Formal Aspects of Computing (1999) 11(1): 56–74

Computing Permutation Encodings

K. Rustan M. Leino[1]

Department of Computer Science, California Institute of Technology,
Pasadena, CA 91125, USA

Keywords: Permutation encodings, inversion table, code of an array, program derivation, program inversion.

Abstract. A permutation can be encoded in several different ways. This paper discusses some relations among some encodings and how one can be computed from others. The paper shows a short proof of an existing efficient algorithm for encoding a permutation and presents two new efficient algorithms. One of the new algorithms is constructed as the inverse of an existing algorithm for decoding, making it the first efficient permutation encoding algorithm obtained in that way.

0. Overview

0.0. Permutation encodings

Let $\mathbf{A}$ be a permutation of the first $\mathbf{N}$ natural numbers. For any $\mathbf{i}$ such that $0 \le \mathbf{i} < \mathbf{N}$, we let $\mathbf{A}.\mathbf{i}$ denote element $\mathbf{i}$ of the permutation. For example, if $\mathbf{N} = 9$ and $\mathbf{A}$ is the permutation

$$4 \quad 8 \quad 0 \quad 7 \quad 1 \quad 5 \quad 3 \quad 6 \quad 2\,,$$

then $\mathbf{A}.0 = 4$ and $\mathbf{A}.2 = 0$.

Since the elements of $\mathbf{A}$ are unique, for all $\mathbf{i}$ and $\mathbf{j}$ such that $0 \le \mathbf{i} \le \mathbf{j} < \mathbf{N}$, we have,

$$\mathbf{A}.\mathbf{i} = \mathbf{A}.\mathbf{j} \;\equiv\; \mathbf{i} = \mathbf{j}\,. \tag{0}$$

Correspondence and offprint requests to: K.R.M. Leino, Compaq Systems Research Center, 130 Lytton Ave., Palo Alto, CA 94301, USA, rustan@pa.dec.com.

[1] Supported in part by a research grant from the Digital Equipment Corporation Systems Research Center, 1993–94.

For the rest of this section, $\mathbf{i}$ is implicitly universally quantified over $0 \le \mathbf{i} < \mathbf{N}$.

Consider the encoding $\mathbf{B}$ of $\mathbf{A}$ defined by

$$\mathbf{B}.\mathbf{i} \;=\; \langle\, \#\,\mathbf{j} \mid 0 \le \mathbf{j} < \mathbf{i} \;::\; \mathbf{A}.\mathbf{j} < \mathbf{A}.\mathbf{i} \,\rangle\ . \tag{1}$$

The right-hand side is a quantified expression, where $\mid$ and $::$ separate the dummy (bound variable), the range of the dummy, and the term of the expression. The formula expresses that $\mathbf{B}.\mathbf{i}$ equals the number of times $\mathbf{A}.\mathbf{j} < \mathbf{A}.\mathbf{i}$ is true as $\mathbf{j}$ ranges over $0 \le \mathbf{j} < \mathbf{i}$. Thinking of the $\mathbf{A}.\mathbf{i}$'s as arranged from left to right, (1) states that $\mathbf{B}.\mathbf{i}$ is the number of values smaller than $\mathbf{A}.\mathbf{i}$ to the left of $\mathbf{A}.\mathbf{i}$ in the permutation, *i.e.*, among the leftmost $\mathbf{i}$ numbers in the permutation. For example, if $\mathbf{N} = 9$ and $\mathbf{A}$ is the permutation shown above, then $\mathbf{B}$ is

$$0 \quad 1 \quad 0 \quad 2 \quad 1 \quad 3 \quad 2 \quad 5 \quad 2\ .$$

We have $\mathbf{B}.5 = 3$ because three of the numbers $\mathbf{A}.0$, $\mathbf{A}.1$, $\mathbf{A}.2$, $\mathbf{A}.3$, and $\mathbf{A}.4$ are less than $\mathbf{A}.5$.

In [Dij78] and [Gri81], $\mathbf{B}$ is called the *code* of $\mathbf{A}$. A variation $\mathbf{C}$ of $\mathbf{B}$ is

$$\mathbf{C}.\mathbf{i} \;=\; \langle\, \#\,\mathbf{j} \mid 0 \le \mathbf{j} < \mathbf{i} \;::\; \mathbf{A}.\mathbf{j} > \mathbf{A}.\mathbf{i} \,\rangle\ . \tag{2}$$

Due to (0), a relationship between $\mathbf{B}$ and $\mathbf{C}$ is

$$\mathbf{B}.\mathbf{i} + \mathbf{C}.\mathbf{i} \;=\; \mathbf{i}\ , \tag{3}$$

from which we see that one can compute $\mathbf{B}$ from $\mathbf{C}$ or $\mathbf{C}$ from $\mathbf{B}$ in time $\mathrm{O}(\mathbf{N})$.

Two other encodings $\mathbf{D}$ and $\mathbf{E}$ of $\mathbf{A}$ are defined by

$$\mathbf{D}.\mathbf{i} \;=\; \langle\, \#\,\mathbf{j},\mathbf{k} \mid \mathbf{A}.\mathbf{k} = \mathbf{i} \;\wedge\; 0 \le \mathbf{j} < \mathbf{k} \le \mathbf{N} \;::\; \mathbf{A}.\mathbf{j} < \mathbf{A}.\mathbf{k} \,\rangle\ , \tag{4}$$

$$\mathbf{E}.\mathbf{i} \;=\; \langle\, \#\,\mathbf{j},\mathbf{k} \mid \mathbf{A}.\mathbf{k} = \mathbf{i} \;\wedge\; 0 \le \mathbf{j} < \mathbf{k} \le \mathbf{N} \;::\; \mathbf{A}.\mathbf{j} > \mathbf{A}.\mathbf{k} \,\rangle\ . \tag{5}$$

Encoding $\mathbf{E}$ is called the *inversion table* of $\mathbf{A}$ (*cf.* [Knu73]). Stated in words, $\mathbf{D}.\mathbf{i}$ is the number of times $\mathbf{A}.\mathbf{j} < \mathbf{A}.\mathbf{k}$ is true as $\mathbf{j}$ and $\mathbf{k}$ range over

$$\mathbf{A}.\mathbf{k} = \mathbf{i} \;\wedge\; 0 \le \mathbf{j} < \mathbf{k} \le \mathbf{N}\ . \tag{6}$$

Because $\mathbf{A}$ is a permutation, only one $\mathbf{k}$ satisfies the first conjunct of (6) for a given $\mathbf{i}$. For example, if $\mathbf{N} = 9$ and $\mathbf{A}$ is the permutation shown above, then $\mathbf{D}$ is

$$0 \quad 1 \quad 2 \quad 2 \quad 0 \quad 3 \quad 5 \quad 2 \quad 1\ .$$

We have $\mathbf{D}.3 = 2$ because $\mathbf{A}.6 = 3$ and two of the numbers $\mathbf{A}.0$, $\mathbf{A}.1$, $\mathbf{A}.2$, $\mathbf{A}.3$, $\mathbf{A}.4$, and $\mathbf{A}.5$ are less than $\mathbf{A}.6$.

By considering $\mathbf{D}$ applied to $\mathbf{A}.\mathbf{i}$ instead of $\mathbf{i}$, we get

$$\begin{array}{ll}
 & \mathbf{D}.(\mathbf{A}.\mathbf{i}) \\
= & \quad \{\ (4) \text{ — def. of } \mathbf{D}\text{, with } \mathbf{i} := \mathbf{A}.\mathbf{i}\ \} \\
 & \langle\, \#\,\mathbf{j},\mathbf{k} \mid \mathbf{A}.\mathbf{k} = \mathbf{A}.\mathbf{i} \;\wedge\; 0 \le \mathbf{j} < \mathbf{k} \le \mathbf{N} \;::\; \mathbf{A}.\mathbf{j} < \mathbf{A}.\mathbf{k} \,\rangle \\
= & \quad \{\ (0)\text{, with } \mathbf{i},\mathbf{j} := \mathbf{k},\mathbf{i}\ \} \\
 & \langle\, \#\,\mathbf{j},\mathbf{k} \mid \mathbf{k} = \mathbf{i} \;\wedge\; 0 \le \mathbf{j} < \mathbf{k} \le \mathbf{N} \;::\; \mathbf{A}.\mathbf{j} < \mathbf{A}.\mathbf{k} \,\rangle \\
= & \quad \{\ \text{one-point rule, and using } \mathbf{i} \le \mathbf{N}\ \} \\
 & \langle\, \#\,\mathbf{j} \mid 0 \le \mathbf{j} < \mathbf{i} \;::\; \mathbf{A}.\mathbf{j} < \mathbf{A}.\mathbf{i} \,\rangle \\
= & \quad \{\ (1) \text{ — def. of } \mathbf{B}\ \} \\
 & \mathbf{B}.\mathbf{i}\ ,
\end{array} \tag{7}$$

and a similar calculation reveals

$$\mathbf{E}.(\mathbf{A}.\mathbf{i}) \;=\; \langle\, \#\,\mathbf{j} \mid 0 \le \mathbf{j} < \mathbf{i} \;::\; \mathbf{A}.\mathbf{j} > \mathbf{A}.\mathbf{i} \,\rangle \;=\; \mathbf{C}.\mathbf{i}\ . \tag{8}$$

From (7) and (8), we conclude that, given $\mathbf{A}$, it is possible to compute $\mathbf{B}$ from $\mathbf{D}$, $\mathbf{C}$ from $\mathbf{E}$, $\mathbf{D}$ from $\mathbf{B}$, and $\mathbf{E}$ from $\mathbf{C}$ in time $\mathrm{O}(\mathbf{N})$.

A relationship between $\mathbf{A}$, $\mathbf{D}$, and $\mathbf{E}$, established by (7), (8), and (3), is

$$\mathbf{D}.(\mathbf{A}.\mathbf{i}) + \mathbf{E}.(\mathbf{A}.\mathbf{i}) \;=\; \mathbf{i}\ , \tag{9}$$

from which we can see that, given $\mathbf{A}$, one can compute $\mathbf{D}$ from $\mathbf{E}$ and $\mathbf{E}$ from $\mathbf{D}$ in time $\mathrm{O}(\mathbf{N})$. Furthermore, (9) lets us infer that, given $\mathbf{D}$ and $\mathbf{E}$, one can compute $\mathbf{A}$ in time $\mathrm{O}(\mathbf{N})$.

