Commutation with codes

Juhani Karhumäki [a] Michel Latteux [b] Ion Petre [c]

[a] *Department of Mathematics, University of Turku and Turku Centre for Computer Science, Turku 20014, Finland*

[b] *LIFL, URA CNRS 369 Université des Sciences et Technologie de Lille F-59655 Villeneuve d'Ascq, France*

[c] *Department of Computer Science, Åbo Akademi University and Turku Centre for Computer Science, Turku 20520, Finland*

Abstract

The centralizer of a set of words X is the largest set of words $\mathcal{C}(X)$ commuting with X: $X\mathcal{C}(X) = \mathcal{C}(X)X$. It has been a long standing open question due to Conway, 1971, whether the centralizer of any rational set is rational. While the answer turned out to be negative in general, see Kunc 2004, we prove here that the situation is different for codes: the centralizer of any rational code is rational and if the code is finite, then the centralizer is finitely generated. This result has been previously proved only for binary and ternary sets of words in a series of papers by the authors and for prefix codes in an ingenious paper by Ratoandromanana 1989 – many of the techniques we use in this paper follow her ideas. We also give in this paper an elementary proof for the prefix case.

Key words: Codes, Commutation, Centralizer, Conway's problem, Prefix codes.

1 Introduction

The centralizer of a set of words X is the largest set of words $\mathcal{C}(X)$ commuting with X: $X\mathcal{C}(X) = \mathcal{C}(X)X$. It is easy to see that the centralizer is well-defined for any language X – indeed, $\mathcal{C}(X)$ is the union of all languages commuting with X. It is important to note that for any language X, $X^* \subseteq \mathcal{C}(X)$ and $\mathcal{C}(X)$ is a monoid. Conway raised the following problem related to centralizers,

Email addresses: karhumak@it.utu.fi (Juhani Karhumäki), michel.latteux@lifl.fr (Michel Latteux), ion.petre@abo.fi (Ion Petre).

see [8], p.55 (note that Conway uses the term "normalizer"), more than thirty years ago:

Conway's Problem: Is it true that the centralizer of any rational language is rational?

This problem has received recently much attention. In a series of papers by the authors and others, see [5,10,11,13–16,26,22,23], it has been proved that the problem has indeed a positive answer for sets with at most three words and for rational prefix codes. It has also been proved in [14] that the centralizer of any recursive language is Co-RE. However, it has recently been proved in a breakthrough paper [18], see also [17] for related issues, that Conway's problem has a negative answer in general: there are finite languages with non-RE centralizer. The surprising power of finite sets of words is also shown in a related result of [12], showing that the equivalence problem for finite substitutions on ab^*c is undecidable!

Ratoandromanana raised a related question in [23] concerning the commutation with codes. In a paper displaying an impressive array of technical results related to codes she proved that the characterization with prefix codes can be characterized as in free monoids: if X is a prefix code, then for any language L commuting with X, $L = \rho(X)^I$, where $I \subseteq \mathbb{N}$ and $\rho(X)$ is the primitive root of X. In particular, this implies that the centralizer of any prefix code X is $\rho(X)^*$ and thus, Conway's problem has a positive answer for rational prefix codes. Two conjectures are stated in [23]:

Conjecture 1 ([23]) *Two codes commute if and only if they have a common root.*

Conjecture 2 ([23]) *Any code has a unique primitive root.*

These two conjectures, remained open until now, provide evidence that the commutation with codes has very special properties, in particular that Conway's problem may have a positive answer for codes. We prove in this paper that this is indeed the case:

Theorem 1 *The centralizer of any rational code is rational.*

We also prove that the centralizer of any finite code is finitely generated.

It is worth mentioning that throughout the paper we essentially use the techniques of [23], at times refined and extended to codes rather than prefixes. We also give in Section 4 an elementary proof for Ratoandromanana's result [23] that $\mathcal{C}(X) = \rho(X)^*$, for any prefix code X.

2 Definitions

For basic notions and results of Combinatorics on Words we refer to [3,19,20] and for those of Theory of Codes to [2]. For details on the notion of centralizer and the commutation of languages we refer to [14,15,22].

In the sequel, Σ denotes a finite alphabet, Σ^* the set of all finite words over Σ and Σ^ω the set of all (right) infinite words over Σ. We denote by 1 the empty word and by $|u|$ the length of $u \in \Sigma^*$. For a word $u \in \Sigma^*$, u^ω denotes the infinite word $uuu\ldots$, while for a language $L \subseteq \Sigma^*$,

$$L^\omega = \{u_1u_2u_3\ldots \mid u_n \in L, n \geq 1\} \subseteq \Sigma^\omega.$$

For a language $L \subseteq \Sigma^*$, we denote by $l(L)$ the length of a shortest word in L and by $L_{min} = \{u \in L \mid |u| = l(A)\}$.

We say that a word u is a *prefix* of a word v, denoted as $u \leq v$, if $v = uw$, for some $w \in \Sigma^*$. We say that u and v are *prefix comparable* if either $u \leq v$, or $v \leq u$. A language L is called a *prefix code* if no two words of L are prefix comparable. The following result is well-known.

