ORIGINAL RESEARCH

Matching points in compositions and words

Margaret Archibald · Aubrey Blecher · Arnold Knopfmacher

Abstract A matching point in compositions and words is an extension to these objects of the well-studied concept of fixed points in permutations. The equivalent of the derangement problem is solved here by providing a formula for the number of compositions of n having no matching points, and showing that the number of words with no matching points tends to zero as (the length of the word) n tends to infinity. We also find formulae for the average number of matching points in both words and compositions.

Keywords Generating function · Fixed point · Derangement · Composition

Mathematics Subject Classification Primary: 05A15 · 05A16; Secondary: 05A05

1 Introduction

In this paper we consider matching points (defined below) in both compositions of n and words over an alphabet $[k]$. "Matching points" is often abbreviated to "matchings".

Definitions and examples:

(1) A *composition* of a positive integer n is a representation of n as an ordered sum of positive integers $n = a_1 + a_2 + \cdots + a_m$ where a_i is called the *i*th part of the composition (see e.g. [5]).
To avoid confusion, we clarify here that we refer to the *i*th value in a compositions as the *i*th part whereas for a general integer sequence the word "part" is avoided. So if a composition of n is understood as a sequence *i*th "part" and *i*th "value" are interchangeable.

(2) A *word of length n over an alphabet* $[k]$ is an element of the set $[k]^n$ where $[k] = \{1, 2, \ldots, k\}$ (see e.g. [5]).

(3) The *i*-th value in a sequence of positive integers is a *matching point* if it is equal to the *i*-th value in this sequence after it has been re-ordered from smallest to largest.

Example 1 The sequence 1211515 (viewed either as a composition of 16 or a word of length 7) written in non-decreasing order is 1111255. It has four matching points (in positions 1, 3, 4 and 7).

The main results in this paper are as follows: For both compositions and words, we find formulae for the average number of matching points. These are given in Theorem 2 and Theorem 10 respectively. In the case

M. Archibald· A. Blecher (✉) · A. Knopfmacher
The John Knopfmacher Centre for Applicable Analysis and Number Theory School of Mathematics, University of the Witwatersrand, Private Bag 3 Wits, Johannesburg 2050, South Africa
E-mail: Margaret.Archibald@wits.ac.za

1	1	1	1	2	5	5
1	2	1	1	5	1	5

of compositions, we show in Theorem 9 that the proportion of these with no matching points tends to zero as $n \to \infty$. In the case of words, a similar result is proved in Theorem 12.

A different extension of fixed points in permutations (see, for instance [3,8,9]) to compositions and words is to define the i-th value in the word to be a fixed point if it has value i. The current authors have recently studied this concept in the paper [1]. A related concept called *alphabetic points* has recently been studied in [2].

For the convenience of the reader, here is a summary of the structure of the sections that follow. Section 2 is about compositions, and the main theorem, namely Theorem 2 is stated at the outset. The rest of section 2 constitutes the proof of this Theorem: Firstly, in subsections 2.1 and 2.2, we describe a construction for creating compositions of $n + 1$ from those of n. The proof works by examining the detail of the many cases of how this construction changes the number of matching points. Finally, all these cases are brought together to construct the generating function for the total number of matching points, $M(x)$, in Equation (6). The rest of this Section constitutes the proof of Theorem 9.

Section 3 deals with words. As already stated, the main results here are Theorems 10 and 12. Both of these are proved prior to the statement of the theorems.

2 Matching points in compositions of n

Inspired by the concept of fixed points in permutations (where the concept of fixed and matching points are identical), we consider matching points in compositions. A matching point occurs if the i-th part in the composition of n is equal to the i-th part in the corresponding partition of n (which is the original composition written in weakly-increasing order). Another way of explaining this which is more helpful for the following computations is that a matching point occurs if the i-th part in the composition is the i-th smallest part in the composition (including repetitions).

Theorem 2 *The average number of matchings in a composition of n is*

$$\frac{1}{18}\left(3n + (-1)^n 2^{3-n} + 19\right)$$

for $n > 0$.

The sequence corresponding to the formula in Theorem 2 is in Sloane's Online Encyclopedia of Integer Sequence [7] as sequence A175656.

We now prove this by enumerating the total number of matching points. To do this, we count the number of new matching points that are created when constructing all compositions of $n + 1$ from those of n. This can be done by applying the following two operations separately to every composition of n, namely

(1) put the element 1 at the beginning of an existing composition, e.g $2113 \to 12113$, or
(2) add 1 to the first part (i.e., element), e.g $2113 \to 3113$.

This construction creates each composition of $n + 1$ from those of n by exactly one of the two processes specified.

2.1 Putting the element 1 at the beginning of an existing composition.

For this process, there will be one additional matching point per composition as the first value will always be a matching point. Because there are precisely 2^{n-1} compositions of n, this results in a total of 2^{n-1} additional matching points over all compositions of n.

2.2 Adding 1 to the first part

In this process, five different scenarios may arise. These will be explained further in what follows:

  (i) Gain two matching points (not possible as explained later),
 (ii) Gain one matching point,
(iii) Neither gain nor lose any matching point (irrelevant to the calculation of the changes caused by adding 1 to the first part),
 (iv) Lose one matching point,
  (v) Lose two matching points.

