Measures of disorder and straight insertion sort with erroneous comparisons

Petros Hadjicostas
Department of Mathematics and Statistics,
Texas Tech University, Box 41042,
Lubbock, TX 79409-1042
and
K. B. Lakshmanan
Department of Computer Science, State University of New York,
SUNY Brockport, 350 New Campus Drive,
Brockport, NY 14420

ABSTRACT: In this paper, we analyze the familiar straight insertion sort algorithm and quantify the deviation of the output from the correct sorted order if the outcomes of one or more comparisons are in error. The disarray in the output sequence is quantified by six measures. For input sequences whose length is large compared to the number of errors, a comparison is made between the robustness to errors of bubble sort and the robustness to errors of straight insertion sort. In addition to analyzing the behaviour of straight insertion sort, we review some inequalities among the various measures of disarray, and prove some new ones.

KEYWORDS: Analysis of algorithms, comparisons, errors, inversions, measures of disarray, runs, sorting, straight insertion sort

1 Introduction

A number of researchers have developed and analyzed variations of search and sort algorithms to cope with possible errors in comparisons. By an error, we mean the outcome of a binary comparison between two data elements is "no" when factually it should be "yes", and vice versa. Pelc [19] provides a comprehensive survey of the literature in this field. In this paper, we analyze the familiar straight insertion sort algorithm in a new light: if the outcomes of one or more comparisons are in error, by how much will the output deviate from the correct sorted order?

Islam and Lakshmanan [12] analyzed several sort algorithms under the assumption that the outcome of exactly one comparison is in error. Had-

ARS COMBINATORIA 98(2011), pp. 259-288

jicostas and Lakshmanan [11] analyzed bubble sort under the assumption that the outcomes of several comparisons are in error. In this paper, we perform a similar analysis for straight insertion sort. Fault-tolerant algorithms for sorting with a worst-case upper bound on the number of erroneous comparisons have been studied, for example, by Bagchi [3] and Lakshmanan et al. [15]. For a discussion of fault-tolerant sorting networks, we refer the reader to the papers by Yao and Yao [20] and Leighton and Ma [16, 17]. For a probabilistic analysis of sorting when some comparisons are unreliable, see for example Alonso *et al.* [1]. For a more thorough review of the literature, see the introduction in Hadjicostas and Lakshmanan [11].

Let $a = (a_1, a_2, \ldots, a_n)$ be a list or finite sequence consisting of n *distinct* integers. Assume that the correct order for sorting is the ascending one. The degree of *disorder* of the list a can be quantified in a variety of ways (e.g., see [5], [9], [10], [14] and [18]): by the number of runs in a; the smallest number of elements in a that should be removed from a to leave it sorted; the number of inversions in a; the smallest number of successive exchanges of elements in a needed to sort a; the sum of squares of the difference in the ranks between a and the sequence $(1, 2, \ldots, n)$; and the sum of the absolute values of the difference in the ranks between a and the sequence $(1, 2, \ldots, n)$.

By a *run* in a we mean a non-descending sublist of consecutive elements in a, say $(a_i, a_{i+1}, \ldots, a_m)$, such that a_i is not preceded by a smaller number, and a_m is not followed by a larger number. For a sorted list a the number of runs is 1, while for a list a with n elements in reverse order the number of runs is n. The smallest number of integers that should be removed from a list a of n elements to leave it sorted is 0 for a sorted list, while this number equals $n - 1$ for a list a in reverse order. By *inversion* in a we mean a pair of integers in a in the wrong order. For a sorted list a the number of inversions is 0, while for a sequence a in reverse order the number of inversions is $n(n-1)/2$.

It can be shown that the smallest number of successive exchanges of elements in a needed to sort a is n minus the number of cycles in a (when a is considered as a permutation of the first n positive integers). See, for example, [14], Ex. 5.2.2-2, pp. 134 and 628. A sorted sequence has smallest number of exchanges equal to 0, while a sequence in reverse order has smallest number of exchanges equal to $\lfloor n/2 \rfloor$. The list $a = (n, 1, 2, \ldots, n-1)$ has smallest number of exchanges equal to $n - 1$, which is the highest possible.

For a sorted sequence of length n, the sum of squares of the difference in the ranks between a and the sequence $(1, 2, \ldots, n)$ is obviously zero, while for a sequence in reverse ordering is $(n^3 - n)/3 = n(n-1)(n+1)/3$ (e.g., see Kendall [13, p. 9]). Similarly, for a sorted sequence of length n, the sum of absolute values of the difference in the ranks between a and the sequence $(1, 2, \ldots, n)$ is obviously zero, while for a sequence in reverse ordering is $\lfloor n^2/2 \rfloor$, where $\lfloor x \rfloor$ is the greatest integer less than or equal to x (e.g., see Kendall [13, p. 32]).

Section 2 gives some inequalities relating these six measures of disorder. Some of these inequalities are useful in later sections, and some of them are interesting in their own right. For a detailed treatment of measures of disorder and their use in generating nearly sorted sequences (i.e., sequences whose value of a given measure of disorder is bounded by a given number), see the excellent paper by Estivill-Castro [8].

The organization of the other sections of the paper is as follows. In Section 3 we briefly describe straight insertion sort, while in Section 4 we introduce the basic notation of the paper, and give some preliminary results. In Section 5 we give some complementarity results. In Sections 6-17 we give results about the maxima and minima of the six measures of disorder for the output sequences obtained when executions of straight insertion sort (with erroneous comparisons) operate on lists of integers with a specified length. Finally, Section 18 contains a comparison between bubble sort and straight insertion sort. In addition, the section contains some suggested future research topics.

2 Inequality relations among measures of disorder

Let $\mathbb{N} = \{0, 1, 2, \ldots\}$ be the set of nonnegative integers. For a finite set A, let $\#A$ denote the number of elements of A. For each $n \in \mathbb{N} \backslash \{0, 1\}$, let $\mathcal{A}_n$ be the set of all lists with n *distinct* integers as elements.

For a list $a \in \mathcal{A}_n$, let $R(a)$ and $I(a)$ be the number of runs and the number of inversions, respectively, in a. Let also $RM(a)$ be the smallest number of integers that should be removed from a to leave it sorted, and $EX(a)$ be the smallest number of exchanges of elements in a to leave it sorted.

If $a \in \mathcal{A}_n$, let ranks$(a) = (r_1, r_2, \ldots, r_n)$ be the list of ranks of the elements of a, where the smallest number gets the smallest rank. In other words, for $1 \leq i \leq n$, r_i is the rank of the i^{th} element of a. Obviously, ranks(a) is a permutation of the integers $1, 2, \ldots, n$. Define $D(a) = \sum_{i=1}^{n} |i - r_i|$, $SQ(a) = \sum_{i=1}^{n} (i - r_i)^2$, and $V(a) = \sum_{1 \leq i < j \leq n} (j - i) d_{ij}$, where $d_{ij} = 1$ when j precedes i in the sequence $(r_1, \ldots, r_n)$, and zero otherwise. Note that $I(a) = \sum_{1 \leq i < j \leq n} d_{ij}$.

The previous six measures of disarray essentially compare the sequence $(1, 2, \ldots, n)$ to ranks$(a) = (r_1, \ldots, r_n)$ when $a \in \mathcal{A}_n$. If one wants to compare ranks(a) to ranks(b) for $a, b \in \mathcal{A}_n$, he or she can consider right-invariant metrics on the symmetric group Ω_n, i.e., the set of all permutations of the numbers $1, 2, \ldots, n$ equipped with the composition of permutations. Diaconis and Graham [5] examine right-invariant metrics that generalize $I(\cdot)$, $EX(\cdot)$, $SQ(\cdot)$, and $D(\cdot)$. We only comment about one of the metrics. Define $SQ^* : \Omega_n \times \Omega_n \to \mathbb{R}$ by $SQ^*(\pi, \sigma) = \sum_{i=1}^{n}(\pi_i - \sigma_i)^2$ for $\pi, \sigma \in \Omega_n$. Diaconis and Graham [5] state that $SQ^*(\cdot,\cdot)$ is a metric on Ω_n. Of course, that is not what they mean: If $n = 4$, $\pi = (2, 1, 4, 3)$, $\sigma = (2, 3, 1, 4)$, and $\mathrm{id}_4 = (1, 2, 3, 4)$ is the identity in Ω_4, then $SQ^*(\pi, \mathrm{id}_4) + SQ^*(\mathrm{id}_4, \sigma) = 4 + 6 = 10 < SQ^*(\pi, \sigma) = 14$. To transform SQ^* into a metric one has to use $\sqrt{SQ^*}$. For more details, see Estivill-Castro [7].

In this section we give some inequalities among the various measures of disarray. Some of them are useful in later sections, and some of them are interesting in their own right. If $a = (a_1, \ldots, a_n) \in \mathcal{A}_n$, we say that a has a *3-inversion* if there are three elements a_i, a_k, a_j such that $1 \le i < k < j \le n$ and $a_i > a_k > a_j$. The inequality in the following lemma is mentioned by [18] and has an obvious proof. The rest of lemma is a consequence of [14, Ex. 5.2.2-1, pp. 134 and 628].

Lemma 2.1 *For $n \in \mathbb{N}\backslash\{0, 1\}$ and $a \in \mathcal{A}_n$, we have*

$$EX(a) \le I(a).$$

If equality holds in the above inequality, then a has no 3-inversions. (For each $n \ge 4$, there is $a \in \mathcal{A}_n$ for which the converse of the last statement is not true. The simplest example is the sequence $a = (3, 4, 1, 2)$.)

The following four inequalities are "classic" and they are related to non-parametric statistics. Some of them are used in later sections of the paper.

Lemma 2.2 *For $n \in \mathbb{N}\backslash\{0, 1\}$ and $a \in \mathcal{A}_n$, we have*

(a) $-\frac{n(n-1)(n-2)}{6} \le SQ(a) - nI(a) \le 0$;

(b) $SQ(a) = 2V(a) \ge \frac{4}{3}I(a)\left(1 + \frac{I(a)}{n}\right)$;

(c) $I(a) + EX(a) \le D(a) \le 2I(a)$;

(d) $\frac{SQ(a)}{n-1} \le D(a) \le \min[SQ(a), (n\, SQ(a))^{1/2}]$.

Proof: The first inequality is equivalent to an inequality due to Daniels [4], while the second one is due to Durbin and Stuart [6]. The last two inequalities are due to Diaconis and Graham [5]. (There is a minor typo for inequality (d) in [5].) □

As Diaconis and Graham [5] note, in inequality (c) in the previous lemma, $D(a) = 2I(a)$ if and only if a has no 3-inversions. This gives another proof of the fact that $EX(a) = I(a)$ implies list a has no 3-inversions; see

Lemma 2.1. (Due to space limitations we omit the proof of Diaconis and Graham's claim.)

Diaconis and Graham [5] also note (without proof) that for $a \in \mathcal{A}_n$, ranks$(a) = (r_1, \ldots, r_n)$, and n even, we have $D(a) = n^2/2$ if and only if $r_i > n/2$ for $i = 1, 2, \ldots, n/2$. Since this result is needed later in the paper, we prove a slightly more general result. First we state the following easy lemma, whose proof is easy and hence is omitted.

