Counting and Matching

Bart Jacobs ✉
Radboud University Nijmegen, The Netherlands

Dario Stein ✉
Radboud University Nijmegen, The Netherlands

Abstract

Lists, multisets and partitions are fundamental datatypes in mathematics and computing. There are basic transformations from lists to multisets (called "accumulation") and also from lists to partitions (called "matching"). We show how these transformations arise systematically by forgetting/abstracting away certain aspects of information, namely order (transposition) and identity (substitution). Our main result is that suitable restrictions of these transformations are isomorphisms: This reveals fundamental correspondences between elementary datatypes. These restrictions involve "incremental" lists/multisets and "non-crossing" partitions/lists. While the process of forgetting information can be precisely spelled out in the language of category theory, the relevant constructions are very combinatorial in nature. The lists, partitions and multisets in these constructions are counted by Bell numbers and Catalan numbers. One side-product of our main result is a (terminating) rewriting system that turns an arbitrary partition into a non-crossing partition, without improper nestings.

2012 ACM Subject Classification Mathematics of computing → Combinatorics

Keywords and phrases List, Multiset, Partition, Crossing

Digital Object Identifier 10.4230/LIPIcs.CSL.2023.28

Acknowledgements We like to thank Dusko Pavlovic for lively discussions on the topic of this paper – and on much more.

1 Introduction

This paper considers three fundamental datatypes in computing and reasoning, namely lists (or sequences), multisets, and set partitions. Two characteristic properties of lists are: (1) elements in a list are ordered, and (2) elements may occur multiple times. Multisets are datatypes where the first property is dropped, but the second one is kept. Thus, a multiset is like a subset, except that elements may occur multiple times. The order of the occurrences does not matter. Finally, set partitions are collections of non-empty pairwise disjoint subsets of a given set, whose union is the whole set.

We distinguish two fundamental operations on lists, called *accumulation* (abbreviated as *acc*) and *matching* (written as *mat*). Accumulation is a function from lists to multisets that counts occurrences of elements in the list. Matching is a function from lists to partitions that registers equality of elements. Both these functions, accumulation and matching, will be used as two orthogonal operations on lists in the following situation, where the number $K \geq 1$ is a parameter.

$$\begin{tikzcd}[column sep=small] & \begin{pmatrix} \text{lists of} \\ \text{length } K \end{pmatrix} \arrow[dl, "acc"'] \arrow[dr, "mat"] & \\ \begin{pmatrix} \text{multisets} \\ \text{of size } K \end{pmatrix} & & \begin{pmatrix} \text{partitions of} \\ \{1, \ldots, K\} \end{pmatrix} \end{tikzcd} \tag{1}$$

31st EACSL Annual Conference on Computer Science Logic (CSL 2023).
Editors: Bartek Klin and Elaine Pimentel; Article No. 28; pp. 28:1–28:15
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Intuitively, the accumulation of a list forgets the order and only considers the elements in the list with their multiplicity (number of occurrences). In matching we look at which positions equal elements occur and we put these positions in the same block (subset) of the resulting partition of $\{1, \ldots, K\}$.

These two operations *acc* and *mat* are orthogonal in the following sense:

- Accumulation is invariant under *transposition*: permuting (transposing) the elements in a list, by swapping places, does not change their accumulation.
- Matching is invariant under *substitution*: applying a permutation to the elements themselves does not change the outcome of matching.

We recall that transposition and substitution are the two basic operations in (symmetric) cryptography, used to encipher a message.

We now describe in abstract terms the main result of this paper. As in Diagram (1) we fix a parameter $K \geq 1$ and use it in a "double" manner, not only for length/size, but also as the set of elements $\{1, \ldots, K\}$ that may be used in the lists and multisets. In that case we can identify certain subsets of lists, multisets and set partitions that are in bijective correspondence, in a situation:

$$\begin{tikzcd}
 & & \left(\begin{array}{c}\text{lists of}\\ \text{length } K\end{array}\right) \arrow[ddll, "acc"'] \arrow[ddrr, "mat"] & & \\
 & & \mathit{NCIL}(K) \arrow[u, hook] \arrow[dl, leftrightarrow, "\cong"'] \arrow[dr, leftrightarrow, "\cong"] & & \\
\left(\begin{array}{c}\text{multisets}\\ \text{of size } K\end{array}\right) & \mathit{IM}(K) \arrow[l, hook'] \arrow[rr, leftrightarrow, "\cong"'] & & \mathit{NCSP}(K) \arrow[r, hook] & \left(\begin{array}{c}\text{partitions of}\\ \{1,\ldots,K\}\end{array}\right)
\end{tikzcd} \tag{2}$$

The abbreviation *NCIL* stands for "non-crossing incremental list, *IM* for "incremental multiset", and *NCSP* for "non-crossing set partition". Details will be provided below. These isomorphisms in the small triangle in the middle capture fundamental ways to relate the basic structures of lists, multisets and partitions. The construction of these isomorphisms is essentially combinatorial. Interestingly, the number of elements in the isomorphic sets in this sub-triangle is given by Catalan numbers – as observed in [12].

Non-crossing partitions have been introduced in the 1970s in the work of Germain Kreweras [12]. The additional property that we call "incremental" is introduced here. There is a wider story to tell about the usage of these basic datatypes in probability theory, see *e.g.* [5, 4], especially for sufficient statistics [10]. Here, however, we concentrate on the datatypes themselves, and their interconnections.

It is a bit unfortunate that two meanings of the word partition have developed in the literature, namely *set* partitions and *multiset* partitions. We only use these multiset partitions in Section 8, and we focus on set partitions first. When we simply write "partition", we mean "set partition".

The paper is organised as follows. It starts with some background information on multisets and set partitions, which allows us to define accumulation and matching in Sections 2 and 3. Subsequently, Section 4 introduces a special subset of "incremental" lists and shows that these sequences correspond bijectively to partitions. Section 5 recalls the notion of non-crossing partition (from [12]) and defines the corresponding notion on sequences. It captures proper nesting. Our main result, about the subtriangle of isomorphisms in Diagram (2) is in Section 6. A consequence of this result is a mapping from arbitrary partitions to non-crossing partitions. Section 7 shows how this map can be obtained via a terminating rewriting system. In the final two sections we put our findings in a wider perspective: in Section 8 we show how to extend Diagram 1 to a commuting diamond via multiset partitions. This gives a wider perspective on the datatypes at hand. Finally, in Section 9 we describe how this diamond can be obtained from a general categorical construction, taking a colimit with respect to an exponent Y^X in the two different variables. This is based on Joyal's approach to combinatorics using the theory of species [11], formulated in terms of presheaves on the category of (finite) sets and bijections.