Lemma 2 ([2,21]) *The set of prefix codes forms a free semigroup. In particular, any prefix code has a unique primitive root.*

For a word u and a language L, we say that $v_1\ldots v_n$ is an *L-factorization* of u if $u = v_1\ldots v_n$ and $v_i \in L$, for all $1 \leq i \leq n$. For an infinite word α, we say that $v_1v_2\ldots v_n\ldots$ is an L-factorization of α if $\alpha = v_1v_2\ldots v_n\ldots$ and $v_i \in L$, for all $i \geq 1$. A *relation* over L is an equality $u_1\ldots u_m = v_1\ldots v_n$, with $u_i, v_j \in L$, for all $1 \leq i \leq m$, $1 \leq j \leq n$; the relation is *trivial* if $m = n$ and $u_i = v_i$, for all $1 \leq i \leq m$.

We say that L is a *code* if any word of Σ^* has at most one L-factorization. Equivalently, L is a code if and only if all relations over L are trivial.

The following simple result is often useful in our considerations.

Lemma 3 *For any language $L \subseteq \Sigma^+$ and any $u \in L$, $z \in \mathcal{C}(L)$, $(zu)^\omega \in L^\omega$.*

PROOF. Let $z_1 = z$, $u_1 = u$ and for all $n \geq 1$, define $z_{n+1} \in \mathcal{C}(L)$ and $u_{n+1} \in L$ such that $z_nu_n = u_{n+1}z_{n+1}$. Then, by induction on n, it follows that

$$(z_1u_1)^n = u_2u_3\ldots u_nu_{n+1}z_{n+1}z_n\ldots z_2,$$

for all $n \geq 1$, and so, $(zu)^\omega = u_2u_3\ldots u_n\ldots \in L^\omega$. Indeed, since $1 \notin L$, the two infinite words have arbitrarily long common prefixes, and so they coincide. □

3 Preliminary results

We prove in this section several results related to the commutation of arbitrary sets of words. We will use these results in the following sections when we discuss the commutation with codes and prefixes.

For any sets $R \subseteq \Sigma^*$, $S \subseteq \Sigma^* \times \Sigma^*$ and any nonnegative integer $n \in \mathbb{N}$, we denote by $R_{<n}$, $S_{<n}$ the sets

$$R_{<n} = \{u \in R \mid |u| < n\}, \qquad S_{<n} = \{(u, v) \in S \mid |uv| < n\}.$$

We also denote by R_n, S_n the sets

$$R_n = \{u \in R \mid |u| = n\}, \qquad S_n = \{(u, v) \in S \mid |uv| = n\}.$$

For two sets of words $X, Y \subseteq \Sigma^*$, we say that the product XY is *unambiguous* if $x_1y_1 = x_2y_2$ implies $x_1 = x_2$ and $y_1 = y_2$, for any $x_1, x_2 \in X$ and $y_1, y_2 \in Y$.

Lemma 4 *Let A, B be some subsets of Σ^+ such that the product AB is unambiguous. Then*

(i) If $AB \subseteq BA$, then $AB = BA$ and the product BA is unambiguous.
(ii) For any $n \geq 1$, if $(AB)_{<n} \subseteq (BA)_{<n}$, then $(AB)_{<n} = (BA)_{<n}$.

PROOF. (i) Let $S = A \times B$ be the direct product of A and B and $T = B \times A$ be the direct product of B and A.

According to the hypothesis, we have that $(AB)_n \subseteq (BA)_n$, for all $n \geq 0$. Since AB is unambiguous, we have that AB is isomorphic with $A \times B$, denoted $AB \simeq A \times B \simeq B \times A$. Thus, $(AB)_n \simeq S_n \simeq T_n$, for all n. Consequently, $\mathrm{card}((BA)_n) \geq \mathrm{card}(T_n)$, for all n.

Clearly, the mapping $\phi : T_n \to (BA)_n$, $\phi(b, a) = ba$ is surjective and so, $\mathrm{card}((BA)_n) \leq \mathrm{card}(T_n)$, proving that

$$\mathrm{card}((BA)_n) = \mathrm{card}(T_n) = \mathrm{card}(S_n) = \mathrm{card}((AB)_n).$$

Consequently, since $(AB)_n \subseteq (BA)_n$, for all n, it follows that $(AB)_n = (BA)_n$, for all $n \geq 0$, i.e., $AB = BA$. Also, $(BA)_n \simeq T_n$, i.e., BA is unambiguous.

