presented in Table 1.

Table 1

	$\mathbf{u}$	$\mathbf{v}$
1.	$\mathrm{wt}(\mathbf{u}) = 1$	0^n
2.	$\mathrm{wt}(\mathbf{u}) = n-1$	1^n
3.	$p^{m-1}qpq^{n-m-1}$	p^mq^{n-m}

We proceed now with the case $t = 2$. Since the case $\mathbf{v} = p^n$ is clear in what follows we assume that $\mathbf{v} \neq p^n$.

For $n = 3$ up to equivalence we have: $\mathbf{v} = ppq$ and $\mathbf{u} \neq p^3, q^3, \mathbf{v}$ or $\mathbf{v} = pqp$ and $\mathbf{u} \neq p^3, q^3, \mathbf{v}$.

For $n = 4$ we have that up to equivalence there exist 5 choices for $\mathbf{v}$, namely: $pppq$, $ppqp$, $ppqq$, $pqpq$ and $pqqp$. For any of these instances it is easy to enumerate all vectors $\mathbf{u}$ that are 2-dominant over $\mathbf{v}$.

Let $n \geq 5$ be a positive integer and $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ be 2-dominant over $\mathbf{v} = (v_1, v_2, \ldots, v_n)$. Denote $k = \min\{i \mid u_i \neq v_i\}$ and $s = \max\{i \mid u_i \neq v_i\}$ where $k \leq s$. Up to equivalence we split the proof in three cases depending on k and s.

Case A. $k = s$, i.e., $d(\mathbf{u}, \mathbf{v}) = 1$;

Case B. $k = 1$ and $s = n$, i.e., $u_1 \neq v_1$ and $u_n \neq v_n$;

Case C. $k \neq s$ and $k > 1$, i.e., $u_1 = v_1$.

Remark. Up to equivalence the case $k = 1$, $s < n$ and $s \neq k$ is a part of **C** by considering $(u_n, u_{n-1}, \ldots, u_1)$ and $(v_n, v_{n-1}, \ldots, v_1)$ instead of $\mathbf{u}$ and $\mathbf{v}$.

Propositions 3, 4 and 5 settle the cases **A**, **B** and **C**, respectively.

Proposition 3. *If* $\mathbf{u}$ *is* 2*-dominant over* $\mathbf{v}$ *and* $d(\mathbf{u}, \mathbf{v}) = 1$ *then up to equivalence* $\mathbf{u} = p^mqp^{n-m-2}q$ *and* $\mathbf{v} = p^{n-1}q$ *or* $\mathbf{u} = p^mqp^{n-m-3}qp$ *and* $\mathbf{v} = p^{n-2}qp$ *for some integer* $m \geq 0$.

P r o o f. Since $d(\mathbf{u}, \mathbf{v}) = 1$ we have that there exists a positive integer k such that $u_i = v_i$ for $i \neq k$ and $u_k = q$, $v_k = p$. The number of elements q in $\mathbf{u}$ is with one more than the corresponding entries in $\mathbf{v}$. Therefore if there exist two or more entries q in $\mathbf{v}$ then the vector $\mathbf{w}$ obtained by deleting two elements q in $\mathbf{v}$ has at least three elements q less than $\mathbf{u}$. Therefore $\mathbf{w} \notin D_2(\mathbf{u})$. Since $\mathbf{v} \neq p^n$ we conclude that $\mathbf{v} = p^bqp^{n-b-1}$ for some integer b for which $0 \leq b \leq n-1$. Up to equivalence $\mathbf{u} = p^{k-1}qp^{b-k}qp^{n-b-1}$. If $n - b - 1 \geq 2$ then the deletion of the last two symbols from $\mathbf{v}$ gives a vector not in $D_2(\mathbf{u})$. Thus, $n - b - 1 = 0$ or 1 and we obtain $\mathbf{u} = p^mqp^{n-m-2}q$ and $\mathbf{v} = p^{n-1}q$ or $\mathbf{u} = p^mqp^{n-m-3}qp$ and $\mathbf{v} = p^{n-2}qp$. It is easy to check that in both cases $\mathbf{u}$ is 2-dominant over $\mathbf{v}$. $\square$