2 Multisets and accumulation

A *multiset*, or a *bag*, is a "subset" in which elements may occur multiple times. We use "ket" notation $|\,-\,\rangle$ for multisets and write for instance $3|R\rangle + 4|G\rangle + 5|B\rangle$ for a multiset with three elements R, four elements G, and five elements B. This multiset may represent an urn with three red, four green, and five blue balls. In general, a multiset over a set X is a formal finite sum $\sum_i n_i|x_i\rangle$ with $x_i \in X$ and $n_i \in \mathbb{N}$. Alternatively, such a multiset may be represented as a function $\varphi\colon X \to \mathbb{N}$ whose support $supp(\varphi) = \{x \in X \mid \varphi(x) \neq 0\}$ is finite. We write $\mathcal{M}(X)$ for the set of such multisets over X.

The *size* $\|\varphi\|$ of a multiset $\varphi \in \mathcal{M}(X)$ is its number of elements, including multiplicities, so $\|\varphi\| \coloneqq \sum_x \varphi(x)$. For a number $K \in \mathbb{N}$ we write $\mathcal{M}[K](X) \subseteq \mathcal{M}(X)$ for the subset of multisets of size K.

As is common, we write $X^K = X \times \cdots \times X$ for the K-fold product of X, containing lists of length K. There is an *accumulation* map $acc\colon X^K \to \mathcal{M}[K](X)$ given by:

$$acc\big(x_1, \ldots, x_K\big) \coloneqq 1|x_1\rangle + \cdots + 1|x_K\rangle.$$

Then, for instance $acc(b, a, c, b, b, a) = 2|a\rangle + 3|b\rangle + 1|c\rangle$. This accumulation map thus counts the multiplicities of each of the elements in the list. Transposing the elements in a list, by interchanging their positions, does not change the accumulation outcome. Indeed, also $acc(c, a, b, a, b, b) = 2|a\rangle + 3|b\rangle + 1|c\rangle = acc(b, a, c, b, b, a)$.

Accumulation is a fairly standard operation which can be used, for instance, to describe multinomial distributions, see [8, 9]. We include two standard combinatorial results.

▶ **Lemma 1.**

1. *For each multiset $\varphi \in \mathcal{M}(X)$, there are $\big(\varphi\big)$ many lists that accumulate to φ, where $\big(\varphi\big)$ is the* multinomial coefficient *of the multiset φ, given by the number:*

$$\big(\varphi\big) \coloneqq \frac{\|\varphi\|!}{\prod_x \varphi(x)!}.$$

2. *When the set X has n elements, written as $|X| = n$, then the number of elements in the set $\mathcal{M}[K](X)$ of multisets over X of size K is given by the* multichoose coefficient*:*

$$\left(\!\!\binom{n}{K}\!\!\right) \coloneqq \binom{n+K-1}{K} = \frac{(n+K-1)!}{(n-1)!\cdot K!}.$$ ⌟

For instance, for the multiset $\varphi = 1|a\rangle + 2|b\rangle + 1|c\rangle \in \mathcal{M}\big(\{a, b, c\}\big)$ of size $\|\varphi\| = 4$ there are $\big(\varphi\big) = \frac{4!}{1!\cdot 2!\cdot 1!} = 12$ lists in $\{a, b, c\}^4$ that accumulate to φ, namely:

$$\begin{array}{cccccc} \langle a,b,b,c\rangle & \langle a,b,c,b\rangle & \langle a,c,b,b\rangle & \langle b,a,b,c\rangle & \langle b,a,c,b\rangle & \langle b,b,a,c\rangle \\ \langle b,b,c,a\rangle & \langle b,c,a,b\rangle & \langle b,c,b,a\rangle & \langle c,a,b,b\rangle & \langle c,b,a,b\rangle & \langle c,b,b,a\rangle. \end{array}$$

3 Set partitions and matching

We shall write $\underline{K} := \{1, 2, \ldots, K\}$ for the set of the first K positive natural numbers. By a (set) partition of K, we mean a (set) partition of the set $\underline{K}$: It consists of a collection of "blocks" $B_i \subseteq \underline{K}$ of non-empty pairwise disjoint subsets B_i with $\bigcup_i B_i = \underline{K}$. Examples of partitions of K are the single block-partition $\{\underline{K}\}$ and the K-element partition $\{\{1\}, \ldots, \{K\}\}$ with singleton blocks. The number of blocks in a partition of K ranges between 1 and K. We shall write $SP(K)$ for the set of set partitions of K. It comes with a size function $|-|: SP(K) \to \underline{K}$.

For each set X there is a *matching* function $mat: X^K \to SP(K)$. It forms blocks out of the positions in a list with equal elements, as in:

$$mat\big(b, a, c, b, b, a\big) \;=\; \Big\{\, \{1, 4, 5\},\; \{2, 6\},\; \{3\} \,\Big\}.$$

In general we define matching as:

$$mat\big(x_1, \ldots, x_K\big) \;\coloneqq\; \bigcup_{1 \le i \le K} \Big\{\{j \in \underline{K} \mid x_j = x_i\}\Big\}.$$

This matching operation, like accumulation, is quite standard. An early source is, for instance, [1]. It is not hard to see that matching is stable under each substitution isomorphism $X \xrightarrow{\cong} X$ that is applied elementwise to the input list.

We see that accumulation of lists in X^K is stable under *transposition* – that is, under permutations $K \xrightarrow{\cong} K$ of the positions – whereas matching is stable under *substitution* – that is, under permutations $X \xrightarrow{\cong} X$ of the elements. A systematic categorical perspective is offered in Section 9.

Again we list two basic results, without proof.