(ii) This follows using completely similar arguments as for (i). □

Lemma 5 *Let X, Y, Z be languages with $XY = YX$, $XZ = ZX$, XZ is unambiguous and $Y \subseteq Z$. Then $X(Z \setminus Y) = (Z \setminus Y)X$.*

PROOF. Clearly, from Lemma 4(i), ZX is also unambiguous. Let $T = Z \backslash Y$. Then $XT + XY = YX + TX$. If $XY \cap TX \neq \emptyset$, then $xy = tx'$ for some $x, x' \in X$, $t \in T$, $y \in Y$. Since $XY = YX$, $xy = y'x''$ with $x'' \in X$ and $y' \in Y$. Thus, $tx' = y'x''$ with $t, y' \in Z$ and ZX unambiguous. Consequently, $t = y' \in Y$, a contradiction. Thus, $TX \subseteq XT$. The other inclusion can be proved similarly. □

Lemma 6 *Let X and Y be two commuting languages. Then $X_{min}Y_{min} = Y_{min}X_{min}$. Also, if $l(X) = kl(Y)$, for some $k \geq 1$, then $X_{min} = Y_{min}^k$.*

PROOF. It is easy to see based on a length argument that $X_{min}Y_{min} = Y_{min}X_{min}$. Since both X_{min} and Y_{min} are prefix codes then by Lemma 2, there is a prefix T such that $X_{min} = T^i$ and $Y_{min} = T^j$, for some $i, j \geq 1$. If $l(X) = kl(Y)$, then $i = kj$. □

Lemma 7 *Let X, Y be two non-empty languages such that $YX \subseteq XY$. For any $x \in X$ and $y \in Y$, there exist $k > 0$ and $\alpha \in X^+$ such that $(xy)^k\alpha \in X^+$.*

PROOF. It follows from Lemma 2, [23] that there exists $\beta \in X^+$ such that $(yx)^k\beta \in X^+$, for some $k \geq 1$. But then $(xy)^kx\beta = x(yx)^k\beta \in X^+$. □

The following are results of Ratoandromanana [23] that we will use often in our considerations.

Lemma 8 (Lemma 3, [23]) *For any code X and any language Y such that $YX \subseteq XY$ or $XY \subseteq YX$, if $X \cap Y \neq \emptyset$, then $X \subseteq Y$.*

Lemma 9 (Proposition 7, [23]) *Two codes X, Y commute if and only if there are positive integers m, n such that $X^m = Y^n$.*

Lemma 10 (Lemma 10, [23]) *For any code X consider the set*

$$C(X) = \{Y \mid Y \text{ is a code commuting with } X\}.$$

Then $C(X)$ is a commutative stable semigroup. In particular, for any two codes Y, Z commuting with X, YZ is a code and $YZ = ZY$.

4 The commutation with prefix codes

We characterize in this section the commutation with prefix codes, proving that for any prefix X, $\mathcal{C}(X) = \rho(X)^*$ and $LX = XL$ implies $L = \rho(X)^I$,

where $\rho(X)$ is the primitive root of X and $I \subseteq \mathbb{N}$. These results were originally proved in Ratoandromanana [23] using ingenious combinatorial techniques on words and prefix codes. Following the ideas in [23], we give here simpler proofs of those results. There are two crucial ingredients in our proof. First, we prove that the products LX and XL are unambiguous for any language L commuting with X. Second, we prove that for any such L, there is a prefix code $\mathcal{P}(L) \subseteq L$ that commutes with X, thus being able to exploit the fact that the set of prefix codes forms a free monoid. We prove several lemmata first.

Lemma 11 *For any prefix code X and any language L commuting with X, both LX and XL are unambiguous.*

PROOF. The result follows from Lemma 4 and the fact that XL is necessarily unambiguous since X is a prefix. □

For a set of words A over the alphabet Σ, let

$$\mathrm{Com}(A) = \{L \subseteq \Sigma^* \mid LA = AL\} \text{ and } \mathcal{P}(A) = A \setminus A\Sigma^+.$$

Note that $\mathcal{P}(A)$ is a prefix code for any A and if $A \neq \emptyset$, then $\mathcal{P}(A) \neq \emptyset$. Indeed, $A_{min} \subseteq \mathcal{P}(A)$.

Lemma 12 *For any prefix code X, if $L \in \mathrm{Com}(X)$, then $\mathcal{P}(L) \in \mathrm{Com}(X)$.*

PROOF. If $\mathcal{P}(L)X \subseteq X\mathcal{P}(L)$, then we are done by Lemma 4. So, let us assume the contrary and let lx be a shortest word in $\mathcal{P}(L)X \setminus X\mathcal{P}(L)$, with $l \in \mathcal{P}(L)$, $x \in X$ and let $n = |lx|$. Then $(\mathcal{P}(L)X)_{<n} \subseteq (X\mathcal{P}(L))_{<n}$ and thus, by Lemma 4, $(\mathcal{P}(L)X)_{<n} = (X\mathcal{P}(L))_{<n}$.

Since $\mathcal{P}(L)X \subseteq LX = XL$, we have $lx = yku$, with $y \in X$, $k \in \mathcal{P}(L)$, and $u \in \Sigma^+$. Then $yk \in (X\mathcal{P}(L))_{<n} = (\mathcal{P}(L)X)_{<n}$. So, $lx = l'x'u$, with $l, l' \in \mathcal{P}(L)$ and $x, x' \in X$, implying that $l = l'$, $x = x'$, and $u = 1$, a contradiction. □

The following result is proved in [23] in the case of codes, using some involved arguments and results. For the sake of completeness, we give here a simple proof in the case of prefix codes, which are the focus of this section. The techniques used here are essentially those of [23].