In Section 1, we describe a linear-time linear-space algorithm for computing $\mathbf{A}$ from $\mathbf{B}$ and $\mathbf{D}$, and one for computing $\mathbf{A}$ from $\mathbf{C}$ and $\mathbf{E}$.

To summarize, given any one of $\mathbf{B}$ and $\mathbf{C}$, one can compute the other in linear time, and given any *two* of $\mathbf{A}$, $\mathbf{B}$, $\mathbf{D}$, and $\mathbf{E}$, one can compute any other in linear time.

0.1. Algorithms

In [Dij78] and [Gri81], we find an *in situ* program that computes $\mathbf{B}$ from $\mathbf{A}$ in time $\mathrm{O}(\mathbf{N}^2)$. This program has the nice property of being *invertible*. That is, without writing a new program from scratch, there is a way to run this program "backwards". The backward rendering of the program, which then can be written as a program in its own right, is called the *inverse* of the forward rendering. The inverse of the program in [Dij78, Gri81] therefore gives an *in situ* algorithm for computing $\mathbf{A}$ from $\mathbf{B}$ in time $\mathrm{O}(\mathbf{N}^2)$.

We present a faster program for computing $\mathbf{B}$ from $\mathbf{A}$ in Section 2. This program runs in time $\mathrm{O}(\mathbf{N} \log \mathbf{N})$ using linear storage. Although invertible, the inverse does not compute $\mathbf{A}$ from $\mathbf{B}$. The reason is that the forward program computes, in addition to $\mathbf{B}$, two auxiliary arrays. So, the inverse program takes as input $\mathbf{B}$ and the auxiliary arrays. The problem is that the auxiliary arrays cannot be efficiently computed directly from $\mathbf{B}$, so the inverse program is not of much use (see Section 2.4).

In [Knu73], Knuth shows a linear space, time $\mathrm{O}(\mathbf{N} \log \mathbf{N})$ algorithm for computing $\mathbf{E}$ from $\mathbf{A}$ (Ex. 6). We show a short proof of this algorithm in Section 3. Using this algorithm and our observed relations between $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{E}$, it is possible to compute $\mathbf{B}$ from $\mathbf{A}$ in time $\mathrm{O}(\mathbf{N} \log \mathbf{N})$ in the following way:

compute $\mathbf{E}$ from $\mathbf{A}$ in time $\mathrm{O}(\mathbf{N} \log \mathbf{N})$;
compute $\mathbf{C}$ from $\mathbf{A}$ and $\mathbf{E}$ in time $\mathrm{O}(\mathbf{N})$;
compute $\mathbf{B}$ from $\mathbf{C}$ in time $\mathrm{O}(\mathbf{N})$.

This method is asymptotically as good as, but differs from, that presented in Section 2.

For the other direction, *i.e.*, computing $\mathbf{A}$ from $\mathbf{E}$, Knuth gives not only an $\mathrm{O}(\mathbf{N}^2)$ algorithm (Ex. 4) but also a linear space, time $\mathrm{O}(\mathbf{N} \log \mathbf{N})$ algorithm (Ex. 5).

Although Knuth shows one efficient algorithm that computes $\mathbf{E}$ from $\mathbf{A}$ and one that computes $\mathbf{A}$ from $\mathbf{E}$, these algorithms are quite different. Regarding the inverse of his programs, Knuth mentions nothing. It is hard to see how the former of these algorithms can be inverted. The latter, however, we can invert, as will be our concern in Section 4.

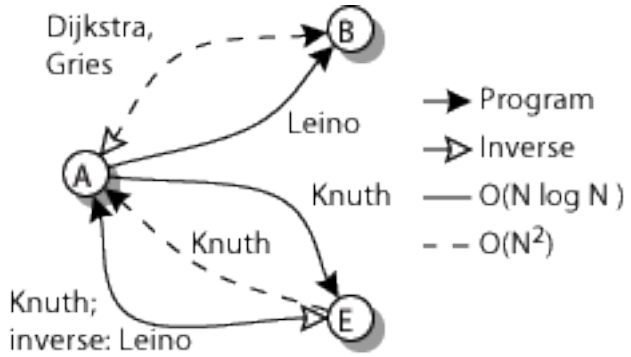

Fig. 0. Algorithms for encoding and decoding permutations

Figure 0 summarizes the individual algorithms to encode and decode permutations and their running times. Each $O(\mathbf{N} \log \mathbf{N})$ algorithm uses linear space.

0.2. Outline of paper

The rest of the paper is organized in the following way. Section 1 demonstrates that $\mathbf{A}$ can be obtained in linear time from $\mathbf{B}$ and $\mathbf{D}$ and from $\mathbf{C}$ and $\mathbf{E}$. Section 2 develops a new, efficient algorithm for computing the code, that is, computing $\mathbf{B}$ from $\mathbf{A}$. Section 3 contains a short proof of an existing efficient algorithm for computing the inversion table of a permutation. An efficient algorithm, due to Knuth, for decoding an inversion table is presented in Section 4. It is then shown that this algorithm can be inverted, yielding a new efficient algorithm for computing the inversion table of a permutation. Finally, Section 5 offers some concluding remarks.

1. Decoding a permutation from two of its encodings

In this section, we show two algorithms: one computes $\mathbf{A}$ from $\mathbf{B}$ and $\mathbf{D}$, the other computes $\mathbf{A}$ from $\mathbf{C}$ and $\mathbf{E}$. Throughout this section, $\mathbf{i}$ is implicitly universally quantified over $0 \leq \mathbf{i} < \mathbf{N}$, and similarly for $\mathbf{k}$.

Recall from (7) that $\mathbf{B.i} = \mathbf{D}.(\mathbf{A.i})$ for any $\mathbf{i}$. Consequently, we have, for any $\mathbf{m}$,

$$\{\, \mathbf{i} \mid \mathbf{B.i} = \mathbf{m} :: \mathbf{i} \,\} \;=\; \{\, \mathbf{i} \mid \mathbf{D}.(\mathbf{A.i}) = \mathbf{m} :: \mathbf{i} \,\} \,. \tag{10}$$

This equation exhibits two set constructors, defined as follows. Let $\mathbf{f.i}$ be any function of $\mathbf{i}$ and $\mathbf{R.i}$ be any boolean function of $\mathbf{i}$. Then, $\{\, \mathbf{i} \mid \mathbf{R.i} :: \mathbf{f.i} \,\}$ is the set of all $\mathbf{f.i}$'s obtained when $\mathbf{i}$ ranges over the values prescribed by $\mathbf{R.i}$. So, (10) states that the set of $\mathbf{i}$'s for which $\mathbf{B.i} = \mathbf{m}$, equals the set of $\mathbf{i}$'s for which $\mathbf{D}.(\mathbf{A.i}) = \mathbf{m}$.

We may rewrite equation (10) as follows.

$$\begin{array}{ll} & \{\, \mathbf{i} \mid \mathbf{B.i} = \mathbf{m} :: \mathbf{i} \,\} \;=\; \{\, \mathbf{i} \mid \mathbf{D}.(\mathbf{A.i}) = \mathbf{m} :: \mathbf{i} \,\} \\ = & \quad \{\;\; \text{apply } \mathbf{A} \text{ to the terms} \;\;\} \end{array} \tag{11}$$

$$
\begin{array}{ll}
 & \{\, \mathbf{i} \mid \mathbf{B.i} = \mathbf{m} \,::\, \mathbf{A.i} \,\} \;=\; \{\, \mathbf{i} \mid \mathbf{D}.(\mathbf{A.i}) = \mathbf{m} \,::\, \mathbf{A.i} \,\} \\
= & \quad \{\; \text{rename dummy } \mathbf{i} := \mathbf{A.i} \text{, since } \mathbf{A} \text{ has an inverse} \;\} \\
 & \{\, \mathbf{i} \mid \mathbf{B.i} = \mathbf{m} \,::\, \mathbf{A.i} \,\} \;=\; \{\, \mathbf{i} \mid \mathbf{D.i} = \mathbf{m} \,::\, \mathbf{i} \,\} \quad . \qquad (12)
\end{array}
$$

As functions of $\mathbf{m}$, we let $\mathbf{S.m}$ denote the set on either side of (11) and $\mathbf{T.m}$ the set on either side of (12). The equality $|\mathbf{S.m}| = |\mathbf{T.m}|$ follows from the hints in the calculation above.

We let $\mathbf{A}.(\mathbf{S.m})$ be the set $\mathbf{S.m}$ in which $\mathbf{A}$ has been applied to every element. With that, we can rewrite (12) as

$$\mathbf{A}.(\mathbf{S.m}) \;=\; \mathbf{T.m} \;. \qquad (13)$$

For $\mathbf{S.m} = \{\mathbf{i}\}$ and $\mathbf{T.m} = \{\mathbf{k}\}$, we have $\{\mathbf{A.i}\} = \{\mathbf{k}\}$, that is, $\mathbf{A.i} = \mathbf{k}$, which gives us a way to reconstruct $\mathbf{A.i}$ given $\mathbf{S.m}$ and $\mathbf{T.m}$. Now, that is only whenever $|\mathbf{S.m}| = 1$, but the idea inspires our algorithm.

We continue our exposition together with an example. Consider the following values for $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{D}$ where $\mathbf{N} = 9$.

