HIROSHIMA MATH. J.
22 (1992), 597–629

An efficient method for searching characteristic patterns of a subset in a large set of character sequences

Kahoru FUTAGAMI

(Received September 3, 1991)

1. Introduction

It is important to search similarities between two character sequences or characteristic patterns of a subset in a large set of sequences, in the areas of molecular biology, computer science and so on. For simplicity, we call sequences instead of character sequences.

The problem of searching similarities between two sequences has been formulated as the one of searching the longest common subsequence of two sequences under certain deletion/insertion constraints. This problem can be modified so as to search an optimum alignment under certain scoring rules, such as $+1$ for a base match and $-g$ for a gap. These problems have been studied by many authors. For global search methods, see Fitch [4], Dayhoff [1], Lipman and Pearson [13], Needleman and Wunsch [17], Sellers [20], Sankoff [19], and Wilbur and Lipman [23, 24]. For local search methods, see Hirschberg [10], Sellers [21], Smith and Waterman [22], and Goad and Kanehisa [9].

With the development of large database of sequences such as genes or images, it is necessary to compare several sequences. Relating to this problem, Korn et al. [12] developed a program for searching subsequences common to all of several sequences. In this paper we consider the problem of searching characteristic patterns of a subset in a large set of sequences. We formulate this problem as follows:

Let Z be a finite set of some alphabet, and let S and P be two finite sets of sequences whose units are composed from Z, such that $S \supsetneqq P$. Then, we are interesting in a sequence $a = (a_1, \ldots, a_k)$ with $a_i \in Z \cup \{0\}$, $i = 1, \ldots, k$ satisfying the following conditions (1) $\sim$ (4):

(1) $a \neq (0, \ldots, 0)$,
(2) $k \leqq \min \{\ell(b) \mid b \in P\}$ ($\ell(b)$ denotes the length of b),
(3) For any $b = (b_1, \ldots, b_h) \in P$, there exists an integer i_0 such that

$$0 \leqq i_0 \leqq h - k \quad \text{and} \quad a_i = b_{i+i_0} \text{ if } a_i \neq 0,$$

(4) For any $b=(b_1,\dots,b_h)\in S-P$, there does not exist an integer i_0 satisfying the above condition,

where 0 denotes an element not to belong to Z.

We call such a a characteristic pattern of a subset P in S or a solution of (S, P). A solution of (S, P) means a sequence which gives a characteristic of P in S. In general, it is difficult to find all solutions of (S, P) when $|S|$ (= the number of elements in S) and $|P|$ are large or the length of elements in S or P is long. The purpose of this paper is to propose an efficient method of finding all solutions of (S, P), which has an application to molecular biology. To do this, we first consider the case when S and P consist of elements with fixed length k, i.e., solutions of (k, S, P). Then, we consider the case when S and P consist of elements with various lengths.

In Section 2, we give a formal setup for the problem of searching all characteristic patterns of P in S, i.e., all solutions of (S, P). Section 3 discusses the problem in the case when S and P consist of elements with fixed length k. We introduce a notion of the maximum common element and a mapping from $Z\cup\{0\}$ to $\{0, 1, 2\}$. It is shown that our problem can be essentially reduced to the one in the case $|P|=1$ and $|Z|=2$. Section 4 treats the problem in a general case. In order to find solutions of (S, P) more effectively, we present an inductive method on $|P|$ and length k of solutions, whose algorithm is given in Section 5. In Section 6, we give an application to analysis of large database of nucleic acid sequences.

2. Definitions and the statement of the problems

Let Z be a set of some alphabet. Without loss of generality, we may denote Z by

$$Z=\{1, 2,\dots,|Z|\},$$

which consists of $|Z|$ integers. Let $\bar{Z}=\{0\}\cup Z$. We denote by 0 an element which is any one in Z. Moreover, for any positive integer k, consider the cartesian product Z^k (resp. $\bar{Z}^k$) of the k copies of Z (resp. $\bar{Z}$), whose element is denoted by

$$a=(a_1\cdots a_k),\ a_i\in Z \text{ (resp. } \bar{Z}) \qquad \text{for } 1\leqq i\leqq k,$$

instead of $a=(a_1,\dots,a_k)$ omitting the commas.

DEFINITION 2.1. For two elements $a=(a_1\cdots a_k)$ and $b=(b_1\cdots b_k)$ in $\bar{Z}^k$, we say that b *contains* a, denoted by

$$a\subset b,$$

$$[P; k] = \bigcup_{b\in P} [b; k] = \bigcup_{b\in P'} [b; k] \quad (P' = \{b\in P \mid \ell(b) \geqq k\});$$

where $J(P; k) = \phi$ when $k > \ell(b)$ for some $b\in P$.

(iii) Moreover, consider the maximum common element $m(P(j; k))\in \bar{Z}^k$ of $P(j; k) \subset Z^k$ in (ii) given by Definition 3.2, and we define

$$M(P; k) = \{m(P(j; k)) \mid j\in J(P; k)\} \subset \bar{Z}^k.$$

By using those notations for $\subset$ in Definition 2.3 and $B^k(P)$ and $C^k(\bar{S})$ in Proposition 4.1, we have the following

LEMMA 4.1. (i) *$a \subset b$ in Definition* 2.3 *means that*

$$a \subset b' \text{ in Definition 2.1 for some } b' = b(i; \ell(a))\in [b; \ell(a)],$$

which is $a \subset b$ in Definition 2.1 *when $\ell(a) = \ell(b)$.*

(ii) *The sets in Proposition* 4.1 *are given as follows*:

$$B^k(P) = \{a\in \bar{Z}^k \mid a \subset m \text{ for some } m\in M(P; k)\},$$

i.e., $B^k(P)$ is the union of $B(k, \{m\}) = \{a\in \bar{Z}^k \mid a \subset m\}$ for $m\in M(P; k)$; and

$$C^k(\bar{S}) = C(S') \quad (S' = [\bar{S}; k], \bar{S} = S - P),$$

where $C(S') = C^k(S') = \{a\in \bar{Z}^k \mid a \subset c' \text{ for some } c'\in S'\}$ for $S' \subset \bar{Z}^k$.

(iii) *$A^k(S, P) = B^k(P) - C^k(\bar{S})$ in Proposition* 4.1 *is the union of*

$$B(k, \{m\}) - C(S') \qquad \text{for } m\in M(P; k) - C(S').$$

PROOF. (i) According to Definition 4.1 (i), it is the restatement of Definitions 2.3 and 2.1.

(ii) According to Definition 4.1 and (i), $a\in B^k(P)$ means $a\in \bar{Z}^k$ and that for any $b\in P$, there exists $j_b\in I(b; k)$ with $a \subset b(j_b; k)$, or equivalently, that there exists $j = (j_b \mid b\in P)\in J(P; k)$ with $a \subset b'$ for any $b'\in P(j; k)$, i.e., with $a\in B(k, P(j; k)) = B(k, \{m(P(j; k))\})$ by Lemma 3.5 in Section 3. Therefore, $a\in B^k(P)$ is equivalent to $a\in B(k, \{m\})$ for some $m\in M(P; k)$ by Definition 4.1 (iii); and (ii) for $B^k(P)$ is proved.

The equality for $C^k(\bar{S})$ follows from its definition, (i) and Definition 4.1 (ii).

(iii) According to (ii), $A^k(S, P)$ is the union of

$$B(k, \{m\}) - C(S') \qquad \text{for } m\in M(P; k);$$

and if $m\in C(S')$, then any $a \subset m$ satisfies $a\in C(S')$, i.e., $B(k, \{m\}) \subset C(S')$. Therefore, we see (iii).

In Lemma 4.1 (iii), the last set is equal to

$$\{a\in \bar{Z}^k \mid a \subset m, \text{ and } a \not\subset c' \text{ for any } c'\in S'\}.$$

Therefore, relating to the extension of Definition 2.2 to the case $P \subset S \subset \bar{Z}^*$, we also use the notation in the following

DEFINITION 4.2. By using the notations in Lemma 4.1, we put

$$N(S, P; k) = M(P; k) - C(S') \subset \bar{Z}^k \quad (S' = [\bar{S}; k], \bar{S} = S - P),$$

$$A(k, S' \cup \{m\}, \{m\}) = B(k, \{m\}) - C(S') \subset \bar{Z}^k \text{ for } m \in N(S, P; k),$$

and call the latter the set of all *solutions of* $(k, S' \cup \{m\}, \{m\})$.

Now, we find all solutions of (S, P) in the following example by using Lemma 4.1 (iii) and Definition 4.2.

EXAMPLE 4.1. Let $S = \{(121), (211), (221), (112), (122), (222)\}$ and $P = \{(121), (211), (221)\}$, then we have $P(j; k)$ and $m(P(j; k))$ for $k = 2$ as follows:

$P(j; 2)$	$m(P(j; 2))$
$\{(12), (21), (22)\}$	(00)
$\{(12), (21)\}$	(00)
$\{(12), (11), (22)\}$	(00)
$\{(12), (11), (21)\}$	(00)
$\{(21), (22)\}$	(20)
$\{(21)\}$	(21)
$\{(21), (11), (22)\}$	(00)
$\{(21), (11)\}$	(01)

Hence, we have

$$\begin{aligned} M &= M(P; 2) = \{(00), (20), (21), (01)\}, \\ S' &= [S - P; 2] = \{(11), (12), (22)\}, \\ N &= M - C(S') = \{(21)\}, \\ A^2(S, P) &= \bigcup_{m \in N} (A(2, S' \cup \{m\}, \{m\}), \\ &= \{(21)\}. \end{aligned}$$

Here, we notice the following lemma, which is seen by the definition according to Lemmas 3.2 and 3.3.

LEMMA 4.2. *Let* $m = (m_1 \cdots m_k) \in \bar{Z}^k$, $m \neq (0 \cdots 0)$ *and* $I = I(m) = \{i \mid m_i = 0\}$. *Then, we have*

$$m = m[I] \text{ and } m' = m[-I] \in Z^{k'} \text{ for } k' = k - |I| \neq 0,$$

and

$$A(k, S' \cup \{m\}, \{m\}) = A(k', S'[-I] \cup \{m'\}, \{m'\})[+I]$$

when $a_i = b_i$ for any $1 \leqq i \leqq k$ with $a_i \neq 0$. If b does not contain a, we denote it by $a \not\subset b$.

DEFINITION 2.2. Let k be a positive integer, and S and P be two sets of Z^k such that

$$Z^k \supset S \supset P \quad \text{and} \quad S \neq P \neq \phi.$$

Then, we say that $a \in \bar{Z}^k$ is a *solution of* (k, S, P) when

$$(*) \qquad a \subset b \text{ for any } b \in P \text{ and } a \not\subset b \text{ for any } b \in S - P.$$

Moreover, the set of all solutions of (k, S, P) is denoted by

$$A(k, S, P) = \{a \in \bar{Z}^k \mid a \text{ satisfies } (*)\}.$$

Our notation $a \subset b$ means that a and b have a similarity. The problem of finding solutions of (k, S, P) can be generalized as follows:

DEFINITION 2.3. For two elements $a = (a_1 \cdots a_k)$ and $b = (b_1 \cdots b_h)$ in $\bar{Z}^* = \bigcup_{k=1}^{\infty} \bar{Z}^k$, we say that b *contains* a, denoted also by

$$a \subset b,$$

when $k \leqq h$ and there exists an integer i_0 such that

$$0 \leqq i_0 \leqq h - k, \text{ and } a_i = b_{i_0 + i} \text{ for any } 1 \leqq i \leqq k \text{ with } a_i \neq 0.$$

If b does not contain a, we denote it also by $a \not\subset b$.

DEFINITION 2.4. Let S and P be two finite sets such that

$$Z^* = \bigcup_{k=1}^{\infty} Z^k \supset S \supset P \quad \text{and} \quad S \neq P \neq \phi.$$

Then, we say that $a \in \bar{Z}^*$ is a *solution of* (S, P) *with length* k when $a \in \bar{Z}^k$ and a satisfies $(*)$ in Definition 2.2, for $\subset$ in Definition 2.3. Also, the set of all solutions of (S, P) is denoted by $A^*(S, P)$ (See Proposition 4.1).

The main purpose in this paper is to study the problems of finding solutions of (k, S, P) and (S, P). It may be noted that such solutions express the characteristic patterns of P in S.

We give simple examples in order to understand our problems.