To understand this, it helps to understand what happens when we add 1 to the first part of a composition (which we then refer to as an augmented composition). In what follows we reserve the letter k to represent the integer value in the first position of the original integer composition and we also reserve the letter i for the value explained below. Furthermore, we let l_k denote any positive integer values in the composition that are less than or equal to k, g_k denote any such values that are greater than or equal to k. Augmentation yields exactly the same composition as before in all positions except position one. Consider what happens when comparing these two sequences after re-ordering. The last value k for the re-ordering of the original sequence will be $k + 1$ for the re-ordering of the augmented sequence. Otherwise the two re-ordered sequences will be identical. The last value k in the re-ordered original sequence occurs in position i where the number of parts of type l_k in the original sequence is i. In this latter number of parts, parts that are repeated are counted once for each repeat. The positive integer i will be reserved to denote this for the rest of this section. The value in position i will change from a k to a $k + 1$ in the re-ordered augmented composition.

This implies that there are only two possible changes to the number of matching points in the augmented composition of $n + 1$: the value in position one (of size $k + 1$) and whether or not there is a value $k + 1$ in the ith position of the original composition before reordering.

Example 3 Consider the composition 2113315, which is re-ordered to 1112335. Here $k = 2$ and $i = 4$. By adding 1 to the first part (to make it a $k + 1$), the augmented composition is 3113315 which, when re-ordered becomes 1113335. We represent this idea in the table below, with the matchings underlined.

There are four matching points originally (in positions 2,3,5 and 7) and after augmentation there are five matching points (in positions 2,3,4,5 and 7). According to our analysis above, the additional candidate positions for matching points have to be in position 1 or 4. And indeed in this example, there is an additional matching point in position 4 which is a gain of one matching point resulting from adding 1 to the first part of the original composition.

	original		augmented
original	21133<u>15</u>	→	31133<u>15</u>
re-ordered:	1112335		1113335

Since the number of matching points is determined by comparing the top row to the bottom row (in the format as seen in the table of Example 3), and the only changes after augmentation are in positions one and four, these are the only possible positions where the value there may contribute to changing the number of matching points in the composition. In Example 3, there was no gain in position one, but there was a gain in position $i = 4$, since the value in this position changed from a 2 ($k = 2$) to a 3 after re-ordering.

Since augmentation can in principle only change matchings in two positions (namely position 1 and what we have called position i), there are five theoretical scenarios for how the number of matchings could change after augmenting the first part of the composition by 1, namely: gain two; gain one; gain none; lose one; lose two. However, gaining none does not contribute to the change in the number of matchings and is therefore not discussed further. Also, gaining two is not possible for the following reason: the only way to gain a matching point in position one is if the first part is not the minimal value in the composition before augmentation but is the minimal value afterwards. This is not possible.

So three of the original scenarios (the only ones which cause changes in the number of matching points) need to be characterised. This will be done in the next three subsections, each with its own characterising proposition(s).

2.2.1 Gain one matching

In this case the number of matching points in the composition of $n + 1$ will be one larger than that in the original composition of n.

Example: The number of matchings in the composition 2131 is one and the number of matchings in the composition 3131 is two.

$$
\begin{array}{ccc}
\underline{2}1\underline{31} & & \underline{3}1\underline{31} \\
1123 & \rightarrow & 1133
\end{array}
$$

We characterise compositions of this type by the following proposition. Recall that the first part has value k.

Proposition 4 *If there is a part in the composition which is strictly smaller than the first part k and there is a $k + 1$ in the original composition located at position i where i is the number of parts in the original composition that are less than or equal to k, then the number of matchings will increase by one as a result of obtaining a new matching in position i, after the first part of the composition is increased by 1.*

The use of the conventions and notation which we adopted above for this Section, allow us to symbolically represent the situation hypothesized in Proposition 4, for a composition before re-ordering and before augmentation, as

$$
\underbrace{k}_{\text{position 1}} \quad \underbrace{l_k^j \, g_{k+1}^{i-2-j}}_{i-2 \text{ parts in total}} \quad \underbrace{k+1}_{\text{position } i} \quad l_k^{i-1-j} g_{k+1}^m \tag{1}
$$

where the parts which are not in positions 1 or i can appear in any order, m is a non-negative integer and at least one of the l_ks is less than k.

What we now do below is provide a generating function f_{+1} for all compositions of n which gain one matching point from the process of augmentation. This is the same as providing a generating function for all compositions that are of type given in Equation (1). In this generating function as per Definition (1) in the Introduction (integer compositions of n), x tracks the size n and y tracks the number of parts in each composition. Note that to ensure that k is not the minimum, we exclude the case where there are i values of size k (which are the first i values after re-ordering). The first binomial $\binom{i-2}{j}$ specifies the j possible positions for values $\leq k$, each themselves of the form $y\frac{x-x^{k+1}}{1-x}$, before position i. The second binomial, $\binom{m+i-2-j}{m}$ specifies the m positions for values $\geq k+1$ from the remaining available positions after i; the generating function for values less than $k+1$ is the geometric series sum $yx + yx^2 + \ldots + yx^k = \left(y\frac{x-x^{k+1}}{1-x}\right)$; and a similar infinite geometric series for values greater than or equal to $k + 1$.