		0	1	2	3	4	5	6	7	8
A	=	4	8	0	7	1	5	3	6	2
B	=	0	1	0	2	1	3	2	5	2
D	=	0	1	2	2	0	3	5	2	1

For $\mathbf{m} = 2$, we find

$$
\begin{array}{lll}
\mathbf{S}.2 & = & \{3, 6, 8\} \\
\mathbf{T}.2 & = & \{2, 3, 7\} \;.
\end{array}
$$

Hence, from (13) we know

$$\{\mathbf{A}.3, \mathbf{A}.6, \mathbf{A}.8\} \;=\; \{2, 3, 7\} \;, \qquad (14)$$

but which value goes with which index? For any $\mathbf{i}$ and $\mathbf{k}$, we have,

$$
\begin{array}{ll}
 & \mathbf{true} \\
= & \quad \{\; \text{for } \mathbf{i} < \mathbf{k} \;\wedge\; \mathbf{A.i} < \mathbf{A.k}\text{, } \mathbf{B.k} \text{ counts at least } \mathbf{A.i} \text{ and} \\
 & \qquad \text{everything counted by } \mathbf{B.i} \;\} \\
 & \mathbf{i} < \mathbf{k} \;\wedge\; \mathbf{A.i} < \mathbf{A.k} \;\Rightarrow\; \mathbf{B.i} < \mathbf{B.k} \\
= & \quad \{\; \text{calculus} \;\} \\
 & \mathbf{i} < \mathbf{k} \;\wedge\; \mathbf{B.i} \geq \mathbf{B.k} \;\Rightarrow\; \mathbf{A.i} \geq \mathbf{A.k} \\
= & \quad \{\; \mathbf{i} \neq \mathbf{k} \;\Rightarrow\; \mathbf{A.i} \neq \mathbf{A.k} \;\} \\
 & \mathbf{i} < \mathbf{k} \;\wedge\; \mathbf{B.i} \geq \mathbf{B.k} \;\Rightarrow\; \mathbf{A.i} > \mathbf{A.k} \\
\Rightarrow & \quad \{\; \text{strengthen antecedent} \;\} \\
 & \mathbf{i} < \mathbf{k} \;\wedge\; \mathbf{B.i} = \mathbf{B.k} \;\Rightarrow\; \mathbf{A.i} > \mathbf{A.k} \quad . \qquad (15)
\end{array}
$$

Our solution is now clear: We treat $\mathbf{S.m}$ and $\mathbf{T.m}$ as ordered lists, $\mathbf{S.m}$ in increasing order and $\mathbf{T.m}$ in decreasing order. Indexing these lists by an $\mathbf{i}$ in $0 \leq \mathbf{i} < |\mathbf{S.m}|$, we can formulate a stronger version of (13) as

$$\mathbf{A}.(\mathbf{S.m.i}) \;=\; \mathbf{T.m.i} \;.$$

Applying this to our example (14), we get

$$\mathbf{A}.3 = 7 \qquad \mathbf{A}.6 = 3 \qquad \mathbf{A}.8 = 2 \;.$$

We write our algorithm in Figure 1. For any $\mathbf{m}$ and $\mathbf{n}$ such that $\mathbf{m} \leq \mathbf{n}$, we use the notation $\mathbf{m}, ..\mathbf{n}$ to mean the $\mathbf{n} - \mathbf{m}$ integers beginning at $\mathbf{m}$, that is, $\mathbf{m}, \mathbf{m}+1, \ldots, \mathbf{n}-1$. (This range notation is taken from [Heh93].) We also

```
forall m in 0,..N do S.m, T.m := ε, ε end;
for i := 0 to N − 1 do S.(B.i), T.(D.i) := S.(B.i)++i, i++T.(D.i) end;
forall m in 0,..N do
  forall i in 0,..|S.m| do A.(S.m.i) := T.m.i end
end
```

Fig. 1. An algorithm for computing **A** given **B** and **D**

use $+\!\!+$ as the operator that concatenates two lists. The empty list is denoted ε, and we do not distinguish between an integer and a list containing one integer.

The iterations of the **forall** loops can be done sequentially in any order or can be done in parallel, whereas the **for** loop, as written, should be done sequentially. The algorithm uses linear space and runs in time $O(\mathbf{N})$.

Computing **A** from **C** and **E** can be done in a similar way. The difference from the above is that (15) becomes

$$\mathbf{i} < \mathbf{k} \ \wedge \ \mathbf{C.i} = \mathbf{B.k} \ \Rightarrow \ \mathbf{A.i} < \mathbf{A.k} \ .$$

Hence, an algorithm for computing **A** from **C** and **E** is the one in Figure 1 with the assignment in the **for** loop replaced by

$$\mathbf{S.(C.i)}, \mathbf{T.(E.i)} := \mathbf{i}+\!\!+\mathbf{S.(C.i)}, \mathbf{i}+\!\!+\mathbf{T.(E.i)} \ .$$

2. Computing the code

In this section, we develop an algorithm for computing the code of an array. Our algorithm will be slightly more general: the array may contain any **N** distinct integers instead of a permutation of the first **N** natural numbers. The algorithm will run in $O(\mathbf{N} \log \mathbf{N})$ time using linear space.

2.0. Specification

For any array **a** and integers **m** and **n** (with $\mathbf{m} \leq \mathbf{n}$), we let $\mathbf{a}[\mathbf{m}, ..\mathbf{n}]$ denote the subarray of the $\mathbf{n}-\mathbf{m}$ elements of **a** indexed by consecutive integers starting at **m**. (Note that **a.n** is not an element of that subarray.) We let the application of the brackets after **a** bind stronger than function application. So, for example, $\mathbf{f.a}[\mathbf{m}, ..\mathbf{n}]$ means $\mathbf{f}.(\mathbf{a}[\mathbf{m}, ..\mathbf{n}])$, not $(\mathbf{f.a})[\mathbf{m}, ..\mathbf{n}]$. The elements in the subarray are referenced using the same indices as those in the whole array. We do not make any distinction between arrays and subarrays and refer to either as arrays. To refer to the element indexed by an appropriate **i** in an array **a**, we write **a.i**. We define equality between two arrays $\mathbf{a}[\mathbf{m}, ..\mathbf{m}+\mathbf{l}]$ and $\mathbf{b}[\mathbf{k}, ..\mathbf{k}+\mathbf{l}]$ of the same size as

$$\langle \forall \mathbf{i} \mid 0 \leq \mathbf{i} < \mathbf{l} :: \ \mathbf{a}.(\mathbf{m}+\mathbf{i}) = \mathbf{b}.(\mathbf{k}+\mathbf{i}) \rangle \ .$$

For any array $\mathbf{a}[\mathbf{m}, ..\mathbf{n}]$, we define $\mathbf{code.a}[\mathbf{m}, ..\mathbf{n}]$ as an array $\mathbf{b}[\mathbf{m}, ..\mathbf{n}]$ where, for each **i**,

$$\mathbf{b.i} = \langle \, \#\mathbf{j} \mid \mathbf{m} \leq \mathbf{j} < \mathbf{i} :: \ \mathbf{a.j} < \mathbf{a.i} \, \rangle \ . \tag{16}$$

Finally, we let $\mathbf{distinct.a}[\mathbf{m}, ..\mathbf{n}]$ denote that $\mathbf{a}[\mathbf{m}, ..\mathbf{n}]$ is an array of distinct integers.

We define our specification as a program that computes $\mathbf{b}$ according to

$$\begin{array}{l} \text{Pre}: \ 0 \le \mathbf{N} \ \wedge \ \mathbf{distinct.a}[0, ..\mathbf{N}] \\ \text{Post}: \ \mathbf{b}[0, ..\mathbf{N}] = \mathbf{code.a}[0, ..\mathbf{N}] \end{array} \tag{17}$$

where $\mathbf{a}$ and $\mathbf{b}$ are global arrays. Under initial condition $\mathbf{a} = \mathbf{A}$, our algorithm then establishes $\mathbf{b} = \mathbf{B}$, where $\mathbf{A}$ and $\mathbf{B}$ are those from Section 0.

In order to make our specifications easier to read, we follow the convention that any array element not mentioned in a postcondition or invariant is unchanged by the procedure or loop, respectively. Moreover, array $\mathbf{a}$ and all **in** parameters are to remain unchanged.

In our program notation, we allow local variables to be declared at any time. The scope of such a variable begins at its declaration and ends at the end of the current block. For purposes of counting how much space is being used, we assume that a local variable exists only within its scope. We count space in terms of number of integers and pledge not to pull any bitpacking stunts.

A special kind of local variable is declared using **const** $\mathbf{k} : \mathbf{P}$. This construct declares a local variable $\mathbf{k}$ whose initial value satisfies condition $\mathbf{P}$. After its declaration, $\mathbf{k}$ is read-only.

In the following sections, we develop the program and enhance its efficiency in stages.

2.1. A divide-and-conquer algorithm

The large range of $\mathbf{j}$ in (16) is an obstacle when attempting to write an efficient solution to the problem. We reduce this range by means of a divide-and-conquer algorithm **Code** with parameters $\mathbf{m}$ and $\mathbf{n}$, according to the specification

$$\begin{array}{l} \text{Pre}: \ \mathbf{m} \le \mathbf{n} \ \wedge \ \mathbf{distinct.a}[\mathbf{m}, ..\mathbf{n}] \\ \text{Post}: \ \mathbf{b}[\mathbf{m}, ..\mathbf{n}] = \mathbf{code.a}[\mathbf{m}, ..\mathbf{n}] \ . \end{array} \tag{18}$$

The program statement that solves the entire problem (17) is then $\mathbf{Code}(0, \mathbf{N})$.

With two base cases and one recursive step, we write our program in Figure 2. We annotate our programs with the crucial ingredients of a correctness proof, but, for brevity, we omit the proofs themselves. Termination is always obvious.

The recursion depth of this algorithm is $\mathrm{O}(\log(\mathbf{n} - \mathbf{m}))$. The quantified expression in the loop takes $\mathrm{O}(\mathbf{h} - \mathbf{m})$ steps to compute using a naïve implementation, and the loop iterates $\mathbf{n} - \mathbf{h}$ times. The total time needed for the loop is then $\mathrm{O}((\mathbf{n} - \mathbf{m})^2)$, so $\mathbf{Code}(0, \mathbf{N})$ has time complexity $\mathrm{O}(\mathbf{N}^2)$. The algorithm requires space proportional to the recursion depth.

Our next objective is to eliminate the unfortunate quantified expression. In the following sections, we maintain the overall structure of the above algorithm but attempt to reduce the time required for the recursive step.

