Conjugacy of finite biprefix codes

Julien Cassaigne a, Juhani Karhumäki b, Petri Salmela b,*

a Institut de Mathématiques de Luminy—CNRS/FRUMAM, Case 907, FR-13288 Marseille Cedex 9, France
b Department of Mathematics and TUCS, University of Turku, FI-20014 University of Turku, Finland

ARTICLE INFO

Keywords:
Conjugacy
Language equation
Biprefix code

ABSTRACT

Two languages X and Y are called conjugates, if they satisfy the conjugacy equation \( XZ = ZY \) for some non-empty language \( Z \). We will compare solutions of this equation with those of the corresponding equation of words and study the case of finite biprefix codes \( X \) and \( Y \). We show that the maximal \( Z \) in this case is rational. We will also characterize \( X \) and \( Y \) in the case where they are both finite biprefix codes. This yields the decidability of the conjugacy of two finite biprefix codes.

1. Introduction

The conjugacy equation \( xz = zy \) is a basic equation for words. Words \( x \) and \( y \) are conjugates, i.e., they satisfy the conjugacy equation for some word \( z \) if and only if \( x \) and \( y \) have factorizations \( x = pq \) and \( y = qp \) with some words \( p \) and \( q \), and then the above \( z \) can be expressed as \( z = (pq)^k p \).

For languages we say that languages \( X \) and \( Y \) are conjugates, if they satisfy the conjugacy equation \( XZ = ZY \) for some non-empty language \( Z \). For empty set \( Z \) the conjugacy equation always holds. We also restrict our research on languages \( X \) and \( Y \) which do not include empty word since we concentrate on finite biprefix codes. We can also note, that not all biprefix codes \( X \) and \( Y \) are conjugates. For example with \( X = \{a\} \) and \( Y = \{b\} \) the conjugacy equation \( aZ = Zb \) does not have any non-empty solution \( Z \). The conjugacy equation on languages is not equally easy to solve as the same equation on words. Formula of general solutions of conjugacy equation on words can be extended to languages simply by replacing words \( x, y, z, p \) and \( q \) in the formula by languages \( X, Y, Z, P \) and \( Q \). However in several cases this formula does not include all possible solutions. For example, as observed in [2], the solution \( X = \{a, ab, abb, ba, babb\}, Y = \{a, ba, bba, bbbab\}, Z = \{a, ba\} \) is not of this type. However, for some special classes of languages all solutions can be obtained essentially with the same formula as for the conjugacy of words. To analyze this is the topic of this note.

In this paper we first define the so-called word type solutions of conjugacy equation on languages. As a starting point, we note that the solutions for words can be expressed as \( x = (pq)^k \), \( y = (qp)^k \) and \( z = (pq)^i p \) with some integers \( i, k \) and primitive word \( pq \). This formulation of solutions is equivalent to the standard one, which was mentioned in the beginning. This formulation, however, has some advantages. For language equations we refer to solutions of form

\[
X = (PQ)^k, \quad Y = (QP)^k \quad \text{and} \quad Z = \bigcup_{i \in I} (PQ)^i p
\]

with primitive (see below) languages \( PQ \) as word type solutions. This notion has been defined in [2], however, our definition in Section 3 is a slight extension.
Now, we describe our four results. First we define and study the conjugator of \(X\) and \(Y\), that is the largest language \(Z\) (with respect to the subset relation) such that \(XZ = ZY\). We show that for finite biprefix codes \(X\) and \(Y\) the conjugator is rational, in fact, even of form \(X^*U\) for some finite language \(U\).

After this we characterize finite biprefix codes \(X\) and \(Y\) satisfying the conjugacy equation with some non-empty language \(Z\). We show that these languages can always be factorized as \(X = UV\) and \(Y = VU\) for some biprefix codes \(U\) and \(V\). This is achieved by rather complicated combinatorial analysis. However, this factorization is not necessarily unique, but we also provide a unique representation.

