Insertion and Deletion for Involution Codes

Lila Kari and Kalpana Mahalingam

University of Western Ontario,
Department of Computer Science,
London, ON N6A5B7
lila, kalpana@csd.uwo.ca

Abstract. *This paper introduces a generalization of the operation of catenation: $u[k]_l v$, the left-k-insertion, is the set of all words obtained by inserting v into u in positions that are at most k letters away from the left extremity of the word u. We define k-suffix codes using the left-k-insertion operation and extend the concept of k-prefix and k-suffix codes to involution k-prefix and involution k-suffix codes. An involution code refers to any of the generalizations of the classical notion of codes in which the identity function is replaced by an involution function. (An involution function θ is such that θ^2 equals the identity). We also extend the notion of k-insertion closure and k-deletion closure of a language to incorporate the notion of an involution function. Thus to an involution map θ and a language L, we associate a set k-θ-ins(L) (k-θ-del(L)) with the property that their k-insertion (k-deletion) into any word of L yields words which belongs to $\theta(L)$. We study the properties of these languages.*

1 Introduction

Catenation and quotient are basic operations in formal language theory. The catenation and quotient operations were generalized to the concept of k-catenation and k-quotient which was studied in [1]. The catenation of two words u and v is just uv. The k-catenation of u and v is the set of all words obtained by inserting v into u in positions that are atmost k letters away from the right extremity of the word u. Similarly the k-quotient of u from v is the deletion of u from v resulting in x_1x_2 where $v = x_1ux_2$ for some words x_1 and x_2 , $|x_2| \leq k$. When $k = 0$ the 0-catenation and 0-quotient are the regular catenation and quotient operation. Similar to the generalization of catenation to k-catenation, [1] has generalized the concept of prefix codes([6, 7]) to k-prefix codes and has discussed various properties of such codes.

The k-catenation and the k-quotient defined in [1] were later called k-insertion and k-deletion in [2]. In [2] the authors have discussed the properties of k-insertion (k-deletion) closure of a given language L. Procedures of constructing the k-insertion and k-deletion closure of a language were also given in [2].

In this paper we follow the approach from [1, 2] and extend these concepts to incorporate the notion of an involution function replacing the identity function. (An involution function θ is such that θ^2 equals the identity). In the following an involution code refers to any of the generalizations of classical notion of codes ([6, 7]) that replace the identity function with the involution function in a way explained later (Definition 5). Involution codes were introduced in [3] in the process of designing DNA strands with certain properties. The operation of k-catenation in [1] allows insertion to take place close to the right extremity of a word. In this paper we introduce a k-catenation that allows insertion to take place at its left. In order to differentiate between the two operations we call the former right-k-catenation and the latter left-k-catenation. We also generalize the concept of suffix codes ([6, 7]) to k-suffix codes using the left-k-catenation. Section 2 discusses the properties of such codes. We extend the concept of k-prefix and k-suffix codes to involution k-prefix and involution k-suffix codes in Section 3.

In Section 4 we define for a language L and an involution θ, the k-θ-insertion closure of a language L denoted by k-θ-ins(L) as the language consisting of the words with the property that their k-insertion into any word of L yields a word in $\theta(L)$. The k-θ-deletion closure of a language L denoted by (k-θ-del(L)) is defined as the language consisting of the words with the property that their k-deletion from any word of $\theta(L)$ yields a word in L. We construct these languages using the dual operation of dipolar k-deletion.

In this paper we use the following notations. By Σ we denote the finite nonempty alphabet set and by Σ^* the free monoid generated by Σ under the catenation operation. Any word over Σ is a finite sequence of letters from Σ and by 1 we denote the empty word. The length of a word $u \in \Sigma^*$ is the number of letters in u and is denoted by $|u|$. Throughout the rest of the paper, we concentrate on sets $L \subseteq \Sigma^+$ that are *codes* meaning that every word in L^+ can be written uniquely as a product of words in L (i.e. L^+ is a free semigroup generated by L). For the background on codes we refer the reader to [6, 7].An involution $\theta : \Sigma \mapsto \Sigma$ is a function such that $\theta^2 = I$ where I is the identity function and can be extended to a morphic involution on Σ^* if for all $u, v \in \Sigma^*$, $\theta(uv) = \theta(u)\theta(v)$ or an antimorphic involution if $\theta(uv) = \theta(v)\theta(u)$. For more on involution codes we refer the reader to [3–5].

2 k-Suffix codes

In this section we introduce a new class of suffix codes called as the k-suffix codes with respect to the left-k-catenation or left-k-insertion as it was called later in [2]. The concept of k-prefix codes was introduced and studied in [1]. Most of the results that hold for k-prefix codes also hold for k-suffix codes.The following is a generalization of the catenation operation. The definition of right-k-insertion (right-k-catenation) was introduced in [1] and was just called as k-catenation. The definition of left-k-insertion (left-k-catenation) is the new concept we introduce here. Throughout the rest of the paper we assume $k \geq 0$ to be an integer.

Definition 1. *Let u, v be the words over the alphabet Σ.*

1. *The right-k-insertion of v into u is defined by:*
   $u[k]_r v = \{u_1 v u_2 : u = u_1 u_2, |u_2| \leq k\}$.