Lemma 13 (cf. Lemma 17, [23]) *For any prefix code X and any language L, if $X^iL = LX^i$, for some nonnegative integer i, then $X^i(L \setminus X^*) = (L \setminus X^*)X^i$.*

PROOF. Let $L_1 = L \cap X^*$, $L_2 = L \setminus X^*$. If $X^i L = L X^i$, then

$$X^i L_1 + X^i L_2 = L_1 X^i + L_2 X^i.$$

Let us assume that $X^i L_2 \cap L_1 X^i \neq \emptyset$. Then there are $x_1, x_2 \in X^i$ and $l_1 \in L_1$, $l_2 \in L_2$ such that $x_2 l_2 = l_1 x_1$. Thus, $x_2 l_2 \in X^*$ and, since X is a prefix code, $l_2 \in X^*$, a contradiction. Thus, $X^i L_2 \subseteq L_2 X^i$. Since $X^i L_2$ is unambiguous, it follows by Lemma 4 that $X^i L_2 = L_2 X^i$. □

We are now ready to characterize the centralizer of a prefix code. Based on this characterization we then answer Conway's problem and characterize the commutation with prefix codes.

Theorem 14 *Let X be a prefix code, $\rho(X)$ its primitive root, and $\mathcal{C}(X)$ its centralizer. Then $\mathcal{C}(X) = \rho(X)^*$.*

PROOF. Assume that $\mathcal{C}(X) \neq \rho(X)^*$. Then, by Lemma 13, the language $L = \mathcal{C}(X) \setminus \rho(X)^* \neq \emptyset$ commutes with X and so, by Lemma 12, $\mathcal{P}(L)$ is a prefix code commuting with X. Thus, $\mathcal{P}(L) = \rho(X)^t$, for some nonnegative integer t. This is a contradiction since $L' \subseteq L$ and $L \cap \rho(X)^* = \emptyset$. □

Corollary 15 *For any prefix code X, if the set of words L commutes with X, then $L = \bigcup_{i \in I} \rho(X)^i$, for some $I \subseteq \mathbb{N}$.*

PROOF. To prove the claim of the theorem, it is enough to prove that for any $n \geq 0$, if $L \cap \rho(X)^n \neq \emptyset$, then $\rho(X)^n \subseteq L$. This follows from [23], Lemma 18, but for the sake of completeness, we include a short proof here.

Let $u_1, \ldots, u_n \in \rho(X)$ such that $u_1 \ldots u_n \in L$ and let $\alpha_1, \ldots, \alpha_n$ be arbitrary elements of $\rho(X)$. Let also $X = \rho(X)^k$, $k \geq 1$. Then, since $X^n L = L X^n$ and $(\alpha_1 \ldots \alpha_n)^k \in \rho(X)^{nk} = X^n$, it follows that $u_1 \ldots u_n (\alpha_1 \ldots \alpha_n)^k \in X^n L = \rho(X)^{kn} L$. Since $L \subseteq \rho(X)^*$ and $\rho(X)$ is a prefix, this can only lead to a trivial $\rho(X)$-relation, i.e., $\alpha_1 \ldots \alpha_n \in L$. Thus, $\rho(X)^n \subseteq L$, proving the claim. □

Corollary 16 *Conway's problem has an affirmative answer for rational prefix codes: for any rational prefix code X, both $\rho(X)$ and $\mathcal{C}(X)$ are rational and $\mathcal{C}(X) = \rho(X)^*$.*

PROOF. It is not difficult, see, e.g., [4] or [24], to prove that for any rational language R such that $R = R_0^n$, for some language R_0 and some positive integer n, there is a rational language R_1 such that $R_0 \subseteq R_1$, and $R = R_1^n$. Using this observation it follows that $\rho(X)$ and thus, also $\mathcal{C}(X)$ must be rational. □

5 The commutation with codes

We describe in this section the form of the centralizer of any code. In particular, we prove that the centralizer of any rational code is rational, thus giving a positive answer to Conway's problem in the case of codes. It also follows that the centralizer of any finite code is finitely generated.

One of the crucial ingredients in our proof is that for any code X and any language L commuting with X, the products LX and XL are unambiguous.

Theorem 17 *For any code X, the products $X\mathcal{C}(X)$ and $\mathcal{C}(X)X$ are unambiguous.*

PROOF. Assume that $X\mathcal{C}(X)$ is ambiguous, i.e., there are $x, y \in X$, $u, v \in \mathcal{C}(X)$ such that $xu = yv$ and $x \neq y$. By Lemma 7, there exists $\alpha \in X^+$ such that $(xu)^k\alpha \in X^+$. Let $z = (xu)^k\alpha$. Then

$$z^\omega = ((xu)^k\alpha)^\omega = (x(ux)^{k-1}u\alpha)^\omega = x((ux)^{k-1}u\alpha x)^\omega = x(wx)^\omega,$$

where $w = (ux)^{k-1}u\alpha \in \mathcal{C}(X)$. As it is easy to see, for any $\delta \in \mathcal{C}(X)$ and any $t \in X$, $(\delta t)^\omega \in X^\omega$ and so, $(wx)^\omega \in X^\omega$. Consequently, $z^\omega \in xX^\omega$.