▶ **Lemma 2.**
1. *The numbers $\big|SP(K)\big|$ of partitions of K, for $K = 1, 2, 3, \ldots$, are given by the Bell numbers:* 1*,* 2*,* 5*,* 15*,* 52*,* 203*,* 877*,* 4140*, . . .*
2. *Let X be a finite set with $|X| = n$ and let $P \in SP(K)$ be a partition of K with $|P| \le n$. The number of lists in X^K that match to P is given by the falling factorial:*

$$(n)_{|P|} \;=\; n(n-1)(n-2)\cdots(n-|P|+1) \;=\; \frac{n!}{(n-|P|)!}.$$ ◀

For instance, consider the four-element set $X = \{a, b, c, d\}$ and the partition $P \in SP(7)$ with two blocks:

$$P \;=\; \Big\{\, \{1, 3, 7\},\; \{2, 4, 5, 6\} \,\Big\}.$$

Then there are $\frac{4!}{(4-|A|)!} = \frac{4!}{2!} = 12$ lists in X^7 that match to P, namely:

$$\begin{array}{cccc}
\langle a, b, a, b, b, b, a\rangle & \langle a, c, a, c, c, c, a\rangle & \langle a, d, a, d, d, d, a\rangle & \langle b, a, b, a, a, a, b\rangle \\
\langle b, c, b, c, c, c, b\rangle & \langle b, d, b, d, d, d, b\rangle & \langle c, a, c, a, a, a, c\rangle & \langle c, b, c, b, b, b, c\rangle \\
\langle c, d, c, d, d, d, c\rangle & \langle d, a, d, a, a, a, d\rangle & \langle d, b, d, b, b, b, d\rangle & \langle d, c, d, c, c, c, d\rangle.
\end{array}$$

4 Incremental lists and multisets

This section introduces what we call "incremental" lists and multisets. As we shall see, the numbers of these lists and multisets are described by the Bell and Catalan numbers, respectively. The main result in this section says that set partitions correspond to incremental lists.

In the previous two sections we have considered partitions, lists and multisets on an arbitrary set X. From now on, we will take the underlying set to be $X = \underline{K} = \{1, \ldots, K\}$. By taking advantage of the order on $\underline{K}$, we can obtain concrete representations of partitions in terms of lists. Of special importance in this context are the minimal elements of each block, which we name *parents*.

▶ **Definition 3.** *Let $P \in SP(K)$ be a set partition of K.*

1. *An element $a \in \underline{K}$ is called a* parent *of the partition P if it is the minimum element of its block, that is, if $a = \min(B)$ where $B \in P$ is the necessarily unique block with $a \in B$. For any element $b \in \underline{K}$, we call the least element of its block the parent of b and denote it by $\mathrm{par}_P(b)$.*
2. *For a partition $P \in SP(K)$, its* parent list *$(s_1, \ldots, s_K) \in \underline{K}^K$ is defined as $s_i = \mathrm{par}_P(i)$. Its* parent multiset *$\varphi \in \mathcal{M}[K](\underline{K})$ is defined as $\varphi = acc(s_1, \ldots, s_K)$.*

For example, in the partition $P \in SP(\underline{8})$ given by

$$P = \Big\{ \{\mathbf{1}, 3, 5\}, \ \{\mathbf{2}, 6\}, \ \{\mathbf{4}, 7, 8\} \Big\}$$

we have highlighted the parents $\mathbf{1}, \mathbf{2}, \mathbf{4}$ in bold. The parent list of P is $\vec{s} = \langle \mathbf{1}, \mathbf{2}, \mathbf{1}, \mathbf{4}, \mathbf{1}, \mathbf{2}, \mathbf{4}, \mathbf{4} \rangle$. For instance, there is a 1 in the fifth entry since $par_P(5) = \mathbf{1}$. The associated parent multiset is $\varphi = acc(\vec{s}) = 3|\mathbf{1}\rangle + 2|\mathbf{2}\rangle + 3|\mathbf{4}\rangle$. From the parent multiset, we can read off the parents of the partition, and the sizes of their respective blocks, but not which elements belong to those blocks. Explicitly, the parent multiset φ satisfies:

$$\varphi(b) = \begin{cases} 0 & \text{if } b \text{ is not a parent} \\ |B| & \text{if } b = \min(B) \text{ for } B \in P. \end{cases}$$

It is easy to see that the parent list uniquely characterises the partition, via matching. The question comes up: which lists arise as parent lists? A characterisation of such lists will be formulated below in terms of an "incremental" property. We then get an isomorphism between incremental lists and set partitions, see Theorem 5 below.

We will give a similar characterisation of parent multisets as "incremental" multisets. A partition is generally *not* uniquely characterised by its parent multiset, however we will show in Section 5 that so-called *noncrossing partitions* are.

▶ **Definition 4.**

1. *A list $\vec{s} = (s_1, \ldots, s_K) \in \underline{K}^K$, with $1 \leq s_i \leq K$, is called* incremental *if for each index $1 \leq i \leq K$, we have $s_i \leq i$ and $s_{s_i} = s_i$.*
   We shall write $\mathrm{IL}(K) \subseteq \underline{K}^K$ for the subset of incremental lists.
2. *A multiset $\varphi \in \mathcal{M}[K](\underline{K})$ of size K with elements from $\underline{K} = \{1, \ldots, K\}$ is called* incremental *if for all $i \in \underline{K}$, we have*

$$\sum_{j \leq i} \varphi(j) \geq i.$$

We write $IM(K) \subseteq \mathcal{M}[K](\underline{K})$ for the subset of incremental multisets.

The two requirements in Definition 4 (1) express that in an incremental list $\vec{s}$, an entry s_i must be in the range $\{1, \ldots, i\}$ and $s_i = j \leq i$ can only happen if $s_j = j$, that is, if j occurs already in $\vec{s}$ at position j. This follows since $s_j = s_{s_i} = s_i = j$.

Let's make a bit more concrete what this means, for incremental lists.

- When $K = 1$, there is only one list $\langle 1\rangle \in \underline{1}^1$, which is incremental.
- For $K = 2$ there are four lists $\langle 1, 1\rangle$, $\langle 1, 2\rangle$, $\langle 2, 1\rangle$ and $\langle 2, 2\rangle$ in $\underline{2}^2$. The last two lists are not incremental because the first requirement $s_i \leq i$ fails for $i = 1$.
- For $K = 3$ we thus have five incremental lists, namely: $\langle 1, 1, 1\rangle$, $\langle 1, 2, 1\rangle$, $\langle 1, 2, 2\rangle$, $\langle 1, 1, 3\rangle$, $\langle 1, 2, 3\rangle$.