Our last result proves that the conjugacy problem for finite biprefix codes, i.e., the problem, whether given finite biprefix codes \(X\) and \(Y\) are conjugates, is decidable. This is shown as corollary of the previous results and the fact that the set of all biprefix codes is the free monoid. In the case of arbitrary finite language the problem is open, and does not seem to be easy, see [8].

2. Preliminaries

Let \(A\) be a finite alphabet, and \(A^*\) the free monoid generated by \(A\). Lowercase letters are used to denote words, i.e., elements of \(A^*\), and uppercase letters languages, i.e., subsets of \(A^*\). The empty word will be denoted by \(1\). For words notation \(|w|\) means the length of word \(w\) and for languages \(|X|\) is the cardinality of \(X\). Language is uniform, if all its elements have the same length.

Notation \(\text{Pref}(X)\) is used for the set of all prefixes of words in \(X\), and similarly \(\text{Suf}(X)\) means all suffixes of words in \(X\). Empty word and words in \(X\) are included. We use also a shorthand \(L^i\) for the union of powers \(\bigcup_{j\geq i} L^j\). Notation \(L^{\leq n}\) is a shorthand for \(\bigcup_{i\leq n} L^i\). The language \(L\) is called primitive, if \(L = K^i\) implies \(L = K\) and \(i = 1\), i.e., if the language \(L\) is a proper power of any other language. If the language is not primitive it is non-primitive or just non-primitive.

We note that the representation \(X = K^i\) with \(K\) primitive is closely related to prime factorizations of languages. Such a research was initiated in [14], and shown to be a rich research topic in [7].

When we say that an element \(w\) in language \(L\) is prefix (resp. suffix) incomparable, we mean that neither \(w\) is a prefix (resp. suffix) of any other word in \(L\) nor any other word in \(L\) is a prefix (resp. suffix) of \(w\). Sometimes this kind of element is also called left (resp. right) singular in \(L\). (see [9, 16] or [13]) The language \(L\) is a prefix (resp. suffix) code or just prefix (resp. suffix), if all elements in \(L\) are left (resp. right) singular.

If the language \(L\) is both prefix and suffix code, we say it is biprefix code or just biprefix. It is known, that the families of prefix, suffix and biprefix codes are free monoids \([1, 15]\). This means that each prefix (resp. suffix or biprefix) code has unique factorization as catenation of indecomposable prefix (resp. suffix or biprefix) codes. This also means that prefix (resp. suffix or biprefix) set can be viewed as a word over a special alphabet of indecomposable prefix (resp. suffix or biprefix) codes. The free base of each of these monoids is infinite, but in many considerations only finite subsets are needed. We also recall that for any prefix (resp. suffix or biprefix) code \(L\) there always exists the unique primitive root \(\rho(L)\), see [1, 15]. For codes the existence of the primitive root is an open problem, see [9], while for arbitrary sets it is not unique, see, e.g., [4].

The following simple fact is needed in many later considerations. Any solution \(Z\) of the conjugacy equation \(XZ = ZY\) satisfies \(Z \subseteq \text{Pref}(X^*) \cap \text{Suf}(Y^*)\). This is clear, since obviously also \(X^nZ = ZY^n\) for any integer \(n\), and so for any words \(z \in Z\) and \(y \in Y\) there exists words \(x, x' \in X\) and \(z' \in Z\) such that \(zy|z| = x_1 \cdots x_{|y|} z'\). This means, since \(|z| < |x_1| + \cdots + |x_{|y|}|\), that \(z\) is a prefix of \(x_1 \cdots x_{|y|} \in X^{|z|}\), i.e., \(z \in \text{Pref}(X^*)\). Dually, \(z\) is also suffix of some word in \(Y^*\).