EXAMPLE 2.1. Let $Z = \{1, 2\}$, $S = \{(11), (12), (21)\}$ and $P = \{(11)\}$, then $A(2, S, P) = A^*(S, P) = \{(11)\}$.

EXAMPLE 2.2. Let $P = \{(12), (21)\}$ in the above example, then $A(2, S, P)$

$= \phi$ and $A^*(S, P) = \{(2)\}$.

3. Solutions of (k, S, P) in Definition 2.2

Let k, S and P be the same ones as in Definition 2.2. Then, by Definition 2.2, we easily have the following properties of solutions of (k, S, P).

LEMMA 3.1. *If $a \in \bar{Z}^k$ is a solution of (k, S, P), then $a \neq (0 \cdots 0)$.*

PROOF. $0 = (0 \cdots 0)$ satisfies $0 \subset b$ for any $b \in Z^k$ by Definition 2.1. Therefore, if $a = 0$ satisfies $(*)$ in Definition 2.2, then $P = S$.

PROPOSITION 3.1. *Let $A(k, S, P)$ be the set of all solutions of (k, S, P). Then*

$$A(k, S, P) = B(k, P) - C(k, S, P),$$

where

$$B(k, P) = \{a \in \bar{Z}^k \mid a \subset b \text{ for any } b \in P\},$$

$$C(k, S, P) = \{a \in B(k, P) \mid a \subset c \text{ for some } c \in S - P\}.$$

PROOF. The result immediately follows from Definition 2.2.

We try to find the set of all solutions of the following example by using the above proposition.

EXAMPLE 3.1. Let S and P be given as follows:

S	P	(1122)
		(2112)
		(1212)
		(2212)
		(1222)

In order to find the set of all solutions of $(4, S, P)$, we find $B(4, P)$ and $C(4, S, P)$ as follows:

$$B(4, P) = \{(0000), (0100), (0002), (0102)\},$$

$$C(4, S, P) = \{(0000), (0002)\}.$$

Then, we have

$$A(4, S, P) = \{(0100), (0102)\}.$$

In general, it is troublesome to list up all elements of $B(k, P)$ and $C(k, S, P)$ as k, $|P|$ or $|Z|$ are large. Therefore, in the rest of this section, we will try to reduce k, $|P|$ and $|Z|$.

First, we prepare some notations.

DEFINITION 3.1. For $b = (b_1 \cdots b_k) \in \bar{Z}^k$, we define

$$b[i] = (b_1 \cdots b_{i-1} 0 b_{i+1} \cdots b_k) \in \bar{Z}^k \qquad (1 \leqq i \leqq k),$$

$$b[-i] = (b_1 \cdots b_{i-1} b_{i+1} \cdots b_k) \in \bar{Z}^{k-1} \qquad (1 \leqq i \leqq k),$$

$$b[+j] = (b_1 \cdots b_{j-1} 0 b_j \cdots b_k) \in \bar{Z}^{k+1} \qquad (1 \leqq j \leqq k+1).$$

More generally, for a subset I of $\{1,\ldots,k\}$ with $|I|$ elements, $b[I] \in \bar{Z}^k$ is the element whose i-th coordinate is given by

$$b[I]_i = \begin{cases} 0 & \text{for } i \in I, \\ b_i & \text{otherwise,} \end{cases}$$

and $b[-I] \in \bar{Z}^{k-|I|}$ ($|I| \neq k$) is the one obtained by removing b_i for $i \in I$ from $(b_1 \cdots b_k)$. Also, for a subset J of $\{1,\ldots,k+|J|\}$, we define $b[+J] \in \bar{Z}^{k+|J|}$ by the equalities

$$(b[+J])[J] = b[+J] \quad \text{and} \quad (b[+J])[-J] = b.$$

Moreover, for a subset B of $\bar{Z}^k$, we define

$$B[K] = \{b[K] \mid b \in B\} \ (K = i,\ -i,\ +j,\ I,\ -I \text{ or } +J).$$

The following lemma is immediately seen by the definition.

LEMMA 3.2. (i) $b[I] = b = (b_1 \cdots b_k)$ *if and only if* $b_i = 0$ *for any* $i \in I$.
(ii) $(b[-I])[+I] = b[I]$ *when* $I \neq \{1,\ldots,k\}$.
(iii) $a[+J] = b[+J]$ *if and only if* $a = b$.

For $a \subset b$ in Definition 2.1, we immediately see the following

LEMMA 3.3. (i) $a \subset b$ *means* $a = b[I(a)]$, *where* $I((a_1 \cdots a_k)) = \{i \mid a_i = 0\}$.
(ii) $b[I] \subset b$ *for any* b *and* I.
(iii) $a \subset b$ *implies* $a[K] \subset b[K]$ ($K = I$, $-I$ *or* $+J$).
(iv) $a[I] \subset b[I]$ *is equivalent to* $a[-I] \subset b[-I]$, *and so is* $a \subset b$ *to* $a[+J] \subset b[+J]$.
(v) *If* $a = a[I]$, *then* $a[I] \subset b[I]$ *or* $a[-I] \subset b[-I]$ *is equivalent to* $a \subset b$.

DEFINITION 3.2. For $B \subset Z^k$, we define the subsets $L(B)$ and $R(B)$ of $\{1,\ldots,k\}$ by

$$L(B) = \{i \mid \text{there is } s_i \text{ such that } b_i = s_i \text{ for any } (b_1 \cdots b_k) \in B\},$$

and

$$R(B) = \{i \mid a_i \neq b_i \text{ for some } (a_1 \cdots a_k),\ (b_1 \cdots b_k) \in B\},$$

respectively. We call $m(B) = (m_1 \cdots m_k)$, defined by $m_i = s_i$ if $i \in L(B)$, $= 0$ otherwise, or $s(B) = m(B)[-R(B)]$ the *maximum common element of* B.

Lemma 3.4. (i) $R(B) = \{1, \dots, k\} - L(B)$.

(ii) *When* $L(B) \neq \phi$, $B[-R(B)]$ *consists of the one element* $s(B)$, *i.e.*

$$b[-R(B)] = s(B) \qquad \textit{for any } b \in B.$$

(iii) *Let* $L(B) \neq \phi$ *and* $a \in \bar{Z}^k$ *satisfy* $a[-R(B)] \subset s(B)$. *Then for any* $b \in B$, $a \subset b$ *is equivalent to* $a[L] \subset b[L]$ *or* $a[-L] \subset b[-L]$ *where* $L = L(B)$.

PROOF. (i) and (ii) hold by the definition.

(iii) Let $a = (a_1 \cdots a_k)$. Then $a_i = s_i$ if $i \in L$ and $a_i \neq 0$ by the assumption. For $b = (b_1 \cdots b_k) \in B$, $a[L] \subset b[L]$ means $a_i = b_i$ if $i \notin L$ and $a_i \neq 0$, which implies $a \subset b$ since $b_i = s_i$ if $i \in L$.

Using these notations, we will obtain a reduction for $B(k, P)$ in Proposition 3.1.

PROPOSITION 3.2. (i) *If* $L(P) = \phi$, *then* $B(k, P) = \{(0 \cdots 0)\}$.

(ii) *If* $|L(P)| = h \neq 0$, *then*

$$\begin{aligned} B(k, P) &= \{a \in \bar{Z}^k \mid a[R] = a \textit{ and } a[-R] \subset s\} \\ &= B(h, \{s\})[+R], \end{aligned}$$

where $R = R(P)$ *and* $s = s(P)$.

PROOF. Take any $a \in B(k, P)$. Then, by the definition,

$$a = (a_1 \cdots a_k) \subset b = (b_1 \cdots b_k) \qquad \text{for any } b \in P.$$

If $a_i \neq 0$, then $b_i = a_i$ by Definition 2.1, and so $i \in L(P)$ by Definition 3.2. Therefore

$$a_i = 0 \qquad \text{for any } i \in R = \{1, \dots, k\} - L(P).$$

Thus we see (i), because $(0 \cdots 0) \subset b$. Also $a[R] = a$ by Lemma 3.2(i). Hereafter assume $h \neq 0$. Then $a[-R] \subset b[-R] = s$ by Lemmas 3.3(iii) and 3.4(ii); hence a belongs to the second set B' in (ii).

Conversely, take any $a \in B'$. Then, for any $b \in P$, we see $a[-R] \subset s = b[-R]$ by Lemma 3.4(ii). Hence $a \subset b$ by Lemma 3.3(v), since $a[R] = a$. Thus $a \in B(k, P)$. On the other hand, let $a' = a[-R] \in \bar{Z}^h$. Thus $a' \subset s$ means $a' \in B(h, \{s\})$ by the definition; and using Lemma 3.2 (ii), we have

$$a = a[R] = (a[-R])[+R] = a'[+R] \in B(h, \{s\})[+R].$$

Finally, if $a = a'[+R]$ for $a' \in B(h, \{s\})$, then

$$a[R]=(a'[+R])[R]=a'[+R]=a$$
$$\text{and}\quad a[-R]=(a'[+R])[-R]=a'\subset s$$

by Definition 3.1, and these mean $a\in B'$. Therefore, (ii) is proved completely.

For $C(k, S, P)$ in Proposition 3.1, we have the following

PROPOSITION 3.3. (i) *If* $L(P)=\phi$, *or* $L(P)\neq\phi$ *and* $s(P)\in(S-P)[-R(P)]$, *then* $C(k, S, P)=B(k, P)$.

(ii) *If* $|L(P)|=h\neq 0$ *and* $s=s(P)\notin(S-P)[-R]$ $(R=R(P))$, *then*

$$(S-P)[-R]=S[-R]-\{s\}\neq\phi,$$

and

$$C(k, S, P)=C(h, S[-R], \{s\})[+R].$$

PROOF. Recall that $C(k, S, P)=\{a\in B(k, P)\,|\,a\subset c \text{ for some } c\in S-P\}$.

(i) When $L(P)=\phi$, we see (i) by Proposition 3.2(i). Assume that $L(P)\neq\phi$ and

$$s=s(P)=c[-R] \text{ for some } c\in S-P \qquad (R=R(P)).$$

Take $a\in B(k, P)$. Then $a[R]=a$ and $a[-R]\subset s$ by Proposition 3.2(ii); hence $a[-R]\subset c[-R]$ and we see $a\subset c$ by Lemma 3.3(v). Thus $a\in C(k, S, P)$; and (i) is proved.

(ii) Assume that $L(P)\neq\phi$ and $s\notin(S-P)[-R]$. Then we see the first desired equality, because $\{s\}=P[-R]$ by Lemma 3.4(ii).

Take any $a\in C(k, S, P)$. Then $a\in B(k, P)$, i.e.,

$$a[R]=a,\quad a'=a[-R]\in B(h, \{s\}) \text{ and } a'[+R]=a$$

by Proposition 3.2 (ii) and its proof. Also $a\subset c$ for some $c\in S-P$. Hence

$$a'=a[-R]\subset c[-R]\in S[-R]-\{s\}$$

by the first equality, and we see $a'\in C(h, S[-R], \{s\})$ by the definition. Thus, $a=a'[+R]\in C(h, S[-R], \{s\})[+R]$.

Conversely, take any $a'\in C(h, S[-R], \{s\})$. Then $a'\in B(h, \{s\})$, and

$$a=a'[+R]\in B(k, P)$$

by Proposition 3.2 (ii). Also $a'\subset c'$ for some $c'\in S[-R]-\{s\}=(S-P)[-R]$. Hence $c'=c[-R]$ for some $c\in S-P$, and

$$a=a'[+R]\subset c'[+R]=(c[-R])[+R]=c[R]\subset c.$$

Therefore $a\in C(k, S, P)$; and the last equality is proved.

Propositions 3.1, 3.2 and 3.3 imply the following theorem for the set $A(k, S, P)$ of all solutions of (k, S, P) in Definition 2.2.

THEOREM 3.1. *Let P and S be the subsets in Z^k such that $P \subset S$ and $P \neq S$.*

(i) *There exists a solution of (k, S, P) if and only if*

$$h = k - |R| \neq 0 \quad and \quad s \notin (S - P)[-R],$$

where $R = R(P) \subset \{1, \dots, k\}$ and the maximum common element $s = s(P) \in Z^h$ of P are defined by Definition 3.2.