$$
\begin{aligned}
&f_{+1}(x, y) \\
&:= \sum_{k=1}^{\infty} yx^k \sum_{i=2}^{\infty} \sum_{j=0}^{i-2} \binom{i-2}{j} \left(y\frac{x-x^{k+1}}{1-x}\right)^j \left(\frac{yx^{k+1}}{1-x}\right)^{i-2-j} yx^{k+1} \\
&\quad \times \sum_{m=0}^{\infty} \binom{m+i-2-j}{m} \left(y\frac{x-x^{k+1}}{1-x}\right)^{i-1-j} \left(\frac{yx^{k+1}}{1-x}\right)^m \\
&\quad - \sum_{k=1}^{\infty} yx^k \sum_{i=2}^{\infty} \sum_{j=0}^{i-2} \binom{i-2}{j} (yx^k)^j \left(\frac{yx^{k+1}}{1-x}\right)^{i-2-j} yx^{k+1} \\
&\quad \times \sum_{m=0}^{\infty} \binom{m+i-2-j}{m} (yx^k)^{i-1-j} \left(\frac{yx^{k+1}}{1-x}\right)^m \\
&= \sum_{k=1}^{\infty} \left(\frac{\left((1-x)y^3x^2\right)\left(x^{2k}\left(1-x^k\right)\right)}{(1-x-xy)\left(1-x-x^{k+1}y\right)} - \frac{yx^{k+1}\left(-\left((1-x)^2x^{2k}y^2\right)\right)}{\left(-1+x+x^{1+k}\right)\left(1-x-x^{1+k}y-x^ky+x^{1+k}y\right)} \right).
\end{aligned}
$$

Setting $y = 1$ gives us

$$f_{+1}(x) := f_{+1}(x, 1)$$
$$= \sum_{k=1}^{\infty} \left(\frac{\left((1-x)x^2\right)\left(x^{2k}\left(1-x^k\right)\right)}{(1-2x)\left(1-x-x^{k+1}\right)} - \frac{x^{k+1}\left(-\left((1-x)^2 x^{2k}\right)\right)}{\left(-1+x+x^{1+k}\right)\left(1-x-x^{1+k}-x^k+x^{1+k}\right)} \right).$$

2.2.2 Lose one matching

To enumerate the compositions which lose one matching point after increasing the first part by 1, we consider two cases.

Case 1. k is the minimum part size but after augmentation it is no longer the minimum. (i.e., there is a matching point in position 1 which we lose after adding 1 to the first part). Moreover, since this is the case where we lose exactly one matching point, the part in position i (recall that i has been set as the number of parts in the composition that are less than or equal to k) is neither a k nor a $k+1$ (i.e., for the re-ordered composition before or after augmentation, there will not be a match in position i, hence there is neither a gain nor loss of a matching point in this position).

Case 2. Now k is not the minimum part size. Hence the loss of one matching point must occur in position i. This happens when there is a k in position i before re-ordering (which we will then lose on adding 1 to the first part, because the value in position i of the re-ordered composition of $n+1$ will be $k+1$).

Case 1 and 2 above are characterised separately in the following two propositions:

Proposition 5 *If the minimum part value in a composition is k and the part in position i is greater than or equal to $k+2$, then the number of matchings will decrease by one after adding 1 to the first part of the composition.*

Example: The number of matchings in the composition 1312 is one and the number of matchings in the composition 2312 is zero.

1̲312	$\rightarrow$	2312
1123		1223

The situation hypothesized in Proposition 5 is represented symbolically by

$$\underbrace{k}_{\text{position 1}} \quad \underbrace{k^j g_{k+1}^{i-2-j}}_{i-2 \text{ parts in total}} \quad \underbrace{g_{k+2}}_{\text{position } i} \quad k^{i-1-j} g_{k+1}^m. \tag{2}$$

This has almost the same explanation as representation (1), except that by hypothesis the only values less than or equal to k are k itself and hence l_k in (1) is replaced by k in (2). Also the part in position i is hypothesized to be of type g_{k+2}. This is another change required in representation (2).

Proposition 6 *If the value of the minimum part in a composition is less than k, and the part in position i is k then the number of matchings will decrease by one after increasing the first part of the composition by 1.*

Example: The number of matchings in the composition 2321 is one and the number of matchings in the composition 3321 is zero.

232̲1	$\rightarrow$	3321
1223		1233

The situation hypothesized in Proposition 6, for a composition before re-ordering and before augmentation, has symbolic representation

$$\underbrace{k}_{\text{position 1}} \quad \underbrace{l_k^j g_{k+1}^{i-2-j}}_{i-2 \text{ parts in total}} \quad \underbrace{k}_{\text{position } i} \quad l_k^{i-1-j} g_{k+1}^m \tag{3}$$

where at least one of the terms of type l_k is strictly less than k.