2. *The left-k-insertion of v into u is defined by:* $u[k]_l v = \{u_1 v u_2 : u = u_1 u_2, |u_1| \leq k\}$.
3. *For $L_1, L_2 \subseteq \Sigma^*$, $L_1[k]_\alpha L_2 = \bigcup_{u_1 \in L_1, u_2 \in L_2} u_1[k]_\alpha u_2$ for $\alpha \in \{l, r\}$.*

Definition 2. *Let u, v be words over the alphabet Σ.*

1. *The relation $\delta_{k,\alpha}$ is defined on Σ^* by: $u\delta_{k,\alpha}v$ iff $v \in u[k]_\alpha \Sigma^*$ for $\alpha \in \{l, r\}$. From now on we will use $\delta_{k,r} = \delta_k$.*
2. *$\delta_k(u) = \{v \in \Sigma^* : u\delta_k v\}$.*
3. *Let $L \subseteq \Sigma^*$, then $\delta_k(L) = \{v \in \Sigma^* : \exists u \in L$ such that $v \in \delta_k(u)\}$. The language $\delta_k(L)$ is called the δ_k closure of L.*
4. *$R \subseteq \Sigma^*$ is a left-k-subsemigroup if $R[k]_l R \subseteq R$.*
5. *$L \subseteq \Sigma^*$ is a left-k-ideal if $L[k]_l \Sigma^* \subseteq L$.*

The relation δ_k is a reflexive and antisymmetric binary relation. The transitive closure $\bar{\delta_k}$ of δ_k is a right compatible partial order. Remark that if $k = 0$, δ_0 is the usual suffix order.

A nonempty subset $R \subseteq \Sigma^*$ such that $u, v \in R$ implies $u[k]_r v \subseteq R$ is called a right-k-subsemigroup. Clearly R is a subsemigroup of Σ^*.

A left-k-ideal $L \subseteq \Sigma^*$ is a nonempty subset of Σ^* such that $u \in L$ implies $u[k]_r x \subseteq L$ for all $x \in \Sigma^*$. This is equivalent to $L[k]_r \Sigma^* \subseteq L$. Every left-$k$-ideal is a left ideal and a right-k-subsemigroup. If L is a left-k-ideal for every $k \geq 0$, then, for all $u = u_1 u_2 \in L$ and $x \in \Sigma^*$, $u_1 x u_2 \in L$.

Definition 3. *If $L \subseteq \Sigma^*$, then define $\delta_k^{[0]}(L) = L$, $\delta_k^{[1]}(L) = \delta_k(L)$, ... , $\delta_k^{[n]}(L) = \delta_k^{[1]}(\delta_k^{[n-1]}(L))$, ... , $\delta_k^*(L) = \cup_{n \geq 0} \delta_k^{[n]}(L)$.*

Clearly $\delta_k(L) = \{v \in \Sigma^* : \exists u \in L, u\bar{\delta_k}v\}$. The language $\delta_k(L)$ is called the δ_k closure of L.

Lemma 1. *If $T \subseteq \Sigma^*$ is a left-k-ideal containing $L \subseteq \Sigma^*$ then $\delta_k^*(L) \subseteq T$.*

Proof. Let T be a left k-ideal containing L. Suppose that $\delta_k^*(L)$ is not contained in T. Then there is an integer n and a word $v \in \delta_k^{[n]}(L) \subseteq \delta_k^*(L)$ such that $v \notin T$. Suppose n is minimal with this property, then $n \geq 1$ and there exists $u \in \delta_k^{[n-1]}(L)$ such that $v = u_1 x u_2$. Because of the minimality of n, then $u \in T$ and, since T is a left k-ideal, $v = u_1 x u_2 \in T$, a contradiction. Therefore $\delta_k^*(L) \subseteq T$.

Proposition 1. *If L is a nonempty language, $\delta_k^*(L)$ is the minimal left k-ideal containing L.*

Proof. Let us see that $\delta_k^*(L)$ is a left-k-ideal. Note that $\delta_k^*(L)[k]_r \Sigma^* = \delta_k(\delta_k^*(L))$ $= \delta_k(\bigcup_{n \geq 0} \delta_k^{[n]}(L)) = \bigcup_{n \geq 0} \delta_k(\delta_k^{[n]}(L)) = \bigcup_{n \geq 1} \delta_k^{[n]}(L) \subseteq \delta_k^*(L)$.

From Lemma 1 we have that $\delta_k^*(L)$ is the minimal ideal containing L.

Note that a language L is a left k-ideal if and only if $L = \delta_k^*(L)$.

Definition 4. *Let $S \subseteq \Sigma^*$ be a nonempty language.*

1. *S is a k-prefix code if $u \in S$ and $u[k]_r x \cap S \neq \emptyset$ then $x = 1$.*
2. *S is a k-suffix code if $u \in S$ and $u[k]_l x \cap S \neq \emptyset$ then $x = 1$.*

Remark that a k-suffix code is also an m-suffix code for $m \leq k$ and that suffix codes are the 0-suffix codes. Every outfix code is a k-suffix code for all $k \geq 0$. An infix code is not in general a k-suffix code. For example, let $L = ba^+b$ over $\Sigma = \{a, b\}$. Then L is an infix code, but not a k-suffix code for $k \geq 1$. Indeed, $ba^k b = ba^{k-1}ab \in L$ and $ba^{k-1}aab \in L$ with $|ba^{k-1}| \leq k$, but $a \neq 1$.