(ii) *In case of* (i), *it holds that*

$$A(k, S, P) = A(h, S[-R], \{s\})[+R],$$

i.e., all solutions of (k, S, P) are $a'[+R]$ (see Definition 3.1) *of those a' of $(h, S[-R], \{s\})$, where the correspondence sending $a' \in \bar{Z}^h$ to $a'[+R] \in \bar{Z}^k$ is one-to-one.*

(iii) *For $s \in S' = S[-R] \subset Z^h$, consider $L = L[S'] \subset \{1, \dots, h\}$ in Definition* 3.2 *and put $m = h - |L| > 0$ and $s' = s[-L]$. Then, it holds that*

$$A(h, S', \{s\}) = \{a \in \bar{Z}^h \,|\, a'[+L] \subset a \subset s \text{ for some } a' \in A(m, S'[-L], \{s'\})\};$$

and this set contains any a with $s[L] \subset a \subset s$.

PROOF. $A(k, S, P) = B(k, P) - C(k, S, P)$ by Proposition 3.1, and the correspondence sending a' to $a'[+R]$ is one-to-one by Lemma 3.2 (iii). Therefore, (i) and (ii) follow from Propositions 3.2 and 3.3, since $A(h, S[-R], \{s\}) \ni s$.

(iii) Take any $a \in A(h, S', \{s\})$. Then, by Definition 2.2,

$$a \subset s, \quad \text{and} \quad b = s \text{ if } a \subset b \in S';$$

and we can prove $a' = a[-L] \in A(m, S'', \{s'\})$ where $S'' = S'[-L]$. In fact, $a' = a[-L] \subset s[-L] = s'$. Assume $a' \subset b' \in S''$. Then $b' = b[-L]$ for some $b \in S'$, and so $a[-L] \subset b[-L]$. This implies $a \subset b$ by Lemma 3.4 (iii), since $a[-R(S')] \subset s[-R(S')] = s(S')$ by Lemma 3.4 (ii). Therefore, $b = s$, and $b' = s[-L] = s'$. Thus $a' \in A(m, S'', \{s'\})$ is proved. Now, $a'[+L] = (a[-L])[+L] = a[L] \subset a \subset s$; and a belongs to the right hand side A' of the desired equality.

Conversely, take any $a \in A'$. Then, there exists $a' \in \bar{Z}^m$ such that

$$a' \subset s', \quad b' = s' \text{ if } a' \subset b' \in S'[-L], \text{ and } a'[+L] \subset a \subset s.$$

Assume $a \subset b \in S'$. Then $a' = (a'[+L])[-L] \subset a[-L] \subset b[-L] \in S'[-L]$, and so $b[-L] = s' = s[-L]$. This and $b \in S'$ imply $b = s$ by the definition of $L = L(S')$. Thus $a \in A(h, S', \{s\})$; and the desired equality is proved.

The last fact holds, because $s'\in A(m, S'[-L], \{s'\})$ and $s'[+L]=s[L]$.

EXAMPLE 3.2. Let S and P be given as follows:

<table>
<tr><td rowspan="4">S</td><td rowspan="2">P</td><td>(112)</td></tr>
<tr><td>(121)</td></tr>
<tr><td rowspan="2"></td><td>(211)</td></tr>
<tr><td>(122)</td></tr>
</table>

Then $R=R(P)=\{2, 3\}$, $s(P)=(1)$ and $(S-P)[-R]=\{(2), (1)\}$. Therefore, by Theorem 3.1 (i) we obtain $A(3, S, P)=\phi$ since $s\in(S-P)[-R]$.

In the following examples, we find all solutions of (k, S, P) by using Theorem 3.1.

EXAMPLE 3.3. In Example 3.1, we easily see that $L(P)=\{2, 4\}$, $R=R(P)=\{1, 3\}$, $s=s(P)=(12)$ and $S[-R]=\{(12), (22)\}$. Then

$$B(2, \{s\})=\{(00), (10), (02), (12)\},$$

$$C(2, S[-R], \{s\})=\{(00), (02)\},$$

$$A(2, S[-R], \{s\})=\{(10), (12)\},$$

and

$$A(4, S, P)=\{(0100), (0102)\}.$$

EXAMPLE 3.4. In the above example, let $S'=S[-R]$. Then $L=L(S')=\{2\}$, $s'=s[-L]=(1)$ and $S'[-L]=\{(1), (2)\}$. Therefore

$$B(1, \{s'\})=\{(0), (1)\},$$

$$C(1, S'[-L], \{s'\})=\{(0)\},$$

$$A(1, S'[-L], \{s'\})=\{(1)\},$$

and from Theorem 3.1 (iii) we have

$$A(2, S', \{s\})=\{(10), (12)\}.$$

For the set $A(m, S'[-L], \{s'\})$ in (iii) of Theorem 3.1, we have the following

THEOREM 3.2. *For* $s'=(s'_1\cdots s'_m)\in Z^m$, *we define the mapping*

$$f_{s'}\colon \bar{Z}^m \longrightarrow \{0, 1, 2\}^m$$

sending $b=(b_1\cdots b_m)\in\bar{Z}^m$ *to* $f_{s'}(b)=(e_1\cdots e_m)\in\{0, 1, 2\}^m$ *given by*

$$e_i=0 \text{ if } b_i=0, \ =1 \text{ if } b_i=s'_i, \ =2 \text{ if } 0\neq b_i\neq s'_i \qquad (1\leqq i\leqq m).$$

Moreover, for $s'\in S'[-L]\subset Z^m$ *in Theorem* 3.1 (iii), *let*

$$1 = (1 \cdots 1) = f_{s'}(s') \in T = f_{s'}(S'[-L]) \subset \{1, 2\}^m.$$

Then

$$a \in A(m, S'[-L], \{s'\}) \text{ if and only if } f_{s'}(a) \in A(m, T, \{1\}).$$

Furthermore, if $d \in A(m, T, \{1\})$, *then we have a solution* $a \in A(m, S'[-L], \{s'\})$ *with* $f_{s'}(a) = d$, *which is uniquely determined by* $a = s'[I]$, *where the subset* I *of* $\{1,\dots,m\}$ *is defined by* $d = 1[I]$.

PROOF. By the definition of $f = f_{s'}$ and Definition 2.1 of $\subset$, we see that $a \subset s'$ (resp. $b = s'$) is equivalent to $f(a) \subset f(s') = 1$ (resp. $f(b) = 1$). Moreover, $a \subset b$ implies $f(a) \subset f(b)$; and when $a \subset s'$, $f(a) \subset f(b)$ implies $a \subset b$. Thus, the conditions that

$$a \subset s', \quad \text{and} \quad b = s' \text{ if } a \subset b \in S'[-L]$$

are equivalent to the ones that

$$f(a) \subset 1, \quad \text{and} \quad b' = 1 \text{ if } f(a) \subset b' \in T = f(S'[-L]);$$

and we see the first half. The second half is seen by the definition.

These two theorems mean the following main result in this section.

THEOREM 3.3. (i) *The problem of finding solutions of* (k, S, P) *in Definition* 2.2 *for subsets*

$$P \subset S \subset Z^k \text{ with } Z = \{1, 2,\dots,|Z|\} \text{ and } \phi \neq P \neq S$$

can be reduced by Theorems 3.1 *and* 3.2 *to the one of finding solutions of* $(m, T, \{1\})$ *for*

$$1 = (1 \cdots 1) \in T \subset \{1, 2\}^m \text{ with } m \leqq k$$

such that for any $i \in \{1,\dots,m\}$, $e_i = 2$ *for some* $(e_1 \cdots e_m) \in T$.

(ii) *Moreover, d is a solution of* $(m, T, \{1\})$ *if and only if there exists a proper subset I of* $\{1,\dots,m\}$ *such that*

$$d = 1[I] \text{ and } e[I] \neq d \text{ for any } e \in T - \{1\}.$$

Here, $((e_1 \cdots e_m)[I])_i = 0$ *if* $i \in I$, $= e_i$ *if* $i \notin I$ (*cf. Definition* 3.1), and (ii) *is the restatement of Definition* 2.2 *in case of* $(m, T, \{1\})$.

This theorem considerably simplifies the problem in the case that $|Z|$ and k are so large.

EXAMPLE 3.5. Let S and P be given as follows:

S	P	(3123)
		(3323)
		(1323)
		(3113)
		(1223)
		(3233)

Then $R = R(P) = \{2\}$, $L = L(P) = \{1, 3, 4\}$ and $s = s(P) = (323)$. Therefore, we apply the mapping $f_{s'}$ in Theorem 3.2 as follows:

$$S' = S[-R] = \{(323), (123), (313), (333)\},$$
$$L(S') = \{3\},$$
$$s' = s[-L(S')] = (32),$$
$$S'' = S'[-L(S')] = \{(32), (12), (31), (33)\},$$
$$T = f_{s'}(S'') = \{(11), (21), (12)\}.$$

Then, the problem of finding solutions of $(4, S, P)$ is reduced to the one of $(2, T, \{1\})$. Then

$$B(2, \{1\}) = \{(00), (01), (10), (11)\},$$
$$C(2, T, \{1\}) = \{(00), (01), (10)\},$$
$$A(2, T, \{1\}) = \{(11)\}.$$

By Theorem 3.2, we obtain $I = \phi$ since $(11) = (11)\ [I]$ and $a = s' = (32)$ as a solution of $A(2, S'', \{s'\})$. Therefore

$$A(3, S', \{s\}) = \{(320), (323)\},$$
$$A(4, S, P) = \{(3020), (3023)\}.$$

The following lemma is used in the next section.

LEMMA 3.5. *In Proposition* 3.2(ii), *it holds also that*

$$B(k, P) = \{a \in \bar{Z}^k \mid a \subset m(P)\} = B(k, \{m(P)\})$$

for the maximum common element $m(P) \in \bar{Z}^k$ *in Definition* 3.2.

PROOF. It is proved in Proposition 3.2(ii) that

$$B(k, P) = B' = \{a \in \bar{Z}^k \mid a[R] = a \text{ and } a[-R] \subset s\},$$

where $s = s(P)$, $m = m(P)$ and $R = R[P]$ in Definition 3.2 satisfy

$$s = m[-R] \text{ and } s[+R] = m = m[R]$$

for $b[K]$ in Definition 3.1. By Definitions 2.1 and 3.1, we see that

$$a \subset m \text{ and } m = m[R] \text{ imply } a = a[R].$$

Therefore, if $a \in B'$, then $a = a[R] = (a[-R])[+R] \subset s[+R] = m$; and if $a \subset m$, then $a = a[R]$ and $a[-R] \subset m[-R] = s$, and so $a \in B'$, by Lemmas 3.2 and 3.3. Thus the first equality is proved. The second one is the definition.

4. Solutions of (S, P) in Definition 2.4

In this section, let S and P be given finite sets satisfying

$$P \subset S \subset Z^* = \bigcup_{k=1}^{\infty} Z^k \quad \text{and} \quad \phi \neq P \neq S,$$

where $Z = \{1, 2, \ldots, |Z|\}$. Then, the following proposition is seen in the same way as Proposition 3.1, where $\bar{Z} = \{0\} \cup Z$ and $\bar{Z}^* = \bigcup_{k=1}^{\infty} \bar{Z}^k$.

PROPOSITION 4.1. *Let $A^k(S, P)$ be the set of all solutions of (S, P) with length k in Definition 2.4 for $1 \leqq k < \infty$ or $k = *$. Then*

$$A^*(S, P) = \bigcup_{k=1}^{\infty} A^k(S, P) \quad \textit{and} \quad A^k(S, P) = B^k(P) - C^k(\bar{S}),$$

where $\bar{S} = S - P$,

$$B^k(P) = \{a \in \bar{Z}^k \mid a \subset b \textit{ for any } b \in P\},$$

$$C^k(\bar{S}) = \{a \in \bar{Z}^k \mid a \subset c \textit{ for some } c \in \bar{S}\},$$

by the notation $\subset$ in Definition 2.3.

To study the above sets, we prepare some notations.

DEFINITION 4.1. (i) For $a = (a_1 \cdots a_h) \in \bar{Z}^*$, let $\ell(a) = h$ and call it the *length* of a. Moreover, when $1 \leqq k \leqq \ell(a)$, we define

$$I(a; k) = \{1, \ldots, \ell(a) - k + 1\},$$

$$a(i; k) = (a_i \cdots a_{i+k-1}) \in \bar{Z}^k \text{ for } i \in I(a; k),$$

$$[a; k] = \{a(i; k) \mid i \in I(a; k)\} \subset \bar{Z}^k;$$

and $I(a; k) = \phi = [a; k]$ when $k > \ell(a)$.