The generating function for the sum of the two cases is given below. For an explanation of the various trackers x and y and the various coefficients of these, we refer the reader back to the case above where we explained the derivation of $f_{+1}(x, y)$. Note that again we need to subtract off the case where there are i parts of size k to ensure that k is not the minimum as required by case two. Hence the generating function which counts the number of compositions in which one matching point is lost after adding 1 to the first part is

$$
\begin{aligned}
f_{-1}(x, y) := & \sum_{k=1}^{\infty} y x^k \sum_{i=2}^{\infty} \sum_{j=0}^{i-2} \binom{i-2}{j} (y x^k)^j \left(\frac{y x^{k+1}}{1-x}\right)^{i-2-j} y \frac{x^{k+2}}{1-x} \\
& \times \sum_{m=0}^{\infty} \binom{m+i-1-j}{m} (y x^k)^{i-1-j} \left(\frac{y x^{k+1}}{1-x}\right)^m \\
+ & \sum_{k=1}^{\infty} y x^k \sum_{i=2}^{\infty} \sum_{j=0}^{i-2} \binom{i-2}{j} \left(y \frac{x-x^{k+1}}{1-x}\right)^j \left(\frac{y x^{k+1}}{1-x}\right)^{i-2-j} y x^k \\
& \times \sum_{m=0}^{\infty} \binom{m+i-2-j}{m} \left(y \frac{x-x^{k+1}}{1-x}\right)^{i-2-j} \left(\frac{y x^{k+1}}{1-x}\right)^m \\
- & \sum_{k=1}^{\infty} y x^k \sum_{i=2}^{\infty} \sum_{j=0}^{i-2} \binom{i-2}{j} (y x^k)^j \left(\frac{y x^{k+1}}{1-x}\right)^{i-2-j} y x^k \\
& \times \sum_{m=0}^{\infty} \binom{m+i-2-j}{m} (y x^k)^{i-2-j} \left(\frac{y x^{k+1}}{1-x}\right)^m .
\end{aligned}
$$

Setting $y = 1$ yields

$$
f_{-1}(x) := f_{-1}(x, 1) = \sum_{k=1}^{\infty} \frac{(x-1) x^{2k} \left(x^{2k+1} + \left(-2x^3 + x - 1\right) x^k + (1-x)x\right)}{(2x-1) \left(x^k + x - 1\right) \left(x^{k+1} + x - 1\right)}.
$$

2.2.3 Lose two matchings

To enumerate the compositions which lose two matchings when 1 is added to the first part, we consider the following proposition.

Proposition 7 *If the value of the minimum part in a composition is k, and there are exactly i parts k, with a k in position i then the number of matchings will decrease by two after adding 1 to the first part of the composition.*

Example: The number of matchings in the composition 1123 is four and the number of matchings in the composition 2123 is two.

$$
\begin{array}{ccc}
\underline{11\underline{23}} & \rightarrow & 21\underline{23} \\
\underline{1123} & & 1\underline{223}
\end{array}
$$

Symbolically we have

$$
\underbrace{k}_{\text{position 1}} \quad \underbrace{k^j \, g_{k+1}^{i-2-j}}_{i-2 \text{ parts in total}} \quad \underbrace{k}_{\text{position } i} \quad k^{i-2-j} g_{k+1}^m \tag{4}
$$

where the parts which are not in position 1 or i can appear in any order, and m is a non-negative integer.

For an explanation of the various trackers x and y and the various coefficients of these in the generating function which counts the number of compositions in which two matching points are lost when adding 1 to the

first part, we once again refer the reader back to the case above where we explained the derivation of $f_{+1}(x, y)$. This generating function is thus

$$f_{-2}(x, y) := \sum_{k=1}^{\infty} \sum_{i=2}^{\infty} \left(yx^k\right)^i \sum_{j=0}^{i-2} \binom{i-2}{j} \left(\frac{yx^{k+1}}{1-x}\right)^{i-2-j} \sum_{m=0}^{\infty} \binom{m+i-2-j}{m} \left(\frac{yx^{k+1}}{1-x}\right)^m,$$

so that with $y = 1$ we have

$$f_{-2}(x) := f_{-2}(x, 1) = \sum_{k=1}^{\infty} \frac{(1-x)x^{2k}}{1-x-x^k}$$

2.3 Number of matching points

We have now computed the generating functions which determine the number of additional matching points when moving from a composition of n to a composition of $n+1$ via the two procedures described above.

By putting a 1 at the beginning of the composition, there is one additional matching point per composition. As is well known, there are 2^{n-1} compositions of n for which the generating function is $\frac{x}{1-2x}$. The generating function for the number of additional matching points obtained by adding 1 to the first part is

$$f_{+1}(x) - f_{-1}(x) - 2f_{-2}(x)$$
$$= \sum_{k\geq 1} \frac{(x-1)x^{2k}\left(-3x^{k+2} + x^{k+3} - x^{2k+1} + x^{2k+2} + x^k - x^3 - 2x^2 + 5x - 2\right)}{(2x-1)\left(x^k + x - 1\right)\left(x^{k+1} + x - 1\right)}$$
$$= \frac{2x^2}{1-x-2x^2}. \tag{5}$$