Lemma 2.3 *If* $\max\{i, a_i\} < \min\{j, a_j\}$, *then*

$$|i - a_i| + |j - a_j| < |i - a_j| + |j - a_i|.$$

Lemma 2.4 *For* $n \in I\!N\backslash\{0,1\}$, $a \in \mathcal{A}_n$, *and* ranks$(a) = (r_1, \ldots, r_n)$, *we have:*

(i) If n *is even,* $D(a) = \lfloor n^2/2 \rfloor = n^2/2$ *if and only if* $r_i > n/2$ *for* $i = 1, 2, \ldots, n/2$.

(ii) If n *is odd,* $D(a) = \lfloor n^2/2 \rfloor = (n^2 - 1)/2$ *if and only if, either* $r_i > (n+1)/2$ *for* $i = 1, \ldots, (n-1)/2$, *or* $r_i \geq (n+1)/2$ *for* $i = 1, \ldots, (n+1)/2$.

Proof: We only prove the "only if" part for (ii). The proof of (i) and the proof for the "if" part for (ii) are left to the reader. Assume n is odd. Then $n \geq 3$. Without loss of generality, assume $a \in \Omega_n$, i.e., $r_i = a_i$ for $i = 1, \ldots, n$.

Assume that it is not the case that, either $a_i > (n+1)/2$ for $i = 1, \ldots, (n-1)/2$, or $a_i \geq (n+1)/2$ for $i = 1, \ldots, (n+1)/2$. Therefore there is $i_1 \in \{1, \ldots, (n-1)/2\}$ such that $a_{i_1} \leq (n+1)/2$ and $i_2 \in \{1, 2, \ldots, (n+1)/2\}$ such that $a_{i_2} < (n+1)/2$. Since $a_{i_1} \leq (n+1)/2$ and $i_1 < (n+1)/2$, it is impossible to have $1 \leq a_k \leq (n+1)/2$ for $k = (n+1)/2, \ldots, n$. Hence there is $j_0 \in \{(n+1)/2, \ldots, n\}$ such that $a_{j_0} > (n+1)/2$. If it were true that $i_2 = (n+1)/2 = j_0$, then $a_{i_2} < (n+1)/2 < a_{j_0}$, a contradiction. Hence either $i_2 < (n+1)/2$ or $j_0 > (n+1)/2$. This implies $\max(i_2, a_{i_2}) < \min(j_0, a_{j_0})$. Define $b \in \mathcal{A}_n$ by $b_k = a_k$ if $k \neq i_2, j_0$; $b_{i_2} = a_{j_0}$; and $b_{j_0} = a_{i_2}$. By Lemma 2.3, $D(b) > D(a)$, and so $D(a) < (n^2 - 1)/2$. □

Lemma 2.5 *For* $n \in I\!N\backslash\{0,1\}$ *and* $a \in \mathcal{A}_n$,

$$RM(a) \leq \left\lfloor \frac{n(R(a) - 1)}{R(a)} \right\rfloor. \tag{1}$$

Proof: Assume that the number of elements in each of the $R(a)$ runs of a is less than $n/R(a)$. Then

$$n < R(a)\left(\frac{n}{R(a)}\right) = n,$$

a contradiction. Therefore, the number of elements in at least one run of a, say run i ($1 \le i \le R(a)$), will be greater than or equal to $n/R(a)$. If we remove all the numbers from all the other runs, the list will be left sorted. The total number of all the elements in all runs, except in run i, is less than or equal to $n - n/R(a) = n(R(a)-1)/R(a)$ and so $RM(a) \le \lfloor n(R(a)-1)/R(a) \rfloor$. □

Lemma 2.6 *For $n \in \mathbb{N}\backslash\{0,1\}$ and $a \in \mathcal{A}_n$,*

$$I(a) \le \left\lfloor \frac{n^2(R(a)-1)}{2R(a)} \right\rfloor.$$

Proof: Assume that the $R(a)$ runs of a have $k_1, k_2, \ldots, k_{R(a)}$ elements, respectively. Then $\sum_{i=1}^{R(a)} k_i = n$. No inversions can exist within a run, and so

$$I(a) \le \sum_{1 \le i < j \le R(a)} k_i k_j = \frac{\left(\sum_{i=1}^{R(a)} k_i\right)^2 - \sum_{i=1}^{R(a)} k_i^2}{2} = \frac{n^2 - \sum_{i=1}^{R(a)} k_i^2}{2}.$$

It is not difficult to show that

$$R(a) \sum_{i=1}^{R(a)} k_i^2 \ge \left(\sum_{i=1}^{R(a)} k_i \right)^2 = n^2.$$

Therefore,

$$I(a) \le \frac{n^2 - \frac{n^2}{R(a)}}{2} = \frac{n^2(R(a)-1)}{2R(a)},$$

from which the lemma follows. □

3 Straight insertion sort

```
for (i = 1; i < n; i = i + 1)
{   insertValue = a_{i+1}; j = i;
    while (j ≥ 1 and insertValue < a_j)
    {       a_{j+1} = a_j;
            j = j − 1; }
    a_{j+1} = insertValue; }
```

Consider the straight insertion sort algorithm given above. The algorithm consists of $n-1$ passes, and pass i (where $1 \le i \le n-1$) consists of at most i comparisons. If there are no errors in any of the comparisons, before pass i, the first i integers of the intermediate output sequence have been sorted. During pass i, the algorithm inserts the $(i+1)^{\text{th}}$ integer in the appropriate place by "shifting" it to the left. The "shifting" is done by comparing the integer to be inserted to some of the first i integers (obtained after pass $i-1$). (We assume that, during the first pass, the first two numbers of the input sequence are compared and placed in the correct order.) The comparisons in pass i stop once the integer to be inserted is compared to a smaller number. Note that straight insertion sort does at most $n(n-1)/2$ comparisons in total. For example, for the sorted list $1, 2, \ldots, n$, we need only $n-1$ comparisons, while for the sequence $n, n-1, \ldots, 2, 1$ (in reverse order) we need $n(n-1)/2$ comparisons.

4 Notation and some preliminary results

For each $a \in \mathcal{A}_n$, let $\mathcal{S}_n(a)$ be the set of all executions of the straight insertion sort algorithm that can sort list a (whose length is n), and can make up to $n(n-1)/2$ errors when making comparisons. This means that, for each $S \in \mathcal{S}_n(a)$, the collection of comparisons where S is erring is uniquely associated with S. For each pass i of S (where $1 \leq i \leq n-1$), the comparisons stop when the $(i+1)^{\text{th}}$ integer is compared to a smaller number and no error occurs, or is compared to a larger number and an error occurs.

For $a \in \mathcal{A}_n$ and $S \in \mathcal{S}_n(a)$, denote by $C(S)$ and $E(S)$ the total number of comparisons and the total number of errors S does, respectively, when it operates on a. Obviously, $E(S) \leq C(S) \leq n(n-1)/2$. Since straight insertion sort has $n-1$ passes and each pass has at least one comparison, $C(S) \geq n-1$.

Let $a, a' \in \mathcal{A}_n$, $S \in \mathcal{S}_n(a)$, and $S' \in \mathcal{S}_n(a')$. We write $S = S'$ if and only if (a) for each pass i (where $1 \leq i \leq n-1$), S and S' make the same number of comparisons; and (b) for each pass i (where $1 \leq i \leq n-1$), and each comparison k of pass i (where $1 \leq k \leq i$), S errs if and only if S' errs. In such a case, $C(S) = C(S')$ and $E(S) = E(S')$.

Lemma 4.1 *Let $n \in \mathbb{N} \backslash \{0, 1\}$, $a \in \mathcal{A}_n$, and $S \in \mathcal{S}_n(a)$ be given. If $E(S) = C(S)$, then the output list, b, from the operation of S on a has* ranks$(b) = (n, n-1, \ldots, 2, 1)$.

Proof: Since all the comparisons of the execution S are in error, it follows that whenever there is a comparison, the larger number goes (or stays) to the left, and the smaller number goes (or stays) to the right. Therefore, after pass k (where $1 \leq k \leq n-1$), the first $k+1$ integers of the sequence obtained are sorted in reverse order (from largest to smallest).

Thus, at the end of the last pass, the output sequence b is in reverse order, i.e., ranks$(b) = (n, n-1, \ldots, 2, 1)$. □

Note that Lemma 4.1 follows also from Lemma 5.2.

Lemma 4.2 *Let $n \in \mathbb{N} \backslash \{0, 1\}$ and $e \in \mathbb{N}$ be such that $0 \leq e \leq n(n-1)/2$. Then:*

$$A(n, e) := \{(a, S) : \; a \in \mathcal{A}_n, \; S \in \mathcal{S}_n(a), \; E(S) = e\} \neq \emptyset. \qquad (2)$$

Proof: Let $k(e)$ be the largest integer k such that $k(k+1)/2 \leq e$. Let $a = (1, 2, \ldots, n)$, and suppose execution $S \in \mathcal{S}_n(a)$ makes errors in all the comparisons of passes $1, 2, \ldots, k(e)$, and in the first $e - k(e)(k(e)+1)/2$ comparisons of pass $k(e) + 1$. Note that S is well-defined for the following reason: In pass i, where $1 \leq i \leq k(e)$, we have exactly i comparisons, and at the end of the pass, the first $i+1$ integers of the sequence obtained are in reverse order (i.e., they are the numbers $i+1, i, i-1, \ldots, 2, 1$ in this order). This means that the total number of comparisons in the first $k(e)$ passes is $1 + 2 + \ldots + k(e) = k(e)(k(e)+1)/2 \leq e = E(S) \leq C(S)$. It is easy to show that $e - k(e)(k(e)+1)/2 < k(e) + 1$, and so it is possible to make exactly $e - k(e)(k(e)+1)/2$ errors in pass $k(e)+1$. (Since $e - k(e)(k(e)+1)/2 < k(e)+1$, pass $k(e)+1$ of S terminates with comparison $1 + e - k(e)(k(e)+1)/2$.) Since S is well-defined and $E(S) = e$, it follows that $A(n, e) \neq \emptyset$. □

Lemma 4.2 allows us to state the following definitions. For each $n \in \mathbb{N} \backslash \{0, 1\}$, $a \in \mathcal{A}_n$, and $S \in \mathcal{S}_n(a)$, let:

(a) $R(a, S)$ be the number of runs in the output list after S operates on a;

(b) $RM(a, S)$ be the smallest number of integers that should be removed from the output sequence, after S operates on a, to leave it sorted;

(c) $I(a, S)$ be the number of inversions in the output list after S operates on a;

(d) $EX(a, S)$ be the smallest number of successive exchanges needed to sort the output list after S operates on a;

(e) $SQ(a, S)$ be the the sum of squares of the difference in the ranks between the output sequence after S operates on a and the sequence $(1, 2, \ldots, n)$;

(f) $D(a, S)$ be the sum of the absolute values of the difference in the ranks between the output sequence after S operates on a and the sequence $(1, 2, \ldots, n)$.

We then have the following inequalities:

(a) $1 \leq R(a, S) \leq n$;	(b) $0 \leq RM(a, S) \leq n - 1$;
(c) $0 \leq I(a, S) \leq n(n-1)/2$;	(d) $0 \leq EX(a, S) \leq n - 1$;
(e) $0 \leq SQ(a, S) \leq (n^3 - n)/3$;	(f) $0 \leq D(a, S) \leq \lfloor n^2/2 \rfloor$.