3. Word type solutions

We recall that the conjugacy equation \(xz = zy\) for non-empty words has the general solution

\[
\exists p, q \in \Sigma^* \quad \text{s.t.} \quad x = pq, \quad y = qp \quad \text{and} \quad z \in (pq)^*p. \tag{1}
\]

This motivates the notion of word type solution of conjugacy equation of the languages. In [2] this has been straightforwardly defined as:

\[
X = PQ, \quad Y = QP \quad \text{and} \quad Z = (PQ)^lP \tag{2}
\]

for languages \(P, Q\) and set \(l \in \mathbb{N}\). We call these solutions word type 1 solutions.

However, there is also a slightly more general way to define word type solution. The condition (1), in the case of words, is equivalent to the condition

\[
\exists p, q \in \Sigma^* \quad \text{s.t.} \quad x = (pq)^k, \quad y = (qp)^k \quad \text{and} \quad z \in (pq)^*p. \tag{3}
\]

where \(pq\) and \(qp\) are primitive words. This motivates to define, word type solution of languages as:

\[
X = (PQ)^k, \quad Y = (QP)^k \quad \text{and} \quad Z = (PQ)^lP \tag{4}
\]

for languages \(P, Q\) such that \(PQ\) and \(QP\) are primitive, integer \(k\) and set \(l \subseteq \mathbb{N}\).
We call such solutions word type 2 solutions, clearly they include all word type 1 solutions. Unlike in the case of words these notions are not equivalent in the case of languages, as shown in the next example.

**Example 1.** Let $X = BCBC$ and $Y = CBBC$ for $B = \{b\}$ and $C = \{c\}$ (or some other biprefix code). Now both solutions

\[ P_1 = B, \quad Q_1 = CBC, \quad X = P_1Q_1, \quad Y = Q_1P_1, \quad Z_1 = P_1Q_1P_1 = (BCBC)B \]

and

\[ P_2 = BCBC, \quad Q_2 = C, \quad X = P_2Q_2, \quad Y = Q_2P_2, \quad Z_2 = P_2Q_2P_2 = (BCBC)CB \]

are of word type in the sense of (2), but their union $Z_1 \cup Z_2 = BCBCB \cup CBBCBCB$ is not. However, if we would use (4) as the definition of word type solution, we would have

\[ P = B, \quad Q = C, \quad X = (PQ)^2, \quad Y = (QP)^2, \quad Z_1 = (PQ)^2P = (BC)^2B, \]

\[ P = B, \quad Q = C, \quad X = (PQ)^2, \quad Y = (QP)^2, \quad Z_2 = (PQ)^2P = (BC)^2B \]

and

\[ Z = Z_1 \cup Z_2 = (PQ)^{(2,3)}P. \]

Based on above, we choose (4) for our definition of word type conjugation of languages.

**4. The conjugator**

For the commutation equation $XY = YX$ there has been active research on the centralizer, that is on the largest language commuting with given language $X$. J.H. Conway asked in [6], whether the centralizer of given rational language is rational as well. This, so-called Conway’s problem, was open for a long time and has been solved negatively in general [12], but has proven to have positive answers in several special cases like sets with at most two elements [5], rational codes [9], three-element sets [10] and languages with certain special elements [13].

For the conjugacy equation $XZ = ZY$ we can similarly study the maximal solution $Z$ for given languages $X$ and $Y$. The maximal solution exists and is the unique largest one. We call this solution the conjugator. In the case that $X$ and $Y$ are not conjugates the maximal (and only) solution is the empty set. If $X$ and $Y$ are conjugates, and conjugated via languages $Z_i$ for $i$ in some index set $I$, then they are, by the distributivity of catenation and union operations, conjugated also via the union $\bigcup_{i \in I} Z_i$. Hence the unique maximal solution is the union of all solutions $Z$. The special case where $X = Y$ gives us the centralizer of $X$.

We can ask the question similar to the Conway’s problem, namely whether the conjugator of given languages $X$ and $Y$ is rational. The general answer is of course negative, since the original Conway’s problem has a negative answer. However, we can again study some special cases. In what follows we use similar reasoning for conjugacy as has been used for commutation in [11]. First we need the following lemma.