In a similar way:

- For $K = 1$, there is only $\varphi = 1|1\rangle \in \mathcal{M}[1]\big(\{1\}\big)$. This φ is incremental.
- For $K = 2$ we $\mathcal{M}[2]\big(\{1, 2\}\big)$ contains three elements $2|1\rangle$, $1|1\rangle + 1|2\rangle$, $2|2\rangle$. Only the last one, $\varphi = 2|2\rangle$ is not incremental since $\sum_{j\leq 1} \varphi(j) = 0 \not\geq 1$.
- For $K = 3$ there are five incremental multisets, namely $3|1\rangle$, $2|1\rangle + 1|2\rangle$, $2|1\rangle + 1|3\rangle$, $1|1\rangle + 2|2\rangle$ and $1|1\rangle + 1|2\rangle + 1|3\rangle$.

It is not hard to see that for $\varphi \in IM(K)$ one has $\varphi(i) \leq K + 1 - i$, for each $i \in \underline{K}$.

▶ **Theorem 5.** *Fix a number $K \geq 1$. There are bijective correspondences between:*

1. *incremental lists $\vec{s} \in IL(K)$;*
2. *"parent" functions $p\colon \underline{K} \to \underline{K}$ forming an interior operation: $p^2 = p$ and $p \leq id$.*
3. *set partitions $P \in SP(K)$.*

Via these correspondences the match function becomes an isomorphism $\mathrm{mat}\colon IL(K) \overset{\cong}{\to} SP(K)$.

Proof. The equivalence between items (1) and (2) in Theorem 5 is obvious because a list $\vec{s} = \langle s_1, \ldots, s_K\rangle$ corresponds to a function $p\colon \underline{K} \to \underline{K}$ via $p(i) = s_i$. The two requirements $s_i \leq i$ and $s_{s_i} = s_i$ in Definition 4 (1) correspond directly to $p(i) \leq i$ and $p(p(i)) = p(i)$, that is, to $p \leq id$ and $p^2 = p$.

The equivalence between (1) and (3) is obtained by sending a partition to its parent list. ◀

A consequence of the isomorphism $IL(K) \cong SP(K)$ in Theorem 5 is that the number of incremental lists in $IL(K)$ is given by the K-th Bell number, see Lemma 2 (1).

As we will show, the accumulation map takes incremental lists to incremental multisets (Lemma 7). Therefore, we obtain a canonical map from set partitions to incremental multisets.

▶ **Proposition 6.** *The parent multiset function can be expressed by the "minimum size" function ms in:*

$$ms \;\coloneqq\; \Big(SP(K) \xrightarrow[\cong]{mat^{-1}} IL(K) \xrightarrow{\;acc\;} IM(K)\Big).$$

As in Definition 3 (2), it is given by the formula:

$$ms(P) \;=\; \sum_{B\in P} |B| \Big|\; \min(B) \Big\rangle \;.$$ ⌟

The sets $IL(K)$ and $IM(K)$ of incremental lists and multisets can also be defined inductively. This gives a better grip and allows us to express that accumulation restricts to "incremental".

▶ **Lemma 7.**

1. *Define for $K \geq 1$ the sets $S_K \subseteq \underline{K}^K$ as:*

$$\begin{aligned} S_1 &\coloneqq \big\{\langle 1\rangle\big\} \\ S_{K+1} &\coloneqq \big\{\langle s_1, \ldots, s_K, K{+}1\rangle \;\big|\; \vec{s} \in S_K\big\} \;\cup \bigcup_{1\leq i\leq K} \big\{\langle s_1, \ldots, s_K, s_i\rangle \;\big|\; \vec{s} \in S_K\big\}. \end{aligned}$$

   Then $S_K = IL(K)$.

2. *We also define sets* $M_K \subseteq \mathcal{M}[K](\underline{K})$ *via:*

$$M_1 \coloneqq \{1|1\rangle\}$$
$$M_{K+1} \coloneqq \{\varphi + 1|K+1\rangle \mid \varphi \in M_K\} \cup \bigcup_{1 \le i \le K,\, \varphi(i)>0} \{\varphi + 1|i\rangle \mid \varphi \in M_K\}.$$

Then $M_K = IM(K)$.

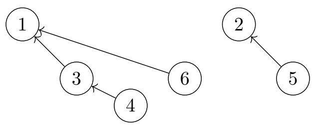
3. *The accumulation map* $acc\colon \underline{K}^K \to \mathcal{M}[K](\underline{K})$ *restricts to* $acc\colon IL(K) \to IM(K)$.

The inductive formulation, especially in item (2), is reminiscent of what is called a *Hoppe urn*, after [7]. One thinks of a multiset $\varphi \in \mathcal{M}[K](\underline{K})$ as an urn with K balls of K-many different colors (in $\underline{K}$). In a "Hoppe" draw an extra ball of the same colour as the drawn ball is returned to the urn but additionally another ball is added with a new, fresh colour, not already occurring in the urn. In item (2) this is represented via the addition of color with number $K+1$ in the sum $\varphi + 1|K+1\rangle$. Biologically, this ball of a new colour $K+1$ can be understood as a (genetic) mutation. Indeed, this structures have first been studied in population biology, see *e.g.* [4, 15].

Proof.

1. Clearly, $S_1 = \{\langle 1\rangle\} = IL(1)$. The inclusions $S_K \subseteq IL(K)$ follow by an easy induction on K. In the other direction, assume $IL(K) \subseteq S_K$; we aim to show $IL(K+1) \subseteq S_{K+1}$. So consider $\langle s_1, \ldots, s_K, n\rangle \in IL(K+1)$. Then $\vec{s} = \langle s_1, \ldots, s_K\rangle \in IL(K)$ and thus $\vec{s} \in S_K$ by induction hypothesis. The element n must be in $\{1, \ldots, K+1\}$. If $n = K+1$ we are done. If $n < K+1$, then $s_n = n$, so that we are also done.
2. Similarly.
3. By induction on K, using the previous two inductive characterisations. ◀

In the end, we add that the name *parent* is motivated by analogy with the disjoint-set forest data structure (sometimes called union-find data structure) [6]. Any list $\vec{p} \in \underline{K}^K$ with $p_i \le i$ represents a forest where p_i is the parent of i. For example, the list $\langle 1, 2, 1, 3, 2, 1\rangle$ represents the forest:



This induces a partition of $\underline{K}$ by taking connected components, or successively taking parents. The forest representing a given partition is not unique. We can make it unique by forcing all trees to have depth al most 1; this corresponds to the condition $p_{p_i} = p_i$ for incremental lists.