For the integers n and e with $n \geq 2$ and $0 \leq e \leq n(n-1)/2$, recall the definition of $A(n, e)$ from (2). We then define:

$$\begin{aligned}
\text{Mruns}(n,e) &= \max\{R(a,S): (a,S) \in A(n,e)\}; \\
\text{mruns}(n,e) &= \min\{R(a,S): (a,S) \in A(n,e)\}; \\
\text{Mrem}(n,e) &= \max\{RM(a,S): (a,S) \in A(n,e)\}; \\
\text{mrem}(n,e) &= \min\{RM(a,S): (a,S) \in A(n,e)\}; \\
\text{Minv}(n,e) &= \max\{I(a,S): (a,S) \in A(n,e)\}; \\
\text{minv}(n,e) &= \min\{I(a,S): (a,S) \in A(n,e)\}; \\
\text{Mexc}(n,e) &= \max\{EX(a,S): (a,S) \in A(n,e)\}; \\
\text{mexc}(n,e) &= \min\{EX(a,S): (a,S) \in A(n,e)\}; \\
\text{Msqr}(n,e) &= \max\{SQ(a,S): (a,S) \in A(n,e)\}; \\
\text{msqr}(n,e) &= \min\{SQ(a,S): (a,S) \in A(n,e)\}; \\
\text{Mabs}(n,e) &= \max\{D(a,S): (a,S) \in A(n,e)\}; \\
\text{mabs}(n,e) &= \min\{D(a,S): (a,S) \in A(n,e)\}.
\end{aligned}$$

For example, Mruns(n, e) and mruns(n, e) represent the worst and the best case scenaria, respectively, for the number of runs in the output list obtained when an execution of straight insertion sort with exactly e errors operates on a list of integers with length n.

5 Complementarity results

Recall that, for $n \in \mathbb{N}\backslash\{0,1\}$ and $a \in \mathcal{A}_n$, ranks$(a) = (r_1, r_2, \ldots, r_n)$ is the list of ranks of the elements of a, where the smallest number gets the smallest rank. Not only ranks(a) is a permutation of the integers $1, 2, \ldots, n$, but also the list $(n+1-r_1, n+1-r_2, \ldots, n+1-r_n)$ is a permutation of the first n positive integers. Consider the list $\overline{a} \in \mathcal{A}_n$, called the *complement* of a, which is created by putting in the i^{th} position the element of a whose rank is $n+1-r_i$. Then ranks$(\overline{a}) = (n+1-r_1, n+1-r_2, \ldots, n+1-r_n)$. Kendall [13, p. 11] calls such sequences a and $\overline{a}$ as *conjugate*. The following lemma gives some complementarity results for the six measures of disarray we study.

Lemma 5.1 *(i)* $R(\overline{a}) = n+1-R(a)$;
(ii) $I(\overline{a}) = n(n-1)/2 - I(a)$;
(iii) $RM(\overline{a}) \geq n-1-RM(a)$.
(iv) $EX(\overline{a}) \geq \lfloor n/2 \rfloor - EX(a)$.
(v) $SQ(\overline{a}) = n(n-1)(n+1)/3 - SQ(a)$.
(vi) $D(\overline{a}) \geq \lfloor n^2/2 \rfloor - D(a)$.
In (iii), (iv), and (vi), when ranks$(a) = (1, 2, \ldots, n)$, *the inequality holds as equality.*

Proof: The proofs of parts (i) and (ii) are easy and hence are omitted. (See also Lemma 4.1 in [11].) To prove part (iii), assume that we have removed $RM(a)$ elements from a and we are left with the numbers $r_1 < r_2 < \ldots < r_k$, where $k = n - RM(a)$. (Without loss of generality we may assume a is a permutation of the first n positive integers.) Then $n+1-r_1 > n+1-r_2 > \ldots > n+1-r_k$. In the complement of a, the latter numbers appear somewhere (not necessarily consecutively) but in the order shown. In other words, $n+1-r_1$ is to the left of $n+1-r_2$, which is to the left of $n+1-r_3$, etc. This means that in order to leave $\overline{a}$ sorted we must remove at least $k-1$ of the numbers $n+1-r_1, n+1-r_2, \ldots, n+1-r_k$. Therefore, $RM(\overline{a}) \geq k-1$, from which the third part of the lemma follows.

To prove (iv), let Ω_n be the symmetric group, i.e., the set of all permutations of the first n positive integers. Diaconis and Graham [5] define the metric T on Ω_n as follows: For $\pi, \sigma \in \Omega_n$, let $T(\pi, \sigma)$ be the minimum number of successive exchanges needed to bring $(\pi_1, \ldots, \pi_n)$ into the order $(\sigma_1, \ldots, \sigma_n)$. If $\mathrm{id}_n = (1, \ldots, n)$ is the identity of the group Ω_n, then $T(\pi, \sigma) = T(\mathrm{id}_n, \sigma\pi^{-1}) = EX(\sigma\pi^{-1})$. Without loss of generality, assume $a \in \Omega_n$. Then (since T is a metric):

$$EX(a) + EX(\overline{a}) = T(a, \mathrm{id}_n) + T(\mathrm{id}_n, \overline{a}) \geq T(a, \overline{a}) = T(\mathrm{id}_n, \overline{a}a^{-1}).$$

Thus $EX(a) + EX(\overline{a}) \geq EX(\overline{a}a^{-1})$. Since $\overline{a} = (n, n-1, \ldots, 1)a$, we have $\overline{a}a^{-1} = (n, n-1, n-2, \ldots, 1)$, which proves part (iv).

Part (v) is proven by Kendall [13, p. 24]. To prove (vi), note that

$$\begin{aligned} D(a) + D(\overline{a}) &= \sum_{i=1}^{n} (|i - r_i| + |i - (n+1-r_i)|) \\ &\geq \sum_{i=1}^{n} |2i - (n+1)| \\ &= D((n, n-1, \ldots, 1)) = \lfloor n^2/2 \rfloor. \end{aligned}$$

The proof of the lemma is complete. □

Lemma 5.2 *Let $a \in \mathcal{A}_n$ and $S \in S_n(a)$. Then there is an execution $\overline{S}$ of straight insertion sort that belongs to $S_n(\overline{a})$ such that:*

(a) In each pass i (where $1 \leq i \leq n-1$), S and $\overline{S}$ make the same number of comparisons.

(b) In each comparison k of pass i, $\overline{S}$ makes an error if and only if S does not make an error.

(c) $E(S) + E(\overline{S}) = C(S) = C(\overline{S})$.

(d) The output list we get when $\overline{S}$ operates on $\overline{a}$ is the complement of the output list we get when S operates on a.

Proof: Without loss of generality we may assume that a is a permutation of the first n positive integers. Let $a(i,k)$ be the output sequence we get after comparison k of pass i when S operates on a. Let C_i be the number of comparisons during pass i of S. We have $1 \leq i \leq n-1$ and $1 \leq k \leq C_i$. Define $a(1,0) = a$ and $a(i,0) = a(i-1, C_{i-1})$ for $i = 2,3,\ldots,n-1$. In other words, let $a(i,0)$ be the output sequence at the beginning of pass i.

For each $i \in \{1,2,\ldots,n-1\}$ and $k \in \{0,1,\ldots,C_i\}$, define $b(i,k) = \overline{a(i,k)}$. Note that $b(1,0) = \overline{a(1,0)} = \overline{a}$, and

$$b(i,0) = \overline{a(i,0)} = \overline{a(i-1,C_{i-1})} = b(i-1,C_{i-1})$$

for $i = 2,3,\ldots,n-1$. We define execution $\overline{S}$ of "sorting lists in $\mathcal{A}_n$ with errors" as follows: We assume that execution $\overline{S}$ has $n-1$ "passes," and "pass" i has exactly C_i "comparisons" ($i = 1,\ldots,n-1$). For each $i \in \{1,\ldots,n-1\}$ and $k \in \{1,\ldots,C_i\}$, we assume that the output sequence we get after "comparison" k of "pass" i is $b(i,k)$. By "pass" of $\overline{S}$ we mean the first $i+1$ integers of $b(i,k)$; by "comparison" k of "pass" i of $\overline{S}$ we mean the pair consisting of the $(i+1-k)^{\text{th}}$ and $(i+2-k)^{\text{th}}$ integers of $b(i,k)$. We say that an "error" has occurred during "comparison" k of "pass" i of $\overline{S}$ if the $(i+1-k)^{\text{th}}$ integer of $b(i,k)$ is greater than the $(i+2-k)^{\text{th}}$ integer of $b(i,k)$.

We will show inductively that $\overline{S}$ is well-defined (i.e., that $\overline{S} \in \mathcal{S}_n(\overline{a})$), and that it satisfies properties (a)-(d) of the lemma. By definition, property (a) holds. Since $b(n-1,C_{n-1}) = \overline{a(n-1,C_{n-1})}$, property (d) is also satisfied.

Since $b(1,0) = \overline{a}$, $\overline{S}$ operates on $\overline{a}$ at the beginning of the first "pass." Assume that after "comparison" k of "pass" i (where $1 \leq i \leq n-1$ and $1 \leq k \leq C_i - 1$), the output sequence of the application $\overline{S}$ on $\overline{a}$ is $(p_1,p_2,\ldots,p_n)$, i.e., assume $b(i,k) = \overline{a(i,k)} = (p_1,p_2,\ldots,p_n)$. Note that after comparison k of pass i of S, the output sequence is $a(i,k) = (n+1-p_1, n+1-p_2,\ldots,n+1-p_n)$.

In pass i of S, the $(i+1)^{\text{th}}$ integer of $a(i,0) = \overline{b(i,0)}$ is shifted to the left and is compared to some of the first i integers of $a(i,0)$. In comparison $k+1$ of pass i of S, integer $n+1-p_{i-k+1}$ is compared to $n+1-p_{i-k}$. In "comparison" $k+1$ of "pass" i of $\overline{S}$, integer p_{i-k+1} is "compared" to p_{i-k}. Note that $p_{i-k+1} < p_{i-k}$ if and only if $n+1-p_{i-k+1} > n+1-p_{i-k}$. If S does not make an error in comparison $k+1$ of pass i, then

$$\begin{aligned} a(i,k+1) = {} & (n+1-p_1,\ldots,n+1-p_{i-k-1}, \\ & \min(n+1-p_{i-k}, n+1-p_{i-k+1}), \\ & \max(n+1-p_{i-k}, n+1-p_{i-k+1}), \\ & n+1-p_{i-k+2},\ldots,n+1-p_n), \end{aligned}$$

which implies

$$b(i,k+1) = \overline{a(i,k+1)} = (p_1,\ldots,p_{i-k-1},\max(p_{i-k},p_{i-k+1}), \min(p_{i-k},p_{i-k+1}),p_{i-k+2},\ldots,p_n).$$

In such a case, in sequence $b(i,k+1)$, integer $\max(p_{i-k},p_{i-k+1})$ is to the left of $\min(p_{i-k},p_{i-k+1})$, and so $\overline{S}$ is in "error" during "comparison" $k+1$ of "pass" i.

It can be similarly proven that, if S makes an error in comparison $k+1$ of pass i, then $\overline{S}$ does not make an "error" during "comparison" $k+1$ of "pass" i. It follows that property (b) is satisfied. Also, the total number of "errors" of $\overline{S}$ is the total number of non-errors of S, and so property (c) holds. The previous arguments show that indeed $\overline{S} \in \mathcal{S}_n(\overline{a})$. □

Given $a \in \mathcal{A}_n$ and $S \in \mathcal{S}_n(a)$, the execution $\overline{S} \in \mathcal{S}_n(\overline{a})$ is called the *complement* of S. Combining Lemmas 5.1 and 5.2, we can prove the following result. It allows us to prove results about the minimum value of a measure of disorder (given n and e) using results about the maximum value of the measure. See Sections 7 and 9.