Recall that if δ is a partial order or a quasiorder (reflexive and antisymmetric relation), then an antichain A is a subset of Σ^* such that $u\delta v, u, v \in A$ implies $u = v$. If $L \subseteq \Sigma^+$, $L \neq \emptyset$, then L is a suffix code iff L is an antichain for the suffix order, that is δ_0. We have a generalized version of the above fact.

Proposition 2. *A nonempty language $L \subseteq \Sigma^+$ is a k-suffix code if and only if L is a δ_k antichain.*

Proof. Immediate from the definitions.

Let Σ be the alphabet set with $|\Sigma| \geq 2$, let $a \in \Sigma$ and $Y = \Sigma \setminus \{a\}$. Then $a^k Y^*$ is a k-suffix code that is not an m-suffix code for all $m > k$. Hence if $S_k(\Sigma)$ denotes the family of the k-suffix codes over Σ, we have the infinite hierarchy:
$S_0(\Sigma) \supset S_1(\Sigma) \supset ...S_k(\Sigma) \supset ...$.

With every nonempty language $L \subset \Sigma^+$ is associated a k-suffix code $Suf_k(L)$ defined in the following way: $Suf_k(L) = \{u \in L : v \in L, v\delta_k u \Rightarrow u = v\}$, (i.e.) $Suf_k(L)$ is the set of words in L that are minimal with respect to the relation δ_k or $\bar{\delta_k}$. Since $1 \notin L$ and $L \neq \emptyset$, then it is clear that $Suf_k(L)$ is a k-suffix code.

Proposition 3. *Let $S \subseteq \Sigma^+$.*

1. *If S is a k-suffix code, then $\delta_k(S)$ is a left k-ideal and $Suf_k(\delta_k(S)) = S$.*
2. *If L is a left k-ideal, $L \neq \Sigma^*$, then there exists a unique k-suffix code S namely $S = Suf_k(L)$, such that $L = \delta_k(S)$.*

A well known property of the suffix code is that it is closed under catenation. We provide a similar result for k-suffix codes.

Proposition 4. *The catenation of k-suffix codes is a k-suffix code.*

Proof. Let S, R be two k-suffix codes and let $\alpha \in S$, $\beta \in R$ such that there is a word in $\alpha\beta[k]_l v$ which belongs to SR. We want to show that $v = 1$. We distinguish two cases:

1. The word v has been inserted into α or catenated to α. This means that $\alpha_1 v \alpha_2 \beta \in SR$, $|\alpha_1| \leq k$, $\alpha_1, \alpha_2 \in \Sigma^*$ and $\alpha = \alpha_1\alpha_2$.
2. The word v has been inserted into β in a similar fashion.

Consider the second case (the other one can be proved in a similar fashion). As $\alpha\beta_1 v\beta_2 \in SR$, it is a catenation of $xy \in SR$ such that $x \in S$, $y \in R$. Then one of the following situations occur:

1. $x = \alpha\beta_1 v\beta_2'$, $y = \beta_2''$. As $|\alpha\beta_1| \leq k$, we have $|\alpha| \leq k$ and since S is a k-suffix code with $\alpha \in S$, $\beta_1 v\beta_2' = 1$ which implies $v = 1$.
2. $x = \alpha\beta_1 v_1$ and $y = v_2\beta_2$ with $v = v_1v_2$. Since $|\alpha| \leq k$ and S is k-suffix code, we have $\beta_1 v_1 = 1$ which implies $v_1 = 1$ and hence $v = v_2$. Therefore $y = 1v_2\beta_2$ with $|1| \leq k$ and since R is a k-suffix code, we have $v_2 = 1$.
3. $x = \alpha\beta_1'$ and $y = \beta_1'' v\beta_2$. As $|\alpha| \leq k$, $\beta_1' = 1$ and hence $\beta_1''\beta_2 \in R$ with $|\beta_1''| \leq k$ and since R is a k-suffix code, $v = 1$.
4. $x = \alpha_1$ and $y = \alpha_2\beta_1 v\beta_2$. Since $\alpha_1, \alpha_1\alpha_2 \in S$ with $|\alpha_1| \leq k$ we have $\alpha_2 = 1$ which implies $\beta_1 v\beta_2$ and hence $v = 1$.

In all cases we obtained $v = 1$ and therefore SR is also a k-suffix code.

3 k-Insertion for involution codes

In this section we generalize the catenation operation to include the notion of an involution function and also generalize the class of k-prefix and k-suffix codes to involution k-prefix(k-θ-prefix) and involution k-suffix (k-θ-suffix) codes. An involution code refers to any of the generalizations of the classical notion of codes that replace the identity function with the involution function as explained in [3–5]. Note that when θ is identity the k-θ-prefix(suffix) code is nothing but k-prefix(suffix) code.