Proposition 4. *Let* $\mathbf{u}$ *be* 2*-dominant over* $\mathbf{v}$ *and* $u_1 \neq v_1$, $u_n \neq v_n$. *Then up to equivalence* $\mathbf{u} = qpq^{n-4}pq$ *and* $\mathbf{v} = pq^{n-2}p$ *or* $\mathbf{u} = qp^{n-3}qp$ *and* $\mathbf{v} = p^{n-1}q$.

Proof. Without loss of generality we may assume $v_1 = p$ and $u_1 = q$.

1. Let $v_n = p$ and $u_n = q$. The deletion of any two elements from $v_2, v_3, \ldots, v_{n-1}$ gives a vector from $D_2(\mathbf{u})$ with first coordinate $v_1 = p$ and last coordinate $v_n = p$. Such a vector can be obtained from $\mathbf{u}$ only if we delete $u_1 = q$ and $u_n = q$. Therefore $u_2 = u_{n-1} = p$ and any two deletions from $(v_2, v_3, \ldots, v_{n-1})$ imply $(u_3, u_4, \ldots, u_{n-2})$. It follows from the remark after Proposition 2 that $v_2 = v_3 = \cdots = v_{n-1} = u_3 = u_4 = \cdots = u_{n-2}$. Since $\mathbf{v} \neq p^n$ we have that $\mathbf{u} = qpq^{n-4}pq$ and $\mathbf{v} = pq^{n-2}p$. A direct verification shows that indeed $\mathbf{u}$ is 2-dominant over $\mathbf{v}$.

2. Let $v_n = q$ and $u_n = p$. As in the previous case we conclude that $u_2 = p$, $u_{n-1} = q$ and $v_2 = v_3 = \cdots = v_{n-1} = u_3 = \cdots = u_{n-2}$. Up to equivalence $\mathbf{u} = qp^{n-3}qp$, $\mathbf{v} = p^{n-1}q$ and it is easy to see that $\mathbf{u}$ is 2-dominant over $\mathbf{v}$. $\square$

Proposition 5. *Let* $\mathbf{u}$ *be* 2*-dominant over* $\mathbf{v}$, $u_1 = v_1$ *and* $k \neq s$ *where* $k = \min\{i \mid u_i \neq v_i\}$ *and* $s = \max\{i \mid u_i \neq v_i\}$. *Then up to equivalence all such vectors* $\mathbf{u}$ *and* $\mathbf{v}$ *are presented in the following table.*

	$\mathbf{u}$	$\mathbf{v}$
1.	$pqp^{m-1}qpq^{n-m-3}$	$pqp^{m}q^{n-m-2}$
2.	$pq^{m}pqp^{n-m-3}$	$pq^{m+1}p^{n-m-2}$
3.	p^2qp^{n-3}	pqp^{n-2}
4.	$p^{m-2}qppqp^{n-m-2}$	$p^{m}qp^{n-m-1}$
5.	$p^{m-1}qpqp^{n-m-2}$	$p^{m}qp^{n-m-1}$
6.	$p^{n-4}qppq$	$p^{n-2}qp$
7.	$p^{n-3}qpq$	$p^{n-2}qp$
8.	$p^{m}qp^{n-m-3}qp$	$p^{n-1}q$
9.	$p^{m-2}qpqpq^{n-m-2}$	$p^{m}q^{n-m}$
10.	$p^{m-2}qppqq^{n-m-2}$	$p^{m}q^{n-m}$
11.	$p^{n-2}qp$	$p^{n-1}q$
12.	$p^{n-3}qp^2$	$p^{n-1}q$
13.	$p^{m-1}qpqpq^{n-m-3}$	$p^{m}qpq^{n-m-2}$
14.	$p^{n-3}q^2p$	$p^{n-2}q^2$
15.	$p^{n-4}qppq$	$p^{n-3}q^2p$
16.	$p^{n-4}qpqp$	$p^{n-3}qpq$
17.	$p^{m-1}q^{n-m-1}pq$	$p^{m}q^{n-m-1}p$
18.	$p^{m-1}q^{n-m}p$	$p^{m}q^{n-m}$

P r o o f. Without loss of generality we may assume $v_1 = u_1 = p$. Note that since $k \neq s$ we have $d(\mathbf{u}, \mathbf{v}) > 1$.