2.2. Adding a sorted permutation of a

The previous version of our program used a lot of time computing the quantified expression

$$\langle \, \#\mathbf{j} \mid \mathbf{m} \le \mathbf{j} < \mathbf{h} :: \ \mathbf{a.j} < \mathbf{a.k} \, \rangle \ ,$$

```
procedure Code(in m, in n) =                          {Specification: (18)}
begin
  if n − m = 0  →  skip
  ‖ n − m = 1  →  b.m := 0
  ‖ n − m ≥ 2  →
    const h : h = (m + n) div 2;
    Code(m, h); Code(h, n);
    var k; k := h;
    {Invariant:  ⟨∀i | m ≤ i < k :: b.i = ⟨#j | m ≤ j < i :: a.j < a.i⟩⟩ ∧
                 ⟨∀i | k ≤ i < n :: b.i = ⟨#j | h ≤ j < i :: a.j < a.i⟩⟩ ∧
                 h ≤ k ≤ n   }
    do k ≠ n →  b.k, k := b.k + ⟨#j | m ≤ j < h :: a.j < a.k⟩, k + 1 od
  fi
end .
```

Fig. 2. A divide-and-conquer algorithm for computing the code

since it did so using a simple loop. We can speed this up if we have a sorted permutation of $\mathbf{a}[\mathbf{m}, ..\mathbf{h}]$, denoted by the array

$$\mathbf{sort}.\mathbf{a}[\mathbf{m}, ..\mathbf{h}] \ .$$

Then we can apply a binary search to the sorted array in order to find the value of the quantified expression. For this purpose, we introduce a new global array $\mathbf{s}$ and modify the specification of **Code** accordingly to

$$\begin{array}{l} \text{Pre}: \ \mathbf{m} \le \mathbf{n} \wedge \mathbf{distinct}.\mathbf{a}[\mathbf{m}, ..\mathbf{n}] \\ \text{Post}: \ \mathbf{b}[\mathbf{m}, ..\mathbf{n}] = \mathbf{code}.\mathbf{a}[\mathbf{m}, ..\mathbf{n}] \wedge \mathbf{s}[\mathbf{m}, ..\mathbf{n}] = \mathbf{sort}.\mathbf{a}[\mathbf{m}, ..\mathbf{n}] \ . \end{array} \tag{19}$$

Updating the base cases of our program is trivial. Updating the recursive step can be done using a merge. We let **min** (**max**) applied to any array $\mathbf{a}[\mathbf{m}, ..\mathbf{n}]$ denote the minimum (maximum) element in $\mathbf{a}[\mathbf{m}, ..\mathbf{n}]$, or ∞ ($-\infty$) if $\mathbf{m} = \mathbf{n}$, and we let the binary infix operator **min** (**max**) denote the minimum (maximum) of its two operands. The loop invariant of the merge is thus

$$\begin{array}{l} \mathbf{ss}[\mathbf{m}, ..\mathbf{k}] = (\mathbf{sort}.\mathbf{a}[\mathbf{m}, ..\mathbf{n}])[\mathbf{m}, ..\mathbf{k}] \wedge \\ \mathbf{s}[\mathbf{m}, ..\mathbf{h}] = \mathbf{sort}.\mathbf{a}[\mathbf{m}, ..\mathbf{h}] \wedge \mathbf{s}[\mathbf{h}, ..\mathbf{n}] = \mathbf{sort}.\mathbf{a}[\mathbf{h}, ..\mathbf{n}] \wedge \\ \mathbf{max}.\mathbf{ss}[\mathbf{m}, ..\mathbf{k}] \ = \ \mathbf{max}.\mathbf{s}[\mathbf{m}, ..\mathbf{x}] \ \mathbf{max} \ \mathbf{max}.\mathbf{s}[\mathbf{h}, ..\mathbf{y}] \\ \qquad < \ \mathbf{min}.\mathbf{s}[\mathbf{x}, ..\mathbf{h}] \ \mathbf{min} \ \mathbf{min}.\mathbf{s}[\mathbf{y}, ..\mathbf{n}] \wedge \\ \mathbf{m} \le \mathbf{x} \le \mathbf{h} \le \mathbf{y} \le \mathbf{n} \wedge \mathbf{m} \le \mathbf{k} \le \mathbf{n} \wedge \mathbf{k} - \mathbf{m} = (\mathbf{x} - \mathbf{m}) + (\mathbf{y} - \mathbf{h}) \ . \end{array} \tag{20}$$

(Actually, the last conjunct is not needed, but it may help explain the relationship between the different variables.) We write the next approximation of our program in Figure 3.

Let us first examine the new space requirements. The new global array $\mathbf{s}$ requires linear space, and so does local array $\mathbf{ss}$. At most one array $\mathbf{ss}$ exists at any one time, so we are within linear space, regardless of the recursion depth.

Looking at the time needed, we know the quantified expression can be replaced by a binary search. The binary search is performed once for each of the $\mathbf{n} - \mathbf{h}$ iterations, requiring a total of $O((\mathbf{n} - \mathbf{h}) \log(\mathbf{h} - \mathbf{m}))$ steps. The merge is faster, needing only $O(\mathbf{n} - \mathbf{m})$ steps. The total time required to compute $\mathbf{Code}(0, \mathbf{N})$ is thus $O(\mathbf{N} \log^2 \mathbf{N})$.

```
procedure Code(in m, in n) =                    {Specification: (19)}
begin
  if n − m = 0   →  skip
  [] n − m = 1   →  b.m, s.m := 0, a.m
  [] n − m ≥ 2   →
     const h : h = (m + n) div 2;
     Code(m, h); Code(h, n);
     var k; k := h;
     do k ≠ n →  b.k, k := b.k + ⟨ #j | m ≤ j < h :: a.j < a.k ⟩, k + 1 od;
     var x, y, ss[m, ..n]; k, x, y := m, m, h;
     {Invariant: (20)}
     do k ≠ n →
        if y = n ∨ (x ≠ h ∧ s.x < s.y) →
              {s.x = min.s[x, ..h] min min.s[y, ..n]}
              x, k, ss.k := x + 1, k + 1, s.x
        [] x = h ∨ (y ≠ n ∧ s.y < s.x) →
              {s.y = min.s[x, ..h] min min.s[y, ..n]}
              y, k, ss.k := y + 1, k + 1, s.y
        fi
     od;
     s[m, ..n] := ss[m, ..n]
  fi
end .
```

Fig. 3. Program with merge

We have already achieved an algorithm that is faster than $O(\mathbf{N}^2)$ using linear space. However, it is unfortunate that a binary search is needed to update each value in the upper half of **b**. Hence, we strive for a faster algorithm.

2.3. Adding a mapping from s to a

Arrays **s** and **a** contain permutations of the same elements. Since the elements of **s** are sorted, it is easy to compute the code of **s**. We can make use of this when computing the code of **a** only if we know how the elements of **s** correspond to those in **a**. We invent a mechanism for just that in this next and final approximation of our program.

We introduce a new global array **t** that satisfies, for each of its indices **i**,

$$\mathbf{a}.(\mathbf{t}.\mathbf{i}) = \mathbf{s}.\mathbf{i} \ . \tag{21}$$

As was the case when adding **s**, updating the base cases is trivial. For the recursive step, we need to permute **t** in the same way we permute **s**. This alters our merge as given in Figure 4, after which **Code** satisfies the specification

$$\begin{array}{ll} \text{Pre}: & \mathbf{m} \le \mathbf{n} \ \wedge \ \mathbf{distinct}.\mathbf{a}[\mathbf{m}, ..\mathbf{n}] \\ \text{Post}: & \mathbf{b}[\mathbf{m}, ..\mathbf{n}] = \mathbf{code}.\mathbf{a}[\mathbf{m}, ..\mathbf{n}] \ \wedge \ \mathbf{s}[\mathbf{m}, ..\mathbf{n}] = \mathbf{sort}.\mathbf{a}[\mathbf{m}, ..\mathbf{n}] \ \wedge \\ & \quad \langle \forall \mathbf{i} \mid \mathbf{m} \le \mathbf{i} < \mathbf{n} :: \ \mathbf{m} \le \mathbf{t}.\mathbf{i} < \mathbf{n} \ \wedge \ \mathbf{a}.(\mathbf{t}.\mathbf{i}) = \mathbf{s}.\mathbf{i} \rangle \ . \end{array} \tag{22}$$

We would like to update the upper half of **b** in this loop as well. In particular, we want the merge loop also to implement

$$
\begin{array}{l}
\mathbf{var}\ \mathbf{x}, \mathbf{y}, \mathbf{ss}[\mathbf{m}, ..\mathbf{n}], \mathbf{tt}[\mathbf{m}, ..\mathbf{n}]; \\
\mathbf{k}, \mathbf{x}, \mathbf{y} := \mathbf{m}, \mathbf{m}, \mathbf{h}; \\
\{\text{Invariant: (20)} \ \ \wedge\ \langle\, \forall\, \mathbf{i} \mid \mathbf{m} \le \mathbf{i} < \mathbf{k} \ ::\ \mathbf{m} \le \mathbf{tt}.\mathbf{i} < \mathbf{n} \wedge \mathbf{a}.(\mathbf{tt}.\mathbf{i}) = \mathbf{ss}.\mathbf{i} \,\rangle \wedge \\
\qquad\qquad\qquad\quad \langle\, \forall\, \mathbf{i} \mid \mathbf{m} \le \mathbf{i} < \mathbf{h} \ ::\ \mathbf{m} \le \mathbf{t}.\mathbf{i} < \mathbf{h} \wedge \mathbf{a}.(\mathbf{t}.\mathbf{i}) = \mathbf{s}.\mathbf{i} \,\rangle \wedge \\
\qquad\qquad\qquad\quad \langle\, \forall\, \mathbf{i} \mid \mathbf{h} \le \mathbf{i} < \mathbf{n} \ ::\ \mathbf{h} \le \mathbf{t}.\mathbf{i} < \mathbf{n} \wedge \mathbf{a}.(\mathbf{t}.\mathbf{i}) = \mathbf{s}.\mathbf{i} \,\rangle\ \} \\
\mathbf{do}\ \mathbf{k} \neq \mathbf{n} \rightarrow \\
\quad \mathbf{if}\ \mathbf{y} = \mathbf{n} \vee (\mathbf{x} \neq \mathbf{h} \wedge \mathbf{s}.\mathbf{x} < \mathbf{s}.\mathbf{y}) \rightarrow\ \mathbf{x}, \mathbf{k}, \mathbf{ss}.\mathbf{k}, \mathbf{tt}.\mathbf{k} := \mathbf{x}+1, \mathbf{k}+1, \mathbf{s}.\mathbf{x}, \mathbf{t}.\mathbf{x} \\
\quad [\!]\ \mathbf{x} = \mathbf{h} \vee (\mathbf{y} \neq \mathbf{n} \wedge \mathbf{s}.\mathbf{y} < \mathbf{s}.\mathbf{x}) \rightarrow\ \mathbf{y}, \mathbf{k}, \mathbf{ss}.\mathbf{k}, \mathbf{tt}.\mathbf{k} := \mathbf{y}+1, \mathbf{k}+1, \mathbf{s}.\mathbf{y}, \mathbf{t}.\mathbf{y} \\
\quad \mathbf{fi} \\
\mathbf{od}; \\
\mathbf{s}[\mathbf{m}, ..\mathbf{n}], \mathbf{t}[\mathbf{m}, ..\mathbf{n}] := \mathbf{ss}[\mathbf{m}, ..\mathbf{n}], \mathbf{tt}[\mathbf{m}, ..\mathbf{n}]
\end{array}
$$