**Lemma 2 (Interchange Lemma).** If $X$ and $Y$ are 1-free languages, such that $Y$ has a suffix incomparable element $y$ and $XZ = ZY$ for some language $Z$, then for each word $z \in Z$ there exist an integer $n$ and a word $u \in \text{Pref}(X) \setminus X$ such that $z = x_1x_2 \cdots x_nu$ for some $x_i \in X$, and moreover $X^nu \subseteq Z$.

**Proof.** Let $X$ and $Y$ be 1-free languages, $y$ a suffix incomparable element in $Y$, and $Z$ such that $XZ = ZY$. Then for each $z \in Z$ there exist an integer $n$ and factorization $z = x_1x_2 \cdots x_nu$ such that $x_i \in X$, $u \in \text{Pref}(X) \setminus X$ and

$$zy^n = x_1x_2 \cdots x_nuy^n \in Zy^n = X^nZ$$

with $uy^n \in Z$. Then again

$$x_1x_2^2 \cdots x_n^uy^n \in X^nZ = Zy^n$$

where $x_i^u$ are arbitrary elements from $X$. This shows that $X^nu \subseteq Z$, since $y$ is suffix incomparable in $Y$. \hfill \Box

**Theorem 3.** For finite languages $X$ and $Y$, such that $Y$ has suffix incomparable element $y$, the conjugator is rational.

**Proof.** Let $X$ and $Y$ be finite languages, $y$ a suffix incomparable element in $Y$, and $Z$ their conjugator. By Lemma 2 for each word $z \in Z$ we have $z \in X^nu \subseteq Z$ for some integer $n$ and word $u \in \text{Pref}(X)$. Since $X^2Z = XZY$ the language $XZ$ is included in the conjugator $Z$. Hence also $X^2Z \subseteq Z$ and $X^2X^nu \subseteq Z$.

Let $U \subseteq \text{Pref}(X)$ be the set of all words $u$ occurring in the above constructions. Since the language $X$ is finite, so is $U$. Now, for each $u \in U$, there exists minimal integer $n_u$ such that $X^X^nu \subseteq Z$ and each word $z \in Z$ is in one of these sets. Hence we conclude that the conjugator of $X$ and $Y$ is

$$Z = X^* \left( \bigcup_{u \in U} X^nu \right).$$

This set is rational, since the set $\bigcup_{u \in U} X^nu$ is finite. Note that if $X$ and $Y$ are not conjugates, then $Z$ is the empty set. \hfill \Box
The proof of this previous theorem is not constructive, since it needs the conjugator to be given. Hence the result is non-effective.

In a suffix set all elements are suffix incomparable, therefore this result holds in the case of finite bifixed codes $X$ and $Y$.

Finally, we make a remark that interchange lemma can also be proven in a sharper form using the primitive root $\rho(X)$ instead of the language $X$. This way we obtain that $u \in \text{Pref}(\rho(X)) \setminus \rho(X)$, $z = r_1 r_2 \cdots r_n u$ for some $r_i \in \rho(X)$ and $\rho(X)^n u \subseteq Z$. This gives us a smaller number of words $u$.

5. Characterization of conjugacy of finite bifixed codes

In this section we characterize, when finite bifixed codes $X$ and $Y$ are conjugates. The fact that set of bifixed codes is a free monoid suggests that this conjugacy would be similar to the conjugacy of words, i.e., of word type. However, we cannot use this freeness property to characterize $X$ and $Y$, since we do not know for sure, if the solution $Z$ is also in this free monoid of bifixed prefixes or even a union of such bifixed solutions. Hence we are tied to a complicated analysis as in the case of determining the centralizer of a prefix code, see [16]. When we have obtained this characterization, we are able, in Section 6, finally to prove that $Z$ indeed is a union of such bifixed solutions.

We can also note, that using looser condition, where $X$ is a prefix code and $Y$ is a suffix code, does not guarantee the conjugacy to be word type. As an example we can have languages $X = \{aba, bda\}$ and $Y = \{aba, ab\}$, which are prefix and suffix respectively. These languages are conjugates for example via language $Z = \{b, ab, ba, aba\}$, but their conjugacy is not word type.