Another common way to encode set partitions as numeric lists are *restricted growth lists* (RGS) [16]. While similar to incremental lists in many aspects, we argue here that incremental lists have convenient properties especially considered in connection with the accumulation map and with multisets.

5 Non-crossing partitions and lists

The notion of a crossing has been introduced in the literature for partitions, see [12]. We recall that definition and formulate a corresponding notion for lists. In this section we introduce the basics of such crossings, especially that they are counted by Catalan numbers. In the next section we make connections to (incremental) multisets.

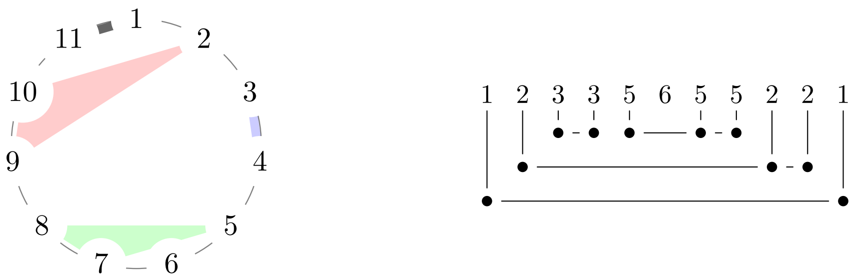
▶ **Definition 8.**

1. *In general, a list $\vec{s} \in X^K$ has a* crossing *if there are indices $n < i < m < j$ with $s_n = s_m \neq s_i = s_j$. The list $\vec{s}$ is called* non-crossing *if it has no crossings.*
   We are especially interested in lists which are both incremental and non-crossing. We write $\mathcal{NCIL}(K) \subseteq \mathcal{IL}(K) \subseteq \underline{K}^K$ for the subset of such lists.
2. *A partition $P \in \mathcal{SP}(K)$ has a* crossing *if there are different blocks $A, B \in P$ with numbers*

$$n < a < m < b \qquad \text{where } n, m \in A \text{ and } a, b \in B.$$

A set partition is called non-crossing *if it has no such crossings. We shall write $\mathcal{NCSP}(K) \subseteq \mathcal{SP}(K)$ for the subset of non-crossing partitions of $\underline{K}$.*

A list is non-crossing when it involves proper nesting, like in nested blocks $\{\ldots\}$ in programming languages, or in nested brackets $(\ldots)$ in expressions. Non-crossing partitions can be visualized nicely by putting the elements of $\underline{K} = \{1, \ldots, K\}$ on a circle and taking convex hulls of the elements in each block. Non-crossing means these hulls don't overlap. Below, we illustrate these definitions for the non-crossing partition $P = \{\{1, 11\}, \{2, 9, 10\}, \{3, 4\}, \{5, 7, 8\}, \{6\}\}$, drawn on a circle on the left, without overlapping regions. The associated non-crossing incremental list $\langle 1, 2, 3, 3, 5, 6, 5, 5, 2, 2, 1\rangle$ is drawn on the right, without improper nestings.



(3)

Non-crossing partitions are counted by the Catalan numbers, as observed in [12, Cor. 4.2].

▶ **Lemma 9.** *The number of non-crossing partitions in $\mathcal{NCSP}(K)$ for $K = 1, 2, 3, \ldots$, is given by the* Catalan *numbers: 1, 2, 5, 14, 42, 132, 429, 1430, 4862, ...* ⌟

The next result is immediate from the formulations of "non-crossing" in Definition 8.

▶ **Proposition 10.** *The match isomorphism $\mathcal{IL}(K) \xrightarrow{\cong} \mathcal{SP}(K)$ restricts to an isomorphism between* non-crossing *incremental lists and set partitions in:*

$$\begin{tikzcd}
\mathcal{NCIL}(K) \arrow[r, dashed, "mat", "\cong"'] \arrow[d, hook] & \mathcal{NCSP}(K) \arrow[d, hook] \\
\mathcal{IL}(K) \arrow[r, "mat", "\cong"'] & \mathcal{SP}(K)
\end{tikzcd}$$

⌟

6 The main sub-triangle result

We go straight to our main result, about the sub-triangle of isomorphisms in Diagram (2).

▶ **Theorem 11.** *For each $K \geq 1$ there are isomorphisms between non-crossing incremental lists, incremental multisets, and non-crossing set partitions, in:*

$$\begin{tikzcd}
& \mathcal{NCIL}(K) \arrow[dl, "acc"', "\cong"] \arrow[dr, "mat", "\cong"'] & \\
\mathcal{IM}(K) & & \mathcal{NCSP}(K) \arrow[ll, "ms"', "\cong"]
\end{tikzcd}$$

As a consequence, the numbers of all these items are given by the Catalan numbers.

Proof. By Proposition 10 it suffices to prove that accumulation restricts to an isomorphism $acc\colon NCIL(K) \to IM(K)$. We first illustrate surjectivity via an exemplaric construction. Then we prove injectivity.

We describe a map $IM(K) \to NCIL(K)$ via a "stack". Let's take $\varphi = 2|1\rangle + 3|2\rangle + 2|3\rangle + 3|5\rangle + 1|6\rangle \in IM(11)$. We proceed in two steps.

1. All elements $1, 2, 3, 5, 6$ in the support of φ must become parents, so we know their required position in the list of size 11 that we need to build:

$$\ell = \langle 1, 2, 3, -, 5, 6, -, -, -, -, -\rangle. \qquad ((*))$$

2. Next we need to complete to a non-crossing incremental list. We go through the above (partial) list ℓ in $(*)$ from left to right using a stack st, which is empty initially.