Definition 5. *Let u, v be words over the alphabet Σ and let θ be a morphic or antimorphic involution.*

1. *A k-θ-prefix-code is a non empty language $P \subseteq \Sigma^+$ such that $u \in P$ and $\theta(u)[k]_r v \cap P \neq \emptyset$ implies $v = 1$.*
2. *A k-θ-suffix-code is a non empty language $S \subseteq \Sigma^+$ such that $u \in S$ and $\theta(u)[k]_l v \cap S \neq \emptyset$ implies $v = 1$.*
3. *A set L is called k-θ-bifix code iff L is both k-θ-prefix and k-θ-suffix code.*

Lemma 2. *Let $L \subseteq \Sigma^+$.*

1. *For a morphic involution θ, L is k-θ-prefix (suffix) iff $\theta(L)$ is k-θ-prefix (suffix).*
2. *For an antimorphic involution θ, L is k-θ-prefix (suffix) iff $\theta(L)$ is k-θ-suffix (prefix).*
3. *L is k-θ-bifix iff $\theta(L)$ is k-θ-bifix.*

Proof. Let θ be morphic involution and L is k-θ-prefix. Suppose there exists $\theta(u) \in \theta(L)$ such that $\theta(\theta(u)) = u_1u_2$ with $|u_2| \leq k$ and $u_1vu_2 \in \theta(L)$ for some $v \in \Sigma^*$. We need to show that $v = 1$. Note that $u_1vu_2 \in \theta(L)$ iff $\theta(u_1vu_2) \in L$ iff $\theta(u_1)\theta(v)\theta(u_2) \in L$ which implies $\theta(v) = 1$ since L is k-θ-prefix. Similarly we can prove the other direction and also the other statements.

Remark that a k-θ-prefix(suffix)-code is also an m-θ-prefix(suffix)-code for $m \le k$ and that θ-prefix(suffix) codes are k-θ-prefix(suffix) codes when $k = 0$.

Proposition 5. *When θ is morphic involution the class of k-θ-prefix(suffix) codes is closed under concatenation.*

Proof. We prove for k-θ-prefix-codes. Let P, Q be two k-θ-prefix codes. Let $a \in P$ and $b \in Q$ such that $\theta(ab)[k]_r v \in PQ$. We need to show that $v = 1$.

We have the two following cases:

(i) $\theta(a_1)v\theta(a_2)\theta(b) \in PQ$ with $|\theta(a_2)\theta(b)| \le k$ and $\theta(a) = \theta(a_1 a_2)$.

(ii) $\theta(a)\theta(b_1)v\theta(b_2) \in PQ$ with $|\theta(b_2)| \le k$ and $\theta(b) = \theta(b_1 b_2)$.

Consider case (i). Let $xy = \theta(a_1)v\theta(a_2)\theta(b) \in PQ$ such that $x \in P$ and $y \in Q$.

Then

1. $x = \theta(a_1')$ and $y = \theta(a_1'')v\theta(a_2)\theta(b)$ with $b, y \in Q$ and $|\theta(a_2)\theta(b)| \le k$. Since $y \in Q$ and Q is k-θ-prefix, we have $|\theta(b) \le k$ and $\theta(a_1'')v\theta(a_2) = 1$.
2. $x = \theta(a_1)v_1$ and $y = v_2\theta(a_2)\theta(b)$ with $b, y \in Q$ and $|\theta(b)| \le k$. Since $y \in Q$ and Q is k-θ-prefix, we have $v_2\theta(a_2) = 1$ which implies $v = v_1$ and $\theta(a) = \theta(a_1)$. Since $x, a \in P$ and P is k-θ-prefix with $|1| \le k$, we have $v_1 = v = 1$.
3. $x = \theta(a_1)v\theta(a_2')$ and $y = \theta(a_2'')\theta(b)$ with $y, b \in Q$ and $|\theta(b)| \le k$. Since Q is k-θ-prefix, we have $\theta(a_2'') = 1$ and since $|\theta(a_2')| \le k$ with $x, a_1 a_2' \in P$, we have $v = 1$.
4. $x = \theta(a_1)v\theta(a_2)\theta(b_1)$ and $y = \theta(b_2)$ with $b, y \in Q$ and $|\theta(b_2)| \le k$ (i.e.) we have $\theta(b_2), b_1 b_2 \in Q$ which implies $b_1\theta(\theta(b_2)) \in Q$ and hence $b_1 = 1$ since $x = \theta(a_1)v\theta(a_2)$ and P is k-θ-prefix with $|\theta(a_2)| \le k$ we have $v = 1$.

Similar proof works for case(ii). Hence PQ is k-θ-prefix. Similarly we can show that k-θ-suffix codes are closed under concatenation when θ is morphic involution.

Proposition 6. *When θ is antimorphic involution, if L is a k-θ-bifix code, then L^n is a k-θ-bifix code for all $n \ge 1$.*

Proof. We prove by induction on n. Given L is k-θ-bifix. Assume that L^m is k-θ-bifix for some $m \ge 1$. Let $a = a_1 ... a_{m+1} \in L^{m+1}$ such that $a_i \in L$ for all $1 \le i \le m+1$ and $\theta(a)[k]_r v \in L^{m+1}$. We need to show that $v = 1$. We have the following $m+1$ cases.

Case (1):
We have $\theta(a_{m+1,1})v\theta(a_{m+1,2})\theta(a_m)...\theta(a_1) \in L^{m+1}$ such that $|\theta(a_{m+1,2})\theta(a_m)...\theta(a_1)| \le k$.
Cases (i) $(2 \le i \le m)$: $\theta(a_{m+1})...\theta(a_{i,1})v\theta(a_{i,2})\theta(a_{i-1})...\theta(a_1) \in L^{m+1}$ with $|\theta(a_{i,2})\theta(a_{i-1})...\theta(a_1)| \le k$ for $2 \le i \le m$.