In what follows, we assume that $X$ and $Y$ are finite bifixed codes such that $XZ = YZ$ for some non-empty language $Z$.

**Lemma 4.** For every integer $n \geq \min|\{x| x \in X\}$ there exist finite bifixed codes $U_n$ and $V_n$ satisfying

$$
X \cap A^{\leq n} = U_n V_n \cap A^{\leq n} \quad \text{and} \quad Y \cap A^{\leq n} = V_n U_n \cap A^{\leq n}.
$$

**Proof.** Let $X_0$, $Y_0$, $Z_0$ be the sets of elements in $X$, $Y$, $Z$ of minimal lengths and $n_0 = \min|\{x| x \in X\}$. Then, since $X_0$, $Y_0$ and $Z_0$ are uniform languages, $X_0 Z_0 = Y_0 Z_0$ holds and the solution is of word type, see [2]. This means that $X_0 = U_{n_0} V_{n_0}$, $Y_0 = V_{n_0} U_{n_0}$ and $Z_0 = (U_{n_0} V_{n_0})^m U_{n_0}$ for some uniform $U_{n_0}$ and $V_{n_0}$ and integer $m \geq 0$. Hence (5) holds for $n = n_0$.

Let us choose $u_0 \in U_{n_0}$, $v_0 \in V_{n_0}$ and $z_0 = (u_0 v_0)^m u_0 \in Z_0$. We assume, inductively, that we have already constructed $U_i$ and $V_i$ for $n_0 \leq i < n$ and construct $U_n$ and $V_n$ for $n > n_0$ satisfying (5), so that $U_{n_1} \subseteq U_n$ and $V_{n_1} \subseteq V_n$.

First we show that $U_{n_1} V_{n_1} \cap A^{\leq n_1} \subseteq X$ and $V_{n_1} U_{n_1} \cap A^{\leq n_1} \subseteq Y$. Let $u \in U_{n_1}$, $v \in V_{n_1}$ such that $|uv| = n$, if such elements exist. Then $|u_0 v_0| < n$ and $|u_0 v_0| < n$, so $u_0 v_0 u \in V$ and $v_0 u_0 v \in Y$. Now $z_0 v_0 u_0 v_0 \in Z = Z^2 = XZ$ and by regrouping elements we have

$$
z_0 v_0 u_0 v_0 (v_0 u_0)^m = (u_0 v_0)^m v_0 z_0 \in Z \quad \text{and}
$$

and hence $X$ is bifixed, we get $u v_0 z_0 \in Z$. Hence $u v_0 z_0 = x z$ with $x \in X$ and $z \in Z$. Here $|z| \geq |Z_0|$, i.e., $x$ is a prefix of $u v$. If $|x| < n$, i.e., $x$ is a proper prefix of $u v$, then also $x \in U_{n_1} V_{n_1} = x \cap Z = \text{Pref}(\rho(X)) \setminus \rho(X)$ and this is a contradiction, since $U_{n_1} V_{n_1}$ is a bifixed code. Therefore $|x| = n$ and $x = u v \in X$. Similarly, $u v \in Y$ and so $U_{n_1} V_{n_1} \cap A^{\leq n_1} \subseteq Y$.

Next we deal with the words in $X \cap A^n \setminus U_{n_1} V_{n_1} = X$, and in $Y \cap A^m \setminus V_{n_1} U_{n_1} = Y$, and show that some words can be added to $U_{n_1}$ and $V_{n_1}$ to form $U_n$ and $V_n$, still satisfying (5).