- We first encounter a 1 in the list in $(*)$. We have $\varphi(1) = 2$, which means that one more 1 needs to be placed somewhere in the list. Hence we push one 1 onto the stack, giving $st = \langle 1\rangle$. This stack serves as an ordered memory that records the elements that we still need to put in the list.
- The next item in the list $(*)$ is a 2, with $\varphi(2) = 3$. We proceed in the same manner and push two numbers 2 onto the stack, giving $st = \langle 1, 2, 2\rangle$.
- There is one more such step when we encounter 3 in $(*)$ with $\varphi(3) = 2$ and turn the stack into $st = \langle 1, 2, 2, 3\rangle$.
- The next thing in $(*)$ is a blank $-$. We pop the last element from the stack and put it at this empty spot, giving new list and stack:

$$\ell = \langle 1, 2, 3, 3, 5, 6, -, -, -, -, -\rangle \qquad st = \langle 1, 2, 2\rangle.$$

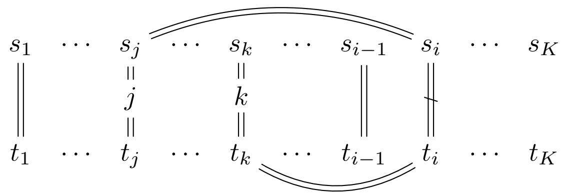
- The next item in $(*)$ is 5, with $\varphi(5) = 3$, so we push two 5's onto the stack, giving $st = \langle 1, 2, 2, 5, 5\rangle$.
- Next we see a 6 in $(*)$ with $\varphi(6) = 1$, so there is no need to put more numbers 6 in the list, and we proceed without any action.
- After 6 in $(*)$ we encounter only blanks. Thus, one-by-one we pop the items from the current stack $st = \langle 1, 2, 2, 5, 5\rangle$ and place them in the list. This gives our outcome $\ell \in NCIL(11)$ of the form:

$$\ell = \langle 1, 2, 3, 3, 5, 6, 5, 5, 2, 2, 1\rangle.$$

By construction, this list is non-crossing, see the diagram on the right in (3). The pushing-and-popping via the stack ensures the proper nesting. This algorithmic description can easily be generalised to an arbitrary incremental multiset.

We turn to injectivity, so let $acc(\vec{s}) = acc(\vec{t})$ for $\vec{s}, \vec{t} \in NCIL(K)$. Our aim is to show $\vec{s} = \vec{t}$. We first note that the assumption $acc(\vec{s}) = acc(\vec{t})$ implies that the sequences $\vec{s}$ and $\vec{t}$ have the same "parent" elements: $s_i = i$ iff $t_i = i$ for each i. Indeed, if $s_i = i$, then $0 < acc(\vec{s})(i) = acc(\vec{t})(i)$. This means that i must occur in $\vec{t}$ and thus $t_i = i$, since $\vec{t}$ is incremental.

Towards a contradiction, let $\vec{s} \neq \vec{t}$, and let i be the least index with $s_i \neq t_i$. Then $s_i \neq i$ and also $t_i \neq i$, by what we just noted. Thus, there is an index $j < i$ with $j = s_j = s_i$ and there is also a $k < i$ with $k = t_k = t_i$. But then $j \neq k$, since $s_i \neq t_i$. Without loss of generality we assume $j < k$. Since i is the least index where s, t differ, we have $t_j = s_j = j = s_i$ and $s_k = t_k = k = t_i$. When we write out the two lists $\vec{s}$ and $\vec{t}$ we get:



The number of occurrences of k in the segment $s_1, \dots s_i$ is one less than in the segment $t_1, \dots, t_i$. We do have $acc(\vec{s})(k) = acc(\vec{t})(k)$, which means that the number of occurrences of k in the whole list $\vec{s}$ is the same as the number of occurrences of k in $\vec{t}$. This means that there is an index $\ell > i$ with $s_\ell = s_k = k$. But now we have a contradiction with the non-crossing assumption for s, since we now have:

$$j < k < i < \ell \qquad \text{with} \qquad s_j = j = s_i \quad \text{and} \quad s_k = t_k = k = s_\ell. \qquad \blacktriangleleft$$

Relationships to other combinatorial families. The three isomorphic structures in Theorem 11 are counted by Catalan numbers. There are dozens of structures in the "Catalan family" which are counted in this way. Famously, sixty-six are listed in a single exercise in [14, Exc 6.19]. While we believe that incremental multisets are a novel addition, we highlight some connections to other family as follows: If $\varphi = acc(s_1, \dots, s_K) \in IM(K)$ then

1. the list of multiplicities $\vec{w} = \langle \varphi(1) - 1, \varphi(2) - 1, \dots, \varphi(K) - 1 \rangle$ satisfies $w_i \geq -1$, has all partial sums nonnegative and $w_1 + \ldots + w_K = 0$; that is Example 6.19.(w) of [14]
2. the increasing rearrangement $\langle s_{(1)}, \dots, s_{(K)} \rangle = sort(s_1, \dots, s_k)$ satisfies $1 \leq s_{(1)} \leq \ldots \leq s_{(K)}$ and $s_{(i)} \leq i$; that is Example 6.19.(s) of [14]. Both of these relationships are invertible.

Elaborating the second bullet point further leads to the well-known combinatorial notion of *parking function* (e.g. the chapter of C. Yan in [2]). A parking function is any list $\langle s_1, \dots, s_k \rangle \in \underline{K}^K$ whose increasing rearrangement $\langle s_{(1)}, \dots, s_{(K)} \rangle$ satisfies $1 \leq s_{(1)} \leq \ldots \leq s_{(K)}$ and $s_{(i)} \leq i$. A parking function is called *increasing* if it furthermore satisfies $s_i \leq s_j$ for $i \leq j$. Increasing parking functions are another prominent member of the Catalan family [13]. Bullet Point 2 states that they are also in bijection with incremental multisets.