Case (m+1): $\theta(a_{m+1})\theta(a_m)...\theta(a_{1,1})v\theta(a_{1,2}) \in L^{m+1}$ with $|\theta(a_{1,2})| \leq k$.

Consider case (i).

Let $xy = \theta(a_{m+1,1})v\theta(a_{m+1,2})\theta(a_m)...\theta(a_1)$ such that $xy \in L^{m+1}$, $x \in L$ and $y \in L^m$ with $|\theta(a_{m+1,2})\theta(a_m)...\theta(a_1)| \leq k$.

Then we have,

1. $x = \theta(a'_{m+1,1})$ and $y = \theta(a''_{m+1,1})v\theta(a_{m+1,2})\theta(a_m)...\theta(a_1)$ with $\theta(a_{m+1,1}) = \theta(a'_{m+1,1})\theta(a''_{m+1,1})$ which implies $v = 1$ since L^m is k-θ-bifix.
2. $x = \theta(a_{m+1,1})v_1$ and $y = v_2\theta(a_{m+1,2})\theta(a_m)...\theta(a_1)$ with $v = v_1v_2$ and $|\theta(a_m)...\theta(a_1)| \leq k$. Since L^m is k-θ-bifix, we have $v_2\theta(a_{m+1,2}) = 1$ and hence $\theta(a_{m+1}) = \theta(a_{m+1,1})$ and $v = v_1$. Since L is k-θ-bifix, we have $v = 1$.
3. $x = \theta(a_{m+1,1})v\theta(a'_{m+1,2})$ and $y = \theta(a''_{m+1,2})\theta(a_m)...\theta(a_1)$. Since L^m is k-θ-bifix we have $\theta(a''_{m+1,2}) = 1$ and hence $v = 1$ since L is k-θ-bifix.
4. $x = \theta(a_{m+1,1})v\theta(a_{m+1,2})\theta(a''_m)$ and $y = \theta(a'_m)...\theta(a_1)$. Since $y = \theta(a'_m)...\theta(a_1)$ which belongs to L^m, we have $\theta(a_1...a'_m) \in L^m$ and hence $\theta(\theta((a_1...a'_m))a''_m \in L^m$ which implies $a''_m = 1$ since L^m is k-θ-bifix. Hence $x = \theta(a_{m+1,1})v\theta(a_{m+1,2})$ with $|\theta(a_{m+1,2})| \leq k$ and since L is k-θ-bifix we have $v = 1$.

The other cases can be proved in a similar fashion and hence L^{m+1} is k-θ-prefix. We can also show that L^{m+1} is k-θ-suffix similarly.

Lemma 3. *Let θ be a morphic involution and let L_1 and L_2 be non empty languages over Σ^+ such that $L_i \cap \theta(L_i) \neq \emptyset$ for $i = 1, 2$. Then the following are true.*

1. *If L_1L_2 is k-θ-prefix code, then L_2 is a k-θ-prefix code.*
2. *If L_1L_2 is k-θ-suffix code, then L_1 is a k-suffix code.*

Proof. Let L_1L_2 be k-θ-prefix code. Let $u \in L_2$ such that $u = u_1u_2$ and $\theta(u_1)v\theta(u_2) \in L_2$ with $|\theta(u_2)| \leq k$. We need to show that $v = 1$. Choose $x \in L_1$ such that $x \in L_1 \cap \theta(L_1)$. Then $x\theta(u_1)v\theta(u_2) \in L_1L_2$ with $x\theta(u_1)\theta(u_2) \in \theta(L_1L_2)$. Since L_1L_2 is k-θ-prefix, we have $v = 1$. Hence L_2 is k-θ-prefix code. Similarly we can show that L_1 is k-θ-suffix codes, when L_1L_2 is a k-θ-suffix code.

Corollary 1. *Let θ be a morphic involution and let $L_i, i = 1, 2, ..., m$ be non empty languages over Σ such that $L_i \cap \theta(L_i) \neq \emptyset$ for all $i = 1, 2, ..., m$. Then the following are true.*

1. *If $L_1L_2...L_m$ is k-θ-prefix code, then $L_2L_3..L_m$, $L_3..L_m$,..., $L_{m-1}L_m$ and L_m are k-θ-prefix codes.*
2. *If $L_1L_2...L_m$ is k-θ-suffix code, then $L_1L_2..L_{m-1}$, $L_1..L_{m-2}$,..., L_1L_2 and L_1 are k-θ-suffix codes.*

Proposition 7. *Let $L \subseteq \Sigma^+$ be such that $L \cap \theta(L) \neq \emptyset$.*

1. *If L^m is k-θ-prefix for $m \geq 1$, then L is k-θ-prefix.*

2. If L^m is k-θ-suffix for $m \geq 1$, then L is k-θ-suffix.
3. If L^m is k-θ-bifix for $m \geq 1$, then L is k-θ-bifix.