If there exists $x \in X \cap A^n \setminus U_{n_1} V_{n_1}$, then

$$
(U_0 V_0)^m x z_0 = z_0 v_0 x u_0 (v_0 u_0)^m x Z \in Z = Z^2 + Z
$$

and hence $Y$ is bifixed, $z_0 v_0 x u_0 \in Z^2$. Therefore $z_0 v_0 x u_0 \in Z^2$. Consequently, $U_0 V_0 x u_0 \in Z^2$ and $u_0 v_0 z_0 = x y z'$ for some $y, y' \in Y$ and $x \in X$, $|z| \geq |z_0|$, see Fig. 1 for illustration. Now $y \in A^m$ and $|u_0 v_0| \leq n_0 \leq n \leq |y| \leq |y_0| - |y'| \leq |y| \leq n$. So $y' = v_0 u_0$, where $v_0 u_0$ is a suffix of $x$. We have two cases:

(i) If $|y'| < n$, then $y' = v_0 u_0 \in V_{n_1} U_{n_1}$ and, since $U_{n_1}$ is a bifixed code, $y' \in V_{n_1}$. Now $x = u' u_0$, where $u' \notin U_{n_1}$, and $y'$ is a suffix of $u_0 v_0 u_0$. For lengths we have now $n_0 \leq |y'| \leq |u_0 v_0 u_0| - |y'| = n_0 - |y'| \leq n$. Hence $u_0 v_0 u_0 \in U_{n_1}$ and, as we justified above, by its length

$$
|u_0 v_0 u_0| \leq |v_0 x u_0| - |u_0| \leq n
$$

also $u_0 v_0 u_0 \in X$. This means that $u_0 v_0 u_0$ and $x = u_0 v_0 u_0$ are both in $X$ and $u_0 v_0 u_0$ is a proper suffix of $x = u_0 v_0 u_0$. This contradicts the fact that $X$ is a bifixed code.

On the other hand, if $|y| = |x|$, then $|y'| = n_0$, $|z'| = |z_0|$, and $y = v_0 u_0$. In this case we add $u_0'$ to $U_n$, so that $x \in U_n V_{n_0}$.

(ii) If $|y| = n$, then $x = u' u_0$ with $|u'| = |u_0|$ and $|y| = |v_0 x u_0| - |y'| = n_0$, so $y = v_0 u_0$. Hence $x \in V_{n_1} U_{n_1}$ and, as we justified above, by its length

$$
|u' u_0| \leq |v_0 x u_0| - |u_0| \leq n
$$

also $u' u_0 \in X$. This means that $u' u_0$ and $x = u' u_0$ are both in $X$ and $u' u_0$ is a proper suffix of $x = u' u_0$. This contradicts the fact that $X$ is a bifixed code.

We proceed similarly for $y \in Y \cap A^n \setminus V_{n_1} U_{n_1}$. Note that by the construction of $U_n$ and $V_n$, $\max_{v \in V_n} |v| + \min_{u \in U_n} |u| \leq n$ and $\max_{u \in U_n} |u| + \min_{v \in V_n} |v| \leq n$. 


Now for each element $u$ in $U_n \setminus U_{n-1}$ there exist elements $v'$ and $v''$ in $V_{n_0}$ such that $uv' \in X \cap A^n$ and $v''u \in Y \cap A^n$. We have to show that $uvu_0 \subseteq X$ and $V_{n_0}u \subseteq Y$.

Let $v \in V_{n_0}$. Then $vu_0 \in Y$ and $u_0v \in X$. Since 

$$ (u_0v_0)^m u_0v''u_0v_0 = z_0(v''u)(vu_0)(v_0u_0)^m \in ZY^{m+2} = X^{m+2}Z, $$

there is $u_0v''u_0v_0 \in X^2Z$. Since $u_0v'' \in U_{n_0}V_{n_0} \subseteq X$ we obtain $uvu_0 \in XZ$, so $uvu_0 = xz$.

If $|x| < n = |uv|$, then $x \in U_{n-1}V_{n-1-1} \subseteq U_nV_n$ and $x$ is proper prefix of $uv \in U_nV_n$. However, this cannot be the case, since $U_n$ and $V_n$ are both biprefix codes (see below).