(ii) For any subset $P \subset Z^*$ and $1 \leqq k < \infty$, we define

$$J(P; k) = \prod_{b \in P} I(b; k) = \{(j_b \mid b \in P) \mid j_b \in I(b; k) \text{ for } b \in P\},$$

$$P(j; k) = \{b(j_b; k) \mid b \in P\} \text{ for } j = (j_b \mid b \in P) \in J(P; k),$$

for the set in Definition 4.2, *where* $B[K]$ *is the notation in Definition* 3.1.

Moreover, $A(k', S'[-I] \cup \{m'\}, \{m'\})$ *of the above is the set of solutions in Definition* 2.2, *and so we can apply Theorems* 3.1 (iii), 3.2 *and* 3.3 *to it.*

If the size of $M(P; k)$ or $N(S, P; k)$ is large, it is difficult to obtain it. Therefore it is important to reduce $M(P; k)$ or $N(S, P; k)$ without any change of solvability.

DEFINITION 4.3. (i) For any subset $M \subset \bar{Z}^k$, we call $m \in M$ a *reduced element in* M when

$$m \subset m' \neq m \qquad \text{for some } m' \in M.$$

(ii) Moreover, we denote by $\hat{M}$ the set of all *non-reduced elements in* M, i.e., by $\hat{M} = \hat{M}(P; k)$, $\hat{N}(S, P; k)$, $\hat{A}^k(S, P)$ or $\hat{B}^k(P)$ when $M = M(P; k)$, $N(S, P; k)$, $A^k(S, P)$ or $B^k(P)$, respectively.

LEMMA 4.3. (i) $M \neq \phi$ *if and only if* $\hat{M} \neq \phi$; *and then any* $m \in M$ *satisfies* $m \subset \hat{m}$ *for some* $\hat{m} \in \hat{M}$, *i.e.,* $M \subset C(\hat{M}) = C(M)$. *Moreover, for* $m \in M$, $m \in \hat{M}$ *holds if and only if* $m \subset m' \in M$ *implies* $m' = m$.

(ii) *If* $\hat{M} \subset N \subset C(M) = C(\hat{M})$, *then* $\hat{N} = \hat{M}$.

(iii) *If* $N = M - C(S')$ $(S' \subset \bar{Z}^k)$, *then* $\hat{N} = \hat{M} - C(S')$.

PROOF. We see (i) by the definition. Let $\hat{M} \subset N \subset C(M)$. If $n \in \hat{N}$, then $n \subset m$ for some $m \in \hat{M} \subset N$, and so $n = m \in \hat{M}$, by (i). Conversely, if $m \in \hat{M}$, then $m \subset n \subset m'$ for some $n \in \hat{N}$ and $m' \in M$; hence $m = m'$ by (i) and $m = n \in \hat{N}$. Thus $\hat{N} = \hat{M}$, and (ii) is proved. (iii) is proved in the same way by noticing that $n \subset m$ and $n \notin C(S')$ imply $m \notin C(S')$.

Now, we see the following results by Proposition 4.1 and Lemma 4.1.

THEOREM 4.1. (i) *For the set* $A^k(S, P)$ *of all solutions of* (S, P) *of length* k *in Definition* 2.4, *and the sets in Definitions* 4.1, 4.2 *and* 4.3, *we have*

$$A^k(S, P) = \bigcup_{m \in \hat{N}(S, P; k)} A(k, S' \cup \{m\}, \{m\}) \quad (S' = [S - P; k]),$$

the union of the sets in Lemma 4.2. *Also we have* $B^k(P) = C(\hat{M}(P; k))$ *and*

$$A^k(S, P) = B^k(P) - C(S') = C(\hat{N}(S, P; k)) - C(S').$$

(ii) *In particular, there exists a solution of* (S, P) *of length* k *if and only if* $N(S, P; k) \neq \phi$ *or equivalently,* $\hat{N}(S, P; k) \neq \phi$; *and then any* $m \in N(S, P; k)$ *is a solution.*

PROOF. By the above statements, we have already seen that $B^k(P)$ (resp. $A^k(S, P)$) is the union of $B(k, \{m\})$ (resp. $A(m) = A(k, S' \cup \{m\}, \{m\}) = B(k, \{m\})$

$-C(S'))$ for $m \in M(P; k)$ (resp. $N(S, P; k)$). Therefore, the same is valid also for $m \in \hat{M}(P; k)$ (resp. $\hat{N}(S, P; k)$) by Lemma 4.3 (i), because $m \subset m'$ implies $B(k, \{m\}) \subset B(k, \{m'\})$ and so $A(m) \subset A(m')$. Thus we see (i), since $\bigcup_{m \in M} B(k, \{m\}) = C(M)$ by the definition.

(ii) follows from (i), because $m \in B(k, \{m\}) - C(S') = A(m)$ when $m \notin C(S')$. We note that $\hat{N}(S, P; k) = \hat{A}^k(S, P) = \hat{M}(P; k) - C(S')$ and $\hat{M}(P; k) = \hat{B}^k(P)$ by Lemma 4.3 (iii).

We shall find all solutions of (S, P) by using Theorem 4.1.

EXAMPLE 4.2. Let $S = \{(121), (211), (122), (222)\}$ and $P = \{(121), (211)\}$, then we have for $k = 2$

$$M(P; 2) = \{(00), (10), (21), (01)\},$$
$$S' = [S - P; 2] = \{(12), (22)\},$$
$$N(S, P; 2) = \{(21), (01)\},$$
$$\hat{N} = \hat{N}(S, P; 2) = \{(21)\}.$$

Hence, we have only to find the set of solutions $A(2, S' \cup \{m\}, \{m\})$ for $m = (21)$, then

$$A^2(S, P) = A(2, S' \cup \{(21)\}, \{(21)\}) = \{(01), (21)\}.$$

Also we have

$$C(\hat{N}) = \{(00), (20), (01), (21)\},$$
$$A^2(S, P) = C(\hat{N}) - C(S') = \{(01), (21)\}.$$

EXAMPLE 4.3. Let S and P be the same as in Example 4.1, then we see that $N(S, P; 2) = \hat{N}(S, P; 2) = \{(21)\}$. Hence

$$A^2(S, P) = A(2, S' \cup \{(21)\}, \{(21)\}) = \{(21)\}.$$

EXAMPLE 4.4. Let the element (112) in S of Example 4.1 be changed to (212). Then

$$[S - P; 2] = \{(21), (12), (22)\},$$
$$M(P; 2) = \{(00), (20), (21), (01)\},$$
$$N(S, P; 2) = \phi.$$

Hence there is no solution.

By Proposition 4.1 and Theorem 4.1, solutions of (S, P) can be obtained by examining the sets of solutions in Lemma 4.2 and the sets

$$N(S, P; k) \supset \hat{N}(S, P; k) \qquad \text{for } 1 \leqq k < \infty;$$

and by Definitions 4.1, 4.2 and 4.3, the latter two sets are obtained by making $J(P; k)$, by taking $m(P(j; k))$ for all $j \in J(P; k)$, and then finding $m(P(j, k)) \notin C(S')$, and so on. However, $J(P; k)$ may contain many elements, e.g.

$$|J(P; k)| = (h-k+1)^{|P|} \text{ when } P \subset Z^h.$$

Therefore, we are concerned hereafter with the problem of finding a non-empty set $\hat{N}(S, P; k)$ more effectively.

In the first place, we present an inductive method on the number $|P|$ of elements in P.

DEFINITION 4.4. For $a, b \in \bar{Z}^k$, we denote $m(\{a, b\})$ in Definition 3.2 by $a \& b$, i.e., for $a = (a_1 \cdots a_k)$ and $b = (b_1 \cdots b_k)$, the i-th coordinate of $a \& b$ is given by

$$(a \& b)_i = a_i \text{ if } a_i = b_i, \ = 0 \text{ if } a_i \neq b_i \qquad (1 \leqq i \leqq k).$$

For $A, B \subset \bar{Z}^k$, we put $A \& B = \{a \& b \mid a \in A \text{ and } b \in B\}$.

LEMMA 4.4. (i) *For $b \in Z^k$, $M(\{b\}; k) = [b; k] = \hat{M}(\{b\}; k)$, which is ϕ if $k > \ell(b)$.*

(ii) *Let $P = P_1 \cup P_2$ and $P_1 \cap P_2 = \phi$. Then*

$$M(P; k) = M(P_1; k) \& M(P_2; k), \text{ and}$$
$$\hat{M}(P; k) = \hat{M} \text{ for } M = \hat{M}(P_1; k) \& \hat{M}(P_2; k).$$

PROOF. (i) We see (i) by the definition.

(ii) $J(P; k) = J(P_1; k) \times J(P_2; k)$, i.e., $j = (j_b \mid b \in P) \in J(P; k)$ gives us $j^\varepsilon = (j_b \mid b \in P_\varepsilon) \in J(P_\varepsilon; k)$ for $\varepsilon = 1, 2$, and also the converse holds, by Definition 4.1; and moreover $P(j; k) = P_1(j^1; k) \cup P_2(j^2; k)$. Therefore, $m(P(j; k)) = m(P_1(j^1; k)) \& m(P_2(j^2; k))$ by Definitions 3.2 and 4.4; and we see the first equality by Definition 4.1 (iii).

Take any $m \in \hat{M}(P; k)$, and $m^\varepsilon \in M_\varepsilon = M(P_\varepsilon; k)$ with $m = m^1 \& m^2$. Then, $m^\varepsilon \subset \hat{m}^\varepsilon$ for some $\hat{m}^\varepsilon \in \hat{M}_\varepsilon$ by Lemma 4.3 (i). Thus $m \subset \hat{m}^1 \& \hat{m}^2 \in M(P; k)$ and so $m = \hat{m}^1 \& \hat{m}^2$ by Lemma 4.3 (i), since $m \in \hat{M}(P; k)$. Therefore, we see $\hat{M}(P; k) \subset \hat{M}_1 \& \hat{M}_2 = M \subset M(P; k)$, and $\hat{M}(P; k) = \hat{M}$ by Lemma 4.3 (ii).

PROPOSITION 4.2. *For $\phi \neq Q \subset P \subset S$, let $\bar{S} = S - P$ and*

$$\hat{N}_k(Q) = \hat{N}(\bar{S} \cup Q, Q; k) = \hat{M}(Q; k) - C([\bar{S}; k]).$$

(i) $$\hat{N}_k(\{b\}) = [b; k] - [\bar{S}; k] \qquad \textit{for any } b \in P;$$

and this is ϕ if $k > \ell(b)$. Also, $\hat{N}_{k'}(\{b\}) = \phi$ if $k' < k \leqq \ell(b)$ and $\hat{N}_k(\{b\}) = \phi$.

(ii) *If $\hat{N}_k(Q) = \phi$ and $Q \subset Q' \subset P$, then $\hat{N}_k(Q') = \phi$. In particular,*

$$\hat{N}(S, P; k) = \hat{N}_k(P) = \phi \text{ if } [b; k] \subset [\bar{S}; k] \qquad \text{for some } b \in P.$$

(iii) *Let $b \in P - Q$. Then*

$$\hat{N}_k(Q \cup \{b\}) = \hat{N} \text{ for } N = (\hat{N}_k(Q) \& \hat{N}_k(\{b\})) - C([\bar{S}; k]).$$

PROOF. (i) We see the first half by Lemma 4.4(i). If $k' < k \leqq \ell(b)$, then we see $[b; k'] = [[b; k]; k']$ by Definition 4.1. Hence, if $\hat{N}_k(\{b\}) = \phi$ in addition, then $[b; k'] \subset [[\bar{S}; k]; k'] \subset [\bar{S}; k']$, and so $\hat{N}_{k'}(\{b\}) = \phi$.

(ii) Hereafter, denote simply by $\hat{N}(Q) = \hat{N}_k(Q)$ and $S' = [\bar{S}; k]$. Put $M' = M(Q'; k)$, $M_1 = M(Q; k)$, $M_2 = M(Q' - Q; k)$, and take $\hat{M}$ of them. Then $\hat{M}' \subset \hat{M}_1 \& \hat{M}_2$ by Lemma 4.4(ii). If m^ε is contained in $C = C(S')$ for $\varepsilon = 1$ or 2, then so is $m^1 \& m^2 \subset m^\varepsilon$. Therefore,

$$(\hat{M}_1 \& \hat{M}_2) - C \subset ((\hat{M}_1 - C) \& (\hat{M}_2 - C)) - C = (\hat{N}(Q) \& \hat{N}(Q' - Q)) - C.$$

Thus, $\hat{N}(Q) = \phi$ implies $\hat{M}' \subset \hat{M}_1 \& \hat{M}_2 \subset C$ and $\hat{N}(Q') = \phi$.