Proof. Assume that L^m is k-θ-prefix for some $m \geq 1$. Suppose there exists a $u \in L$ such that $\theta(u)[k]_r v \cap L \neq \emptyset$ for some $v \in \Sigma^*$. Then we need to show that $v = 1$. The case when θ is a morphic involution is a special case of proposition 1 when $L_i = L$ for all i. When θ is antimorphism, let $u = u_1u_2$ then $\theta(u) = \theta(u_2)\theta(u_1)$ and $\theta(u_2)v\theta(u_1) \in L$ with $|\theta(u_1)| \leq k$. Let $z_1, z_2, ..., z_{m-1} \in L \cap \theta(L)$ then $z_1...z_{m-1}\theta(u_2)v\theta(u_1) \in L^m$ which implies $v = 1$ since L^m is k-θ-prefix. Similar proof works when L^m is k-θ-suffix.

4 k-Involution insertion and deletion of languages

Let $L \subseteq \Sigma^+$. To the language L, a set k-ins(L) can be associated consisting of all the words with the following property: their k-insertion into any word of L yields a word belonging to L [2]. Formally k-ins(L) was defined by :
$k\text{-ins}(L) = \{x \in \Sigma^* : \forall u \in L, u = u_1u_2, |u_2| \leq k \Longrightarrow u_1xu_2 \in L\}$. In a similar fashion, for a moprhic or antimorphic involution θ, we associate two sets left-k-θ-ins(L) and right-k-θ-ins(L) consisting of all words with the following property: their k-insertion into any word of L yields a word belonging to $\theta(L)$. Formally right-k-θ-ins(L) and left-k-θ-ins(L) are defined by:

1. right-k-θ-ins$(L) = \{x \in \Sigma^* : \forall u \in L, u = u_1u_2, |u_2| \leq k \Longrightarrow u_1xu_2 \in \theta(L)\}$.
2. left-k-θ-ins$(L) = \{x \in \Sigma^* : \forall u \in L, u = u_1u_2, |u_1| \leq k \Longrightarrow u_1xu_2 \in \theta(L)\}$.

Note that throughout the rest of this section we use $\star$-k-θ-ins(L), where $\star$ either denotes left or right.

Lemma 4. *For a language $L \subseteq \Sigma^+$ we have :*

1. *L is k-θ-prefix code iff right-k-θ-ins$(L) = \{1\}$.*
2. *L is k-θ-suffix code iff left-k-θ-ins$(L) = \{1\}$.*

Proposition 8. *If L is a commutative language, then $\star$-k-θ-ins(L) is also a commutative language.*

Proof. It is sufficient to show that $xuvy \in \star$-k-θ-ins(L) implies $xvuy \in \star$-k-θ-ins(L). If $w \in L$, such that $w = w_1w_2$, $|w_2| \leq k$, then $w_1xuvyw_2 \in \theta(L)$, hence $w_1xvuyw_2 \in \theta(L)$ (Note that L is commutative iff $\theta(L)$ is commutative.). Therefore $xvuy \in k$-θ-ins(L).

Definition 6. *For u, v words over the alphabet set Σ, the right and the left dipolar k-deletion $u \rightleftharpoons^k v$ is defined respectively by:*
$u \rightleftharpoons^k_r v = \{x \in \Sigma^* : u = v_1xv_2, v = v_1v_2, |v_2| \leq k\}$ *and*
$u \rightleftharpoons^k_l v = \{x \in \Sigma^* : u = v_1xv_2, v = v_1v_2, |v_1| \leq k\}$.

In [1], the operation $u \rightleftharpoons_r^k v$ has been introduced under the name of k-deletion and was later called as dipolar k-deletion in [2]. In other words, the right(left) dipolar-k-deletion erases from u a prefix(suffix) v_1 of any length and a suffix(prefix) v_2 of length $\leq k$ whose catenation v_1v_2 (v_2v_1) equals v. The operation can be extended to languages in the natural fashion. If L_1 and L_2 are languages over the alphabet Σ, then the $\star$-dipolar k-deletion of L_2 into L_1 is the language
$L_1 \rightleftharpoons_\star^k L_2 = \bigcup_{u \in L_1, v \in L_2} u \rightleftharpoons_\star^k v$, where $\star$ =left or right.

Now we construct the set $\star$-k-θ-ins(L) using the $\star$-dipolar k-deletion.

Proposition 9. *$\star$-k-θ-ins(L)* $= ((\theta(L))^c \rightleftharpoons_\star^k L)^c$.

Proof. Take $x \in$ right-k-θ-ins(L). Suppose, $x \in ((\theta(L))^c \rightleftharpoons_\star^k L)$ then there exists $u_1xu_2 \in (\theta(L))^c$, $u_1u_2 \in L$, $|u_2| \leq k$ such that $x \in u_1xu_2 \rightleftharpoons_r^k u_1u_2$ which is a contradiction as $x \in$ right-k-θ-ins(L) and $u_1u_2 \in L$, $|u_2| \leq k$, but the right-k-θ-insertion of x into u_1u_2 belongs to $(\theta(L))^c$. Conversely, let $x \in ((\theta(L))^c \rightleftharpoons_r^k L)^c$. If $x \notin$ right-k-θ-ins(L), then there exists $u_1u_2 \in L$, $|u_2| \leq k$ such that $u_1xu_2 \notin \theta(L)$ which implies $u_1xu_2 \in (\theta(L))^c$ and hence $x \in ((\theta(L))^c \rightleftharpoons_r^k L)$ which is a contradiction.