It is easy to conclude that a sequence $\vec{s}$ satisfies $acc(\vec{s}) \in IM(K)$ if and only if $\vec{s}$ is a parking function. The sets $IPF(K) \subseteq PF(K)$ of increasing parking functions and parking functions thus fit into the following diagram, which becomes a pullback:

$$\begin{tikzcd}
PF(K) \arrow[r, two heads, "sort"] \arrow[rr, bend left, two heads, "acc"] \arrow[d, hook] \arrow[dr, phantom, "\lrcorner", very near start] & IPF(K) \arrow[r, "acc", "\cong"'] & IM(K) \arrow[d, hook] \\
\underline{K}^K \arrow[rr, "acc"] & & M[K](\underline{K})
\end{tikzcd}$$

7 Un-crossing via term rewriting

The constructions in the previous sections give us a mapping from arbitrary partitions to non-crossing partitions, namely:

$$SP(K) \xrightarrow[\cong]{mat^{-1}} IL(K) \xrightarrow{acc} IM(K) \xrightarrow[\cong]{ms^{-1}} NCSP(K).$$

In this section we show how this mapping can also be obtained via rewriting. Concretely, we start with an arbitrary partition $P_0 \in SP(K)$ and successively eliminate single crossings via rewrite steps $P_0 \to P_1 \to P_2 \to \ldots$ while in each step preserving the parent multiset $ms(P_i) = ms(P_0)$. We show that this reduction system strongly terminates, that is any choice of reduction sequence leads to a crossing-free normal form. By Theorem 11 this normal form must then be unique.

We begin with a preliminary observation: An *increasing crossing* in a partition $P \in SP(K)$ is a crossing $a < b < c < d$ such that $par_P(a) < par_P(b)$. It is easy to see that if a partition has a crossing, it also has an increasing crossing.

▶ **Definition 12.** *We define the "uncrossing" reduction relation ($\to$) on $SP(K)$ as follows. If $P \in SP(K)$ and $a < b < c < d$ is an increasing crossing with $a, c \in A$ and $b, d \in B$, we define P' as the partition in which c, d switch block, giving a new partition:*

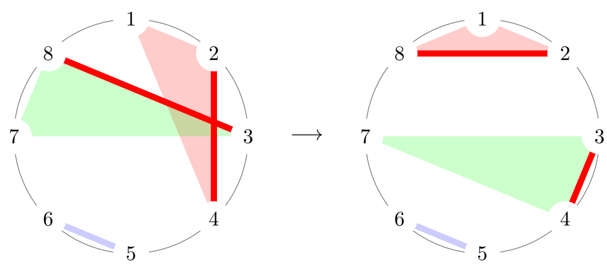
$$P' = P \setminus \{A, B\} \cup \Big\{ A \setminus \{c\} \cup \{d\},\ B \setminus \{d\} \cup \{c\} \Big\}.$$

This defines a single reduction step $P \to P'$.

There is a corresponding rewrite system on incremental lists: an increasing crossing here looks like a length 2-repetition, and we define reductions

$$\langle \ldots a, \ldots, b \ldots, a, \ldots, b, \ldots \rangle \longrightarrow \langle \ldots a, \ldots, b \ldots, b, \ldots, a, \ldots \rangle \quad \text{for} \quad a < b. \tag{4}$$

We illustrate a rewrite step on partitions and on the corresponding incremental lists:



$\langle 1, 1, 3, 1, 5, 5, 3, 3 \rangle \longrightarrow \langle 1, 1, 3, 3, 5, 5, 3, 1 \rangle$

It is easy to see that every uncrossing reduction preserves parents as well as block sizes, so it preserves parent multisets: If $P \to P'$ then $ms(P) = ms(P')$.

▶ **Proposition 13.** *The rewriting system of Definition 12 is strongly terminating, that is every list of reductions $P_0 \to P_1 \to \ldots$ is finite and terminates with a non-crossing partition P^*.*

Proof. This is elegantly expressed in terms of incremental lists (4): every reduction $\vec{s_i} \to \vec{s_{i+1}}$ is a strict up-step in the lexicographic order on lists, *i.e.* satisfies $\vec{s_i} < \vec{s_{i+1}}$. Because the set $IL(K)$ is finite, there can only be finitely many of such steps. ◀

▶ **Corollary 14.** *The reduction system of Definition 12 has the Church-Rosser property (i.e. is confluent).* ⌟

This follows abstractly from strong termination and uniqueness of normal forms. We conjecture that (local) confluence can also be established directly, in the following form: For all P, if $Q_1 \leftarrow P \to Q_2$ then there exists an R and lists of reductions *of length at most 2* with $Q_1 \to^* R \;{}^*\!\!\leftarrow Q_2$.

As an aside, note that we can apply the reduction rule (4) not just to incremental lists but also to arbitrary lists of natural numbers $\mathbb{N}^*$. This extended rewriting system is still strongly terminating, but no longer confluent. For example, we have reductions $\langle 1,3,2,3,2,1\rangle \; {}^*\!\leftarrow \langle 1,3,2,1,3,2\rangle \rightarrow \langle 1,3,2,2,3,1\rangle$ where the left and right lists are both irreducible.

8 A wider picture: adding multiset partitions

Having presented our main results, we step back and put things in a wider perspective. We started in Diagram (1) with lists, multisets and set partitions. The two legs in this diagram can be completed to a diamond of the form:

$$\begin{tikzcd} & X^K \arrow[dl, "acc"'] \arrow[dr, "mat"] & \\ \mathcal{M}[K](X) \arrow[dr, "mc"'] & & SP(K) \arrow[dl, "sc"] \\ & MP(K) & \end{tikzcd} \tag{5}$$

The set $MP(K)$ contains the *multiset partitions* with total K. It is defined as:

$$MP(K) \coloneqq \big\{\sigma \in \mathcal{M}\big(\underline{K}\big) \;\big|\; \textstyle\sum_i \sigma(i)\cdot i = K\big\}.$$

Multiset partitions represent unlabelled partitions where the underlying elements of X are not distinguishable anymore. Such multiset partitions are commonly considered in number theory. The sizes $\big|MP(K)\big|$, for $K \geq 1$ are given by the partition function $p(K)$ with values $1,2,3,5,7,11,15,22,30,42,56,\ldots$. For example $p(4) = 5$ counts the 5 number-theoretic partitions of the number 4,

$$1+1+1+1=4 \qquad 1+1+2=4 \qquad 1+3=4 \qquad 2+2=4 \qquad 4=4.$$

Alternatively, in multiset notation, we get the elements of $MP(4)$, namely:

$$4|1\rangle \qquad 2|1\rangle + 1|2\rangle \qquad 1|1\rangle + 1|3\rangle \qquad 2|2\rangle \qquad 1|4\rangle.$$