(iii) By the above proof for $Q' = Q \cup \{b\}$ and Lemma 4.4(ii), we have

$$\begin{aligned}\hat{N}(Q') = \hat{M}' - C &\subset (\hat{N}(Q) \& \hat{N}(\{b\})) - C = N \\ &\subset (M_1 \& M_2) - C = M' - C = N(Q'),\end{aligned}$$

and the desired equality $\hat{N}(Q') = \hat{N}$ by Lemma 4.3(ii).

According to (i) and (ii) in Proposition 4.2, we put $k_0 = \min\{\ell(b) \mid b \in P\}$ and take the smallest $k_1 \leqq k_0$ such that

$$\hat{N}_{k_1}(\{b\}) = [b; k_1] - [\bar{S}; k_1] \neq \phi \qquad \text{for any } b \in P,$$

finding $b(j; k_1) \notin [\bar{S}; k_1]$ by their definition. Moreover, we put $P = \{b^1, b^2, \ldots, b^p\}$ by giving some order of its elements, and put $P^q = \{b^1, \ldots, b^q\}$. Then, for $k_1 \leqq k \leqq k_0$, $\hat{N}_k(P^q) = \hat{N}(\bar{S} \cup P^q, P^q; k)$ is seen inductively on q by the equality in Proposition 4.2(iii) for $Q = P^{q-1}$ and $b = b^q$; and we obtain $\hat{N}(S, P; k) = \hat{N}_k(P^q)$, which is ϕ if so is $\hat{N}_k(P^q)$ for some $q \leqq p$.

Therefore, to find a non-empty set $\hat{N}(S, P; k)$, we may examine it for $k = k_1, k_1 + 1, \ldots$ successively by the above way; and when k_1 does not exist or when $\hat{N}(S, P; k) = \phi$ for $k_1 \leqq k \leqq k_0$, we have no solution of (S, P) by Theorem 4.1.

EXAMPLE 4.5. We obtain $\hat{N}_2(P)$ for S and P of Example 4.1 by Proposition 4.2 as follows:

q	1	2	3
$M(\{b^q\}; 2)$	$\{(12), (21)\}$	$\{(21), (11)\}$	$\{(22), (21)\}$
$N_2(\{b^q\})$	$\{(21)\}$	$\{(21)\}$	$\{(21)\}$
$\hat{N}_2(P^{q-1}) \& \hat{N}_2(\{b^q\})$	$-$	$\{(21)\}$	$\{(21)\}$
$(\hat{N}_2(P^{q-1}) \& \hat{N}_2(\{b^q\})) - C$	$-$	$\{(21)\}$	$\{(21)\}$
$\hat{N}_2(P^q)$	$\{(21)\}$	$\{(21)\}$	$\{(21)\}$

Hence $\hat{N}_2(P) = \{(21)\}$.

EXAMPLE 4.6. Let $S = \{(22112), (22212), (21211), (11212), (22122), (12122), (12121), (21221)\}$ and $P = \{(22112), (22212), (21211), (11212)\}$, then $[S - P; 3] = \{(221), (212), (122), (121)\}$. Similarly We obtain $\hat{N}_3(P)$ as follows:

q	1	2
$M(\{b^q\}; 3)$	$\{(221), (211), (112)\}$	$\{(222), (221), (212)\}$
$N_3(\{b^q\})$	$\{(211), (112)\}$	$\{(222)\}$
$\hat{N}_3(P^{q-1}) \& \hat{N}_3(\{b^q\})$	$-$	$\{(200), (002)\}$
$(\hat{N}_3(P^{q-1}) \& \hat{N}_3(\{b^q\})) - C$	$-$	ϕ
$\hat{N}_3(P^q)$	$\{(211), (112)\}$	ϕ

Hence $\hat{N}_3(P) = \phi$.

In the rest of this section, we consider an inductive method on k.

DEFINITION 4.5. (i) We put $(a_1 \cdots a_k) \oplus (b_1 \cdots b_h) = (a_1 \cdots a_k b_1 \cdots b_h) \in \bar{Z}^{k+h}$, and $A \oplus B = \{a \oplus b \mid a \in A \text{ and } b \in B\}$.

(ii) For $a = (a_1 \cdots a_{k+h})$ $(k, h \geqq 1)$, we put $\partial_h a = (a_1 \cdots a_k)$ and $\partial^k a = (a_{k+1} \cdots a_{k+h})$ so that $a = \partial_h a \oplus \partial^k a$.

(iii) In Definition 4.1, we consider $j = (j_b \mid b \in P) \in J(P; k)$ satisfying

$$1 \leqq j_b \leqq \ell(b) - k + 1 - h \text{ (resp. } h + 1 \leqq j_b \leqq \ell(b) - k + 1) \text{ for any } b \in P,$$

and denote the set of $m(P(j; k))$ of all such j by

$$M_h(P; k) \text{ (resp. } M^h(P; k)) \subset M(P; k).$$

LEMMA 4.5. (i) *$M(P; t) \subset M_h(P; k) \oplus M^k(P; h)$ for $t = k + h$.*

(ii) *Consider the set*

$$B^t(P) = \{a \in \bar{Z}^t \mid a \subset b \text{ (in Definition 2.3) for any } b \in P\} = C(\hat{M}(P; t))$$

in Proposition 4.1 and Theorem 4.1 (i). Then

$$\hat{M}(P; t) = \hat{M}' \text{ for } M' = M \cap B^t(P), \text{ when } M(P; t) \subset M \subset \bar{Z}^t.$$

(iii) $[\bar{S}; k + h] \subset [\partial_h \bar{S}; k] \oplus [\partial^k \bar{S}; h] \subset [\bar{S}; k] \oplus [\bar{S}; h]$.

PROOF. (i) Take any $m=m(P(j;t))\in M(P;t)$ for $j=(j_b\,|\,b\in P)\in J(P;t)$. Then, Definitions 4.5 and 4.1 show that $m=\partial_h m\oplus\partial^k m$, $\partial_h m=m(P(j;k))\in M_h(P;k)$ and $\partial^k m=m(P(j+k;h))\in M^k(P;h)$ by regarding as $j\in J(P;k)$ and $j+k=(j_b+k\,|\,b\in P)\in J(P;h)$, because $\partial_h b'=b(j_b;k)$ and $\partial^k b'=b(j_b+k;h)$ for $b'=b(j_b;t)$. Hence $m\in M_h(P;k)\oplus M^k(P;h)$; and we see (i).

(ii) Put $M_t=M(P;t)$. Then $M_t\subset C(\hat{M}_t)=B^t(P)$. Therefore, if $M_t\subset M\subset\bar{Z}^t$, then $\hat{M}_t\subset M_t\subset M\cap B^t(P)=M'\subset C(\hat{M}_t)$, and so $\hat{M}_t=\hat{M}'$ by Lemma 4.3(ii).

(iii) In the same way as the proof of (i), we see $\partial_h b(i;t)=b(i;k)=(\partial_h b)(i;k)$, $\partial^k b(i;t)=b(i+k;h)=(\partial^k b)(i;h)$, and (iii).

We see the following lemma in the same way as Lemma 4.4.

LEMMA 4.6. *We have* $M_h(\{b\};k)=[\partial_h b;k]$, *and*

$$M_h(P_1\cup P_2;k)=M_h(P_1;k)\,\&\,M_h(P_2;k)\ \textit{if}\ P_1\cap P_2=\phi.$$

These are also valid for M^h *and* ∂^h *instead of* M_h *and* ∂_h.

PROPOSITION 4.3. (i) *$M(P;1)=B^1(P)$ and $\hat{N}(S,P;1)=\hat{B}^1(P)-C([\bar{S};1])$.*
(ii) *$M_k(P;1)=B^1(\partial_k P)$ and $M^k(P;1)=B^1(\partial^k P)$.*
(iii) *By starting from k, we take any $M'(P;k)\supset M_1(P;k)$ and put*

$$M'(P;h)=(M'(P;h-1)\oplus M^{h-1}(P;1))\cap B^h(P)\qquad\textit{for}\ h\geqq k+1$$

inductively. Then $M(P;h)\subset M'(P;h)$, $\hat{M}(P;h)=\hat{M}'(P;h)$ and

$$\hat{N}(S,P;h)=\hat{M}'(P;h)-C([\bar{S};h])\quad(\bar{S}=S-P)\ \textit{for}\ h>k.$$

(iv) *In* (iii), *$m\oplus b\notin C([\bar{S};h])$ holds if $m\notin C([\partial_1\bar{S};h-1])$ or $b\notin C([\partial^{h-1}\bar{S};1])$.*

PROOF. (i) and (ii) are seen by the definition.

(iii) (i) and (ii) of Lemma 4.5 show (iii) for $h=k+1$. Moreover, if $M(P;h)\subset M'(P;h)$, then

$$M(P;h+1)\subset(M(P;h)\oplus M^h(P,1))\cap B^{h+1}(P)\subset M'(P;h+1),$$

and we see $\hat{M}(P;h+1)=\hat{M}'(P;h+1)$ by Lemma 4.5(ii). Therefore (iii) is proved inductively.

(iv) is a consequence of Lemma 4.5(iii).

In (iii) of this proposition, we may start from $k=1$ by using $M'(P;1)=B^1(\partial_1 P)$ in (ii). However, we know that $\hat{N}(S,P;k)=\phi$ for $k<k_1$ in the statement given after Proposition 4.2. Therefore, it is reasonable to start from $k=k_1$, where $P=P^p$ and $M_1(P^q;k)=[\partial_1 b^1;k]\,\&\cdots\&\,[\partial_1 b^q;k]$ by Lemma 4.6.

Moreover, we take the first $p_1 > 1$ with $\hat{N}_{k_1}(P^{p_1}) = \phi$ and $p_1 \leqq p+1$, where $p_1 = p+1$ means $\hat{N}_{k_1}(P) \neq \phi$. When $p_1 \leqq p$, we apply Proposition 4.3 for $P = P^{p_1} \subset S = \bar{S} \cup P^{p_1}$, $k = k_1$ and $M'(P^{p_1}; k_1) = M_1(P^{p_1}; k_1)$ to find the first $k_2 > k_1$ with $k_2 \leqq k_0$ and $\hat{N}_{k_2}(P^{p_1}) \neq \phi$. When such k_2 exists, we take the first $p_2 > p_1$ with $\hat{N}_{k_2}(P^{p_2}) = \phi$ and $p_2 \leqq p+1$. When $p_2 \leqq p$, we find k_3 from p_2 and k_2 in the same way as k_2 finding from p_1 and k_1, where we can take $M'(P^{p_2}; k_2) = M' \,\&\, M_1(P^{p_2} - P^{p_1}; k_2)$ for $M' = M'(P^{p_1}; k_2)$ obtained already to find k_2; and so on.

Then, we reach to the case $p_i = p+1$ or the case $p_i \leqq p$ and k_i does not exist; and $\hat{N}(S, P; k_i) = \hat{N}_{k_i}(P) \neq \phi$ in the first case, and $\hat{N}(S, P; k) = \phi$ for any k in the second case. Moreover, in the first case, $\hat{N}(S, P; k)$ is seen by Proposition 4.3 for any $k > k_i$ with $k \leqq k_0$.

According to the above statements, we find all solutions of (S, P) in the following example.

EXAMPLE 4.7. In example 4.6, we find $A^*(S, P)$ illustrated as follows:

(1) Since $k_0 = 5$ and $k_1 = 3$ are seen, we start from $k = 3$ and find first $p_1 > 1$ with $\hat{N}_3(P^{p_1}) = \phi$ by using Proposition 4.2. Then $p_1 = 2$.

(2) We find the first $k_2 > 3$ with $\hat{N}_{k_2}(P^2) \neq \phi$ by Proposition 4.3(iii) as follows:

$$\begin{aligned}
M'(P^2; 3) &= M_1(P^2; 3) \\
&= \{(200), (220), (201), (221)\}, \\
M^3(P^2; 1) &= \{(0), (1), (2)\}, \\
B^4(P^2) &= C(\hat{M}(P^2; 4)) \\
&= C(\{(2201), (2210), (2012)\}), \\
M'(P^2; 4) &= (M'(P^2; 3) \oplus M^3(P^2; 1)) \cap B^4(P^2) \\
&= \{(2000), (2001), (2002), (2010), (2012), (2200), \\
&\qquad (2201), (2210)\}, \\
[\bar{S}; 4] &= \{(2212), (2122), (1212), (2121), (1221)\}, \\
\hat{N}_4(P^2) &= \hat{M}'(P^2; 4) - C([\bar{S}; 4]) \\
&= \{(2201)\} \neq \phi,
\end{aligned}$$

then $k_2 = 4$.