Since half of the compositions of $n+1$ come from inserting 1 at the beginning of a composition of n and the other half come from adding 1 to the first part, we obtain the following recursion. Let $m(n)$ be the number of matchings over all compositions of n. Then

$$m(n+1) = 2m(n) + \text{ all additional matchings.}$$

We define the generating function $M(x) := \sum_{n\geq 1} m(n)x^n$; multiply the recursion by x^n and sum on n to get

$$\sum_{n\geq 1} m(n+1)x^n = \sum_{n\geq 1} 2m(n)x^n + \frac{x}{1-2x} + \frac{2x^2}{1-x-2x^2}$$
$$\Leftrightarrow \frac{1}{x}(M(x) - x) = 2M(x) + \frac{x}{1-2x} + \frac{2x^2}{1-x-2x^2}$$
$$\Leftrightarrow M(x) - x = 2xM(x) + \frac{x^2}{1-2x} + \frac{2x^3}{1-x-2x^2}$$
$$\Leftrightarrow M(x) = \frac{x - 3x^3}{4x^3 - 3x + 1}. \tag{6}$$

The coefficient of x^n in the above gives the total number of matching points over all compositions of $n > 0$ as

$$[x^n]M(x) = \frac{1}{36}\left(3 \cdot 2^n n + 8(-1)^n + 19 \cdot 2^n\right).$$

The result in Theorem 2 follows after dividing by 2^{n-1} to obtain an average. This concludes the proof of Theorem 2.

$$i-1 \text{ parts} \geq 2 \quad | \quad \text{at least } i-1 \text{ parts of size } 1$$

$$i$$

(Position of first 1)
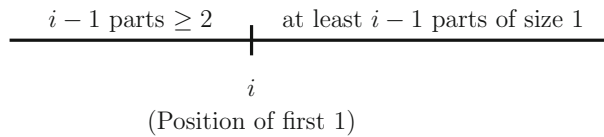
Fig. 1 Diagram for subclass $\mathcal{A}$ compositions having at least one matching point 1 in the ith position

2.4 Compositions of n with no matching points

The aim of this section is to show that as n tends to ∞, the ratio of compositions with no matching points to all compositions tends to 0. In order to do this we will find a subclass $\mathcal{A}$ of compositions of n which have at least one matching point. This is a lower bound for all such compositions and as n tends to ∞, we will show that the proportion of these to all compositions tends to 1. The proportion of the complementary set in which we are interested will therefore tend to 0. Below we illustrate a decomposition for the subclass $\mathcal{A}$: namely all compositions having the first 1 in the ith position followed by at least $i-1$ parts 1. To the right of position i, there may also be other points. Such compositions have at least one matching point in the ith position.

Let $t \geq i-1$ be the number of parts after the first occurrence of 1, i.e., to the right of position i, and let u track the number $j \leq t$ of 1s after the initial 1. The generating function for these translates from Figure 1 as

$$\left(\frac{x^2}{1-x}\right)^{i-1} x \left(xu + \frac{x^2}{1-x}\right)^t.$$

where the first bracketed term tracks the leftmost $i-1$ parts each of size at least 2, x on its own tracks the 1 in the ith position and the rightmost bracket tracks the t parts to the right of position i. These latter have u which tracks the number that are 1s (marked by the x) or do not have a u in the case that they are at least 2, marked by $\frac{x^2}{1-x}$.

The generating function for the subclass $\mathcal{A}$ is therefore

$$x \sum_{i\geq 1} \left(\frac{x^2}{1-x}\right)^{i-1} \sum_{t\geq i-1} \sum_{j=i-1}^{t} [u^j]\left(xu + \frac{x^2}{1-x}\right)^t$$

$$= x \sum_{i\geq 1} \left(\frac{x^2}{1-x}\right)^{i-1} \sum_{t\geq i-1} \sum_{j=i-1}^{t} \binom{t}{j} x^j \left(\frac{x^2}{1-x}\right)^{t-j}$$

$$= x \sum_{i\geq 1} \left(\frac{x^2}{1-x}\right)^{i-1} \sum_{t\geq i-1} \left(\sum_{j=0}^{t} \binom{t}{j} x^j \left(\frac{x^2}{1-x}\right)^{t-j} - \sum_{j=0}^{i-2} \binom{t}{j} x^j \left(\frac{x^2}{1-x}\right)^{t-j} \right)$$

$$= x \sum_{i\geq 1} \left(\frac{x^2}{1-x}\right)^{i-1} \left(\sum_{t\geq i-1} \left(\frac{x}{1-x}\right)^t - \sum_{j=0}^{i-2} \sum_{t\geq i-1} \binom{t}{j} x^j \left(\frac{x^2}{1-x}\right)^{t-j} \right)$$

$$= \frac{x(1-x)^3}{(1-2x)((1-x)^2 - x^3)} - x \sum_{i\geq 1} \sum_{j=0}^{i-2} \sum_{t\geq i-1} \binom{t}{j} x^j \left(\frac{x^2}{1-x}\right)^{t-j+i-1}$$

$$= \frac{x(1-x)^3}{(1-2x)((1-x)^2 - x^3)} - x \sum_{j=0}^{\infty} \left(\frac{1-x}{x}\right)^j \sum_{t\geq j+1} \binom{t}{j} \left(\frac{x^2}{1-x}\right)^t \sum_{i=j+2}^{t+1} \left(\frac{x^2}{1-x}\right)^{i-1}$$