Corollary 2. *If L is regular, then $\star$-k-θ-ins(L) is regular.*

Proof. It has been proven in [1] that if a language L is regular, then $L \rightleftharpoons_\star^k R$ is regular. Since L is regular, $\theta(L)$ is regular and hence $(\theta(L))^c$ is regular which implies $((\theta(L))^c \rightleftharpoons_\star^k L)$ is regular and hence $((\theta(L))^c \rightleftharpoons_\star^k L)^c$ is regular.

Given two words $u, v \in \Sigma^*$, the insertion of v in to u is defined as $u \longleftarrow v = \{u_1vu_2 : u = u_1u_2\}$. The k-insertion was introduced in [1] under the name of k-catenation. The operation of k-insertion restricts the generality of insertion by allowing words to be inserted only in at most $k+1$ positions. The left and the right k-insertions of v into u are respectively the right and the left k-catenation of v in to u
$u \longleftarrow_r^k v = \{u_1vu_2 : u = u_1u_2, |u_2| \leq k\} = u[k]_l v$
$u \longleftarrow_l^k v = \{u_1vu_2 : u = u_1u_2, |u_1| \leq k\} = u[k]_r v$.
The left and the right insertion of a language L_2 in to L_1 can be defined in a natural fashion.

Definition 7. *A language L is $\star$-k-θ-ins-closed iff $L \subseteq \star$-k-θ-ins(L).*

Proposition 10. *L is $\star$-k-θ-ins-closed iff $L \longleftarrow_\star^k L \subseteq \theta(L)$.*

Proof. Let L be right-k-θ-ins-closed. Take $x \in L$ and let $u = u_1u_2 \in L$ such that $|u_2| \leq k$. Then as $x \in L \subseteq$ right-k-θ-ins(L), $u_1xu_2 \in \theta(L)$ which implies $L \longleftarrow_r^k L \subseteq \theta(L)$. Conversely, let $L \longleftarrow_r^k L \subseteq \theta(L)$ and let $x \in L$. To show that $x \in$ right-k-θ-ins(L). Let $u_1u_2 \in L$, $|u_2| \leq k$. Then $L \longleftarrow_r^k L \subseteq \theta(L)$ implies that $u_1xu_2 \in \theta(L)$ which implies $x \in$ right-k-θ-ins(L).

Lemma 5. *For a language $L \subseteq \Sigma^+$ we have :*

1. *When θ is morphic involution, L is $\star$-k-θ-ins-closed iff $\theta(L)$ is $\star$-k-θ-ins-closed.*
2. *When θ is antimorphic involution, L is left(right)-k-θ-ins-closed iff $\theta(L)$ is right(left)-k-θ-ins-closed.*
3. *For $k = 0$, if L is $\star$-k-θ-ins-closed then L^n, $n \geq 1$ is $\star$-k-θ-ins-closed.*
4. *L is $\star$-k-θ-ins-closed and $L^i(L \xleftarrow{k}_{\star} L^n)L^j \subseteq \theta(L^n)$ for all $i, j \geq 0$ such that $i + j = n - 1$, $n \geq 1$ iff L^n is $\star$-k-θ-ins-closed for all $n \geq 1$.*

Given two words $u, v \in \Sigma^*$, the deletion of v in to u is defined as $u \longrightarrow v = \{u_1u_2 : u = u_1vu_2\}$. The notion of k-deletion was introduced in [1] under the name of k-quotient. The operation of k-deletion restricts the generality of deletion by allowing words to be deleted only in at most $k + 1$ positions. The right and left k-deletions of v from u is defined respectively by :

$u \longrightarrow_r^k v = \{u_1u_2 : u = u_1vu_2, |u_2| \leq k\}$
$u \longrightarrow_l^k v = \{u_1u_2 : u = u_1vu_2, |u_1| \leq k\}$.
If $k = 0$, we get the right and the left quotient respectively. The left and the right deletion of a language L_2 in to L_1 can be defined in a natural fashion. The right-k-deletion was called as k-deletion in [1]. We extend these concepts to incorporate the notion of an involution function and hence we define left-k-θ-deletion and right-k-θ-deletion of a given language.

Let $L \subseteq \Sigma^*$ and let right-k-Sub$(L) = \{u \in \Sigma^* : xuy \in L, |y| \leq k\}$ and left-k-Sub$(L) = \{u \in \Sigma^* : xuy \in L, |x| \leq k\}$. The elements of left(right)-k-Sub(L) are called the left(right)-k-subwords. To the language L, one can associate a language $\star$-k-θ-del(L) consisting of all the words with the following property: x is a $\star$-k-subword of atleast one of the word of $\theta(L)$, and the $\star$-k-deletion of x from any word of $\theta(L)$ containing x as a $\star$-k-subword yields word belonging to L. Formally,
right-k-θ-del$(L) = \{x \in$ right-k-Sub$(\theta(L))$: $\forall u \in \theta(L), u = u_1xu_2, |u_2| \leq k, u_1u_2 \in L\}$
left-k-θ-del$(L) = \{x \in$ left-k-Sub$(\theta(L))$: $\forall u \in \theta(L), u = u_1xu_2, |u_1| \leq k, u_1u_2 \in L\}$.

Proposition 11. *If L is a commutative language, then $\star$-k-θ-del(L) is also commutative.*

Proof. It is sufficient to show that $xuvy \in \star$-k-θ-del(L) implies $xvuy \in \star$-k-θ-del(L). If $w \in \theta(L)$, $w = w_1xuvyw_2$ then $w_1w_2 \in L$, but $w_1xvuyw_2 \in \theta(L)$ since L is commutative which implies $xvuy \in \star$-k-θ-del(L).