Fig. 4. Merge extended to also calculate $\mathbf{t}$

$$
\begin{array}{ll}
\text{Pre}: & \langle\, \forall\, \mathbf{i} \mid \mathbf{m} \le \mathbf{i} < \mathbf{h} \ ::\ \mathbf{b}.\mathbf{i} = \langle\, \#\, \mathbf{j} \mid \mathbf{m} \le \mathbf{j} < \mathbf{i} \ ::\ \mathbf{a}.\mathbf{j} < \mathbf{a}.\mathbf{i} \,\rangle \,\rangle \wedge \\
& \langle\, \forall\, \mathbf{i} \mid \mathbf{h} \le \mathbf{i} < \mathbf{n} \ ::\ \mathbf{b}.\mathbf{i} = \langle\, \#\, \mathbf{j} \mid \mathbf{h} \le \mathbf{j} < \mathbf{i} \ ::\ \mathbf{a}.\mathbf{j} < \mathbf{a}.\mathbf{i} \,\rangle \,\rangle \\
\text{Post}: & \langle\, \forall\, \mathbf{i} \mid \mathbf{m} \le \mathbf{i} < \mathbf{n} \ ::\ \mathbf{b}.\mathbf{i} = \langle\, \#\, \mathbf{j} \mid \mathbf{m} \le \mathbf{j} < \mathbf{i} \ ::\ \mathbf{a}.\mathbf{j} < \mathbf{a}.\mathbf{i} \,\rangle \,\rangle\ ,
\end{array}
$$

which would free us from the need for the first loop. Since $\mathbf{t}[\mathbf{m}, ..\mathbf{h}]$ is a permutation of the integers in the range $\mathbf{m}, ..\mathbf{h}$ and $\mathbf{t}[\mathbf{h}, ..\mathbf{n}]$ is a permutation of the integers in $\mathbf{h}, ..\mathbf{n}$, we can restate the above conditions as

$$
\begin{array}{ll}
\text{Pre}: & \langle\, \forall\, \mathbf{i} \mid \mathbf{m} \le \mathbf{i} < \mathbf{h} \ ::\ \mathbf{b}.(\mathbf{t}.\mathbf{i}) = \langle\, \#\, \mathbf{j} \mid \mathbf{m} \le \mathbf{j} < \mathbf{t}.\mathbf{i} \ ::\ \mathbf{a}.\mathbf{j} < \mathbf{a}.(\mathbf{t}.\mathbf{i}) \,\rangle \,\rangle \\
& \quad \wedge\ \langle\, \forall\, \mathbf{i} \mid \mathbf{h} \le \mathbf{i} < \mathbf{n} \ ::\ \mathbf{b}.(\mathbf{t}.\mathbf{i}) = \langle\, \#\, \mathbf{j} \mid \mathbf{h} \le \mathbf{j} < \mathbf{t}.\mathbf{i} \ ::\ \mathbf{a}.\mathbf{j} < \mathbf{a}.(\mathbf{t}.\mathbf{i}) \,\rangle \,\rangle \\
\text{Post}: & \langle\, \forall\, \mathbf{i} \mid \mathbf{m} \le \mathbf{i} < \mathbf{n} \ ::\ \mathbf{b}.(\mathbf{t}.\mathbf{i}) = \langle\, \#\, \mathbf{j} \mid \mathbf{m} \le \mathbf{j} < \mathbf{t}.\mathbf{i} \ ::\ \mathbf{a}.\mathbf{j} < \mathbf{a}.(\mathbf{t}.\mathbf{i}) \,\rangle \,\rangle.
\end{array}
$$

For the loop invariant, we can replace a constant, *viz.*, $\mathbf{h}$, by a variable in the range from $\mathbf{h}$ to $\mathbf{n}$. We have a variable just like that, *viz.*, $\mathbf{y}$. Thus, we add to the merge loop invariant

$$
\begin{array}{l}
\langle\, \forall\, \mathbf{i} \mid \mathbf{m} \le \mathbf{i} < \mathbf{y} \ ::\ \mathbf{b}.(\mathbf{t}.\mathbf{i}) = \langle\, \#\, \mathbf{j} \mid \mathbf{m} \le \mathbf{j} < \mathbf{t}.\mathbf{i} \ ::\ \mathbf{a}.\mathbf{j} < \mathbf{a}.(\mathbf{t}.\mathbf{i}) \,\rangle \,\rangle \wedge \\
\langle\, \forall\, \mathbf{i} \mid \mathbf{y} \le \mathbf{i} < \mathbf{n} \ ::\ \mathbf{b}.(\mathbf{t}.\mathbf{i}) = \langle\, \#\, \mathbf{j} \mid \mathbf{h} \le \mathbf{j} < \mathbf{t}.\mathbf{i} \ ::\ \mathbf{a}.\mathbf{j} < \mathbf{a}.(\mathbf{t}.\mathbf{i}) \,\rangle \,\rangle\ .
\end{array}
$$

One way to satisfy this invariant is to use a statement like

$$
\mathbf{b}.(\mathbf{t}.\mathbf{y}) := \mathbf{b}.(\mathbf{t}.\mathbf{y}) + \langle\, \#\, \mathbf{j} \mid \mathbf{m} \le \mathbf{j} < \mathbf{h} \ ::\ \mathbf{a}.\mathbf{j} < \mathbf{a}.(\mathbf{t}.\mathbf{y}) \,\rangle
$$

in the second alternative of the conditional inside the loop. We see the unwanted quantified expression crop up once more, but this time we are armed to eliminate it. We know that the value of the quantified expression is at most $\mathbf{h} - \mathbf{m}$. The simplest expression we can think of that has a good chance of "being right" is $\mathbf{x} - \mathbf{m}$. We calculate, under the new invariant and the guards leading to execution of the second alternative of the conditional (see Figure 3),

$$
\begin{array}{ll}
& \mathbf{x} - \mathbf{m} = \langle\, \#\, \mathbf{j} \mid \mathbf{m} \le \mathbf{j} < \mathbf{h} \ ::\ \mathbf{a}.\mathbf{j} < \mathbf{a}.(\mathbf{t}.\mathbf{y}) \,\rangle \\
= & \quad \{\ (21) \text{ — property of } \mathbf{t}\ \} \\
& \mathbf{x} - \mathbf{m} = \langle\, \#\, \mathbf{j} \mid \mathbf{m} \le \mathbf{j} < \mathbf{h} \ ::\ \mathbf{a}.\mathbf{j} < \mathbf{s}.\mathbf{y} \,\rangle \\
= & \quad \{\ \mathbf{s}[\mathbf{m}, ..\mathbf{h}] = \mathbf{sort}.\mathbf{a}[\mathbf{m}, ..\mathbf{h}]\text{, so } \mathbf{s}[\mathbf{m}, ..\mathbf{h}] \text{ is a permutation of } \mathbf{a}[\mathbf{m}, ..\mathbf{h}]\ \} \\
& \mathbf{x} - \mathbf{m} = \langle\, \#\, \mathbf{j} \mid \mathbf{m} \le \mathbf{j} < \mathbf{h} \ ::\ \mathbf{s}.\mathbf{j} < \mathbf{s}.\mathbf{y} \,\rangle \\
= & \quad \{\ \text{split range, since } \mathbf{m} \le \mathbf{x} \le \mathbf{h}\ \} \\
& \mathbf{x} - \mathbf{m} = \langle\, \#\, \mathbf{j} \mid \mathbf{m} \le \mathbf{j} < \mathbf{x} \ ::\ \mathbf{s}.\mathbf{j} < \mathbf{s}.\mathbf{y} \,\rangle + \langle\, \#\, \mathbf{j} \mid \mathbf{x} \le \mathbf{j} < \mathbf{h} \ ::\ \mathbf{s}.\mathbf{j} < \mathbf{s}.\mathbf{y} \,\rangle \\
\Leftarrow & \quad \{\ \text{all/none of the terms being } \mathbf{true}\ \}
\end{array}
$$

```
procedure Code(in m, in n) =                    {Specification: (22)}
begin
  if n − m = 0 →  skip
  [] n − m = 1 →  b.m, s.m, t.m := 0, a.m, m
  [] n − m ≥ 2 →
    const h : h = (m + n) div 2;
    Code(m, h);  Code(h, n);
    var k, x, y, ss[m, ..n], tt[m, ..n];
    k, x, y := m, m, h;
    do k ≠ n →
      if y = n ∨ (x ≠ h ∧ s.x < s.y) →
          x, k, ss.k, tt.k := x + 1, k + 1, s.x, t.x
       [] x = h ∨ (y ≠ n ∧ s.y < s.x) →
          y, k, ss.k, tt.k, b.(t.y) := y + 1, k + 1, s.y, t.y, b.(t.y) + x − m
      fi
    od;
    s[m, ..n], t[m, ..n] := ss[m, ..n], tt[m, ..n]
  fi
end .
```

Fig. 5. The final program

```
    x − m = ⟨ # j | m ≤ j < x :: s.j < s.y ⟩ ∧
    0 = ⟨ # j | x ≤ j < h :: s.j < s.y ⟩
=       { # and ∀ }
    ⟨ ∀ j | m ≤ j < x :: s.j < s.y ⟩ ∧ ⟨ ∀ j | x ≤ j < h :: s.y ≤ s.j ⟩
=       { max, min }
    max.s[m, ..x] < s.y ∧ s.y ≤ min.s[x, ..h]
=       { s.y = min.s[x, ..h] min min.s[y, ..n] , twice }
    max.s[m, ..x] < min.s[x, ..h] min min.s[y, ..n]
⇐       { max }
    max.s[m, ..x] max max.s[h, ..y] < min.s[x, ..h] min min.s[y, ..n]
=       { Invariant }
    true .
```

We write our final program in Figure 5. This program uses only $O(\mathbf{N} \log \mathbf{N})$ time with $O(\mathbf{N})$ space!