1. If $v_2 = q$ and $u_2 = q$ then the deletion of v_1 and an arbitrary v_i for $i \geq 3$ implies the deletion of u_1 in $\mathbf{u}$. Thus $\mathbf{u}$ without its first 2 entries is 1-dominant over $\mathbf{v}$ without its first 2 entries. Hence, $\mathbf{u} = pq\mathbf{u_1}$ and $\mathbf{v} = pq\mathbf{v_1}$ where $\mathbf{u_1}$ is 1-dominant over $\mathbf{v_1}$ and $d(\mathbf{u_1}, \mathbf{v_1}) > 1$. Therefore the pair $(\mathbf{u_1}, \mathbf{v_1})$ is equivalent to one of the pairs from Table 1 and $d(\mathbf{u_1}, \mathbf{v_1}) > 1$. Only the third entry in Table 1 satisfies $d(\mathbf{u_1}, \mathbf{v_1}) > 1$. Hence, we obtain the following pairs: $\mathbf{u} = pqp^{m-1}qpq^{n-m-3}$ and $\mathbf{v} = pqp^{m}q^{n-m-2}$; $\mathbf{u} = pq^{m}pqp^{n-m-3}$ and $\mathbf{v} = pq^{m+1}p^{n-m-2}$. In both cases we have that $\mathbf{u}$ is 2-dominant over $\mathbf{v}$.

2. If $v_2 = q$ and $u_2 = p$ then the deletion of v_1 and arbitrary v_i for $i \geq 3$ results a vector $\mathbf{w} \in D_2(\mathbf{u})$ with first coordinate $v_2 = q$. To obtain $\mathbf{w}$ from $\mathbf{u}$ by two deletions we should have $u_3 = q$ and we have to delete u_1 and u_2. We obtain that a single deletion in $(v_3, v_4, \ldots, v_n)$ gives $(u_4, u_5, \ldots, u_n)$. Proposition 1 implies that $v_3 = \cdots = v_n = u_4 = \cdots = u_n$. Thus, $\mathbf{u} = ppqp^{n-3}$ and $\mathbf{v} = pqp^{n-2}$ or $\mathbf{u} = ppq^{n-2}$ and $\mathbf{v} = pq^{n-1}$. For both pairs $\mathbf{u}$ is 2-dominant over $\mathbf{v}$ but only for the first pair we have $d(\mathbf{u}, \mathbf{v}) > 1$.

Let $v_2 = p$ and assume $p = v_1 = v_2 = \cdots = v_m \neq v_{m+1} = q$ for some $m \geq 2$. If $u_1 = u_2 = \cdots = u_m = p$ the deletion of v_1 and v_2 implies the deletion of two of the first m elements in $\mathbf{u}$. Thus, $\mathbf{u} = \mathbf{v}$, a contradiction.

We conclude that $k \leq m$ and then $u_1 = \cdots = u_{k-1} = p$ and $u_k = q$.

For $p \in \{0, 1\}$ and a vector $\mathbf{w}$ denote by $n_p(\mathbf{w})$ the number of entries p in the vector $\mathbf{w}$.

- If $n_q(\mathbf{v}) > n_q(\mathbf{u})$ then delete two elements p from $\mathbf{v}$ and let $\mathbf{w}$ be the resulting vector. Since $n_q(\mathbf{w}) > n_q(\mathbf{u})$ we have that $\mathbf{w} \notin D_2(\mathbf{u})$.