Corollary 5.3 *For $a \in \mathcal{A}_n$ and $S \in \mathcal{S}_n(a)$:*
(i) $R(\overline{a},\overline{S}) = n+1-R(a,S)$;
(ii) $I(\overline{a},\overline{S}) = n(n-1)/2 - I(a,S)$;
(iii) $RM(\overline{a},\overline{S}) \geq n-1-RM(a,S)$;
(iv) $EX(\overline{a},\overline{S}) \geq \lfloor n/2 \rfloor - EX(a,S)$.
(iv) $SQ(\overline{a},\overline{S}) = n(n-1)(n+1)/3 - SQ(a,S)$;
(v) $D(\overline{a},\overline{S}) \geq \lfloor n^2/2 \rfloor - D(a,S)$.

6 Maximum number of runs

It was proven in [12] that $\text{Mruns}(n,1) = 2$ for $n \in \mathbb{N}\backslash\{0,1\}$. The following theorem gives some more general results regarding $\text{Mruns}(n,e)$.

Theorem 6.1 *Let $n \in \mathbb{N}\backslash\{0,1\}$ and e be an integer with $0 \leq e \leq n(n-1)/2$. Then:*
(a) For $0 \leq e \leq n-1$, $\text{Mruns}(n,e) = e+1$.
(b) For $n-1 \leq e \leq n(n-1)/2$, $\text{Mruns}(n,e) = n$.

Proof: (a) We first use finite induction on e to prove that

$$\text{Mruns}(n,e) \leq e+1 \tag{3}$$

for $e \leq n(n-1)/2$ (and thus for $e \leq n-2$). If $e=0$, then $\text{Mruns}(n,e) = 1 = 0+1$. If $e=1$, then it follows from [12] that $\text{Mruns}(n,e) = 2 = 1+1$.

Let ϵ be an integer with $2 \le \epsilon \le n-2$, and assume that for any integer e with $0 \le e < \epsilon$, we have $\text{Mruns}(m, e) \le e+1$ for all $m \in \mathbb{N}\backslash\{0,1\}$ with $e \le m(m-1)/2$. Let $a \in \mathcal{A}_n$ and $S \in \mathcal{S}_n(a)$ be such that $E(S) = \epsilon$ and $\text{Mruns}(n, \epsilon) = R(a, S)$. Assume that, when S operates on a, the last error occurs during pass i of the algorithm, where $1 \le i \le n-1$. Since $\epsilon \ge 2$ and since there is only one comparison in the first pass, it follows that $i \ge 2$.

For each integer k (where $1 \le k \le n-1$), let β_k be the sequence obtained after k passes of S operating on a, i.e.,

$$a \xrightarrow{\text{1st pass}} \beta_1 \xrightarrow{\text{2nd pass}} \beta_2 \xrightarrow{\text{3rd pass}} \cdots \xrightarrow{(n-2)\text{th pass}} \beta_{n-2} \xrightarrow{(n-1)\text{th pass}} \beta_{n-1}.$$

Note that β_k is a permutation of a. For $1 \le k \le n-1$, let α_k be the subsequence of β_k consisting of the first $k+1$ integers of β_k, i.e.,

$$\alpha_k = (\beta_{k1}, \ldots, \beta_{k(k+1)}).$$

Let also γ_k be the subsequence of a consisting of the first $k+1$ integers of a, i.e.,

$$\gamma_k = (a_1, \ldots, a_{k+1}).$$

Because of the way straight insertion sort works, γ_k is a permutation of α_k. In addition, let S_k be the execution of straight insertion sort that consists of the first k passes of S and operates on γ_k. Then $\alpha_k, \gamma_k \in \mathcal{A}_{k+1}$ and $S_k \in \mathcal{S}_{k+1}(\gamma_k)$. Also $\gamma_{n-1} = a$ and $S_{n-1} = S$.

Since $E(S_{i-1}) < E(S_i) = \epsilon$, it follows from the induction hypothesis that

$$R(\gamma_{i-1}, S_{i-1}) \le \text{Mruns}(i, E(S_{i-1})) \le E(S_{i-1}) + 1 \le (\epsilon - 1) + 1 = \epsilon.$$

Consider the insertion of the $(i+1)^{\text{th}}$ integer of β_{i-1} into α_{i-1} (which consists of the first i integers of β_{i-1}). We obtain β_i, whose first $i+1$ integers make up α_i. No matter how many errors there are in pass i of S, the insertion of the $(i+1)^{\text{th}}$ integer of β_{i-1} into α_{i-1} introduces at most one extra run, so $R(\gamma_i, S_i) \le R(\gamma_{i-1}, S_{i-1}) + 1$.

Consider now the insertion of the remaining $n-i-1$ integers of β_i into α_i. Since there are no further errors in comparisons, each new integer will be shifted to the left to an appropriate place in either one of the $R(\gamma_i, S_i)$ runs. No new runs are created by the insertion of the last $n-i-1$ integers of β_i into α_i. In other words, for each integer k with $i \le k \le n-1$, $R(\gamma_k, S_k) = R(\gamma_i, S_i)$. Indeed, let j be an arbitrary integer such that $i \le j \le n-2$, and let t be the integer inserted in α_j after pass $j+1$ (thus giving rise to α_{j+1}). The integer t is inserted in one of the $R(\gamma_j, S_j)$ runs, in the appropriate position, extending the length of that run by one, but creating no new runs.

This means $R(\gamma_i, S_i) = R(\gamma_j, S_j) = R(\gamma_{j+1}, S_{j+1}) = R(\gamma_{n-1}, S_{n-1})$, and so

$$\text{Mruns}(n,e) = R(a,S) = R(\gamma_i, S_i) \leq R(\gamma_{i-1}, S_{i-1}) + 1 \leq \epsilon + 1.$$

Inequality (3) has thus been proven.

To prove that equality holds in part (a) when $0 \leq e \leq n-1$, let $n = (e+1)k + u$, where k, u are integers with $0 \leq u \leq e$. Then $k = \lfloor n/(e+1) \rfloor$. (Since $e \leq n-1$, we have $k \geq 1$.) Let

$$\begin{aligned} a \;=\; & (n-k+1, n-k+2, \ldots, n, n-2k+1, n-2k+2, \ldots, n-k, \\ & n-3k+1, \ldots, n-2k, \ldots, n-(e-1)k, 1, 2, \ldots, n-(e+1)k, \\ & n-(e+1)k+1, \ldots, n-ek). \end{aligned} \tag{4}$$

In other words, a consists of the k largest integers between 1 and n (in ascending order); the next k largest integer (in ascending order), etc. The last part of a consists of the remaining $n - ke$ integers $1, 2, \ldots, n - ke$ (in ascending order). Let $S \in \mathcal{S}_n(a)$ be the execution of insertion sort that errs in the first comparison of each of the passes $k, 2k, \ldots, ek$. (Note that $ek \leq n-1$.) The output of the operation of S on a is identical to the input sequence a. Then $R(a,S) = e+1$, which completes the proof of part (a).

(b) By Lemma B.1 in Appendix B, there is $a \in \mathcal{A}_n$ and $S \in \mathcal{S}_n(a)$ such that $E(S) = 0$ and $C(S) = e$. The output of the operation of S on a is $(1, 2, \ldots, n-1, n)$. By Lemma 5.2, the complement $\overline{S}$ of S satisfies $\overline{S} \in \mathcal{S}_n(\overline{a})$, $C(\overline{S}) = C(S) = e$, and $E(\overline{S}) = C(S) - E(S) = e - 0 = e$. Also, the output of the operation of $\overline{S}$ on $\overline{a}$ is $(n, n-1, \ldots, 2, 1)$, and so $R(\overline{a}, \overline{S}) = n$, which implies $\text{Mruns}(n,e) = n$. □

7 Minimum number of runs

The following theorem gives some results regarding $\text{mruns}(n,e)$.

Theorem 7.1 *Let $n \in \mathbb{N} \setminus \{0,1\}$ and e be an integer with $0 \leq e \leq n(n-1)/2$. Then:*

(a) $\text{mruns}(n,0) = 1$.

(b) If $e \geq 1$ then $\text{mruns}(n,e) \geq 2$.

(c) If $1 \leq e \leq n-1$, then $\text{mruns}(n,e) = 2$.

(d) If $n(n-1)/2 - (n-1) \leq e \leq n(n-1)/2$, then $\text{mruns}(n,e) \geq n + e - n(n-1)/2$.

(e) If $n(n-1)/2 - e \in \{0,1,2\}$, then $\text{mruns}(n,e) = n + e - n(n-1)/2$. *(If $n(n-1)/2 - e = 2$, we need to assume $n \geq 4$.)*

Proof: (a) The proof of this part is obvious.

(b) If $a \in \mathcal{A}_n$ and $S \in \mathcal{S}_n(a)$ with $E(S) = e \geq 1$ and $R(a,S) = \text{mruns}(n,e)$, then the output of the operation of S on a contains at least one inversion, and so $\text{mruns}(n,e) = R(a,S) \geq 2$.

(c) Assume $1 \leq e \leq n-1$. By part (b), $\text{mruns}(n,e) \geq 2$. To show equality in this case, let $a_1 = (1,2,\ldots,n)$ and $S_1 \in \mathcal{S}_n(a_1)$ be such that it errs only in the first comparison of each of the passes $1,2,\ldots,e$. Then $E(S_1) = e$, and the output of the operation of S_1 on a_1 is $(2,3,\ldots,e+1,1,e+2,\ldots,n)$. (If $e = n-1$, then the output is $(2,3,\ldots,n,1)$.) Then $R(a_1,S_1) = 2$, which proves the equality $\text{mruns}(n,e) = 2$.

(d) Choose $a \in \mathcal{A}_n$ and $S \in \mathcal{S}_n(a)$ such that $\text{mruns}(n,e) = R(a,S)$ and $E(S) = e$. By Corollary 5.3,

$$\text{mruns}(n,e) = R(a,S) = n+1-R(\overline{a},\overline{S}) \geq n+1-\text{Mruns}(n,C(S)-e).$$

However, $C(S) - e \leq n(n-1)/2 - e \leq n-1$, and by Theorem 6.1(a), $\text{Mruns}(n,C(S)-e) = C(S)-e+1$. Hence,

$$\text{mruns}(n,e) \geq n - C(S) + e \geq n + e - n(n-1)/2.$$

(e) If $n(n-1)/2-e = 0$, then $n \geq \text{mruns}(n,e) \geq n$, and so $\text{mruns}(n,e) = n$.

If $n(n-1)/2 - e = 1$, then $\text{mruns}(n,e) \geq n-1$. Let $a = (1,2,\ldots,n)$, and $S \in \mathcal{S}_n(a)$ be an execution of straight insertion sort that errs in all comparisons of all passes, except in the last comparison of the last pass. Then the output is $(n-1,n,n-2,n-3,\ldots,2,1)$ if $n \geq 4$, and $(2,3,1)$ if $n=3$, which has $n-1$ runs. Since $E(S) = n(n-1)/2-1$, we have shown that $\text{mruns}(n,e) = n-1$.