$$= \frac{x(1-x)^3}{(1-2x)((1-x)^2 - x^3)} - \frac{x^3}{1-x-x^2} \sum_{j=0}^{\infty} x^j \sum_{t\geq j+1} \binom{t}{j} \left(\frac{x^2}{1-x}\right)^t$$

$$+ \frac{x^3}{1-x-x^2} \sum_{j=0}^{\infty} \left(\frac{1-x}{x}\right)^j \sum_{t\geq j+1} \binom{t}{j} \left(\frac{x^2}{1-x}\right)^{2t}$$
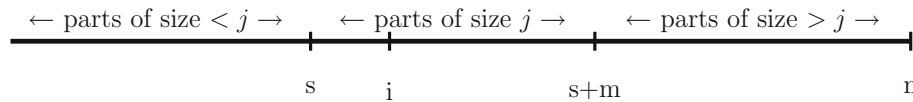
Fig. 2 Diagram for a word of length n with a matching point in position i after sorting

$$= \frac{x(1-x)^3}{(1-2x)((1-x)^2-x^3)}$$
$$- \frac{x^3}{1-x-x^2} \sum_{j=0}^{\infty} x^j \frac{(x-1)^{j+1}x^{2j}}{(1-x)^{j+1}(x^2+x-1)^{j+1}} \left(\frac{(x^2+x-1)^{j+1}}{(x-1)^j} - x + 1 \right)$$
$$+ \frac{x^3}{1-x-x^2} \sum_{j=0}^{\infty} \left(\frac{1-x}{x} \right)^j \frac{x^{4j}\left(-\left(-x^4+x^2-2x+1\right)^{j+1} + (x-1)^{2(j+1)} \right)}{(1-x)^{2j}\left(-x^4+x^2-2x+1\right)^{j+1}}$$
$$= \frac{x(1-x)^3}{(1-2x)((1-x)^2-x^3)} + \frac{x^5\left(x^3-2x+1\right)}{\left(-x^2-x+1\right)\left(x^2+x-1\right)\left(x^3+x-1\right)\left(x^3+x^2+x-1\right)}$$
$$+ \frac{x^3\left(x^9-x^8-x^7+3x^6-3x^5+x^4\right)}{\left(-x^2-x+1\right)\left(x^3+x-1\right)\left(x^3-x^2+2x-1\right)\left(-x^4+x^2-2x+1\right)}$$
$$= \frac{x\left(1-2x+x^3\right)}{(1-2x)(1-x-x^2-x^3)}$$
$$= \frac{1}{2} + \frac{1}{2}\frac{1}{1-2x} - \frac{1-x}{1-x-x^2-x^3}.$$

Hence we have the following lemma.

Lemma 8 *The generating function for the subclass $\mathcal{A}$ of compositions of n with at least one 1 as a matching point is*

$$g(x) := \frac{1}{2} + \frac{1}{2}\frac{1}{1-2x} - \frac{1-x}{1-x-x^2-x^3}.$$

As a consequence, we can subtract this from the generating function for all compositions to get a generating function for an upper bound on the number of compositions of n with no matching points:

$$\frac{1-x}{1-2x} - \left(\frac{1}{2} + \frac{1}{2}\frac{1}{1-2x} - \frac{1-x}{1-x-x^2-x^3} \right) = \frac{1-x}{1-x-x^2-x^3}.$$

The dominant root of this rational function is $x = 0.543689\cdots$. Hence the coefficients of x^n in the rational function are $O((1.83929\cdots)^n)$ (see [4]). Dividing by 2^{n-1} we obtain

Theorem 9 *The proportion of compositions of n with no matching points is $O((0.919643\cdots)^n)$ which tends to zero at an exponential rate as $n \to \infty$.*

Open problem: Find an exact generating function for compositions of n with no matching points.

3 Matching points in words of length n over alphabet k

We now study the above statistics for words of length n over alphabet $[k] = \{1, 2, \ldots, k\}$.

3.1 Total number of matching points in words

To compute the total number of matching points for all words of length n over alphabet $[k]$, we illustrate in the diagram below, any such word with matching point j in position i after it has been sorted.

In this diagram, since j in position i before the sort is a matching point, it remains in position i after being sorted. Out of the remaining $n-1$ positions before the sort, we let s be the number of positions for the parts that are less than j; $m-1$ be the number of positions for the other parts of size j and then $n-m-s$ is the number of possible positions for parts of size more than j. This yields the total number of words which have a matching point j in position i as

$$\sum_{s=0}^{i-1}\sum_{m=i-s}^{n-s}\binom{n-1}{s,\,m-1,\,n-m-s}\left(\frac{j-1}{k}\right)^s\left(\frac{1}{k}\right)^{m-1}\left(\frac{k-j}{k}\right)^{n-1-(m-1+s)}\frac{1}{k}. \tag{7}$$