Analogously, $z^\omega = ((yv)^k\alpha)^\omega \in yX^\omega$ and so, z^ω has two different X-factorizations. It is not difficult now to see that this leads to a contradiction. For the sake of completeness, we give here a simple argument on how to conclude it, but note that the same follows also from a result of [9] stating that X is a code if and only if for any $\gamma \in X^+$, γ^ω has exactly one X-factorization.

Assume that there is a word $z \in X^+$ such that z^ω has a second X-factorization $z^\omega = \alpha_1\alpha_2\ldots$, $\alpha_i \in X$, $\alpha_1 \neq z$. By the pigeon hole principle, it follows that there are $i < j$ such that $\alpha_1 \ldots \alpha_i = z^{n_i}\gamma$ and $\alpha_1 \ldots \alpha_j = z^{n_j}\gamma$, for some nonnegative integers $n_i < n_j$ and a proper prefix γ of z. It is easy to see then that $\alpha_1 \ldots \alpha_j = z^{n_j - n_i}\alpha_1 \ldots \alpha_i$, a contradiction since X is a code. □

Corollary 18 *For any code X and any language L commuting with X, the products LX and XL are unambiguous.*

PROOF. If XL were ambiguous, then necessarily $X\mathcal{C}(X)$ would be ambiguous since $L \subseteq \mathcal{C}(X)$. □

Lemma 19 *Let X be a code, n a positive integer, and L a language commuting with X^n, with $l(L) = l(X)$. Then $X \subseteq L$.*

PROOF. Clearly, L_{min} and X_{min} are two commuting prefix codes and since $l(L) = l(X)$, it follows that $L_{min} = X_{min}$.

Since X^n is a code, the product X^nL is unambiguous by Corollary 18. Assume now that there exists a word $x \in X \setminus L$. Let us consider $u = xs^n$ with $s \in L_{min} = X_{min}$. Then $u = xs^{n-1}s \in X^nL = LX^n$. Since $x \notin L$, $u \in (L \setminus X^*)X^n = X^n(L \setminus X^*)$. This is a contradiction since $u = xs^n \in X^n(L \cap X)$. □

The following result was proved in Lemma 24, [23] for prefix codes. We extend it here to arbitrary codes, using essentially the techniques in [23].

Lemma 20 *Let X be a code and L a language commuting with X. If $l(x) = kl(L)$, for some $k > 1$, then there exists a code Y such that $X = Y^k$.*

PROOF. Clearly, $X_{min}L_{min} = L_{min}X_{min}$ and since $l(X) = kl(L)$, it follows that $X_{min} = L_{min}^k$. Thus, $X \cap L^k \neq \emptyset$ and it follows from Lemma 8 that $X \subseteq L^k$.

Let $l_0 \in L_{min}$ and $Y = \{y \in L \mid l_0^{k-1}y, yl_0^{k-1} \in X\}$. We prove that Y is a code and $X = Y^k$. Clearly, $Y \neq \emptyset$, e.g., $L_{min} \subseteq Y$.

Claim 1. If $x = l_1 \dots l_k \in X$, with $l_i \in L$, then $l_2 \dots l_k l_0,\ l_0 l_1 \dots l_{k-1} \in X$.

Proof of Claim 1. We have $u = l_2 \dots l_k (l_0)^k \in L^{k-1}X = XL^{k-1}$, so $u = wy$, with $w \in X$ and $y \in L^{k-1}$. Note that $x(l_0)^k = l_1u = l_1wy$. Since $l_1w \in LX = XL$, we deduce that $l_1w = x'l'$, for some $x' \in X$, $l' \in L$. Consequently, $x(l_0)^k = x'(l'y)$, with $x, x' \in X$, $(l_0)^k, l'y \in L^k$. Since XL^k is unambiguous by Corollary 18, it follows that $l_0^k = l'y$. Now, $l_0 \in L_{min}$ and so, $l' = l_0$ and $y = l_0^{k-1}$. Then, since $l_2 \dots l_k (l_0)^k = wy$, it follows that $w = l_2 \dots l_k l_0 \in X$. The second part of Claim 1 is proved analogously.

Using Claim 1, we can deduce easily Claim 2.

Claim 2. If $x = l_1 \dots l_k \in X$, with $l_i \in L$, then for any $i \in \{1, \dots, k\}$, $l_i l_0^{k-1}, l_0^{k-1} l_i \in X$.

Claim 3. Any word $x \in X \subseteq L^k$ has a unique L-factorization in L^k.