(3) From (2), we find the first $p_2 > 2$ with $\hat{N}_4(P^{p_2}) = \phi$ as follows:

$$\begin{aligned}
&(\hat{N}_4(P^2) \,\&\, \hat{N}_4(\{b^3\})) - C([\bar{S}; 4]) \\
&\qquad = \{(0201)\} - C([\bar{S}; 4]) \\
&\qquad = \phi,
\end{aligned}$$

then $p_2 = 3$.

(4) We find the first $k_3 > 4$ with $\hat{N}_{k_3}(P^3) \neq \phi$ as follows:

$$\begin{aligned}
M'(P^3; 4) &= M'(P^2; 4) \& M_1(P^3 - P^2; 4) \\
&= \{(2000), (2001)\}, \\
M^4(P^3; 1) &= \{(0)\}, \\
B^5(P^3) &= C(\hat{M}(P^3; 5)) \\
&= C(\{(20010)\}), \\
M'(P^3; 5) &= (M'(P^3; 4) \oplus M^4(P^3; 1)) \cap B^5(P^3) \\
&= \{(20000), (20010)\}, \\
[\bar{S}; 5] &= \{(22122), (12122), (12121), (21221)\}, \\
\hat{N}_5(P^3) &= \hat{M}'(P^3; 5) - C([\bar{S}; 5]) \\
&= (20010) \neq \phi,
\end{aligned}$$

then $k_3 = 5$.

(5) In the same way as (3), we find the first $p_3 > 3$ with $\hat{N}_5(P^{p_3}) = \phi$ as follows:

$$\begin{aligned}
&(\hat{N}_5(P^3) \& \hat{N}_5(\{b^4\})) - C([\bar{S}; 5]) \\
&\quad = \{(00010)\} \\
&\quad \neq \phi,
\end{aligned}$$

then $p_3 = 5$ and $\hat{N}(S, P; 5) = \hat{N}_5(P) = \{(00010)\}$.

(6) We find

$$\begin{aligned}
A^5(S, P) &= C(\hat{N}(S, P; 5)) - C([\bar{S}; 5]) \\
&= \{(00010)\}.
\end{aligned}$$

Therefore, (00010) is a solution of the given (S, P).

5. An algorithm for finding solutions of (S, P)

In this section, we propose an algorithm for finding solutions of (S, P) in Definition 2.4. Let k_0 be the minimum length of P. Then we can write the set of all solutions of (S, P) as

$$A^*(S, P) = \bigcup_{k=1}^{k_0} A^k(S, P),$$

where $A^k(S, P)$ denotes the set of all solutions of (S, P) with length k. Further we have shown in Theorem 4.1 (i) that

$$A^k(S, P) = C(\hat{N}_k(P)) - C(S'),$$

where $S' = [\bar{S}; k]$, $S' = S - P$,

$$\hat{N}_k(P) = \hat{M}(P; k) - C(S').$$

For the definitions of $M(P; k)$, $\hat{M}(P; k)$ and $C(S')$, see Definitions 4.1, 4.3 and Lemma 4.1 (ii), respectively.

Our algorithm is essentially based on the above expression for $A^k(S, P)$. However, in order to reduce its computational time, we employed several devices stated bellow. Let $P = \{b^1, \ldots, b^p\}$ and $P^q = \{b^1, \ldots, b^q\}$ for $q \leqq p$. Here we choose b^1 whose length is k_0. Consider the set of sequences with length k constructed from b^1, i.e., $[b^1; k] = \{b^1(i; k) \mid i \in I(b^1; k)\}$. Let $P_{(i)} = \{b^1(i; k), b^2, \ldots, b^p\}$ and $P^q_{(i)} = \{b^1(i; k), b^2, \ldots, b^q\}$ for $q \leqq p$. Then it is easily seen that

$$\begin{aligned} A^k(S, P) &= \bigcup_{i \in I(b^1; k)} A^k(S, P_{(i)}) \\ &= \bigcup_{i \in I(b^1; k)} (C(\hat{N}_k(P_{(i)})) - C(S')). \end{aligned}$$

In our algorithm, we obtain $A^k(S, P_{(i)})$, $i \in I(b^1; k)$ inductively on k. As being noted in the statement given after Proposition 4.2, we may start from $k = k_1$, where k_1 is the minimum number of k satisfying $[b; k] \not\subset [\bar{S}; k]$ for any $b \in P$. Also we start from P^1, i,e., $q = 1$. The whole computation method, which is based on the statement given after the proof of Proposition 4.3, is given in Algorithm 1.

Relating to redundant solutions, we give the following

DEFINITION 5.1. For any $a \in A^*(S, P)$, we call a a *redundant solution* when there exists $a' \neq a$ in $A^*(S, P)$ such that

$$a' \subset a \quad \text{and} \quad a'[-I'] = a[-I],$$

where $a' = a'[I']$ and $a = a[I]$. We denote by $\underline{A}^*(S, P)$ the set of all *non-redundant solutions*. Further, let $\underline{A}(S, P_{(i)}) = A(S, P_{(i)}) \cap \underline{A}^*(S, P)$.

In this algorithm, we obtain only the set of non-redundant solutions, since the set of all redundant solutions can be constructed from the set of non-redundant solutions by Definition 5.1.

Here we note that we employ the following devices inside a computer:

(i) Let M be a subset of $\bar{Z}^k$. We use the mapping

$$\tilde{f}_{b^1(i;k)} : M \longrightarrow T(M, i) = \tilde{f}_{b^1(i;k)}(M) \subset \{0, 1\}^k$$

sending $c = (c_1 \cdots c_k) \in M$ to

$$\tilde{f}_{b^1(i;k)}(c) = (d_1 \cdots d_k),$$

where

$$d_j = \begin{cases} 1 & \text{if } c_j = (b^1(i;k))_j, \\ 0 & \text{otherwise.} \end{cases}$$

Then we can use AND/OR binary operations. Moreover, it is possible to convert any element in $M(P^q_{(i)}; k)$ (resp. $\hat{N}_k(P^q_{(i)})$) into an integer between 0 and $2^k - 1$. This enables us to check easily whether an element matches with the other in $M(P^q_{(i)}; k)$ (resp. $\hat{N}_k(P^q_{(i)})$). Moreover $M^{k-1}(P; 1)$ can be replaced by a subset of $\{0, 1\}$ in Proposition 4.3(iii).

(ii) In order to obtain S' rapidly, we prepare a hashing table in the main memory which gives a correspondence from the first few components of the elements of S' to a set of integers.

Some of the computations used in Algorithm 1 are given in Algorithms $2 \sim 6$. In Algorithm 2, we obtain the set $D_{(i)}$ of all elements $a = (a_1 \cdots a_k)$ satisfying

$$a \subset (1 \cdots 1) = \tilde{f}_{b^1(i,k)}(b^1(i,k)) \quad \text{and} \quad a \notin T = T(C(S'), i)$$

for $b^1(i;k)$. Let $T = \{d^1, \dots, d^t\}$. We define

$$c^i_j = \begin{cases} 0 & \text{if } d^i_j = 1 \\ 1 & \text{if } d^i_j = 0 \end{cases} \qquad \text{for } 1 \leqq j \leqq k,\ 1 \leqq i \leqq t.$$

Then, $a \in D_{(i)}$ if and only if

$$\sum_{j=1}^{k} c^i_j a_j \geqq 1 \qquad \text{for } 1 \leqq i \leqq t.$$

Therefore, this problem is related to the set covering problem whose computation methods have been studied in [7, 18]. We define

$$\underline{D}_{(i)} = \{d \in D_{(i)} \mid d' \not\subset d \text{ for any } d' \in D_{(i)} - \{d\}\},$$

since $d \in D_{(i)}$ with $d' \in D_{(i)}$ and $d' \subset d$. For $m \in T(\hat{M}(P^q_{(i)}; k), i) \subset C(\{1\})$, $m \notin T$ implies $d \subset m$ for some $d \in \underline{D}_{(i)}$ by the definition. Our algorithm uses this property in obtaining $T(\hat{M}(P^q_{(i)}; k), i) - T$. These are also valid for $T(\hat{N}_k(P^q_{(i)}), i)$ or $T(C(\hat{N}_k(P_{(i)})), i)$ instead of $T(\hat{M}(P^q_{(i)}; k), i)$. Let

$$I = \{i \mid b^1(i;k) \in M_1(\{b^1\}; k)\},$$
$$J = \{i \in I(b^1; k) \mid \hat{N}_k(P^q_{(i)}) \neq \phi\},$$

where $k_1 \leqq k \leqq k_0$ and $1 \leqq q \leqq p$.

In Algorithm 3, we obtain the first integer q' such that $T(\hat{N}_k(P^{q'}_{(i)}), i) = \phi$

for any $i \in J$ and $q < q' \leqq p$. Then, let $q = q'$ if such q' exists, $= p+1$ otherwise. Further, the sets $T(M(P^q_{(i)}; k), i)$, $i \in I$ and $T(\hat{N}_k(P^q_{(i)}), i)$, $i \in J$ are obtained by applying Lemma 4.4 (ii) and Proposition 4.2 (iii), respectively.

In Algorithm 4, we apply Proposition 4.3 (iii) for $T(M(P^q_{(i)}; k), i)$, obtain the first k' such that $T(\hat{N}_{k'}(P^q_{(i)}), i) \neq \phi$ for some $i \in J$ and $k < k' \leqq k_0$. Then, let $k = k'$ if such k' exists, $= k_0 + 1$ otherwise. All non-redundant solutions with length k are found in Algorithm 5 by using $\underline{D}_{(i)}$ in Algorithm 2. In Algorithm 6, the computation removing all elements in $T(C(S'), i)$ from $T(\hat{M}(P^q_{(i)}; k), i)$ or $T(\hat{N}_k(P^{q-1}_{(i)}), i) \,\&\, T(N_k(\{b^q\}), i)$ is also done by using $\underline{D}_{(i)}$ in Algorithm 2.

We list some additional notations used in Algorithms $1 \sim 6$.

(i) $T(M^q_k, i) = T(M(P^q_{(i)}; k), i)$ for $i \in I$.

(ii) $T(\hat{N}^q_k, i) = T(\hat{N}_k(P^q_{(i)}), i)$ for $i \in J$.

(iii) $(T(M, i), I)$ or $(\underline{D}_{(i)}, I)$ denote $T(M, i)$ or $\underline{D}_{(i)}$ for each $i \in I$, respectively.

(vi) Let e^h be the element whose i-th coordinate is given by

$$e^h_i = \begin{cases} 1 & \text{if } i = h, \\ 0 & \text{otherwise.} \end{cases}$$

ALGORITHM 1.

Function: To find the set $\underline{A}^*(S, P)$ of all non-redundant solutions of (S, P).

(S1) Find b^1 whose length is $k_0 = \min\{\ell(b) \mid b \in P\}$.

(S2) Let $k := 1$, $p := |P|$, $\bar{S} := S - P$, $S' := [\bar{S}; k]$ and $\underline{A}^*(S, P) := \phi$.

(S3) For each $b \in P$, do steps S4 and S5.

(S4) If $[b; k] - S' = \phi$, then do step S5.

(S5) If $k < k_0$, then let $k := k + 1$,
$S' := [\bar{S}; k]$,
go to step S4,
else exit: "there is no solution".

(S6) Let $I := \{i \mid b^1(i; k) \in M_1(\{b^1\}; k)\}$,
$J := \{i \mid b^1(i; k) \in \hat{N}_k(\{b^1\})\}$.

(S7) Let $T(M^1_k, i) := T(M(P^1_{(i)}; k), i)$ for each $i \in I$,
$T(\hat{N}^1_k, i) := T(\hat{N}_k(P^1_{(i)}), i)$ for each $i \in J$.

(S8) For each $i \in J$, do step S9.

(S9) Obtain $\underline{D}_{(i)}$ by Algorithm 2 with parameters k, $b^1(i; k)$ and S'.

(S10) Let $q := 1$.