As a little optimization, if **a** is not needed at the end of this computation, we can eliminate array **b** by storing **b**[**m**, ..**n**] in **a**[**m**, ..**n**] . This is because the only reference to an element of **a** occurs in one of the base cases.

2.4. Inverting the program

In this section, we touch on the rules for program inversion (see [Dij78, Gri81, CU90]) and show how the program of Figure 5 can be inverted. We do so by explaining how to invert an assignment statement, sequential composition, **if** statements, and **do** statements. By inverting our program, we hope to arrive at a program that computes **A** from **B** .

We begin with the assignment statement. Consider, for example, the state-

ment that increments a variable $\mathbf{n}$ by 1, $\mathbf{n} := \mathbf{n} + 1$. The inverse of it is the statement that decreases the same variable by 1, $\mathbf{n} := \mathbf{n} - 1$.

So what about a statement like $\mathbf{n} := 3$? We cannot invert this statement by itself, because there is no information about the previous value of $\mathbf{n}$. However, the program

$$\{\mathbf{n} = 5\}\ \mathbf{n} := 3\ ,$$

that is, the assignment $\mathbf{n} := 3$ known to start in a state where $\mathbf{n} = 5$, can be inverted. Its inverse is

$$\{\mathbf{n} = 3\}\ \mathbf{n} := 5\ .$$

The point of inverting a program is often to reverse the rôles of input and output. If variable **out** is part of the output of a program, we can view a statement $\mathbf{out} := \mathbf{x}$ as $\mathbf{write}(\mathbf{out}, \mathbf{x})$, that is, writing $\mathbf{x}$ to (the end of) stream **out**. Thus, the inverse of such a statement is $\mathbf{read}(\mathbf{out}, \mathbf{x})$, that is, reading $\mathbf{x}$ from (the end of) stream **out**. Rather than the **read** statement, we may use $\mathbf{x} := \mathbf{out}$. Similarly, if **in** is a variable that is part of the input of a program, we invert $\mathbf{x} := \mathbf{in}$ as $\mathbf{in} := \mathbf{x}$.

So much for assignment. Sequential composition is easy: For statements $\mathbf{S}$ and $\mathbf{T}$,

$$(\mathbf{S}; \mathbf{T})^{-1} \ = \ \mathbf{T}^{-1}; \mathbf{S}^{-1}\ .$$

Executing an **if** statement in reverse requires that it be known which alternative to take. So, to invert an **if** statement like

$$\mathbf{if}\ \mathbf{B}0 \to \mathbf{S}0\ [\!]\ \mathbf{B}1 \to \mathbf{S}1\ \mathbf{fi}\ ,$$

we need to find two mutually exclusive conditions $\mathbf{C}0$ and $\mathbf{C}1$ such that one holds after execution of $\mathbf{S}0$ and the other after execution of $\mathbf{S}1$. We restrict our attention to inverting only deterministic programs, so we assume $\mathbf{B}0$ and $\mathbf{B}1$ are mutually exclusive, too. With that, we state the rule for inverting an **if** statement as

$$\begin{array}{l} \quad (\mathbf{if}\ \mathbf{B}0 \to \mathbf{S}0\{\mathbf{C}0\}\ [\!]\ \mathbf{B}1 \to \mathbf{S}1\{\mathbf{C}1\}\ \mathbf{fi})^{-1} \\ = \\ \quad \mathbf{if}\ \mathbf{C}0 \to \mathbf{S}0^{-1}\{\mathbf{B}0\}\ [\!]\ \mathbf{C}1 \to \mathbf{S}1^{-1}\{\mathbf{B}1\}\ \mathbf{fi}\ . \end{array}$$

Similar to the **if** statement is the **do** loop. Here, too, there is a choice of whether to iterate the loop backwards once more. For this, we find a condition $\mathbf{C}$ that holds after every execution of the loop body and that does not hold prior to executing the loop. The rule for inverting a loop can then be stated as

$$\begin{array}{l} \quad (\{\neg\mathbf{C}\}\ \mathbf{do}\ \mathbf{B} \to \mathbf{S}\{\mathbf{C}\}\ \mathbf{od}\ \{\neg\mathbf{B}\})^{-1} \\ = \\ \quad \{\neg\mathbf{B}\}\ \mathbf{do}\ \mathbf{C} \to \mathbf{S}^{-1}\{\mathbf{B}\}\ \mathbf{od}\ \{\neg\mathbf{C}\}\ . \end{array}$$

Now that we have presented what program inversion is all about, we return to the task of inverting our final program from Figure 5. The following conditions show that the program is indeed invertible.

Outer **if** alternatives:	$\mathbf{n} - \mathbf{m} = 0$	$\mathbf{n} - \mathbf{m} = 1$	$\mathbf{n} - \mathbf{m} \geq 2$
Before and after **do** loop:	$\mathbf{m} = \mathbf{k}$	$\mathbf{m} \neq \mathbf{k}$	
Inner **if** alternatives:	$\mathbf{m} \leq \mathbf{tt}.(\mathbf{k} - 1)$	$\mathbf{tt}.(\mathbf{k} - 1) < \mathbf{m}$	

```
{a = A}
forall k in 0,..N do e.k := 0 end;
for k := ⌊log N⌋ downto 0 do
  forall s in 0..⌊N/2^(k+1)⌋ do x.s := 0 end;
  var j; j := 0;
  do j ≠ N →
    const r, s : r = ⌊a.j/2^k⌋ mod 2 ∧ s = ⌊a.j/2^(k+1)⌋ ;
    if r = 0 → e.(a.j) := e.(a.j) + x.s [] r = 1 → x.s := x.s + 1 fi;
    j := j + 1
  od
end
{e = E}
```

Fig. 6. Knuth's algorithm for computing the inversion table

The inverse computes array **a** given its code **b** and the arrays **s** and **t**. It is reasonable for **s** to be part of the input to a program that computes an array from its code because the code contains only ordering information, not the array numbers themselves. The programs in [Dij78, Gri81] require that the given array consist of a permutation of the first **N** natural numbers; hence, **s** is implied and can be omitted. Our program, on the other hand, allows **a** to be any array of distinct integers.

Array **t**, however, is generally not available when wanting to compute an array from its code. In fact, by letting **t** into the picture, **a** can be computed directly from **s** and **t** without using **b** at all, as seen from (21).

So, although it exists, the inverse of **Code** is a program that is probably not useful in practice.

3. A proof of Knuth's encoding algorithm

In this section, we present a proof of an algorithm that computes **E** from **A**. The algorithm, shown in Figure 6, is from Ex. 6 of [Knu73].

For any **j** in the range $0,..\mathbf{N}$, we have $\mathbf{a.j} < \mathbf{N}$, so constant **s** satisfies $\mathbf{s} \leq \lfloor \mathbf{N}/2^{\mathbf{k}+1} \rfloor$. Hence, every **x.s** used in the inner loop has been set to 0 in the **forall s** loop, in the same **for k** iteration. We then see that the only information kept between iterations of the outer loop is **e**. Since the only operation on **e** within the inner loop is an increment, the order in which the iterations of the outer loop are performed does not affect the program output (despite the fact that Knuth specifies the iterations to be done in decreasing order of **k**).

In order to understand the program, we write an invariant for the inner loop. For convenience, we let **bit.k.y** denote $\lfloor \mathbf{y}/2^{\mathbf{k}} \rfloor$ **mod** 2, for natural **k** and **y**. We then write the invariant as

$$\begin{array}{l} 0 \leq \mathbf{j} \leq \mathbf{N} \ \wedge \\ \langle\, \forall\, \mathbf{s} \mid 0 \leq \mathbf{s} \leq \lfloor \mathbf{n}/2^{\mathbf{k}+1} \rfloor \ ::\ \mathbf{x.s} \ = \ \langle\, \#\, \mathbf{t} \mid 0 \leq \mathbf{t} < \mathbf{j} \ ::\ \lfloor \mathbf{a.t}/2^{\mathbf{k}+1} \rfloor = \mathbf{s} \wedge \mathbf{bit.k.(a.t)} = 1 \,\rangle\rangle \ . \end{array} \tag{23}$$

We leave it to the reader to verify this invariant.