To obtain the total number of matching points in all words of length n over alphabet [k], we sum over all possibilities for i and j. Call this function $f(n)$. So, we obtain

$$f(n)=\sum_{i=1}^{n}\sum_{j=1}^{k}\sum_{s=0}^{i-1}\sum_{m=i-s}^{n-s}\binom{n-1}{s,\,m-1,\,n-m-s}\left(\frac{j-1}{k}\right)^s\left(\frac{1}{k}\right)^{m-1}\left(\frac{k-j}{k}\right)^{n-1-(m-1+s)}\frac{1}{k}$$

$$=\frac{1}{k^n}\sum_{i=1}^{n}\sum_{j=1}^{k}\sum_{s=0}^{i-1}\sum_{m=i-s}^{n-s}\frac{(n-1)!}{s!(m-1)!(n-1-(m-1+s))!}(j-1)^s(k-j)^{n-1-(m-1+s)}$$

$$=\frac{1}{k^n}\sum_{i=1}^{n}\sum_{j=1}^{k}\sum_{s=0}^{i-1}\sum_{m=i-s}^{n-s}\frac{(n-1)!}{s!(m-1)!(n-m-s)!}(j-1)^s(k-j)^{n-m-s}$$

$$=\frac{1}{k^n}\sum_{i=1}^{n}\sum_{j=1}^{k}\sum_{s=0}^{i-1}\sum_{t=i}^{n}\frac{(n-1)!}{s!(t-s-1)!(n-t)!}(j-1)^s(k-j)^{n-t}\quad\text{by letting }m=t-s\text{ (or }m+s=t)$$

$$=\frac{1}{k^n}\sum_{i=1}^{n}\sum_{j=1}^{k}\sum_{s=0}^{i-1}\sum_{t=i}^{n}\frac{(n-1)!(t-1)!}{s!(t-s-1)!(n-t)!(t-1)!}(j-1)^s(k-j)^{n-t}$$

$$=\frac{1}{k^n}\sum_{i=1}^{n}\sum_{j=1}^{k}\sum_{s=0}^{i-1}\sum_{t=i}^{n}\frac{(n-1)!}{(n-t)!(t-1)!}\frac{(t-1)!}{s!(t-s-1)!}(j-1)^s(k-j)^{n-t}$$

$$=\frac{1}{k^n}\sum_{i=1}^{n}\sum_{j=1}^{k}\sum_{s=0}^{i-1}\sum_{t=i}^{n}\binom{n-1}{t-1}\binom{t-1}{s}(j-1)^s(k-j)^{n-t}$$

$$=\frac{1}{k^n}\sum_{j=1}^{k}\sum_{s=0}^{n-1}\sum_{i=1}^{s+1}\sum_{t=i}^{n}\binom{n-1}{t-1}\binom{t-1}{s}(j-1)^s(k-j)^{n-t}$$

$$=\frac{1}{k^n}\sum_{j=1}^{k}\sum_{s=0}^{n-1}\sum_{t=1}^{n}\sum_{i=s+1}^{t}\binom{n-1}{t-1}\binom{t-1}{s}(j-1)^s(k-j)^{n-t}$$

$$=\frac{1}{k^n}\sum_{j=1}^{k}\sum_{s=0}^{n-1}\sum_{t=1}^{n}(t-(s+1)+1)\binom{n-1}{t-1}\binom{t-1}{s}(j-1)^s(k-j)^{n-t}$$

$$=\frac{1}{k^n}\sum_{j=1}^{k}\sum_{s=0}^{n-1}\sum_{t=1}^{n}(t-s)\binom{n-1}{t-1}\binom{t-1}{s}(j-1)^s(k-j)^{n-t}$$

$$=\frac{1}{k^n}\sum_{j=1}^{k}\sum_{s=0}^{n-1}\sum_{t=1}^{n}(t-s)\frac{(n-1)!}{(t-1)!(n-1-(t-1))!}\frac{(t-1)!}{s!(t-1-s)!}(j-1)^s(k-j)^{n-t}$$

$$
= \frac{1}{k^n} \sum_{j=1}^{k} \sum_{s=0}^{n-1} \sum_{t=1}^{n} (t-s) \frac{(n-1)!}{(n-t)!} \frac{1}{s!(t-1-s)!} (j-1)^s (k-j)^{n-t}
$$

$$
= \frac{1}{k^n} \sum_{j=1}^{k} \sum_{s=0}^{n-1} (j-1)^s \sum_{t=1}^{n} (t-s) \binom{n-1}{s, t-s-1, n-t} (k-j)^{n-t}. \tag{8}
$$

Now we create the following generating function out of

$$
f(n) := \sum_{j=1}^{k} \sum_{s=0}^{n-1} \sum_{t=1}^{n} (t-s) \binom{n-1}{t-1} \binom{t-1}{s} (j-1)^s (k-j)^{n-t},
$$

namely:

$$
F(x) := \sum_{n=1}^{\infty} f(n) x^n.
$$

By interchanging sums as follows

$$
\sum_{n=1}^{\infty} \sum_{j=1}^{k} \sum_{s=0}^{n-1} \sum_{t=1}^{n} \to \sum_{j=1}^{k} \sum_{s=0}^{\infty} \sum_{t=s+1}^{\infty} \sum_{n=t}^{\infty}
$$

and considering the inner sum on n with only terms involving n we obtain

$$
\sum_{n=t}^{\infty} \binom{n-1}{t-1} (k-j)^{n-1-(t-1)} x^n = x^t (1 + jx - kx)^{-t}.
$$