Proof of Claim 3. Assume that $x = l_1 l_2 \dots l_k = l'_1 l'_2 \dots l'_k \in X$, with $l_i, l'_i \in L$, for all $i = 1, 2, \dots, k$. Then $((l_0)^{k-1}l_1)(l_2 \dots l_k) = ((l_0)^{k-1}l'_1)(l'_2 \dots l'_k)$, with $(l_0)^{k-1}l_1, (l_0)^{k-1}l'_1 \in X$ according to Claim 2. Since XL^{k-1} is unambiguous, we

obtain that $(l_0)^{k-1}l_1 = (l_0)^{k-1}l'_1$ and so $l_1 = l'_1$. Then, according to Claim 1, $l_2 \dots l_k l_0 = l'_2 \dots l'_k l_0 \in X$, etc.

Claim 4. If $y \in Y$ and $x = l_1 l_2 \dots l_k \in X$, with $l_i \in L$, then $l_2 \dots l_k y \in X$.

Proof of Claim 4. Since $Y \subseteq L$, $xy \in XL = LX$ and so, $xy = l'_1 x'$, with $x' \in X$ and $l'_1 \in L$. We will prove that $l_1 = l'_1$. Then, $x' = l_2 \dots l_k y \in X$, proving the claim.

Clearly, $x' l_0^{k-1} \in XL^{k-1} = L^{k-1}X$ and so, $x' l_0^{k-1} = l'_2 \dots l'_k x''$, with $x'' \in X$ and $l'_i \in L$. It follows from the definition of Y that $u = y l_0^{k-1} \in X$. Consequently,

$$(l_1 l_2 \dots l_k)(y l_0^{k-1}) = xy l_0^{k-1} = l'_1 x' l_0^{k-1} = (l'_1 l'_2 \dots l'_k)x''.$$

Since $L^k X$ is unambiguous by Corollary 18, it follows that $l_1 l_2 \dots l_k = l'_1 l'_2 \dots l'_k$ Now, $l_1 l_2 \dots l_k = x \in X$ and it follows by Claim 3 that $l_1 = l'_1$, concluding the proof of Claim 4.

We can prove now that $X \subseteq Y^k$. For this, let $x \in X$. As observed in the beginning of the proof, $X \subseteq L^k$ and so, $x = l_1 \dots l_k$, with $l_i \in L$. From Claim 2 it follows that $l_i l_0^{k-1}, l_0^{k-1} l_i \in X$ for all $i = 1, 2, \dots, k$ and so, $l_i \in Y$, for all i. Consequently, $X \subseteq Y^k$.

For the reverse inclusion, consider $y_1, \dots, y_k \in Y$ and $x = l_1 \dots l_k \in X$, with $l_i \in L$. It follows from Claim 4 by induction that $l_i \dots l_k y_1 \dots y_{i-1} \in X$, for all $i = 2, 3, \dots, k$. Thus, $y_1 \dots y_k \in X$, i.e., $Y^k \subseteq X$. It follows then by Claim 3 that $X = Y^k$. It also follows that Y is a code, concluding the proof. □

Lemma 21 *Let X be a code and $L \subseteq \Sigma^+$ be a language commuting with X. Then there exists a code Y commuting with X such that $L_{min} = Y_{min}$ and $Y \subseteq L$. Moreover, if X is rational, then Y is rational.*

PROOF. Set $t = l(X)$ and $s = l(L)$. Since $LX^s = X^s L$, X^s is a code and $l(X^s) = tl(L)$, it follows from Lemma 20 that there exists a code Y such that $Y^t = X^s$. Then $LY^t = Y^t L$, with $l(Y) = l(L)$ and so, $L_{min} = Y_{min} \subseteq Y$ implying by Lemma 19 that $Y \subseteq L$. Moreover, from Lemma 9 we also obtain that Y is commuting with X.

Observe now that if X is a rational code, then X^s and so, Y^t, is a rational code. It follows then that Y is rational. □

The following result describes the form of all monoids commuting with a given code.

Theorem 22 *For any code X and any monoid M commuting with X, there exist codes $C_1, \ldots, C_k$ commuting with X such that $M = (C_1 \cup \ldots \cup C_k)^*$. Moreover, if X is rational, then M is rational.*

PROOF. Let $M_0 = M \setminus \{1\}$. It is a result of [23] (Lemma 4, [23]) that $M_0X = XM_0$. Thus, by Lemma 21, there exists a code $C_1 \subseteq M_0$ commuting with X with $(C_1)_{min} = (M_0)_{min}$. Let $B_1 = C_1$.

For all $i \geq 1$ consider $B_i = C_1 \cup \ldots \cup C_i \subseteq M_0$ and $M_i = M \setminus B_i^*$. Since M is a monoid, $B_i^* \subseteq M$ and so, by Lemma 5, we have that $M_iX = XM_i$. If $M_i \neq \emptyset$, then by Lemma 21 there exists a code $C_{i+1} \subseteq M_i$ commuting with X such that $(C_{i+1})_{min} = (M_i)_{min}$.