If $n(n-1)/2 - e = 2$ and $n \geq 4$, then $\text{mruns}(n,e) \geq n-2$. Let $a = (1,2,\ldots,n)$, and $S \in \mathcal{S}_n(a)$ be an execution of straight insertion sort that errs in all comparisons of all passes, except in the first comparison of the first pass and in the last comparison of the last pass. Then the output is $(n-1,n,n-2,n-3,\ldots,3,1,2)$ if $n \geq 5$, and $(3,4,1,2)$ if $n=4$, which has $n-2$ runs. Since $E(S) = n(n-1)/2-2$, we have shown that $\text{mruns}(n,e) = n-2$. □

8 Maximum number of inversions

In [12] it was proven that $\text{Minv}(n,1) = \lfloor n^2/4 \rfloor$ for all $n \in \mathbb{N}\backslash\{0,1\}$. The following theorem gives some more general results regarding $\text{Minv}(n,e)$.

Theorem 8.1 *Let $n \in \mathbb{N}\backslash\{0,1\}$ and e be an integer with $0 \leq e \leq n(n-1)/2$. Then:*

(a) For $0 \le e \le n-1$,

$$\left\lfloor \frac{n}{e+1} \right\rfloor \left\{ ne - \left\lfloor \frac{n}{e+1} \right\rfloor \frac{e(e+1)}{2} \right\} \le \mathrm{Minv}(n,e) \le \left\lfloor \frac{n^2 e}{2(e+1)} \right\rfloor .$$

(b) If either $n \equiv 0 (\mathrm{mod}\ e+1)$ *or* $n \equiv 1 (\mathrm{mod}\ e+1)$, *and* $0 \le e \le n-1$, *then*

$$\mathrm{Minv}(n,e) = \left\lfloor \frac{n^2 e}{2(e+1)} \right\rfloor .$$

(c) For $n-1 \le e \le n(n-1)/2$, $\mathrm{Minv}(n,e) = n(n-1)/2$.

Proof: (a) The right inequality follows from Lemma 2.6 and Theorem 6.1(a). To prove the left inequality, define $a \in \mathcal{A}_n$ and $S \in \mathcal{S}_n(a)$ as in the second part of the proof of part (a) of Theorem 6.1 (see (4)). The output of the operation of S on a is identical to the input sequence a. Then

$$I(a,S) = k(n-k) + k(n-2k) + \ldots + k(n-ek) = k \left[ne - k\frac{e(e+1)}{2} \right]$$

with $k = \lfloor n/(e+1) \rfloor$, which proves the left inequality of (a).

(b) It follows from part (a) of this theorem.

(c) It follows from Theorem 6.1(b). □

9 Minimum number of inversions

The following theorem gives some results regarding $\mathrm{minv}(n,e)$.

Theorem 9.1 *Let* $n \in \mathbb{N} \backslash \{0,1\}$ *and* e *be an integer with* $0 \le e \le n(n-1)/2$. *Then:*

(a) $\mathrm{minv}(n,0) = 0$.

(b) $\mathrm{minv}(n,e) \ge e$.

(c) If $0 \le e \le n-1$, *then* $\mathrm{minv}(n,e) = e$.

(d) If $n(n-1)/2 - (n-1) \le e \le n(n-1)/2$, *then*

$$\mathrm{minv}(n,e) \ge \frac{n(n-1)}{2} - \left\lfloor \frac{n^2\{n(n-1)-2e\}}{2\{n(n-1)-2e+2\}} \right\rfloor .$$

Proof: (a) The proof of this part is obvious.

(b) We use induction on e. For $e = 0$ or $e = 1$, the inequality is obvious. Let $\epsilon \ge 2$ be an integer and assume that $\mathrm{minv}(n,e) \ge e$ for all integers e and n with $0 \le e < \epsilon$, $n \ge 2$, and $e \le n(n-1)/2$. Let m be an integer such that $m \ge 3$ and $\epsilon \le m(m-1)/2$. By definition, there is $a_0 \in \mathcal{A}_m$ and $S_0 \in \mathcal{S}_m(a_0)$ with $E(S_0) = \epsilon$ such that $\mathrm{minv}(m,\epsilon) = I(a_0, S_0)$. Write

$a_0 = (a_{01}, a_{02}, \ldots, a_{0m})$. Let j be the pass where the last error of S_0 occurs. Since $\epsilon \geq 2$, we have $j \geq 2$. Let $b = (a_{01}, a_{02}, \ldots, a_{0j})$ and S_b be the part of S that consists of the first $j-1$ passes of S. Then $b \in \mathcal{A}_j$, $S_b \in \mathcal{S}_j(b)$, and $E(S_b) < \epsilon$. Obviously, $E(S_b) \leq j(j-1)/2$.

By the induction hypothesis, $I(b, S_b) \geq \text{minv}(j, E(S_b)) \geq E(S_b)$. In pass j, we have $\epsilon - E(S_b) \geq 1$ errors. If α is the number in position $j+1$ in list a_0 at the beginning of pass j, then for the first $\epsilon - E(S_b) - 1$ erroneous comparisons, α has moved to the left bypassing smaller numbers at least $\epsilon - E(S_b) - 1$ times. During the final erroneous comparison of pass j, an inversion was created no matter whether α moved one position to the left. Therefore, during pass j at least $\epsilon - E(S_b)$ inversions were created, and no inversions from the previous passes have been destroyed. At the end of pass $n-1$, the other $n-j-1$ numbers of a_0 (if there are any) will be placed somewhere in the list, but no previous inversions will be destroyed (even though some new ones may be created). Therefore, $\text{minv}(m, \epsilon) = I(a_0, S_0) \geq I(b, S_b) + \epsilon - E(S_b) \geq E(S_b) + \epsilon - E(S_b) = \epsilon$. The induction step is complete.

(c) Assume $0 \leq e \leq n-1$. By part (b) of this theorem, $\text{minv}(n, e) \geq e$. To prove equality, consider list $a_1 = (1, 2, \ldots, n)$ and execution $S_1 \in \mathcal{S}_n(a_1)$ of the example in the proof of Theorem 7.1, part (c). Then $E(S_1) = e$ and the output of the operation of S_1 on a_1 is $(2, 3, \ldots, e+1, 1, e+2, \ldots, n)$, and so $I(a_1, S_1) = e$.

(d) Choose $a \in \mathcal{A}_n$ and $S \in \mathcal{S}_n(a)$ such that $\text{minv}(n, e) = I(a, S)$ and $E(S) = e$. By Corollary 5.3,

$$\text{minv}(n, e) = I(a, S) = n(n-1)/2 - I(\overline{a}, \overline{S}) \geq n(n-1)/2 - \text{Minv}(n, C(S) - e).$$

However, $C(S) - e \leq n(n-1)/2 - e \leq n-1$, and by Theorem 8.1,

$$\text{Minv}(n, C(S) - e) \leq \left\lfloor \frac{n^2(C(S) - e)}{2(C(S) - e + 1)} \right\rfloor .$$

Hence,

$$\begin{aligned} \text{minv}(n, e) &\geq \frac{n(n-1)}{2} - \left\lfloor \frac{n^2(C(S) - e)}{2(C(S) - e + 1)} \right\rfloor \\ &\geq \frac{n(n-1)}{2} - \left\lfloor \frac{n^2\left\{\frac{n(n-1)}{2} - e\right\}}{2\left\{\frac{n(n-1)}{2} - e + 1\right\}} \right\rfloor , \end{aligned}$$

from which the result follows. □

10 Maximum number of smallest number of removals

It was proven in [12] that $\mathrm{Mrem}(n,1) = \lfloor n/2 \rfloor$ for $n \in \mathbb{N}\backslash\{0,1\}$. The following theorem gives some more general results regarding $\mathrm{Mrem}(n,e)$.

Theorem 10.1 *Let $n \in \mathbb{N}\backslash\{0,1\}$ and e be an integer with $0 \le e \le n(n-1)/2$. Then:*

(a) For $0 \le e \le n-1$,

$$e\left\lfloor \frac{n}{e+1} \right\rfloor \le \mathrm{Mrem}(n,e) \le \left\lfloor \frac{ne}{e+1} \right\rfloor .$$

(b) If either $n \equiv 0(\mathrm{mod}\ e+1)$ or $n \equiv 1(\mathrm{mod}\ e+1)$, and $0 \le e \le n-2$, then

$$\mathrm{Mrem}(n,e) = \left\lfloor \frac{ne}{e+1} \right\rfloor .$$

(c) For $n-1 \le e \le n(n-1)/2$, $\mathrm{Mrem}(n,e) = n-1$.

Proof: (a) The right inequality follows from Lemma 2.5 and Theorem 6.1(a). To prove the left inequality, let a and $S \in \mathcal{S}_n(a)$ be defined as in the proof of the second part of part (a) of Theorem 6.1. Since the output of the operation of S on a is again a (see equation (4)), it is easy to see that $RM(a,S) = ek = e\lfloor n/(e+1) \rfloor$, which proves the left inequality.

(b) It follows from part (a) of this theorem.

(c) It follows from Theorem 6.1(b). □

11 Minimum number of smallest number of removals

The following theorem gives some results regarding $\mathrm{mrem}(n,e)$.

Theorem 11.1 *Let $n \in \mathbb{N}\backslash\{0,1\}$ and e be an integer with $0 \le e \le n(n-1)/2$. Then:*

(a) $\mathrm{mrems}(n,0) = 0$.

(b) For $e \ge 1$, $\mathrm{mrem}(n,e) \ge 1$.

(c) For $1 \le e \le n-1$, $\mathrm{mrem}(n,e) = 1$.

Proof: (a) The proof of the first part of the theorem is obvious.

(b) Let $e \ge 1$. Choose $a \in \mathcal{A}_n$ and $S \in \mathcal{S}_n(a)$ such that $E(S) = e$ and $RM(a,S) = \mathrm{mrem}(n,e)$. By Theorem 7.1(b), $R(a,S) \ge \mathrm{mruns}(n,e) \ge 2$, i.e., the output of the operation of S on a contains at least two runs, and so $\mathrm{mrem}(n,e) = RM(a,S) \ge 1$.

(c) Assume $1 \le e \le n-1$. By part (b) of this theorem, $\mathrm{mruns}(n,e) \ge 1$. To show equality, let $a_1 = (1,2,\ldots,n)$ and assume $S_1 \in \mathcal{S}_n$ is the same as in the proof of part (c) in Theorem 7.1. Then $E(S_1) = e$, and the output of the operation of S_1 on a is $(2,3,\ldots,e+1,1,e+2,\ldots,n)$, which implies that $RM(a_1,S_1) = 1$. □

12 Maximum number of smallest number of successive exchanges

The following theorem gives some inequalities regarding $\mathrm{Mexc}(n, e)$.

Theorem 12.1 *Let $n \in \mathbb{N} \backslash \{0, 1\}$ and e be an integer with $0 \leq e \leq n(n-1)/2$. Then:*

(a) $\mathrm{Mexc}(n, 0) = 0$.

(b) If $1 \leq e \leq n-1$, then $\mathrm{Mexc}(n, e) = n - 1$.

(c) If $e = n(n-1)/2$, then $\mathrm{Mexc}(n, e) = \lfloor n/2 \rfloor$.

Proof: (a) The proof of this part is obvious.