If $|x| > n$, then $|z| < |z_0|$ which contradicts the minimality of $|z_0|$. Hence $|x| = n = |uv|$ and $x = uv \in X$. The proof for $V_{n_0}$ is obtained dually.

Similarly, for each element $v$ in $V_n \setminus V_{n-1}$ there exist elements $u'$ and $u''$ in $U_{n_0}$ such that $u'v \in X \cap A^n$ and $v'u \in Y \cap A^n$ and we can prove that $U_{n_0}v \subseteq X$ and $vU_{n_0} \subseteq Y$.

By now we have constructed sets $U_n$ and $V_n$ satisfying (5). Hence it remains to conclude that they are biprefix codes. If $u' \in U_n$ is a proper prefix of $u \in U_n$, we can assume that $|w| = n - |v_0|$ (otherwise we are in $U_{n-1}$, which is a biprefix) and $u' \in U_{n-1}$. Then there exists such $v'' \in V_n$ that $u''u \in U$, but then also $v''u' \in V_nU_{n-1} \subseteq Y$. Since $Y$ is biprefix, we have a contradiction.

Similar reasoning applies also, if $u' \in U_n$ is a proper suffix of $u \in U_n$. Hence $U_n$ is also a suffix code and therefore it is a biprefix.

Similarly $V_n$ is a biprefix code. $\Box$

**Theorem 5.** If finite biprefix codes $X$ and $Y$ are conjugates, then $X = UV$ and $Y = VU$ for some biprefix codes $U$ and $V$.

**Proof.** Applying Lemma 4 for $n = \max_{x \in X} |x| + \max_{y \in Y} |y| - n_0$, we obtain:

- for all $u \in U_n$, $uv_0 \in X$, so $|u| \leq \max_{x \in X} |x| - |v_0|$,
- for all $v \in V_n$, $vu_0 \in Y$, so $|v| \leq \max_{y \in Y} |y| - |u_0|$,

so that $|uv| \leq n$. Hence we obtain:

- $U_nV_n \cap A^{\leq n} = U_nV_n$
- $V_nU_n \cap A^{\leq n} = V_nU_n$
- $X \cap A^{\leq n} = X$
- $Y \cap A^{\leq n} = Y$

implying that $X = U_nV_n$ and $Y = V_nU_n$. $\Box$

**Theorem 5** deserves a few comments. It shows that if finite prefixes $X$ and $Y$ are conjugates, that is satisfy the conjugacy equation $XZ = ZY$ with non-empty $Z$, they can be decomposed into the form

$$ X = PQ \quad \text{and} \quad Y = QP \quad \text{for some biprefixes} \ P \text{ and } Q. $$

Of course, the reverse holds as well, namely they satisfy the conjugacy equation, e.g., for $Z = P(QP)^l$, with $l \in \mathbb{N}$. Hence the conjugacy in the case of finite prefixes can be defined equivalently in the above two ways. In general, these definitions are not equivalent as discussed in [3].

To continue our analysis let us see what happens if the prefixes $X$ and $Y$ have two different factorizations

$$ X = UV, \quad Y = VU \quad \text{and} \quad X = U'V', \ Y = V'U'. $$

This indeed is possible, if $X$ and $Y$ are not primitive, as pointed out in the Example 1. We show that unique factorization for $X$ and $Y$ can be given. For this we need the following simple lemma on words.

**Lemma 6.** All solutions of the pair of word equations

\[
\begin{align*}
xy &= uv \\
yx &= vu
\end{align*}
\]

over the alphabet $A$ are of the form $x = \beta(\alpha \beta)^i$, $y = (\alpha \beta)^j\alpha$, $u = \beta(\alpha \beta)^k$ and $v = (\alpha \beta)^l\alpha$ with $i + j = k + l$ for integers $i, j, k, l$ and $\alpha, \beta \in A^*$. 
Let $X$ be a prefix code, for given finite biprefix codes $X$ and $Y$ the conjugator, i.e., the largest solution $Z$ of equation $XZ = ZY$ is $Z = (PQ)^*P$, where $P$ and $Q$ are biprefix codes such that $\rho(X) = PQ$ and $\rho(Y) = QP$.

