MEASURES OF PRESORTEDNESS AND
OPTIMAL SORTING ALGORITHMS

Extended abstract

Heikki Mannila
Department of Computer Science, University of Helsinki
Tukholmankatu 2, SF-00250 Helsinki 25, Finland

Abstract

The concept of presortedness and its use in sorting are studied. Natural ways to measure presortedness are given and some general properties necessary for a measure are proposed. A concept of a sorting algorithm optimal with respect to a measure of presortedness is defined, and examples of such algorithms are given. An insertion sort is shown to be optimal with respect to three natural measures. The problem of finding an optimal algorithm for an arbitrary measure is studied and partial results are proven.

1. Introduction

The question of identifying in some sense "easy" cases of a computational problem and utilizing this easiness has considerable interest. In sorting, easiness is can be identified with existing order. Indeed, when discussing sorting, it is customary to note that the input can be almost in order or at least have some existing order (see e.g. /Knu73, p. 339/, /Sed75, p.126/, /Dij82, p. 223/ and /Her83, p. 165/).

In this paper we study the use of presortedness in sorting. We do this by trying to answer three questions:

- How can the existing order (presortedness) of a sequence be measured?

- What does it mean that an algorithm utilizes the presortedness of input (measured in some way)?
Do there exist algorithms utilizing presortedness (in the sense of the answers to the previous questions)?

Our questions can be seen as special cases of a more general problem: how can the structure of the input be used in sorting? For structure different from presortedness, this problem has been analyzed in e.g. /HPS75/, /Fre75/ and /ElS81/.

We start in Section 2 by discussing the first question. We present four natural ways to measure presortedness, review briefly their properties and then give general conditions which any measure of presortedness should satisfy.

In Section 3 we tackle the second question. We give a definition of an m-optimal algorithm, where m is a measure of presortedness. The definition is similar to the optimality criteria used in /Me79b/ and /GMPR77/ in the case of a particular measure of presortedness. We give two justifications for the definition: one information-theoretic, and the other based on the behaviour of m-optimal algorithms under various probability distributions.

Section 4 moves to the third question and gives examples of m-optimal algorithms for various natural choices of the measure m. In particular, we are able to exhibit an algorithm which is optimal with respect to three natural measures, is intuitively simple and has a reasonably straightforward implementation.

Section 5 studies the existence of optimal algorithms for arbitrary measures. While we cannot exhibit concrete algorithms, we can still show that for any measure m there exists an almost m-optimal sequence of comparison trees; i.e., we are able to give non-uniform algorithms. If the measure is computable in linear time, then we get truly optimal trees. The result is based on a search technique of Fredman (/Fre76/).

In Section 6 we discuss the results and outline some possible extensions.

If A is a set, then |A| denotes its cardinality; for a sequence X the notation |X| means the length of X. The notation log means logarithm in base 2.