Proposition 12. *$\star$-k-θ-del(L) $= (\theta(L) \rightleftharpoons_{\star}^{k} L^c)^c \cap \star$-$k$-*Sub$(\theta(L))$.

Proof. Take $x \in \star$-k-θ-del(L). Then $x \in \star$-k-Sub$(\theta(L))$ which implies for every $u \in \theta(L)$, $u = u_1xu_2$, $u_1u_2 \in L$. Suppose , $x \in (\theta(L) \rightleftharpoons_{\star}^{k} L^c)$, then there exists $u \in \theta(L)$ such that $u = u_1xu_2$ with $u_1u_2 \in L^c$ which is a contradiction. Conversely let $x \in \star$-k-Sub$(\theta(L)) \cap (\theta(L) \rightleftharpoons_{\star}^{k} L^c)^c$. Suppose $x \notin \star$-k-θ-del(L) then there exists $u \in \theta(L)$ such that $u = u_1xu_2 \in \theta(L)$ and $u_1u_2 \notin L$ which implies $u_1u_2 \in L^c$ and hence $x \in \theta(L) \rightleftharpoons_{\star}^{k} L^c$ which is a contradiction. Therefore $x \in \star$-k-θ-del(L).

Definition 8. *A language L is called $\star$-k-θ-del closed if $v \in L$, $u_1vu_2 \in \theta(L)$ then $u_1u_2 \in L$. (Note that when $\star$ = =left, then $|u_2| \leq k$ and when $\star$ =right, $|u_1| \leq k$).*

Lemma 6. *Let $L \subset \Sigma^*$.*

1. *When θ is morphic involution, then L is $\star$-k-θ-del-closed iff $\theta(L)$ is $\star$-k-θ-del-closed.*
2. *When θ is antimorphic involution, then L is left(right)-k-θ-del-closed iff $\theta(L)$ is right(left)-k-θ-del-closed.*

Proposition 13. *Let L be such that L is $\star$-k-θ-ins-closed. Then L is $\star$-k-θ-del-closed iff $L = (\theta(L) \longrightarrow^k_\star L)$.*

Proof. Let L be $\star$-k-θ-del-closed. Let $x \in (\theta(L) \longrightarrow^k_\star L)$. To show that $u \in L$. Since $u \in (\theta(L) \longrightarrow^k_\star L)$, $u = u_1u_2$ such that $u_1xu_2 \in \theta(L)$ with $x \in L$. Since L is $\star$-k-θ-del-closed, $u_1u_2 \in L$ which implies $(\theta(L) \longrightarrow^k_\star L) \subseteq L$. To prove the other inclusion, let $u \in L$ and since L is $\star$-k-θ-ins-closed, $u \in L \subseteq \star$-k-θ-ins(L) and $u = u_1u_2$ such that $u_1xu_2 \in \theta(L)$ which implies $u \in \theta(L) \longrightarrow^k_\star L)$. Hence $L \subseteq \theta(L) \longrightarrow^k_\star L)$. Therefore $L = \theta(L) \longrightarrow^k_\star L)$. Conversely, let $L = \theta(L) \longrightarrow^k_\star L)$. Let $v \in L$ with $u_1vu_2 \in \theta(L)$, then $u_1u_2 \in \theta(L) \longrightarrow^k_\star L) = L$ which implies $u_1u_2 \in L$ and hence L is $\star$-k-θ-del-closed.

5 Conclusion

In this paper we have introduced a generalization of the catenation operation and hence have defined the concept of k-suffix code and k-prefix code. We have extended the concept of these codes to involution k-suffix and involution k-prefix codes and have investigated the theoretical properties of these codes in Section 3. We have also extended the notion of k-insertion closure and k-deletion closure of a language to incorporate the notion of an involution function. In Section 4 we have constructed these languages using the dual operation of dipolar k-deletion. As future work, we would like to investigate the algebraic characterizations of these involution codes through their syntactic monoid. The role of such codes in the design of DNA strands with certain properties (see [3–5]) also needs to be further investigated.

Acknowledgment
Research supported by NSERC and Canada Research Chair grants for Lila Kari.

References

1. L.Kari and G.Thierrin, *K-catenation and applications k-prefix codes*, Journal of Information and Optimization Sciences, Vol 16-2 (1995), 263-276.
2. L.Kari and G.Thierrin, *K-Insertion and K-Deletion Closure of Languages*, Soochow Journal of Mathematics, Vol 21-4 (1995), 479-495.

3. S.Hussini, L.Kari and S.Konstantinidis, *Coding Properties of DNA Languages*, Theoretical Computer Science 290 (2003), 1557-1579.
4. L.Kari and K.Mahalingam, *More on Involution Codes*, preprint.
5. N.Jonoska and K.Mahalingam, *Involution Codes: With Application to DNA Coded Languages*, Natural Computing, Vol 4-2(2005), 141-162.
6. H.J.Shyr, *Free Monoids and Languages*, Hon-Min Book Company, Taichung, Taiwan, (1991).
7. J. Berstel, D. Perrin, *Theory of Codes,* Academis Press, Inc. Orlando Florida, 1985.