- If $n_q(\mathbf{v}) < n_q(\mathbf{u})$ then $n_p(\mathbf{v}) > n_p(\mathbf{u})$. If $n_q(\mathbf{v}) \geq 2$, i.e., there exist at least two entries q in $\mathbf{v}$ then we delete two elements q from $\mathbf{v}$ and obtain a contradiction as above. Therefore $\mathbf{v} = p^{m}qp^{n-m-1}$ and $n_q(\mathbf{u}) = 2$ (if $n_q(\mathbf{u}) \geq 3$ the deletion of a symbol p and the symbol q in $\mathbf{v}$ gives a contradiction). If $u_{m+1} = q$ then $k = s$, a contradiction. Assume first that $n - m - 1 \geq 2$. If $u_i = q$ for $i \neq k$ and $i \leq m$ then, as above, the deletion of $v_{n-1} = p$ and $v_n = p$ gives a contradiction. Thus, for some $i > m + 1$ we have $u_i = q$. It is easy to see that up to equivalence there exist two choices for $\mathbf{u}$: $\mathbf{u} = p^{m-2}qppqp^{n-m-2}$ and $\mathbf{u} = p^{m-1}qpqp^{n-m-2}$. If $n - m - 1 = 1$ then $\mathbf{v} = p^{n-2}qp$, $\mathbf{u} = p^{n-4}qppq$ or $\mathbf{u} = p^{n-3}qpq$ and if $n - m - 1 = 0$ then $\mathbf{v} = p^{n-1}q$ and $\mathbf{u} = p^{m}qp^{n-m-3}qp$.

- Let $n_q(\mathbf{v}) = n_q(\mathbf{u})$. Note that in this case the deleted symbols from $\mathbf{v}$ are identical to the deleted symbols from $\mathbf{u}$. If $m \geq k+2$ the deletion of the first two entries in $\mathbf{v}$ gives a contradiction. If $m = k+1$ then if there exists $i > m+1$ such that $v_i = p$ then the deletion of v_1 and v_i implies a contradiction. Thus, $\mathbf{v} = p^m q^{n-m}$ and then for $n - m \geq 2$ we have $\mathbf{u} = p^{m-2}qpqpq^{n-m-2}$ or $\mathbf{u} = p^{m-2}qppqq^{n-m-2}$. For $n - m = 1$ we have $\mathbf{v} = p^{n-1}q$ and $\mathbf{u} = p^{n-2}qp$ or $\mathbf{u} = p^{n-3}qp^2$.

Let $m = k$. If there exist at least two entries p in $(v_{m+2}, \ldots, v_n)$ then the deletion of these two elements gives a contradiction.

If $u_n = v_n = p$ then we may show as above that $v_{n-1} = p$ and we have at least two entries p in $(v_{m+2}, \ldots, v_n)$, a contradiction.

If $u_n = v_n = q$ then the same observations as above but starting from the right to the left imply that $\mathbf{u} = p^{m-1}q\mathbf{w}pq^b$ and $\mathbf{v} = p^m q\mathbf{h}pq^{b+1}$. If $\mathbf{h}$ is not empty then the deletion of any two elements from $q\mathbf{h}p$ implies the deletion of u_m and u_{n-b}. Proposition 1 gives that all entries in $q\mathbf{h}p$ are equal which is not true. Therefore $\mathbf{h}$ is empty and then $\mathbf{v} = p^m qpq^{n-m-2}$ and $\mathbf{u} = p^{m-1}qu_{m+1}u_{m+2}pq^{n-m-3}$. Since $n_q(\mathbf{v}) = n_q(\mathbf{u})$ we have that $\{u_{m+1}, u_{m+2}\} = \{p, q\}$. Only one of the two cases gives 2-dominant vectors, namely $\mathbf{u} = p^{m-1}qpqpq^{n-m-3}$ and $\mathbf{v} = p^m qpq^{n-m-2}$.