(S11) Obtain q, $(T(M^q_k, i), I)$ and $(T(\hat{N}^q_k, i), J)$ by Algorithm 3 with parameters k_0, k, p, q, S', P, $(T(M^q_k, i), I)$, $(T(\hat{N}^q_k, i), J)$ and $(\underline{D}_{(i)}, J)$.

(S12) If $q \leqq p$, then do steps S13 to S14.

(S13) Obtain k, S', $(T(M_k^q, i), I)$, $(T(\hat{N}_k^q, i), J)$ and $(\underline{D}_{(i)}, J)$ by Algorithm 4 with parameters k_0, k, q, $\bar{S}$, P and $(T(M_k^q, i), I)$.

(S14) If $k \leqq k_0$, then go to step S11, else exit.

(S15) Obtain $\underline{A}^*(S, P)$ by Algorithm 5 with parameters $(T(\hat{N}_k^p, i), J)$, $(\underline{D}_{(i)}, J)$ and $\underline{A}^*(S, P)$.

(S16) Let $q := p$.

(S17) Obtain k, S', $(T(M_k^q, i), I)$, $(T(\hat{N}_k^q, i), J)$ and $(\underline{D}_{(i)}, J)$ by Algorithm 4 with parameters k_0, k, q, $\bar{S}$, P and $(T(M_k^q, i), I)$.

(S18) If $k \leqq k_0$, then obtain $\underline{A}^*(S, P)$ by Algorithm 5 with parameters $(T(\hat{N}_k^p, i), J)$, $(\underline{D}_{(i)}, J)$ and $\underline{A}^*(S, P)$, and go to step S17.

(S19) End ALGORITHM 1.

ALGORITHM 2.
Function: To find all solutions of the set covering problem.
Input parameters: k, $b^1(i; k)$ and S'.
Output parameters: $\underline{D}_{(i)}$.

(S20) Let $m := (1 \cdots 1)$, $s := b^1(i; k)$, $n := (0 \cdots 0)$, $R := \phi$ and $\underline{D}_{(i)} := \phi$.

(S21) For each $b \in S'$, do steps S22 to S26.

(S22) If $b = b^1(i; k)$, then go to step S35.

(S23) Let c be the element whose j-th coordinate is given by

$$c_j = \begin{cases} 0 & \text{if } b_j = (b^1(i; k))_j, \\ 1 & \text{otherwise.} \end{cases}$$

(S24) Let $m := m \& c$,
$s := s \& b$.

(S25) If $c \& n = (0 \cdots 0)$, then do step S26.

(S26) If $c = e^h$ for some $h \in \{1, \ldots, k\}$, then let
$n_h := 1$,
$R := R - \{r \in R \mid r_h = 1\}$,
else if $r \not\subset c$ for any $r \in R$, then let
$R := R - \{r \in R \mid c \subset r\}$,
$R := R \cup \{c\}$.

(S27) If $R = \phi$, then do step S28.

(S28) If $n \neq (0 \cdots 0)$, let $\underline{D}_{(i)} := \{n\}$.
Go to step S34.

(S29) Let $L := \{j \in \{1, \ldots, k\} \mid s_j \neq 0$ or $n_j = 1$ or $m_j = 1\}$,
$C := \{c \mid c \subset (1 \cdots 1)[L]\}$.

(S30) For each $c \in C$, do steps S31 and S32.

(S31) If $d \not\subset c$ for any $d \in \underline{D}_{(i)}$, then do step S32.

(S32) If $c \& r \neq (0 \cdots 0)$ for any $r \in R$, then let $\underline{D}_{(i)} := \underline{D}_{(i)} \cup \{c\}$.

(S33) For each $d \in \underline{D}_{(i)}$, let
$d_j := 1$ if $n_j = 1$ for $1 \leqq j \leqq k$.

(S34) For j from 1 to k,
if $m_j = 1$ and $n_j \neq 1$, then let $\underline{D}_{(i)} := \underline{D}_{(i)} \cup \{e^j\}$.

(S35) End ALGORITHM 2.

ALGORITHM 3.

Function: To obtain the first integer q' such that $q < q' \leqq p$ and $T(\hat{N}_k^{q'}, i) = \phi$ for any $i \in J$.

Input parameters: $k_0, k, p, q, S', P, (T(M_k^q, i), I), (T(\hat{N}_k^q, i), J)$ and $(\underline{D}_{(i)}, J)$.

Output parameters: $q, (T(M_k^q, i), I)$ and $(T(\hat{N}_k^q, i), J)$.

(S36) If $q \geqq p$, then let $q := p + 1$ and go to step S43.

(S37) Let $q' := q + 1$.

(S38) If $k < k_0$, then let $T(M_k^{q'}, i) := T(M_k^q, i) \,\&\, T(M_1(\{b^{q'}\}; k), i)$ for each $i \in I$.

(S39) Let $T(N_k^{q'}, i) := T(\hat{N}_k^q, i) \,\&\, T(N_k(\{b^{q'}\}), i)$ for each $i \in J$,
$q := q'$.

(S40) Obtain $(T(N_k^q, i), J)$ by Algorithm 6 with parameters $(T(N_k^q, i), J)$ and $(\underline{D}_{(i)}, J)$.

(S41) If $J = \phi$, then go to step S43.

(S42) Obtain $T(\hat{N}_k^q, i)$ from $T(N_k^q, i)$ for each $i \in J$ and go to step S36.

(S43) End ALGORITHM 3.

ALGORITHM 4.

Function: To obtain the first integer k' such that $k < k' \leqq k_0$ and $T(\hat{N}_{k'}^q, i) \neq \phi$ for some $i \in J$.

Input parameters: $k_0, k, q, \bar{S}, P$ and $(T(M_k^q, i), I)$.

Output parameters: $k, S', (T(M_k^q, i), I), (T(\hat{N}_k^q, i), J)$ and $(\underline{D}_{(i)}, J)$.

(S44) If $k \geqq k_0$, then let $k := k_0 + 1$ and go to step S57.

(S45) Let $k' := k + 1$,
$S' := [\bar{S}; k']$,
$J := \{i \mid b^1(i; k') \in M(\{b^1\}; k')\}$.

(S46) For each $i \in J$, do step S47.

(S47) If $i \notin I$,
then find j such that $j \in I$ and $b^1(i; k) = b^1(j; k)$ and let $T(M_k^q, i) := T(M_k^q, j)$.

(S48) Let $T(M_{k'}^q, i) := T(M_k^q, i) \oplus \{0, 1\}$ for each $i \in J$,
$k := k'$.

(S49) For each $b \in p^q$, do steps S50 to S51.

(S50) For each $i \in J$, do step S51.

(S51) Let $T(M_k^q, i) := T(M_k^q, i) \cap T(B^k([b; k]), i)$ and if $T(M_k^q, i) = \phi$, then let $T(M_k^q, i) := \{0\}$.

(S52) If $T(M_k^q, i) = \{0\}$ for any $i \in J$, then go to step S44.

(S53) Obtain $T(\hat{M}_k^q, i)$ from $T(M_k^q, i)$ for each $i \in J$.

(S54) For each $i \in J$, do step S55.

(S55) Obtain $\underline{D}_{(i)}$ by Algorithm 2 with parameters k, $b^1(i; k)$ and S'.

(S56) Obtain $(T(\hat{N}_k^q, i), J)$ by Algorithm 6 with parameters $(T(\hat{M}_k^q, i), J)$ and $(\underline{D}_{(i)}, J)$.

(S57) If $J = \phi$, then go to step S44.

(S58) Let $I := \{i \mid b^1(i; k) \in M_1(\{b^1\}; k)\}$.

(S59) End ALGORITHM 4.

ALGORITHM 5.

Function: To obtain all non-redundant solutions of (S, P) with length k.

Input parameters: $(T(\hat{N}_k^q, i), J)$, $(\underline{D}_{(i)}, J)$ and $\underline{A}^*(S, P)$.

Output parameters: $\underline{A}^*(S, P)$.

(S60) For each $i \in J$, do steps S61 to S65.

(S61) Let $C := \{c \mid c \subset n$ and $n \in T(\hat{N}_k^q, i)\}$.

(S62) For each $c \in C$, do steps S63 to S65.

(S63) If $d \subset c$ for some $d \in \underline{D}_{(i)}$, then do steps S64 and S65.

(S64) Let $L := \{i \mid c_i = 0\}$,
$a := (b^1(i; k))[L]$.

(S65) If $a' \not\subset a$ or $a[-L] \neq a'[-L']$ for any $a' = a'[L'] \in \underline{A}^*(S, P)$, then add a to $\underline{A}^*(S, P)$.

(S66) End ALGORITHM 5.

ALGORITHM 6.

Function: To obtain $T(N_k^q, i) - T(C(S'), i)$ (resp. $T(\hat{M}_k^q, i) - T(C(S'), i)$) for each $i \in J$.

Input parameters: $(T(N_k^q, i), J)$ (resp. $(T(\hat{M}_k^q, i), J)$) and $(\underline{D}_{(i)}, J)$.

Output parameters: $(T(N_k^q, i), J)$ (resp. $(T(\hat{N}_k^q, i), J)$).

(S67) For each $i \in J$, do steps S68 to S71.

(S68) Let $W := \phi$.

(S69) For each $m \in T(\hat{M}_k^q, i)$, do step S70.

(S70) If $d \subset m$ for some $d \in \underline{D}_{(i)}$, then add m to W.

(S71) Let $T(\hat{N}_k^q, i) := W$.

(S72) Let $J := \{i \mid T(N_k^q, i) \neq \phi\}$.

(S73) End ALGORITHM 6.

EXAMPLE 5.1. We show the solutions of (S, P) can be constructed through Algorithms $1 \sim 6$ for the sets S and P of Example 4.6.

1) We have $b^1 = (22112)$ and $k_0 = 5$ in S1.

2) We have $k=1$, $p=4$, $\bar{S}=\{(22122), (12122), (12121), (21221)\}$ and $S'=\{1, 2\}$ in S2.
3) We have $k=3$ and $S'=\{(221), (212), (122), (121)\}$ in S3–S5.
4) We have $I=\{1, 2\}$ and $J=\{2, 3\}$ in S6.
5) We have $T(M_3^1, 1)=\{(111)\}$, $T(M_3^1, 2)=\{(111)\}$, $T(\hat{N}_3^1, 2)=\{(111)\}$ and $T(\hat{N}_3^1, 3)=\{(111)\}$ in S7.
6) By Algorithm 2 with parameters k, $b^1(2; k)$ and S' in S9, $m=(000)$, $s=(000)$, $n=(011)$ and $R=\phi$ are obtained in S20–S26. $\underline{D}_{(2)}=\{(011)\}$ is obtained as $\underline{D}_{(i)}$ in S28.
7) By Algorithm 2 with parameters k, $b^1(3; k)$ and S' in S9, $m=(000)$, $s=(000)$, $n=(110)$ and $R=\phi$ are obtained in S20–S26. $\underline{D}_{(3)}=\{(110)\}$ is obtained as $\underline{D}_{(i)}$ in S28.
8) We have $q=1$ in S10.
9) By Algorithm 3 with parameters k_0, k, p, q, S', P, $(T(M_k^q, i), I)$, $(T(\hat{N}_k^q, i), J)$ and $(\underline{D}_{(i)}, J)$ in S11, we have 10) to 11).
10) We have $q'=2$, $T(M_3^2, 1)=\{(110), (111)\}$, $T(M_3^2, 2)=\{(100), (101)\}$, $T(N_3^2, 2)=\{(100)\}$ and $T(N_3^2, 3)=\{(001)\}$ in S37, S38 and S39, respectively. We have $q=2$ in S39.
11) By Algorithm 6 with parameters $(T(N_k^q, i), J)$ and $(\underline{D}_{(i)}, J)$ in S40, $T(N_3^2, 2)=\phi$, $T(N_3^2, 3)=\phi$ and $J=\phi$ are obtained in S68–S71 and S72, respectively.
12) By Algorithm 4 with parameters k_0, k, q, $\bar{S}$, P and $(T(M_k^q, i), I)$ in S13, we have 13) to 18).
13) We have $k'=4$, $S'=\{(2212), (2122), (1212), (2121), (1221)\}$ and $J=\{1, 2\}$ in S45.
14) We have $T(\hat{M}_4^2, 1)=\{(1101), (1110)\}$ and $T(\hat{M}_4^2, 2)=\{(1011)\}$ in S53, since $T(M_4^2, 1)=\{(1100), (1101), (1110)\}$ and $T(M_4^2, 2)=\{(1000), (1001), (1010), (1011)\}$ in S48–S51. We have $k=4$ in S48.
15) By Algorithm 2 with parameters k, $b^1(1; k)$ and S' in S55, $m=(0000)$, $s=(0000)$, $n=(0001)$ and $R=\{(0110), (1010)\}$ are obtained in S20–S26. $L=\phi$ and $C=\{c \mid c \subset (1111)\}$ are obtained in S29. $\underline{D}_{(1)}=\{(1100), (0100)\}$ is obtained as $\underline{D}_{(i)}$ in S30–S32. $\underline{D}_{(1)}=\{(1101), (0011)\}$ is obtained as $\underline{D}_{(i)}$ in S33.
16) By Algorithm 2 with parameters k, $b^1(2; k)$ and S' in S55, $m=(0000)$, $s=(0000)$, $n=(0110)$ and $R=\phi$ are obtained in S20–S26. $\underline{D}_{(2)}=\{(0110)\}$ is obtained as $\underline{D}_{(i)}$ in S28.
17) By Algorithm 6 with parameters $(T(\hat{M}_k^q, i), J)$ and $(\underline{D}_{(i)}, J)$ in S56, $T(\hat{N}_4^2, 1)=\{(1101)\}$, $T(\hat{N}_4^2, 2)=\phi$ and $J=\{1\}$ are obtained in S68–S71 and S72, respectively.
18) We have $I=\{1\}$ in S58.
19) By Algorithm 3 with parameters k_0, k, p, q, S', P, $(T(M_k^q, i), I)$, $(T(\hat{N}_k^q, i)$,

J) and $(\underline{D}_{(i)}, J)$ in S11, we have 20) to 21).