(b) Assume $1 \leq e \leq n-1$, and let $a = (1, 2, \ldots, e, n, e+1, \ldots, n-1)$. (If $e = n-1$, then n is the last element of a.) Let $S \in \mathcal{S}_n(a)$ be the execution of straight insertion sort that errs in all the comparisons of pass e (only). Then the output of the operation of S on a is $(n, 1, 2, \ldots, n-1)$, which has only one cycle. Thus $\mathrm{Mexc}(n, e) = n - 1$.

(c) Define $A(n, e)$ as in the statement of Lemma 4.2. For each $(a, S) \in A(n, e)$ with $e = n(n-1)/2$, we have $E(S) = n(n-1)/2 = C(S)$, and by Lemma 4.1, the ranks of the output of the application of S on a is $(n, n-1, \ldots, 2, 1)$. This means $\mathrm{Mexc}(n, n(n-1)/2) = \lfloor n/2 \rfloor$. □

13 Minimum number of smallest number of successive exchanges

The following theorem gives some inequalities regarding $\mathrm{mexc}(n, e)$.

Theorem 13.1 *Let $n \in \mathbb{N} \backslash \{0, 1\}$ and e be an integer with $0 \leq e \leq n(n-1)/2$. Then:*

(a) $\mathrm{mexc}(n, 0) = 0$.

(b) $\mathrm{mexc}(n+1, e) \leq \mathrm{mexc}(n, e)$.

(c) If $e \geq 1$, then $\mathrm{mexc}(n, e) \geq 1$.

(d) If $1 + 8e$ is a perfect square (i.e., $e = m(m+1)/2$ for some non-negative integer m) and $n \geq \frac{1+\sqrt{1+8e}}{2}$, then

$$\mathrm{mexc}(n, e) \leq \left\lfloor \frac{1 + \sqrt{1 + 8e}}{4} \right\rfloor.$$

(e) If $1 \leq e \leq n-1$, then $\mathrm{mexc}(n, e) \leq \lfloor (e+1)/2 \rfloor$.

(f) $\mathrm{mexc}(n, 1) = 1$ *for $n \geq 2$, and* $\mathrm{mexc}(n, 2) = \mathrm{mexc}(n, 3) = 1$ *for $n \geq 3$.*

Proof: (a) The proof of this part is obvious.

(b) Choose $a \in \mathcal{A}_n$ and $S \in \mathcal{S}_n(a)$ such that $E(S) = e$ and $\mathrm{mexc}(n, e) = EX(a, S)$. Without loss of generality we may assume that a is a permutation of the first n positive integers. Let b be the output sequence of the operation of S on a. Let $a' = (a, n+1)$, and let $S' \in \mathcal{S}_{n+1}(a')$ be the execution of straight insertion sort operating on a' that consists of S followed by a single error-free comparison (in the last pass). Then $E(S') = E(S) = e$, and the output sequence is $(b, n+1)$. It is easy then to check that the number of cycles in $(b, n+1)$ is the number of cycles in b plus one. Therefore,

$$\text{mexc}(n+1, e) \leq EX(a', S') = EX(a, S) = \text{mexc}(n, e).$$

(c) Assume $e \geq 1$. Choose $a \in \mathcal{A}_n$ and $S \in \mathcal{S}_n(a)$ with $E(S) = e$ and $EX(a, S) = \text{mexc}(n, e)$. By Theorem 7.1, part (b), $R(a, S) \geq \text{mruns}(n, e) \geq 2$, and so $a \neq (1, 2, \ldots, n)$, which means $\text{mexc}(n, e) = EX(a, S) \geq 1$.

(d) It follows from part (b) of this theorem and Theorem 12.1(c).

(e) Assume $1 \leq e \leq n-1$, and let $a = (e+1, e, e-1, \ldots, 1) \in \mathcal{A}_{e+1}$ and $S \in \mathcal{S}_{e+1}(a)$ be the execution of straight insertion sort that errs in the first (and only) comparison of each of the e passes. Then $E(S) = e$, and the output of the operation of S on a is again a, and so $\text{mexc}(e+1, e) \leq EX(a, S) = \lfloor (e+1)/2 \rfloor$. Since $n \geq e+1$, it follows from part (b) of this theorem that $\text{mexc}(n, e) \leq \text{mexc}(e+1, e) \leq \lfloor (e+1)/2 \rfloor$.

(f) Equality $\text{mexc}(n, 1) = 1$ for $n \geq 2$ follows from parts (c) and (d) of this theorem. By part (c), $\text{mexc}(n, 2) \geq 1$ for $n \geq 3$. By part (e) of this theorem $\text{mexc}(n, 2) \leq \lfloor (2+1)/2 \rfloor = 1$ for $n \geq 3$, and so $\text{mexc}(n, 2) = 1$. By part (c), $\text{mexc}(n, 3) \geq 1$ for $n \geq 3$. By part (d) of this theorem, $\text{mexc}(n, 3) \leq \lfloor 6/4 \rfloor = 1$ for $n \geq (1 + \sqrt{1 + 8(3)})/2 = 3$. Thus, $\text{mexc}(n, 3) = 1$. □

14 Maximum value of the sum of squares

The following theorem gives some results regarding $\text{Msqr}(n, e)$.

Theorem 14.1 *Let $n \in \mathbb{N} \backslash \{0, 1\}$ and e be an integer with $0 \leq e \leq n(n-1)/2$. Then:*

(a) $\text{Msqr}(n, 0) = 0$.

(b) $\text{Msqr}(n, 1) = n \lfloor n/2 \rfloor (n - \lfloor n/2 \rfloor) = n \lfloor n^2/4 \rfloor$.

(c) $\text{Msqr}(n, e) \geq n \left\lfloor \frac{n^2}{4} \right\rfloor + 2$ *for $e \in \{2, 3\}$ and $n \geq 5$.*

(d) If $0 \leq e \leq n-1$, then

$$\text{Msqr}(n, e) \geq \frac{n}{6(e+1)} [n^2(2e-1) - 3n(e^2-1) - 2(e+1)].$$

(e) For $n - 1 \leq e \leq n(n-1)/2$, $\text{Msqr}(n, e) = (n^3 - n)/3$.

Proof: (a) The proof of the first equality is trivial.

(b) Choose (a, S) such that $a \in \mathcal{A}_n$, $S \in \mathcal{S}_n(a)$, $E(S) = 1$, and $SQ(a, S) = \mathrm{Msqr}(n, 1)$. From Lemma 2.2, part (a), we have $SQ(a, S) \leq nI(a, S)$ (an inequality due to Daniels [4]). From Theorem 8.1 we have $I(a, S) \leq \mathrm{Minv}(n, 1) \leq \lfloor n^2/4 \rfloor$. Therefore

$$\mathrm{Msqr}(n, 1) = SQ(a, S) \leq n \left\lfloor \frac{n^2}{4} \right\rfloor .$$

To prove equality, let $m = \lfloor n/2 \rfloor$, and let $a = (n - m + 1, n - m + 2, \ldots, n, 1, 2, \ldots, n - m)$ if $n > 2$, and $a = (2, 1)$ if $n = 2$. Let $S \in \mathcal{S}_n(a)$ be the execution of straight insertion sort that errs in the first comparison of pass m. Then the output of the operation of S on a is again a, and then it is easy to show that $SQ(a, S) = n\lfloor n/2 \rfloor (n - \lfloor n/2 \rfloor)$, thus proving part (b) of the theorem.

(c) To prove the inequality for $e = 2$, assume $n \geq 5$, and let $m = \lfloor n/2 \rfloor$ and $a = (n - m + 1, n - m + 2, \ldots, n, 2, 1, 3, \ldots, n - m)$. Let $S \in \mathcal{S}_n(a)$ be the execution of straight insertion sort that errs only in the first comparison of passes m and $m + 1$. Then the output of the operation of S on a is again a, and then it is easy to show that $SQ(a, S) = n\lfloor n^2/4 \rfloor + 2$.

To prove the inequality for $e = 3$, assume $n \geq 5$, and let $a = (n - m + 1, n - m + 2, \ldots, n, 1, 2, 3, \ldots, n - m)$. Let $S \in \mathcal{S}_n(a)$ be the execution of straight insertion sort that errs only in the first comparison of pass m and the first two comparisons of pass $m + 1$. Then the output of the operation of S on a is

$$(n - m + 1, n - m + 2, \ldots, n, 2, 1, 3, \ldots, n - m).$$

It is then easy to show that $SQ(a, S) = n\lfloor n^2/4 \rfloor + 2$.

(d) Assume $0 \leq e \leq n - 1$. Choose (a, S) such that $a \in \mathcal{A}_n$, $S \in \mathcal{S}_n(a)$, $E(S) = e$, and $I(a, S) = \mathrm{Minv}(n, e)$. By Daniels' inequality, Lemma 2.2(a),

$$\mathrm{Msqr}(n, e) \geq SQ(a, S) \geq nI(a, S) - \frac{n(n-1)(n-2)}{6}. \tag{5}$$

By Theorem 8.1(a),

$$\begin{aligned} I(a, S) &= \mathrm{Minv}(n, e) \geq \left\lfloor \frac{n}{e+1} \right\rfloor \left\{ ne - \left\lfloor \frac{n}{e+1} \right\rfloor \frac{e(e+1)}{2} \right\} \\ &\geq \left(\frac{n}{e+1} - 1 \right) \frac{ne}{2}. \end{aligned} \tag{6}$$

Combining inequalities (5) and (6), we can prove the inequality in part (e) of the theorem.

(e) It follows from Theorem 6.1(b). □

15 Minimum value of the sum of squares

The following theorem gives some results regarding $\text{msqr}(n, e)$. (The notation $\lceil x \rceil$ denotes the smallest integer greater than or equal to x.)

Theorem 15.1 *Let $n \in \mathbb{N}\backslash\{0,1\}$ and e be an integer with $0 \le e \le n(n-1)/2$. Then:*

(a) $\text{msqr}(n, 0) = 0$.

(b) $\text{msqr}(n, e) \ge \lceil \frac{4}{3} e \left(1 + \frac{e}{n}\right) \rceil$.

(c) If $n \ge 2e$, then $\text{msqr}(n, e) \le 2e$.

(d) If $0 \le e \le n-1$, then $\text{msqr}(n, e) \le e(e+1)$.

(e) $\text{msqr}(n, 1) = 2$ *and* $\text{msqr}(n, 2) = 4$. *(For the last equality we need to assume that $n \ge 4$ because* $\text{msqr}(3, 2) = 6$.*)*

Proof: (a) The proof of this part is obvious.

(b) It follows from the Durbin-Stuart inequality (see Lemma 2.2(b)) and Theorem 9.1(b).

(c) If $e = 0$, the inequality is obviously true (for $n \ge 2$). Assume $n \ge 2e \ge 2$, and $a = (1, 2, \ldots, n)$ and assume $S \in \mathcal{S}_n(a)$ errs in the first comparisons of passes $1, 3, \ldots, 2e-1$ only. Then $E(S) = e$ and the output of the operation of S on a is $(2, 1, 4, 3, \ldots, 2e, 2e-1, 2e+1, \ldots, n)$. Hence $\text{msqr}(n, e) \le SQ(a, S) = 2e$.

(d) For $e = 0$ the inequality is obvious. Assume $1 \le e \le n-1$. Let $a_1 = (1, 2, \ldots, n)$ and assume $S_1 \in \mathcal{S}_n(a_1)$ is the same as in the proof of part (c) in Theorem 7.1. Then $E(S_1) = e$, and the output of the operation of S_1 on a is $(2, 3, \ldots, e+1, 1, e+2, \ldots, n)$, which implies that $\text{msqr}(n, e) \le SQ(a_1, S_1) = e(e+1)$.