So, at the time $\mathbf{e.(a.j)}$ is incremented by **x.s**,

$$(23) \ \wedge \ \mathbf{r} = \mathbf{bit.k.(a.j)} = 0 \ \wedge \ \mathbf{s} = \lfloor \mathbf{a.j}/2^{\mathbf{k}+1} \rfloor$$

holds. Therefore, the program computes each $\mathbf{e.(a.j)}$ to be

$$\langle \Sigma\, \mathbf{k} \mid 0 \le \mathbf{k} \le \lfloor \log \mathbf{N} \rfloor \ \wedge \ \mathbf{bit.k.(a.j)} = 0 \ :: \\ \langle \#\, \mathbf{t} \mid 0 \le \mathbf{t} < \mathbf{j} \ :: \ \lfloor \mathbf{a.t}/2^{\mathbf{k}+1} \rfloor = \lfloor \mathbf{a.j}/2^{\mathbf{k}+1} \rfloor \ \wedge \ \mathbf{bit.k.(a.t)} = 1 \rangle \rangle \ .$$

We calculate,

$$\begin{array}{ll} & \mathbf{e.(a.j)} \\ = & \quad \{ \ \text{observations made above} \ \} \\ & \langle \Sigma\, \mathbf{k} \mid 0 \le \mathbf{k} \le \lfloor \log \mathbf{N} \rfloor \ \wedge \ \mathbf{bit.k.(a.j)} = 0 \ :: \\ & \quad \langle \#\, \mathbf{t} \mid 0 \le \mathbf{t} < \mathbf{j} \ :: \ \lfloor \mathbf{a.t}/2^{\mathbf{k}+1} \rfloor = \lfloor \mathbf{a.j}/2^{\mathbf{k}+1} \rfloor \ \wedge \ \mathbf{bit.k.(a.t)} = 1 \rangle \rangle \\ = & \quad \{ \ \text{term is 0 for } \mathbf{k} > \lfloor \log \mathbf{N} \rfloor \text{, because } \mathbf{a.t} < \mathbf{N} \text{ and} \\ & \quad \ \ \text{therefore } \mathbf{bit.k.(a.t)} = 0 \ \} \\ & \langle \Sigma\, \mathbf{k} \mid \mathbf{bit.k.(a.j)} = 0 \ :: \\ & \quad \langle \#\, \mathbf{t} \mid 0 \le \mathbf{t} < \mathbf{j} \ :: \ \lfloor \mathbf{a.t}/2^{\mathbf{k}+1} \rfloor = \lfloor \mathbf{a.j}/2^{\mathbf{k}+1} \rfloor \ \wedge \ \mathbf{bit.k.(a.t)} = 1 \rangle \rangle \\ = & \quad \{ \ \text{nesting} \ \} \\ & \langle \Sigma\, \mathbf{t} \mid 0 \le \mathbf{t} < \mathbf{j} \ :: \\ & \quad \langle \#\, \mathbf{k} \mid \mathbf{bit.k.(a.j)} = 0 \ :: \ \lfloor \mathbf{a.t}/2^{\mathbf{k}+1} \rfloor = \lfloor \mathbf{a.j}/2^{\mathbf{k}+1} \rfloor \ \wedge \ \mathbf{bit.k.(a.t)} = 1 \rangle \rangle \\ = & \quad \{ \ \text{for each } \mathbf{t} \text{ and } \mathbf{j} \text{, at most one } \mathbf{k} \text{ satisfies the range and term} \ \} \\ & \langle \#\, \mathbf{t} \mid 0 \le \mathbf{t} < \mathbf{j} \ :: \\ & \quad \langle \exists\, \mathbf{k} \mid \mathbf{bit.k.(a.j)} = 0 \ :: \ \lfloor \mathbf{a.t}/2^{\mathbf{k}+1} \rfloor = \lfloor \mathbf{a.j}/2^{\mathbf{k}+1} \rfloor \ \wedge \ \mathbf{bit.k.(a.t)} = 1 \rangle \rangle \\ = & \quad \{ \ \text{lexicographical bit-order} \ \} \\ & \langle \#\, \mathbf{t} \mid 0 \le \mathbf{t} < \mathbf{j} \ :: \ \mathbf{a.t} > \mathbf{a.j} \rangle \quad , \end{array}$$

which shows the correctness of the program.

4. Inverting an algorithm that decodes an inversion table

In this section, we present Knuth's fast algorithm for decoding an inversion table, or, to use the notation introduced in Section 0, for computing $\mathbf{A}$ from $\mathbf{E}$. Using the rules for program inversion from Section 2.4, we then invert Knuth's algorithm to arrive at a new efficient algorithm for computing $\mathbf{E}$ from $\mathbf{A}$.

We consider strings of pairs of natural numbers and use the following notational conventions.

ε	the empty string
juxtaposition	string catenation
$[\mathbf{u}, \mathbf{U}]$	a pair consisting of the natural numbers $\mathbf{u}$ and $\mathbf{U}$
Greek letters, $\mathbf{x}, \mathbf{y}, \mathbf{z}$	strings

So, for example, $[\mathbf{u}, \mathbf{U}][\mathbf{v}, \mathbf{V}][\mathbf{w}, \mathbf{W}]$ is a string of length 3, and $[\mathbf{u}, \mathbf{U}]\alpha$ is a string that begins with the pair $[\mathbf{u}, \mathbf{U}]$. We define binary operator $\circ$, binding weaker than catenation, on strings as follows.

$$\varepsilon \circ \alpha \ = \ \alpha \ = \ \alpha \circ \varepsilon$$

$$[\mathbf{u}, \mathbf{U}]\alpha \circ [\mathbf{v}, \mathbf{V}]\beta \ = \ \begin{cases} [\mathbf{u}, \mathbf{U}](\alpha \circ [\mathbf{v} - \mathbf{u}, \mathbf{V}]\beta) & \text{if } \mathbf{u} \le \mathbf{v} \\ [\mathbf{v}, \mathbf{V}]([\mathbf{u} - \mathbf{v} - 1, \mathbf{U}]\alpha \circ \beta) & \text{if } \mathbf{v} < \mathbf{u} \end{cases}$$

This definition is reminiscent of a merge, in that the recursive step selects either the first pair from $[\mathbf{u}, \mathbf{U}]\alpha$ or the first pair from $[\mathbf{v}, \mathbf{V}]\beta$, after which $\circ$ is

```
procedure circ(in α, in β, out σ) =
{Post: σ = α ∘ β}
begin
  var z, x, y;
  z, x, y := ε, α, β;
  {Invariant: z(x ∘ y) = α ∘ β}
  do x ≠ ε ∧ y ≠ ε →
    const u, U, γ : x = [u, U]γ;
    const v, V, δ : y = [v, V]δ;
    if u ≤ v → z, x, y := z[u, U], γ, [v − u, V]δ
    [] v < u → z, x, y := z[v, V], [u − v − 1, U]γ, δ
    fi
  od;   {Invariant ∧ (x = ε ∨ y = ε)}
  if x = ε → z, y := z y, ε
  [] y = ε → z, x := z x, ε
  fi;   {x = y = ε ∧ z = α ∘ β}
  σ := z
end .
```

Fig. 7. Program that computes $\alpha \circ \beta$

applied to a smaller problem. The time needed to compute $\alpha \circ \beta$ is thus linear in the length of $\alpha\beta$. We display this algorithm, called **circ**, in Figure 7.

Knuth leaves to the reader, and so do we, to verify that $\circ$ is associative and that

$$\begin{array}{l}[\mathbf{E}.0, 0] \circ [\mathbf{E}.1, 1] \circ \cdots \circ [\mathbf{E}.(\mathbf{N}-1), \mathbf{N}-1] \;= \\ \qquad [0, \mathbf{A}.0][0, \mathbf{A}.1] \cdots [0, \mathbf{A}.(\mathbf{N}-1)] \;,\end{array} \tag{24}$$

where $\mathbf{A}$ and $\mathbf{E}$ are those from Section 0. Because $\circ$ is associative, we can compute the $\circ$ operators of the left-hand side of (24) in any order. In particular, we can compute (RHS-24) from (LHS-24) by breaking (LHS-24) up into two halves, computing each of the halves, and then computing the middle $\circ$ operator. This is the description of a divide-and-conquer algorithm, which then runs in time $\mathrm{O}(\mathbf{N} \log \mathbf{N})$. We show the algorithm, named **Decode**, in Figure 8. Its specification is

$$\begin{array}{l}\text{Pre}: \;\; \mathbf{m} \leq \mathbf{n} \\ \text{Post}: \; \mathbf{aa} = \mathbf{ee}[\mathbf{m}] \circ \mathbf{ee}[\mathbf{m}+1] \circ \cdots \circ \mathbf{ee}[\mathbf{n}-1] \;.\end{array} \tag{25}$$

When writing specifications, we use the conventions introduced in Section 2.0. The statement that solves the entire problem is then $\mathbf{Decode}(0, \mathbf{N}, \mathbf{E}', \mathbf{A}')$, where $\mathbf{E}'$ and $\mathbf{A}'$ are $\mathbf{E}$ and $\mathbf{A}$ modified to fit the left- and right-hand sides of (24).

Constructing the inverse of **Decode**, call it **Encode**, is easy. Since $\mathbf{m}$ and $\mathbf{n}$ are not modified, the guards of **Encode** are the same as those of **Decode**. Instead of two recursive calls to **Decode** followed by a call to **circ**, the inverse program has a call to the inverse of **circ** followed by two recursive calls to **Encode**.

So what about the inverse of **circ**? This program, call it **cric**, needs to decompose a given string σ into two strings α and β such that $\sigma = \alpha \circ \beta$. The problem is, however, that such α and β are not, in general, uniquely

```
procedure Decode(in m, in n,
                 in ee : array m, ..n of string,
                 out aa : string) =
{Specification: (25)}
begin
  if n − m = 0 → aa := ε
  [] n − m = 1 → aa := ee[m]
  [] n − m ≥ 2 →
    const p : p = (m + n) div 2;
    var a0, a1 : string;
    Decode(m, p, ee[m, ..p], a0);
    Decode(p, n, ee[p, ..n], a1);
    circ(a0, a1, aa)
  fi
end .
```

Fig. 8. Program that decodes an inversion table

determined. Indeed, inspecting the first **if** statement of **circ**, we find that there is no information available after the **if** statement that allows us to conclude which alternative was executed.

To find the solution, we return to the definition of $\circ$ and (24). From the former, we see that the left-components of the pairs are modified in the recursive step, but the right-components are not. From the latter, we see that regardless of the order in which the $\circ$ operations are carried out when computing (LHS-24), the relative positioning of the operands of $\circ$ is the same. Moreover, the right-components of (LHS-24) are arranged in ascending order.