Assume that for all $j \geq 1$, $M_j \neq \emptyset$ and set $d = \gcd\{l(C_j) \mid j \geq 1\}$. Then $d = \gcd\{l(C_1), l(C_2), \ldots, l(C_n)\}$, for some $n \geq 1$. Clearly, by construction, $l(C_p) < l(C_{p+1})$, for all $p \geq 1$. Thus, there is $h > n$ such that $l(C_h) = t_1 l(C_1) + \ldots + t_n l(C_n)$, for some nonnegative integers $t_1, \ldots, t_n$. Let us consider $Y = (C_1)^{t_1} \ldots (C_n)^{t_n}$. From Lemma 10 it follows that Y is a code commuting with C_h. Since $l(C_h) = l(Y)$, we get that $Y_{min} = (C_h)_{min}$, hence $C_h \cap Y \neq \emptyset$. Consequently, $C_h = Y \subseteq B_n^*$, a contradiction since $C_h \subseteq M_n = M \setminus B_n^*$.

Let now k be the least integer such that $M_k = \emptyset$. Then $M = (C_1 \cup \ldots C_k)^*$.

The second part of the claim follows from Lemma 21: $C_1, \ldots, C_k$ are rational and so, M is rational. □

The main result of this paper follows now as a simple consequence of Theorem 22 since the centralizer of any language is a monoid.

Theorem 23 *The centralizer of any rational code is rational.*

The following result also follows from Theorem 22 in the case of finite codes.

Theorem 24 *Any monoid commuting with a finite code is finitely generated. In particular, the centralizer of a finite code is a finitely generated monoid.*

PROOF. Let X be a finite code and M a monoid commuting with X. Then by Theorem 22 $M = (C_1 \cup \ldots \cup C_k)^*$ with C_i codes commuting with X, for all $i = 1, 2, \ldots, k$. Thus, by Lemma 9, $C_i^{t_i} = X^{s_i}$, for some positive integers t_i, s_i. Thus, each C_i is finite, proving the claim. □

6 Conclusions

The behavior of codes under commutation is special. While the centralizer of a finite set is not necessarily recursively enumerable, we describe here the form of the centralizer of a code and prove that it is necessarily rational if the code is rational. Moreover, if the code is finite, then the centralizer is finitely generated. The crucial difference between codes and arbitrary sets of words seems to be in the fact that for a code X, the product $X\mathcal{C}(X)$ is unambiguous, as proved in Theorem 17.

We also give in this paper a simple, self-contained proof for the case of prefix codes, proving that for any prefix code X, $\mathcal{C}(X) = \rho(X)^*$, a result originally proved in [23].

In proving our results, we exploited a series of deep results on commutation proved in Ratoandromanana [23]. Two conjectures proposed in [23], related to commutation with codes, remain however open.

Conjecture 1 (Conjecture 1, [23]) *Two codes commute if and only if they have a common root.*

Conjecture 2 (Conjecture 2, [23]) *Any code has a unique primitive root.*

Two other conjectures has been given in the literature in connection with the commutation of codes, see, e.g., [11,14,15].

Conjecture 3 *The centralizer of a code is a free monoid.*

Conjecture 4 *For any code X, if $LX = XL$, then there is a code R such that $X = R^m$ and $L = R^I$, for some $m \geq 1$, $I \subseteq \mathbb{N}$.*

Note that the characterization conjectured above holds for the commutation of polynomials and formal power series with coefficients in a field, see [1,6,7,25].

We prove here that in fact Conjectures 1-4 are equivalent.

Theorem 25 *Conjectures 1-4 are equivalent.*

PROOF. Let X be a code.

We prove first that Conjectures 1 and 2 are equivalent. Considering that Conjecture 1 holds, assume that the code X has two distinct primitive roots Y and Z, $X = Y^i = Z^j$. It then follows from Lemma 9 that Y and Z commute and according to Conjecture 1, they have a common root. Since they

are primitive, it follows that $Y = Z$, a contradiction. To prove the reverse implication, assume that Conjecture 2 holds and consider now two commuting codes X, Y and their unique primitive roots U, V: $X = U^s$, $Y = V^t$. Then, by Lemma 9, $X^i = Y^j$, for some $i, j > 0$ and so, U, V are primitive roots of the code $U^{si} = V^{tj}$. It follows then from Conjecture 2 that $U = V$, i.e., X, Y have a common root.

We prove now that Conjectures 1 and 2 imply Conjecture 3. Let Z be the primitive root of the code X. Then Z^* commutes with X and so, $Z^* \subseteq \mathcal{C}(X)$. Assume that $\mathcal{C}(X) \setminus Z^* \neq \emptyset$. Then, by Lemma 5, $\mathcal{C}(X) \setminus Z^*$ commutes with X and then, by Lemma 21, it follows that there is a code $Y \subseteq \mathcal{C}(X) \setminus Z^*$ such that $XY = YX$. Thus, by Lemma 10, $YZ = ZY$ and so, from Conjecture 1 it follows that there is a code R such that $Y = R^m$, $Z = R^n$. Since Z is primitive, we have $Y = Z^m$, contradicting the fact that $Y \cap Z^* = \emptyset$.