(e) The equality $\text{msqr}(n, 1) = 2$ follows from parts (b) and (c) of this theorem. The inequality $\text{msqr}(n, 2) \le 4$ for $n \ge 4$ follows from part (c) of this theorem. By part (b), $\text{msqr}(n, 2) \ge 3$. To prove $\text{msqr}(n, 2) = 4$ for $n \ge 4$, we show that it is impossible for $a \in \mathcal{A}_n$ and $S \in \mathcal{S}_n(a)$ to have $E(S) = 2$ and $SQ(a, S) = 3$. Assume otherwise. Assume without loss of generality that a is a permutation of $(1, 2, \ldots, n)$. Write $a = (a_1, a_2, \ldots, a_n)$. Since $SQ(a, S) = 3$, there are integers i, j, k such that $1 \le i < j < k \le n$, and (a_i, a_j, a_k) is a permutation of (i, j, k) with $a_s \ne s$ and $(s - a_s)^2 = 1$ for $s = i, j, k$. It is very easy to show that this is impossible. □

16 Maximum value of the sum of absolute values

In this section we give some results regarding $\text{Mabs}(n, e)$.

Theorem 16.1 *Let $n \in \mathbb{N}\backslash\{0,1\}$ and e be an integer with $0 \le e \le n(n-1)/2$. Then:*

(a) $\mathrm{Mabs}(n,0) = 0$.
(b) $\mathrm{Mabs}(n,e) = \lfloor n^2/2 \rfloor$ *if* $e = 1$ *or* 2.
(c) *If* $n \geq 3$ *and* $3 \leq e \leq \frac{\lfloor \frac{n}{2} \rfloor (\lfloor \frac{n}{2} \rfloor - 1)}{2} + 3$, *then* $\mathrm{Mabs}(n,e) = \lfloor n^2/2 \rfloor$.
(d) *If* $n - 1 \leq e \leq n(n-1)/2$, *then* $\mathrm{Mabs}(n,e) = \lfloor n^2/2 \rfloor$.

Proof: (a) The proof of this part is obvious.

(b) It is clear that $\mathrm{Mabs}(n,e) \leq \lfloor n^2/2 \rfloor$, because the maximum value of $D(a,S)$ for $a \in \mathcal{A}_n$ and $S \in \mathcal{S}_n(a)$ is $\lfloor n^2/2 \rfloor$. To prove equality when $e = 1$, let $m = \lfloor n/2 \rfloor$, and let $a = (n-m+1, n-m+2, \ldots, n, 1, 2, \ldots, n-m)$ if $n > 2$, and $a = (2,1)$ if $n = 2$. Let $S \in \mathcal{S}_n(a)$ be the execution of straight insertion sort that errs in the first comparison of pass m. Then the output of the operation of S on a is again a, and then $D(a,S) = \lfloor n^2/2 \rfloor$.

To prove equality for $e = 2$, assume $n \geq 3$, and let $a = (n-m+1, n-m+2, \ldots, n, 2, 1, 3, \ldots, n-m)$ if $n > 4$, $a = (3,4,2,1)$ if $n = 4$, and $a = (3,2,1)$ if $n = 3$. Let $S \in \mathcal{S}_n(a)$ be the execution of straight insertion sort that errs only in the first comparison of passes m and $m+1$. Then the output of the operation of S on a is again a, and then $D(a,S) = \lfloor n^2/2 \rfloor$.

(c) As in part (b), $\mathrm{Mabs}(n,e) \leq \lfloor n^2/2 \rfloor$. To prove equality, let $m = \lfloor n/2 \rfloor$. Since $0 \leq e-3 \leq m(m-1)/2$, by Lemma 4.2 there is $a_1 \in \mathcal{A}_m$ and $S_1 \in \mathcal{S}_m(a_1)$ such that $E(S_1) = e-3$. Define

$$a = (a_{11} + n - m, a_{12} + n - m, \ldots, a_{1m} + n - m, 1, 2, \ldots, n-m).$$

Note that $\mathrm{ranks}(a_1) = \mathrm{ranks}(a_{11} + n - m, a_{12} + n - m, \ldots, a_{1m} + n - m)$. Let $S \in \mathcal{S}_n(a)$ be the execution of straight insertion sort that consists of S_1 during its first $m-1$ passes; followed by a single erroneous comparison in pass m; two erroneous comparisons in pass $m+1$, and no erroneous comparisons thereafter.

Note that $E(S) = (e-3) + 1 + 2 = e$. The output of the operation of S on a is a permutation of the numbers $n-m+1, n-m+2, \ldots, n$ followed by the sequence $(2,1,3,\ldots,n-m)$ if $n \geq 5$, and $(2,1)$ if $n = 3$ or 4. By Lemma 2.4, $D(a,S) = \lfloor n^2/2 \rfloor$, and part (c) of the theorem has been proven.

(d) It follows from Theorem 6.1(b). □

17 Minimum value of the sum of absolute values

In this section we give some results regarding $\mathrm{mabs}(n,e)$.

Theorem 17.1 *Let* $n \in I\!N \backslash \{0,1\}$ *and* e *be an integer with* $0 \leq e \leq n(n-1)/2$. *Then:*
(a) $\mathrm{mabs}(n,0) = 0$.
(b) *If* $e \geq 1$, *then* $\mathrm{mabs}(n,e) \geq e + 1$.

(c) If $0 \le e \le n-1$, *then* $\text{mabs}(n,e) \le 2e$.
(d) $\text{mabs}(n,1) = 2$.

Proof: (a) The proof of the first part is obvious.

(b) It follows from the Diaconis-Graham inequality (Lemma 2.2, part (c)), Theorem 9.1(b), and Theorem 13.1(c).

(c) The proof of this part is similar to the proof of Theorem 15.1(d), or to the proof of Theorem 7.1(c).

(d) It follows from parts (b) and (c) of this theorem. □

The following example shows that when $n = 4$ and $e = 3$, we have $4 = e + 1 = \text{mabs}(4,3) < 2e = 6$. Let $a = (1,2,3,4)$ and let $S \in \mathcal{S}_4(a)$ be the execution of straight insertion sort that errs in all the comparisons of passes 1 and 2 (but not in pass 3). Then the output of the operation of S on a is $(3,2,1,4)$, and so $\text{mabs}(4,3) \le D(a,S) = 4$. By Theorem 17.1(b), $\text{mabs}(4,3) \ge 3+1 = 4$. This proves our claim.

18 Concluding remarks and future research

From Theorems 6.1, 8.1, 10.1, 12.1, 14.1, and 16.1 of the paper, we can easily deduce the following corollary about the asymptotic behaviour of straight insertion sort when the number of errors is small compared to the length of the input list:

Corollary 18.1 *Let e be a fixed nonnegative integer. Then*
(a) $\lim_{n\to\infty} \text{Mruns}(n,e) = e+1$;
(b) $\lim_{n\to\infty} \frac{\text{Minv}(n,e)}{n^2} = \frac{e}{2(e+1)}$;
(c) $\lim_{n\to\infty} \frac{\text{Mrem}(n,e)}{n} = \frac{e}{e+1}$.
(d) $\lim_{n\to\infty} \frac{\text{Mexc}(n,e)}{n} = 1$ *(if* $e \ge 1$*).*
(e) $\limsup_{n\to\infty} \frac{\text{Msqr}(n,e)}{n^3} \ge \frac{2e-1}{6(e+1)}$ *(if* $e \ge 1$*).*
(f) $\lim_{n\to\infty} \frac{\text{Mabs}(n,e)}{n^2} = \frac{1}{2}$ *(if* $e \ge 1$*).*

Table 1 gives an asymptotic comparison (i.e., for large n) of the worst case scenario for bubble sort (with no boolean flag) with the worst case for straight insertion sort for a fixed number of errors $e \ge 1$. The six measures of disarray are compared. For bubble sort, the results for Mruns, Minv, and Mrem have been proven in Hadjicostas and Lakshmanan (2005), whereas the results for Mexc, Msqr, and Mabs are proven in Appendix A of this paper.

Table 2 gives an asymptotic comparison of the best case scenaria for bubble sort and straight insertion sort for a fixed number of errors $e \ge 1$. The six measures of disarray are compared. The results for bubble sort follow from Theorem A.3 in Appendix A.

Table 1: Worst case scenario for bubble and straight insertion sort (large n with fixed $e \geq 1$)

Algorithm	Mruns	Minv	Mrem
Bubble sort	$e+1$	$ne + O(1)$	e
Straight insertion sort	$e+1$	$\frac{n^2 e}{2(e+1)} + O(n)$	$\frac{ne}{e+1} + O(1)$
Algorithm	**Mexc**	**Msqr**	**Mabs**
Bubble sort	$n-1$	$n^2 e + O(n)$	$2en + O(1)$
Straight insertion sort	$n-1$	$\geq \frac{n^3(2e-1)}{6(e+1)} + O(n^2)$	$\lfloor \frac{n^2}{2} \rfloor$

Table 2: Best case scenario for bubble and straight insertion sort (large n with fixed $e \geq 1$)

Algorithm	mruns	minv	mrem	mexc	msqr	mabs
Bubble sort	1	0	0	0	0	0
Straight insertion sort	2	e	1	≥ 1	$\geq \lfloor \frac{4e}{3} \rfloor$	$\geq e+1$

Tables 1 and 2 indicate that bubble sort is perhaps more robust to errors than straight insertion sort for sufficiently large n, but very small e (which is not a function of n). However, bubble sort is known to be inefficient requiring a lot more comparisons even for a reasonably sorted input, i.e., bubble sort lacks adaptability to pre-sortedness. It is perhaps this redundancy in the number of comparisons performed that manifests itself in lower disorder due to errors in the output. The anonymous referee for this paper has suggested that it will be interesting to propose and study a measure of robustness of sorting algorithms taking into account the extent of disarray left in the output per error relative to the number of comparisons performed. We leave this direction of study for future work.

A Appendix

In this appendix, we state and prove some more results related to bubble sort (when some comparisons are erroneous) that did not appear in Hadjicostas and Lakshmanan [11]. We are dealing with bubble sort with no boolean flag:

for $(i = 1; i < n; i = i + 1)$

```
for (j = n; j > i; j = j - 1)
    if (a_{j-1} > a_j)
        swap a_{j-1} and a_j
```

During bubble sort, if there are no errors in comparisons, at the beginning of the i^{th} pass, the smallest $i-1$ integers occupy their correct positions. During the i^{th} pass over the input list a, the algorithm "bubbles" the i^{th} smallest element in a to the i^{th} position (from the left) in the output sequence. The algorithm makes $n-1$ passes over the input list, with $n-i$ comparisons during the i^{th} pass. Hence it does exactly $n(n-1)/2$ comparisons in total.

For each $n \in \mathbb{N}\backslash\{0,1\}$, let $\mathcal{B}_n$ be the set of all executions of the bubble sort algorithm, $B : \mathcal{A}_n \to \mathcal{A}_n$, that can sort lists of length n, and can make up to $n(n-1)/2$ errors when making comparisons. This means that, for each $B \in \mathcal{B}_n$, the collection of comparisons where B is erring is uniquely associated with B. Let also $E(B)$ be the total number of errors B does.