20) We have $q' = 3$, $T(M_4^3, 1) = \{(1000), (1001)\}$ and $T(N_4^3, 1) = \{(0101)\}$ in S37, S38 and S39, respectively. We have $q = 3$ in S39.

21) By Algorithm 6 with parameters $(T(N_k^q, i), J)$ and $(\underline{D}_{(i)}, J)$ in S40, $T(N_4^3, 1) = \phi$ and $J = \phi$ are obtained in S68–S71 and S72, respectively.

22) By Algorithm 4 with parameters k_0, k, q, $\bar{S}$, P and $(T(M_k^q, i), I)$ in S13, we have 23) to 27).

23) We have $k' = 5$, $S' = \{(22122), (12122), (12121), (21221)\}$ and $J = \{1\}$ in S45.

24) We have $T(\hat{M}_5^3, 1) = \{(10010)\}$ in S53, since $T(M_5^3, 1) = \{(10000), (10010)\}$ in S48–S51. We have $k = 5$ in S48.

25) By Algorithm 2 with parameters k, $b^1(1; k)$ and S' in S55, $m = (00010)$, $s = (00000)$, $n = (00010)$ and $R = \phi$ are obtained in S20–S26. $\underline{D}_{(1)} = \{(00010)\}$ is obtained as $\underline{D}_{(i)}$ in S28.

26) By Algorithm 6 with parameters $(T(\hat{M}_k^q, i), J)$ and $(\underline{D}_{(i)}, J)$ in S56, $T(\hat{N}_5^3, 1) = \{(10010)\}$ and $J = \{1\}$ are obtained in S68–S71 and S72, respectively.

27) We have $I = \phi$ in S58.

28) By Algorithm 3 with parameters k_0, k, p, q, S', P, $(T(M_k^q, i), I)$, $(T(\hat{N}_k^q, i)$, $J)$ and $(\underline{D}_{(i)}, J)$ in S11, we have 29) to 32).

29) We have $q' = 4$, $T(N_5^4, 1) = \{(00010)\}$ in S37 and S39, respectively. We have $q = 4$ in S39.

30) By Algorithm 6 with parameters $(T(N_k^q, i), J)$ and $(\underline{D}_{(i)}, J)$ in S40, $T(N_5^4, 1) = \{(00010)\}$ and $J = \{1\}$ are obtained in S68–S71 and S72, respectively.

31) We have $T(\hat{N}_5^4, 1) = \{(00010)\}$ in S42.

32) $q = 5$ is obtained in S36.

33) By Algorithm 5 with parameters $(T(\hat{N}_k^q, i), J)$, $(\underline{D}_{(i)}, J)$ and $\underline{A}^*(S, P)$ in S15, $\underline{A}^*(S, P) = \{(00010)\}$ is obtained in S60–S65.

6. An application to molecular biology

In this section, we show an application to analysis of nucleic acid and protein sequences in molecular biology by using our algorithm. These sequences are basic in the biochemical activity of all living things. In case of nucleic acids, the units (or elements) are any one of four nucleotides, and the lengths of the sequences are typically from tens to millions. In case of amino acids, the units are any one of twenty amino acids, and the lengths of the sequences are typically a few hundred. All these sequences can be taken from Nucleic Acid Sequence Databases; DDBJ [2], EMBL [3], GenBank [8] and NBRF [15], and Protein Sequence Database; NBRF/PIR [16] stored in Hiroshima University Database Management System HDM [5, 6, 11]. Here we consider nucleic acid sequences taken from GanBank Database whose records (entries) are stored in HDM as Table 6.1.

Table 6.1 Number of records and groups in GenBank Database

GenBank release 67.0

Group	No. of entries	No. of bases
Primates	8206	9814969
Rodent	8400	8546574
Other mammalian	1638	2077398
Other vertebrate	1965	2249342
Invertebrate	3383	4307294
Plant	3187	5069146
Organelle	1402	2021305
Bacterial	4616	7572518
Structural RNA	1735	518967
Viral	4032	7019564
Phage	602	705269
Synthetic	1053	550086
Unannotated	3684	4716844
TOTAL	43903	55169276

As an attempt of seeing the usefulness of our computation method, we performed to find solutions of (S, P) for the case given in Table 6.2.

Table 6.2 The contents of P and S-P

	P	S−P
Retrieved group	Structural RNA	Viral
No. of records	25	25
Minimum record length	100	195
Maximum record length	1486	4675

Table 6.3 shows the numbers of elements in $M(\{b^1\}; k)$, $\hat{N}_k(\{b^1\})$ and S', the sum of the numbers of elements in $\underline{D}_{(i)}$ for all $b^1(i; k)$ in $\hat{N}_k(\{b^1\})$, $T(M(P; k), i)$ for all $b^1(i; k)$ in $M(\{b^1\}; k)$ and $T(\hat{N}_k(P), i)$ for all $b^1(i; k)$ in $\hat{N}_k(\{b^1\})$, $\underline{A}^k(S, P)$, and the time needed to find solutions, for the case $k = 2 \sim 11$. We note that each of the elements of $\underline{A}^k(S, P)$ or $A^k(S, P)$ given in Table 6.3 can be considered as the characteristic patterns of the 25 elements in Structural RNA in the 50 elements in Structural RNA and Viral.

Table 6.3 The characteristic patterns of P in S

k	2	3	4	5	6	7	8	9	10	11
$M(\{b^1\}; k)$	18									
$\hat{N}_k(\{b^1\})$	2									
S'	20	70	264	1034	3956	10577	16499	19275	20241	20574
$\underline{D}_{(i)}$	2	6	12	18	31	72	253	877	2651	6863
$T(M(P; k), i)$	79	401	1159	2357	4198	6870	10649	15919	23138	32818
$T(\hat{N}_k(P), i)$	1	2	4	6	9	12	15	19	24	32
$\underline{A}^k(S, P)$	cn	—	c00n	—	cn000g gaaggc ttaagt	—	—	t0gg00taa	cg000taa0t	a00agg0tt0g a00agg0t00g gaa0gc000gg g00t0aag00c
Time*	7.48	25.65	49.89	75.98	126.62	242.74	396.36	524.94	675.53	968.83

*) All times are expressed in second.
The set Z consists of 4 alphabet a, c, g, t.
The leter n denotes an element of Z in Database.

The computer program for storing and retrieving sequences was written in the Model 204 Database Management System User Language [14]. The program consists of 3380 statements organized 15 programs and required 560 K bytes basic storage and 6 M bytes extended storage. On the other hand, the program for finding solutions was written in the FORTRAN programming language and consists of 1530 statements organized into 32 subroutines. All computing was done on the HITAC M-680 H computer at Information Processing Center of Hiroshima University.

Acknowledgements

I would like to thank Professor Hideto Ikeda, Hiroshima University, for suggesting me this problem and his valuable guidance all over this work. I am much indebted to Professor Masahiro Sugawara, Hiroshima University, for his encouragement and his valuable directions during the preparation of this paper. I am also grateful to Professor Yasunori Fujikoshi, Hiroshima University, for his helpful advice and constructive criticism.

References

[1] M. O. Dayhoff, Survey of new data and computer methods of analysis, Atlas of Protein Sequence and Structure, M. O. Dayhoff, ed., (National Biomedical Research Foundation, Silver Spring, Md.), **5** (1978), 1–8.

[2] DDBJ, DNA Data Bank of Japan, National Institute of Genetics, Mishima.

[3] EMBL, Nucleotide Sequence Database, European Molecular Biology Laboratory, Heidelberg.

[4] W. M. Fitch, An improved method of testing for evolutionary homology, J. Mol. Biol., **16** (1966), 9–16.

[5] K. Futagami, W. Ishihara and H. Ikeda, An extended functions of database management system, Proc. 24th Conf. on Information Processing Society of Japan (IPSJ) (1982), 493–494 (Japanese).

[6] K. Futagami and H. Ikeda, An estimation of integrated data directory dictionary system, 32nd IPSJ (1986), 833–834 (Japanese).

[7] R. S. Garfinkel and G. L. Nemhauser, Integer Programming, John Wiley & Sons, 1972.

[8] Genbank, Genetic Sequence Data Bank, IntelliGenetics, Inc., Los Alamos National Larboratory.

[9] W. B. Goad and M. Kanehisa, Pattern recognition in nucleic acid sequences. I. A general method for finding local homologies and symmetries, Nucleic Acids Research, **10** (1982), 247–263.

[10] D. S. Hirschberg, A linear space algorithm for computing maximal common subsequences, Comm. ACM, **18** (1975), 341–343.

[11] H. Ikeda, K. Futagami, W. Ishihara and Y. Kobayashi, Modification of a database management system HDM to a relational system, Technical Report Series No. 2, Information Processing Center, Hiroshima University, (1982), 1–20.

[12] L. J. Korn, C. L. Queen and M. N. Wegman, Computer analysis of nucleic acid regulatory sequences, Proc. Nat. Acad. Sci. U.S.A, **74** (1977), 4401–4405.

[13] D. J. Lipman and W. R. Pearson, Rapid and sensitive protein similarity searches, Science, **227** (1985), 1435–1441.

[14] Model 204, the proprietary software product developed and maintained by Computer Corporation of America, Cambridge, Massachusetts.

[15] NBRF, Nucleic Acid Sequence Database, National Biomedical Research Foundation, Georgetown University Medical Center, Washington, D.C..

[16] NBRF/PIR, Protein Sequence Database, Protein Identification Resource (PIR), National Biomedical Research Foundation, Georgetown University Medical Center, Washington, D.C..

[17] S. B. Needleman and C. D. Wunsch, A general method applicable to the search for similarities in the amino acid sequence of two proteins, J. Mol. Biol., **48** (1970), 443–453.

[18] H. M. Salkin, Integer Programming, Addison-Wesley, 1975.

[19] D. Sankoff, Matching sequences under deletion/insertion constraints, Proc. Nat. Acad. Sci. U.S.A., **69** (1972), 4–6.

[20] P. H. Sellers, On the theory and computation of evolutionary distances, SIAM J. Appl. Math., **26** (1974), 787–793.

[21] P. H. Sellers, Pattern recognition in genetic sequences, Proc. Nat. Acad. Sci. U.S.A., **76** (1979), 3041.

[22] T. F. Smith and M. S. Waterman, Identification of common molecular subsequences, J. Mol. Biol., **147** (1981), 195–197.

[23] W. J. Wilbur and D. J. Lipman, Rapid similarity searches of nucleic acid and protein data banks, Proc. Nat. Acad. Sci. U.S.A., **80** (1983), 726–730.

[24] W. J. Wilbur and D. J. Lipman, The context dependent comparison of biological sequences, SIAM J. Appl. Math., **44** (1984), 557–567.

Information Processing Center,
Hiroshima University