We prove now that Conjecture 3 implies Conjecture 4. It follows from Conjecture 3 that $\mathcal{C}(X) = Z^*$, for some code Z. It follows then from Theorem 22 that $XZ = ZX$ and then from Lemma 9 that $X^i = Z^j$, for some $i, j > 0$. Consider now a language L commuting with X. Then L commutes also with X^i, i.e., with Z^j. Since $L \subseteq \mathcal{C}(X) = Z^*$, it follows from Lemma 18 of [23] that $L = Z^I$, where $I = \{i \geq 0 \mid Z^i \subseteq L \neq \emptyset\}$, concluding Conjecture 4. Note that this also implies that Z is the unique primitive root of X.

We prove now that Conjecture 4 implies Conjecture 1. Consider two codes X, Y such that $XY = YX$. It then follows from Conjecture 4 that there is a set V such that $X = V^I$, $Y = V^J$, for some $I, J \subseteq \mathbb{N}$. Then necessarily V is a code, I, J are singletons and V is a common root of X and Y. □

Acknowledgements

The authors gratefully acknowledge the detailed referee reports that helped to improve the presentation of the paper. Juhani Karhumäki was supported by Academy of Finland under grant 44087. Ion Petre was supported by Academy of Finland under grant 203667.

References

[1] Bergman, G.: Centralizers in free associative algebras, *Transactions of the American Mathematical Society* 137 (1969) 327–344.

[2] Berstel, J., Perrin, D.: *Theory of Codes*, Academic Press, New York (1985).

[3] Choffrut, C., Karhumäki, J.: Combinatorics of Words. In Rozenberg, G., Salomaa, A. (eds.), *Handbook of Formal Languages*, Vol. 1, Springer-Verlag (1997) 329-438.

[4] Choffrut, C., Karhumäki, J.: On Fatou properties of rational languages, in Martin-Vide, C., Mitrana, V. (eds.), *Where mathematics, Computer Science, Linguistics and Biology Meet*, Kluwer, Dordrecht (2000).

[5] Choffrut, C., Karhumäki, J., Ollinger, N.: The commutation of finite sets: a challenging problem, *Theoret. Comput. Sci.*, 273 (1-2) (2002) 69–79.

[6] Cohn, P.M.: Factorization in noncommuting power series rings, *Proc. Cambridge Philos. Soc.* 58 (1962) 452–464.

[7] Cohn, P.M.: Centralisateurs dans les corps libres, in Berstel, J. (ed.), *Séries formelles*, Paris, (1978) 45–54.

[8] Conway, J.H.: *Regular Algebra and Finite Machines*, Chapman Hall (1971).

[9] Devolder, J., Latteux, M., Litovsky, I., Staiger, L.: Codes and infinite words, *Acta Cybernetica* 11 (1994) 241–256.

[10] Karhumäki, J.: Challenges of commutation: an advertisement, in *Proc. of FCT 2001*, LNCS 2138, Springer (2001) 15–23.

[11] Karhumäki, J., Latteux, M., Petre, I., The commutation with ternary sets of words, Theory of Computing Systems, to appear, 2005.

[12] Karhumäki, J., Lisovik, L., The equivalence problem for finite substitutions on ab^*c, with applications, IJFCS 14 (2003), 699-710; preliminary version in Springer Lecture Notes in Computer Science 2380, (2002), 812-820.

[13] Karhumäki, J., Petre, I.: On the centralizer of a finite set, in *Proc. of ICALP 2000*, LNCS 1853, Springer (2000) 536–546.

[14] Karhumäki, J., Petre, I.: Conway's Problem for three-word sets, *Theoret. Comput. Sci.*, 289/1 (2002) 705–725.

[15] Karhumäki, J., Petre, I.: Conway's problem and the commutation of languages, *Bulletin of EATCS* 74 (2001) 171–177.

[16] Karhumäki, J., Petre, I.: The branching point approach to Conway's problem, LNCS 2300, Springer (2002) 69–76.

[17] Kunc, M., Regular solutions of language inequalities and well quasi-orders, in *Proc. of ICALP 2004*, LNCS 3142, 870–881, Springer, 2004.

[18] Kunc, M., The power of commuting with finite sets of words, to appear in Proc. STACS 2005.

[19] Lothaire, M.: *Combinatorics on Words* (Addison-Wesley, Reading, MA., (1983).

[20] Lothaire, M.: *Algebraic Combinatorics on Words* (Cambridge University Press), (2002).

[21] Perrin, D., Codes conjugués, Infromation and Control 20, 222-231 (1972).

[22] Petre, I.: *Commutation Problems on Sets of Words and Formal Power Series*, PhD Thesis, University of Turku (2002).

[23] Ratoandromanana, B.: Codes et motifs, *RAIRO Inform. Theor.*, 23(4) (1989) 425-444.

[24] Restivo, A.: Some decision results for recognizable sets in arbitrary monoids, in *Proc. of ICALP 1978*, LNCS 62 Springer (1978) 363–371.

[25] Reutenauer, C.: Centralisers of noncommutative series and polynomials. In: Lothaire, M. (ed.), *Algebraic Combinatorics on Word*, Cambridge University Press, Cambridge, USA (2002), 312–329.

[26] Salomaa, A., Yu, S.: On the decomposition of finite languages. In G. Rozenberg and W. Thomas (eds.), *Developments in Language Theory*, World Scientific, 22-31 (2000).