Now include all the other terms and sums:

$$
\sum_{j=1}^{k} \sum_{s=0}^{\infty} \sum_{t=s+1}^{\infty} (j-1)^s (t-s) x^t (1 + jx - kx)^{-t} \binom{t-1}{s}
$$

$$
= \sum_{j=1}^{k} \sum_{s=0}^{\infty} \frac{(j-1)^s x^{1+s} (1 + jx - kx)^{-s} \left(\frac{1-x+jx-kx}{1+jx-kx} \right)^{-s} (1 + jx - kx + sx)}{(1 - x + jx - kx)^2}
$$

$$
= \sum_{j=1}^{k} \frac{x(1 + x - kx)}{(1 - kx)^2}
$$

$$
= \frac{kx(1 + x - kx)}{(1 - kx)^2}.
$$

The series coefficient of x^n in the above yields the total number of matchings over all words of length n over alphabet $[k]$, namely

$$
(n + k - 1) k^{n-1}, \tag{9}
$$

and dividing by k^n we obtain:

Theorem 10 *The average number of matchings in words of length n over alphabet* $[k]$ *is*

$$
\frac{n + (k - 1)}{k}. \tag{10}
$$

3.2 Words with no matching points

We wish to show that the probability that a word of length n over alphabet $[k]$ has no matchings tends to zero as $n \to \infty$.

In order to count words with no matchings, we first show that almost all words have a significant number of 1s.

Let $0 < \alpha < 1$. We will show that the probability that a word of length n over alphabet $[k]$ has fewer than $\alpha n / k$ 1s tends to zero as $n \to \infty$.

For this we apply a Chernoff bound as in the following theorem which appears in [6].

Theorem 11 *(Mitzenmacher and Upfal Theorem 4.5 [6])*
Let $X_1, \ldots, X_n$ be independent Poisson variables and let $X = \sum_{i=1}^n X_i$ and $\mu = E[X]$. Then for $0 < \delta < 1$,

$$Pr(X \leq (1 - \delta)\mu) \leq e^{-\mu\delta^2/2}.$$

In our case we define $p_i := Pr(X_i = 1)$ and we take $p_i = 1/k$, so $\mu = E[X] = n/k$ is the expected number of 1s in a uniformly chosen random word of length n over alphabet $[k]$, and $\alpha = 1 - \delta$. Hence we find that the probability that such a word has fewer than $\alpha n / k$ 1s is less than or equal to

$$\exp(-n(1 - \alpha)^2/(2k))$$

which tends to 0 as $n \to \infty$.

Thus to find the proportion of words with no matchings we may restrict our attention to words of length n with at least $\alpha n / k$ 1s.

The probability that there isn't a 1 in any of the $\frac{\alpha n}{k}$ positions is $\left(\frac{k-1}{k}\right)^{\frac{\alpha n}{k}}$. We want the complement of this which implies that the probability that there is at least one 1 in the first $\frac{\alpha n}{k}$ positions is equal to $1 - \left(\frac{k-1}{k}\right)^{\frac{\alpha n}{k}}$, which tends to one as $n \to \infty$.

The probability that we have a matching point in these positions is the product of these two probabilities.

So the probability that there is a matching point in the first $\frac{\alpha n}{k}$ positions is $(1 - \exp(-n(1 - \alpha)^2/(2k)))\left(1 - \left(\frac{k-1}{k}\right)^{\frac{\alpha n}{k}}\right)$. This tends to 1 as $n \to \infty$.

Therefore the probability that a word has no matching points is bounded above by

$$1 - \left(1 - \exp(-n(1 - \alpha)^2/(2k))\right)\left(1 - \left(\frac{k-1}{k}\right)^{\frac{\alpha n}{k}}\right)$$

which tends to 0 as $n \to \infty$. Thus we have proved

Theorem 12 *The probability that a word of length n over alphabet $[k]$ has no matching points tends to 0 as $n \to \infty$.*

In conclusion we present an open problem.

Conjecture 13 *The proportion of words with no matchings is asymptotic to $\left(\frac{k-1}{k}\right)^{n+k-1}$.*

References

1. M. Archibald, A. Blecher, A. Knopfmacher, Fixed points in compositions and words, *J. Integer Sequences* **23** (2020), Article 20.11.1.
2. M. Archibald, A. Blecher, A. Knopfmacher, Alphabetic points in compositions and words, *Discrete Mathematics and Applications* **33** (2021), 20–30.
3. M. Bóna, Combinatorics of Permutations, CRC Press, Second edition, 2012.
4. P. Flajolet, R. Sedgewick, Analytic Combinatorics, Cambridge University Press 2008. (Web edition available at http://algo.inria.fr/flajolet/Publications/books.html.)
5. T. Mansour, Combinatorics of compositions and words, CRC Press, 2010.
6. M. Mitzenmacher, E. Upfal, *Probability and Computing*, Second Edition, Cambridge University Press 2017.
7. N. Sloane, The On-line Encyclopedia of Integer Sequences (https://oeis.org/).

8. R. P. Stanley, Enumerative Combinatorics, Volume 1, Cambridge University Press, 1986.
9. H. S. Wilf, generatingfunctionology, A.K. Peters (now CRC press), 1990.