Proof. From previous theorems we know, that $X = (PQ)^k$ and $Y = (QP)^k$ for some $P$ and $Q$ such that $\rho(X) = PQ$ and $\rho(Y) = QP$. Lemma 8 shows us, that the centralizer of $X$ is $C(X) = (PQ)^*$.

Let $Z$ be the conjugator of $X$ and $Y$. When we catenate the language $Q$ to both sides of equation $XZ = ZY$ and notice that $YQ = (QP)^kQ = Q(PQ)^k = QX$, we obtain $XZQ = ZYQ = ZQX$.

This means, that language $ZQ$ commutes with $X$. Now, Lemma 8 implies that $ZQ \subseteq C(X) = \rho(X)^* = (PQ)^*$. Since clearly the empty word is not in $ZQ$, we can write $ZQ \subseteq (PQ)^*$. 

6. The conjugator of finite prefix codes

Now it is rather easy to show that the conjugacy of finite prefix codes $X$ and $Y$ is always of word type 2, i.e., of form (4). This proof is based on some nontrivial results originally proved in [16], see also [9].
The language Q is a biprefix code, so we can eliminate the right factor Q, since the semigroup of biprefix codes is free, and hence obtain:

\[ Z \subseteq (PQ)^+P. \]

On the other hand, we know that \((PQ)^+P\) clearly is a solution of \(XZ = ZY\), and hence \((PQ)^+P \subseteq Z\). As a conclusion we see that the conjugator Z is

\[ Z = (PQ)^+P. \]

More generally we can characterize all conjucators of finite biprefix codes as follows.

**Theorem 11.** If a non-empty solution of the conjugacy equation \(XZ = ZY\) for finite biprefix codes \(X\) and \(Y\) exists, it is of word type, i.e.,

\[ X = (PQ)^k, \quad Y = (QP)^k \quad \text{ and } \quad Z = (PQ)^lP, \]

for languages \(P, Q\) and some set \(I \subseteq \mathbb{N}\).

**Proof.** As in the previous proof, we know that \(X = (PQ)^k\) and \((QP)^k\) and \(PQ\) and \(Y = QP\) are primitive. Let \(Z\) be an arbitrary language such that \(XZ = YZ\). Now again \(XZQ = ZQX\) and, by Lemma 9, we have \(ZQ = (PQ)^j\) for some \(j \subseteq \mathbb{N}\). Clearly \(0 \notin J\) and we can again eliminate the right factor, biprefix code \(Q\), from the equation. This gives us the conjugator \(Z = (PQ)^jP\) with some index set \(I = \{i \in \mathbb{N} \mid i + 1 \in J\}\).

7. The conjugacy problem for finite biprefix codes

We will refer to the problem “Are given finite languages \(X\) and \(Y\) conjucates?” as the conjugacy problem [8]. In general, the decidability status of this problem is not known, and it is expected to be hard. Our results allow to answer it in the case of biprefix codes.

**Theorem 12.** The conjugacy problem for finite biprefix codes is decidable.

**Proof.** Let \(X\) and \(Y\) be finite biprefix codes. Languages \(X\) and \(Y\) have unique factorizations as the catenation of indecomposable biprefix codes. These factorizations can be found, for example, by finding the minimal DFA for these biprefixes [1]. Theorem 5 shows that if \(X\) and \(Y\) are conjucates, then \(X = UV\) and \(Y = VU\) for some biprefix codes \(U\) and \(V\). Since the prime factorizations of \(X\) and \(Y\) are finite, there are only a finite number of candidates for \(U\) and \(V\). If \(U\) and \(V\) can be found, then equation \(XZ = YZ\) has at least word type solutions with given \(X\) and \(Y\). If on the other hand, suitable \(U\) and \(V\) cannot be found, then \(X\) and \(Y\) are not conjucates.

References