The solution now stares us in the face. Our observations show that every right-component of $\mathbf{ee}[\mathbf{m}, ..\mathbf{p}]$ is less than every right-component of $\mathbf{ee}[\mathbf{p}, ..\mathbf{n}]$ in **Decode**. In fact, every right-component of $\mathbf{ee}[\mathbf{m}, ..\mathbf{p}]$ is less than $\mathbf{p}$, and every right-component of $\mathbf{ee}[\mathbf{p}, ..\mathbf{n}]$ is at least $\mathbf{p}$. By passing this $\mathbf{p}$ to **circ**, there is enough information in **circ** for it to be inverted.

Formally, let $\mathbf{R}.\alpha$ be the string of right-components of α, that is, α projected onto its right-components. We define **ascend**, for any string α, by

$$\mathbf{ascend}.\alpha \;=\; \mathbf{R}.\alpha \text{ is strictly ascending .}$$

We also define $\alpha \sqsubset_{\mathbf{p}} \beta$ by

$$\langle\, \forall\, \mathbf{U} \mid \mathbf{U} \in \mathbf{R}.\alpha \;::\; \mathbf{U} < \mathbf{p} \,\rangle \;\wedge\; \langle\, \forall\, \mathbf{V} \mid \mathbf{V} \in \mathbf{R}.\beta \;::\; \mathbf{p} \leq \mathbf{V} \,\rangle \;.$$

We now add the precondition

$$\mathbf{ascend}.(\mathbf{ee}[\mathbf{m}]\,\mathbf{ee}[\mathbf{m}+1] \cdots \mathbf{ee}[\mathbf{n}-1]) \;\wedge\; \langle\, \forall\, \mathbf{i} \mid \mathbf{m} \leq \mathbf{i} < \mathbf{n} \;::\; \mathbf{first}.(\mathbf{R}.\mathbf{ee}[\mathbf{i}]) = \mathbf{i} \,\rangle$$

to **Decode**. We leave it to the reader to verify that this condition holds for (LHS-24) and for each recursive call. We add to **circ** new parameter **in p**, change **circ**'s precondition to $\alpha \sqsubset_{\mathbf{p}} \beta$, and change the call of **circ** from within **Decode** to

$$\mathbf{circ}(\mathbf{a}0, \mathbf{a}1, \mathbf{aa}, \mathbf{p}) \;.$$

Next, we extend the loop invariant in **circ** with

$\mathbf{x} \sqsubset_{\mathbf{p}} \mathbf{y} \;\wedge\; (\mathbf{x} \text{ is suffix of } \alpha) \;\wedge\; (\mathbf{y} \text{ is suffix of } \beta)$
$(\mathbf{z} \neq \varepsilon \;\wedge\; \mathbf{x} = \varepsilon \;\Rightarrow\; \mathbf{last}.(\mathbf{R}.\mathbf{z}) < \mathbf{p}) \;\wedge$
$(\mathbf{z} \neq \varepsilon \;\wedge\; \mathbf{y} = \varepsilon \;\Rightarrow\; \mathbf{p} \leq \mathbf{last}.(\mathbf{R}.\mathbf{z}))$.

If $\alpha = \beta = \varepsilon$ holds initially in **circ**, then the loop will not iterate, so both guards of the second **if** statement will be true, but we want to invert only deterministic statements. Since $\mathbf{m} < \mathbf{p} < \mathbf{n}$ holds in **Decode**, both $\mathbf{a}0$ and $\mathbf{a}1$ will have positive length in the call to **circ**. Realizing the opportunity, we add

$$\alpha \neq \varepsilon \;\wedge\; \beta \neq \varepsilon$$

as another precondition of **circ**. Using this, the invariant, and the negation of the guards, we conclude that

$$\mathbf{z} \neq \varepsilon \;\wedge\; (\mathbf{x} = \varepsilon \;\not\equiv\; \mathbf{y} = \varepsilon)$$

is a precondition of the second **if** statement.

Now, the following annotations reveal that the **if** and **do** statements of **circ** can be inverted.

Before and after **do** loop:	$\mathbf{z} = \varepsilon$	$\mathbf{z} \neq \varepsilon$
Alternatives of first **if** statement:	$\mathbf{z} \neq \varepsilon \wedge \mathbf{y} \neq \varepsilon \wedge \mathbf{last}.(\mathbf{R}.\mathbf{z}) < \mathbf{p}$	
	$\mathbf{z} \neq \varepsilon \wedge \mathbf{x} \neq \varepsilon \wedge \mathbf{p} \leq \mathbf{last}.(\mathbf{R}.\mathbf{z})$	
Alternatives of second **if** statement:	$\mathbf{p} \leq \mathbf{last}.(\mathbf{R}.\mathbf{z})$	$\mathbf{last}.(\mathbf{R}.\mathbf{z}) < \mathbf{p}$

Only one task remains: inverting the assignment statements. These are all straightforward, except the ones in the second **if** statement. We show one alternative; the other is symmetric. Our task is to invert

$\{\mathbf{z} \neq \varepsilon \;\wedge\; \mathbf{x} \neq \varepsilon \;\wedge\; \mathbf{y} = \varepsilon \;\wedge\; \mathbf{p} \leq \mathbf{last}.(\mathbf{R}.\mathbf{z})\}$
$\mathbf{z}, \mathbf{x} := \mathbf{z}\,\mathbf{x}, \varepsilon$
$\{\mathbf{last}.(\mathbf{R}.\mathbf{z}) < \mathbf{p}\}$.

(The last conjunct of the precondition comes from the loop invariant.) The inverse program then moves pairs from the tail of $\mathbf{z}$, putting them on $\mathbf{x}$, until $\mathbf{p} \leq \mathbf{last}.(\mathbf{R}.\mathbf{z})$ holds. This can be done using a loop.

We are now done and show the inverse programs **Encode** and **cric** in Figures 9 and 10. With that, we have shown a new linear space, time $\mathrm{O}(\mathbf{N} \log \mathbf{N})$ algorithm for computing $\mathbf{E}$ from $\mathbf{A}$. This algorithm, however, distinguishes itself from the others mentioned as being efficient *and* having been obtained as the inverse of another algorithm.

5. Conclusion

We showed four encodings, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$, and $\mathbf{E}$, of a permutation $\mathbf{A}$ and how these encodings are related. In particular, we showed that one can compute between $\mathbf{B}$ and $\mathbf{C}$ in linear time, and, given two of $\mathbf{A}$, $\mathbf{B}$, $\mathbf{D}$, and $\mathbf{E}$, one can compute any other in linear time.

We gave an overview of previously known algorithms for computing to and from two of these encodings, the *code* ($\mathbf{B}$) and the *inversion table* ($\mathbf{E}$). Some of these algorithms have time complexity $\mathrm{O}(\mathbf{N}^2)$, whereas the efficient ones run in time $\mathrm{O}(\mathbf{N} \log \mathbf{N})$. We presented a proof for one of the existing efficient algorithms.

```
procedure Encode(in m, in n,
                 out ee : array m, ..n of string,
                 in aa : string) =
begin
  if n − m = 0 → skip
  [] n − m = 1 → ee[m] := aa
  [] n − m ≥ 2 →
    const p : p = (m + n) div 2;
    var a0, a1 : string;
    cric(a0, a1, aa, p);
    Encode(p, n, ee[p, ..n], a1);
    Encode(m, p, ee[m, ..p], a0)
  fi
end .
```

Fig. 9. Program for computing an inversion table

```
procedure cric(out α, out β, in σ, in p) =
begin
  var z, x, y;
  z, x, y := σ, ε, ε;
  if p ≤ last.(R.z) →
    do p ≤ last.(R.z) → const w, W, θ : z = θ[w, W]; z, y := θ, [w, W]y od
  [] last.(R.z) < p →
    do last.(R.z) < p → const w, W, θ : z = θ[w, W]; z, x := θ, [w, W]x od
  fi;
  do z ≠ ε →
    const w, W, θ : z = θ[w, W];
    if W < p → const v, V, δ : y = [v, V]δ;
                z, x, y := θ, [w, W]x, [v + w, V]δ
    [] p ≤ W → const u, U, γ : x = [u, U]γ;
                z, x, y := θ, [u + w + 1, U]γ, [w, W]y
    fi
  od;
  α, β := x, y
end .
```

Fig. 10. Program that decomposes $\alpha \circ \beta$

We also presented two new efficient algorithms, one for computing the code and one for computing the inversion table. Both algorithms have running time $O(\mathbf{N} \log \mathbf{N})$ and use linear space. The concept of *program inversion* has been applied to several problems, including computing the code of a permutation, which was the example application used in the introduction of the concept [Dij78]. We obtained one of our algorithms as an inverse of an existing algorithm, hence introducing the first *efficient* algorithm obtained in that way for a permutation encoding algorithm.

Acknowledgements

The two algorithms in Section 1 were developed jointly with Paolo A.G. Sivilotti. I am grateful for comments on the programs, proofs, and presentation provided by Mani Chandy, David Gries, Robert Harley, Allan Heydon, Peter Hofstee, Rajit Manohar, Adam Rifkin, Paul Sivilotti, and Jan van de Snepscheut (who also introduced me to the problem).

References

[CU90] Wei Chen and Jan Tijmen Udding. Program inversion: More than fun! *Science of Computer Programming*, 15:1–13, 1990.

[Dij78] Edsger W. Dijkstra. Program inversion. EWD 671, Technological University Eindhoven, 1978. Published in *Selected Writings on Computing: A Personal Perspective*, pages 351–354. Texts and Monographs in Computer Science. Springer-Verlag, 1982.

[Gri81] David Gries. *The Science of Programming*, chapter 21. Texts and Monographs in Computer Science. Springer-Verlag, 1981.

[Heh93] Eric C. R. Hehner. *A Practical Theory of Programming.* Texts and Monographs in Computer Science. Springer-Verlag, 1993.

[Knu73] Donald E. Knuth. *The Art of Computer Programming, Vol. 3: Sorting and searching*, section 5.1.1. Addison-Wesley, 1973.