It remains to consider the case $u_n \neq v_n$. If $n \leq m+3$ then an easy enumeration gives:

- $\mathbf{u} = p^{n-2}qp$ and $\mathbf{v} = p^{n-1}q$ for $n = m+1$;
- $\mathbf{u} = p^{n-3}q^2p$ and $\mathbf{v} = p^{n-2}q^2$ for $n = m+2$;
- $\mathbf{u} = p^{n-4}qppq$ and $\mathbf{v} = p^{n-3}q^2p$; $\mathbf{u} = p^{n-4}qpqp$ and $\mathbf{v} = p^{n-3}qpq$ for $n = m+3$.

If $n \geq m+4$ then any two deletions in $(v_{m+1}, \ldots, v_{n-1})$ imply $u_{n-1} = v_n$ and the deletion of u_m and u_n. By Proposition 1 we obtain that $q = v_{m+1} = \cdots = v_{n-1} = u_{m+2} = \cdots = u_{n-1}$, thus $\mathbf{v} = p^m q^{n-m-1} v_n$ and $\mathbf{u} = p^{m-1}q^{n-m-1}v_n u_n$.

Both choices of $u_n \neq v_n$ give 2-dominant pair. $\square$

REFERENCES

[1] L. Calabi. On the computation of Levenshtein distances. Report TM-9-0030, ParkeMath. Labs., Inc. Carlisle, MA, 1967.

[2] A. S. J. Helberg, H. C. Fereira. On multiple insertion/deletion correcting codes, *IEEE Trans. Inf. Theory* **48**, *1*, (2002), 305–308.

[3] I. Landjev, Kr. Haralambiev. On multiple deletion codes, *Serdica J. Computing* **1**, *1* (2007), 13–26.

[4] K. Landjev. Codes Correcting Deletions, MSc. Thesis, Sofia University, 2015 (in Bulgarian).

[5] V. I. Levenshtein. Binary codes capable of correcting, deletions, insertions and reversals. *Doklady Akad. Nauk SSSR* **163** (1965), 845–848 (in Russian); English translation in: *Sov. Physics Dokl.* **10** (1965), 707–710 (1966).

[6] V. I. Levenshtein. Binary codes with correction for deletions and insertions of the symbol 1. *Probl. Peredchi Inf.* **1**, *1* (1965), 12–25 (in Russian); English translation in *Problems Inform. Transmission* **1**, *1* (1965), 8–17.

[7] V. I. Levenshtein. On perfect codes in the metric of appearences and insertions. *Diskretn. Mat.* **3**, *1* (1991), 3–20 (in Russian); English translation in Discrete Math. Appl. 2 (1992), no. 3, 241–258.

[8] V. I. Levenshtein. Efficient reconstruction of sequences from their subsequences or supersequences. *J. Comb. Theory, Ser. A* **93**, *2* (2001), 310–332.

[9] V. I. Levenshtein. Efficient reconstruction of sequences. *IEEE Trans Inf. Theory* **47**, *1* **2001**, 2–22.

[10] N. J. A. Sloane. On single-deletion-coirrecting codes. In: Codes and Designs, Proceedings of a Conference Honoring Professor Dijen K. Ray-Chaudhuri on the Occasion of his 65th Birthday, The Ohio State University, Columbus, OH, USA, May 18–21, 2000. Berlin, de Gruyter. Ohio State Univ. Math. Res. Inst. Publ. **10** (2002), 273–291.

[11] T. G. Swart, H. C. Fereira. A note on double insertion/deletion correcting codes. *IEEE Trans. Inf. Theory* **49**, *1* (2003), 269–273.

[12] L. Tolhuizen. Upper bounds on the size of insertion/deletion codes. Proc. 8th Int. Workshop on ACCT, Tsarskoe selo, Russia, September 2002, 242–246.

Institute of Mathematics and Informatics
Bulgarian Academy of Sciences
Acad. G. Bonchev Str., Bl. 8
1113 Sofia, Bulgaria
e-mail: `emil@math.bas.bg`

Received October 13, 2015