For each $n \in \mathbb{N}\backslash\{0,1\}$, $a \in \mathcal{A}_n$, and $B \in \mathcal{B}_n$, define $R(a,B)$, $I(a,B)$, $RM(a,B)$, $EX(a,B)$, $SQ(a,B)$, and $D(a,B)$ in a way similar to that in Section 4 of this paper. Similarly, for integers e with $0 \leq e \leq n(n-1)/2$, one can define $\text{Mexc}(n,e)$, $\text{Msqr}(n,e)$, $\text{Mabs}(n,e)$, $\text{mrem}(n,e)$, $\text{mexc}(n,e)$ etc. For example, $\text{Msqr}(n,e) = \max\{SQ(a,B) : a \in \mathcal{A}_n,\ B \in \mathcal{B}_n,\ E(B) = e\}$.

Theorem A.1 *Let $n \in \mathbb{N}\backslash\{0,1\}$. Then for bubble sort:*
(a) $\text{Mexc}(n,0) = 0$.
(b) If e is an integer with $1 \leq e \leq n-1$, then $\text{Mexc}(n,e) = n-1$.

Proof: (a) The proof of this part is obvious.

(b) Assume $1 \leq e \leq n-1$, and let $a = (1,2,\ldots,e-1,n,e,\ldots,n-1)$. (If $e=1$, then n is the first element of a.) Let $B \in \mathcal{B}_n$ be the execution of bubble sort that errs only in the last e comparisons of the first pass. Then the output of the operation of B on a is $(n,1,2,\ldots,n-1)$, which has only one cycle. Thus $\text{Mexc}(n,e) = n-1$. □

Theorem A.2 *Let $n \in \mathbb{N}\backslash\{0,1\}$ and $e \in \mathbb{N}$ be such that $e \leq n(n-1)/2$. Then for bubble sort,*
(a) $\max\{ne(n-e),0\} \leq \text{Msqr}(n,e) \leq n(en-2e+1)$.
(b) $\max\{2e(n-e),0\} \leq \text{Mabs}(n,e) \leq 2(en-2e+1)$.

Proof: (i) By Theorem 6.1, part (c), in Hadjicostas and Lakshmanan (2005), we know that for bubble sort, $\text{Minv}(n,e) \leq e(n-2)+1$. For all $a \in \mathcal{A}_n$ and $B \in \mathcal{B}_n$ let $I(a,B)$ be the number of inversions in the output sequence after execution B of bubble sort operates on a. By Lemma 2.2, parts (a) and (c), of this paper we have $SQ(a,B) \leq n\,I(a,B) \leq n\,\text{Minv}(n,e)$

and $D(a,B) \leq 2\,I(a,B) \leq 2\,\mathrm{Minv}(n,e)$ for all $a \in \mathcal{A}_n$ and $B \in \mathcal{B}_n$ with $E(B) = e$. Thus

$$\mathrm{Msqr}(n,e) \leq n\,\mathrm{Minv}(n,e) \leq n(en - 2e + 1),$$

and

$$\mathrm{Mabs}(n,e) \leq 2\,\mathrm{Minv}(n,e) \leq 2(en - 2e + 1).$$

(ii) The left inequalities obviously hold for $e = 0$ or $e \geq n$. To prove them for the case $1 \leq e \leq n-1$, let $a_0 = (n, n-1, \ldots, n-e+1, 1, 2, \ldots, n-e)$, and B_0 be the execution of bubble sort that errs in comparison $n - e$ of pass i for $i = 1, 2, \ldots, e$. Then the output list is $(n - e + 1, \ldots, n - 1, n, 1, 2, \ldots, n - e)$, and so $\mathrm{Msqr}(n,e) \geq SQ(a_0, B_0) = n(n-e)e$ and $\mathrm{Mabs}(n,e) \geq D(a_0, B_0) = 2e(n-e)$. This concludes the proof of the theorem. □

The following theorem deals with the best case scenarios for all six measures of disarray in the case of bubble sort with erroneous comparisons.

Theorem A.3 *Assume $n \in \mathbb{N}\backslash\{0,1\}$ and e is an integer with $0 \leq e \leq n-2$. Then for bubble sort,* $\mathrm{mruns}(n,e) = 1$, *and* $\mathrm{minv}(n,e) = \mathrm{mrem}(n,e) = \mathrm{mexc}(n,e) = \mathrm{msqr}(n,e) = \mathrm{mabs}(n,e) = 0$.

Proof: Let $a = (1, 2, \ldots, n)$ and assume $B \in \mathcal{B}_n$ is such that all errors occur in the first $n - 2$ comparisons of the first pass. (This is possible since $e \leq n - 2$.) At the end of the first pass, number 1 has been placed in position 1. Since there are no further errors, the output of the execution of B on a is again a, and the theorem follows immediately. □

B Appendix

In this appendix we prove an auxiliary result about insertion sort when all the comparisons are error-free. Recall that $A(n,e)$ is defined by equation (2) in the statement of Lemma 4.2. This lemma is used in the proof of Theorem 6.1(b).

Lemma B.1 *Let $n \in \mathbb{N}\backslash\{0,1\}$. Given an integer c with $n - 1 \leq c \leq n(n-1)/2$, there is $(a, S) \in A(n, e = 0)$ such that $C(S) = c$.*

Proof: We proceed by finite induction on n. For $n = 2$, the claim of the lemma is obviously true. Let m be a positive integer such that $m \geq 3$, and assume the claim of the lemma is true for all integers n such that $2 \leq n < m$.

Let c be an integer such that $m-1 \leq c \leq m(m-1)/2$. We consider two cases: (a) $m-1 \leq c \leq (m-1)(m-2)/2+1$; and (b) $(m-1)(m-2)/2+2 \leq c \leq m(m-1)/2$.

(a) Assume $m-1 \leq c \leq (m-1)(m-2)/2+1$. Then $m-2 \leq c-1 \leq (m-1)(m-2)/2$. By assumption, we may choose $(a, S) \in A(m-1, 0)$ such that $C(S) = c-1$. Then $(\text{ranks}(a), S) \in A(m-1, 0)$. It follows that ranks($a$) is a permutation of the first $m-1$ positive integers and $S \in \mathcal{S}_{m-1}(a)$. Let b be the output list after S operates on ranks(a). Let $a' = (\text{ranks}(a), m)$ and define $S' \in \mathcal{S}_m(a')$ to be S followed by a comparison of m with the last integer of b with no error occuring. Then $E(S') = E(S) = 0$, $C(S') = C(S)+1 = c$, and $(a', S') \in A(m, 0)$.

(b) Assume $(m-1)(m-2)/2+2 \leq c \leq m(m-1)/2$. Then

$$m-2 \leq (m^2-5m+8)/2 \leq c-(m-1) \leq (m-1)(m-2)/2.$$

By assumption, we may choose $(a_1, S_1) \in A(m-1, 0)$ such that $C(S_1) = c-(m-1)$. Then $(\text{ranks}(a_1), S_1) \in A(m-1, 0)$. It follows that ranks($a_1$) is a permutation of the first $m-1$ positive integers and $S_1 \in \mathcal{S}_{m-1}(a_1)$. For each integer i with $1 \leq i \leq m-2$, let b_i be the output obtained after pass i of S_1 operating on ranks(a_1). Let $a_1' = (m, \text{ranks}(a_1))$, and let $S_1' \in \mathcal{S}_m(a_1')$ with $E(S_1') = 0$ be defined as follows: In the single comparison of the first pass of S_1', the first element of ranks(a_1) (which is the second element of a_1') is compared to m, and m goes to the second position. For $2 \leq j \leq m-1$, we assume that the output of the operation of S_1' on a_1' after pass j is the list obtained after inserting m on the $(j+1)^{\text{th}}$ position of b_{j-1}. It is easy to show that S_1' is well-defined, $(a_1', S_1') \in A(m, 0)$, and $C(S') = c$.

Puting the two cases together, we conclude that the claim of the lemma is true for $n = m$. □

Acknowledgments

We would like to thank an anonymous scientist for simplifying the proof of Theorem 7.1(b), and for suggesting the statement and the proof of Theorem 6.1(b). We would also like to thank an anonymous referee for helping us improve the presentation of Section 18 of the paper.

References

[1] L. Alonso, P. Chassaing, F. Gillet, S. Janson, E.M. Reingold, and R. Schott (2004). Quicksort with unreliable comparisons: a probabilistic analysis. *Combin. Probab. Comput.*, 13, 419-449.

[2] S. Baase and A. van Gelder, Computer algorithms: Introduction to design and analysis, 3rd edition, Addison-Wesley, Reading, MA, 2000.

[3] A. Bagchi (1992). On sorting in the presence of erroneous information. *Inform. Process. Lett.*, 43, 213–215.

[4] H.E. Daniels (1950). Rank correlation and population models. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 12, 171-191.

[5] P. Diaconis and R.L. Graham (1977). Spearman's footrule as a measure of disarray. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 39, 262-268.

[6] J. Durbin and A. Stuart (1951). Inversions and rank correlation coefficients. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 13, 303-309.

[7] V. Estivill-Castro (1991). Sorting and measures of disorder, *Ph.D. thesis.* University of Waterloo, Waterloo, Ontario, Canada.

[8] V. Estivill-Castro (2004). Generating nearly sorted sequences – The use of measures of disorder. *Electronic Notes in Theoretical Computer Science*, 91 , 56-95.

[9] V. Estivill-Castro, H. Mannila, and D. Wood (1993). Right invariant metrics and measures of presortedness. *Discrete Appl. Math.*, 42, 1-16.

[10] V. Estivill-Castro and D. Wood (1992). A survey of adaptive sorting algorithms. *ACM Computing Surveys*, 24, 441–476.

[11] P. Hadjicostas and K.B. Lakshmanan (2005). Bubble sort with erroneous comparisons. *Australas. J. Combin.*, 31, 85-106.

[12] T. Islam and K. B. Lakshmanan (1990). On the error-sensitivity of sort algorithms, in: S.G. Akl, F. Fiala, W.W. Koczkodaj (Eds.), Advances in Computing and Information, Canadian Scholars' Press, Toronto, Canada, 81-85.

[13] M. Kendall, *Rank Correlation Methods,* 4th edition (2nd impression), Oxford University Press, New York, NY, 1975.

[14] D.E. Knuth, *The art of computer programming, Vol. 3: Sorting and Searching,* 2nd edition, Addison-Wesley, Reading, MA, 1998.

[15] K.B. Lakshmanan, B. Ravikumar, and K. Ganesan (1991). Coping with erroneous information while sorting. *IEEE Trans. Comp.*, 40, 1081–1084.

[16] T. Leighton and Y. Ma (1997). Breaking the $\Theta(n \log^2 n)$ barrier for sorting with faults. *J. Computer and System Sci.*, 54, 265-304.

[17] T. Leighton and Y. Ma (1999). Tight bounds on the size of the fault-tolerant merging and sorting networks with destructive faults. *SIAM J. Comput.*, 29, 258-273.

[18] H. Mannila (1985). Measures of presortedness and optimal sorting algorithms. *IEEE Trans. Comp.*, 34, 318-325.

[19] A. Pelc (2002). Searching games with errors – Fifty years of coping with liars. *Theoret. Comput. Sci.*, 270, 71-109.

[20] A.C. Yao and F.F. Yao (1985). On fault-tolerant networks for sorting. *SIAM J. Computing*, 14, 120-128.