The function $mc\colon \mathcal{M}[K](X) \to MP(K)$ is defined in [10] and is called *multiplicity count*. It counts the multiplicities in a multiset $\varphi \in \mathcal{M}[K](X)$ via:

$$mc(\varphi) \coloneqq \sum_{x\in supp(\varphi)} 1\big|\varphi(x)\big\rangle.$$

The fourth function $sc\colon SP(K) \to MP(K)$ in (5), from set partitions to multiset partitions, will be called *size count*. It keeps track of the sizes of blocks in a set partition:

$$sc(P) \coloneqq \sum_{B\in P} 1\big|\,|B|\,\big\rangle.$$

The following instantiation illustrates how the operations in Diagram (5) work.

$$\begin{tikzcd} & \langle c,b,a,a,c,c,b\rangle \arrow[dl, mapsto, "acc"'] \arrow[dr, mapsto, "mat"] & \\ 2|a\rangle + 2|b\rangle + 3|c\rangle \arrow[dr, mapsto, "mc"'] & & \big\{\{1,5,6\},\{2,7\},\{3,4\}\big\} \arrow[dl, mapsto, "sc"] \\ & 2|2\rangle + 1|3\rangle & \end{tikzcd}$$

It is not hard to show in general that Diagram (5) commutes.

Set and multiset partitions are studied in mathematical biology [5, 4], to capture mutations, and more recently also in clustering in machine learning, to handle possible extension of the numbers of clusters, see *e.g.* [3]. Here we study the underlying datatypes and their relations (5).

Finally, we remark that the diamond (5) is a *weak pullback*. That is for every pair (φ, P) of a multiset and a partition which induce the same multiset partition $mc(\varphi) = sc(P)$, we can find some list $\vec{x}$ with $\varphi = acc(\vec{x})$ and $P = mat(\vec{x})$. For example, the multiset $\varphi = 3|a\rangle + 2|b\rangle + 2|c\rangle$ and the partition $P = \{\{1,3,4\},\{2,6\},\{5,7\}\}$ induce the same multiset partition $1|3\rangle + 2|2\rangle \in MP(7)$. They are themselves induced by the list $\vec{x} = \langle a,b,a,a,c,b,c\rangle$. This list $\vec{x}$ is not unique however; in our example, $\vec{x} = \langle a,c,a,a,b,c,b\rangle$ also works.

9 An even wider categorical picture

In this section, we give a more high-level, structural view on the diamond (5), returning to general finite sets X, Y. Category theory is an abstract language of structure and datatypes. The idea of forgetting information can be formalized using *coequalizers*, an abstract form of quotient. In [9] it is shown that the accumulation map is the coequaliser of all transposition maps:

$$
\begin{tikzcd}
X^K \arrow[r, shift left, "\tau_\pi"] \arrow[r, shift right, "\tau_{\pi'}"'] & X^K \arrow[r, "acc"] \arrow[dr, "\forall"'] & M[K](X) \arrow[d, dashed, "\exists!"] \\
& & Y
\end{tikzcd}
$$

Any bijection $\pi\colon K \xrightarrow{\cong} K$ induces a reordering (transposition) map $\tau_\pi\colon X^K \to X^K$. Accumulation is invariant under all those transpositions, and furthermore universal with that property: any map $f\colon X^K \to Y$ which is invariant under all transpositions factors uniquely through acc. This can be understood as a proof principle about multisets.

We wish to understand partitions $SP(X)$ in the same way. An immediate obstacle is that unlike multisets, set partitions do not push forward under arbitrary functions $X \to Y$ in an obvious way. They do however push forward under *bijections* $X \to Y$. This naturally leads us to consider *combinatorial species*, which are functors $Bij \to Set$ where Bij is the category of finite sets X, Y and bijections between them. Since their invention [11], species have been a staple in combining combinatorial and categorical reasoning. Using their framework, we can precisely state the process of forgetting information by quotienting out the action of permutations.

▶ **Definition 15** (Anonymisation). *Let $\mathbb{C}$ be any category, and write $\mathbb{N}$ for the discrete category whose objects are the natural numbers. If $\mathcal{F}\colon Bij \times \mathbb{C} \to Set$ is a functor, we define a functor $\mathcal{F}_1\colon \mathbb{N} \times \mathbb{C} \to Set$ as the coequalizer*

$$
\begin{tikzcd}
\mathcal{F}(\underline{K}, C) \arrow[r, shift left, "\mathcal{F}(\pi,\mathrm{id}_C)"] \arrow[r, shift right, "\mathcal{F}(\pi',\mathrm{id}_C)"'] & \mathcal{F}(\underline{K}, C) \arrow[r] & \mathcal{F}_1(K, C)
\end{tikzcd}
\tag{6}
$$

More concretely, the set $\mathcal{F}(\underline{K}, C)$ admits an action of the group of bijections $Aut(\underline{K})$ and we let $\mathcal{F}_1(K, C) = \mathcal{F}(\underline{K}, C)/Aut(\underline{K})$. This canonical construction is already present in the first section of [11]. We will also write $\mathcal{F}_1(K, C) = \int^K \mathcal{F}(X, C)\mathrm{d}X$ in analogy with coend calculus or the category of elements.

▶ **Proposition 16.** *Let $\mathcal{H}\colon Bij \times Bij \to Set$ be a functor in two variables. Then we can anonymise variables in different orders and obtain the same result*

$$\int^{n}\int^{m} \mathcal{H}(X,Y)\,\mathrm{d}X\,\mathrm{d}Y \cong \int^{m}\int^{n} \mathcal{H}(X,Y)\,\mathrm{d}Y\,\mathrm{d}X.$$ ⌟

For the formally minded, the cardinality functor $|-|\colon Bij \to \mathbb{N}$ induces a functor $\Delta\colon [\mathbb{N}, Set^{\mathbb{C}}] \to [Bij, Set^{\mathbb{C}}]$ between functor categories by precomposition; it has a left adjoint which is given by "anonymisation". The unit of this adjunction is a natural transformation $\mathcal{F}(X,C) \to \mathcal{F}_1(|X|,C)$. Thus any functor $\mathcal{H}\colon Bij \times Bij \to Set$ gives rise to a commuting diamond of natural transformations

$$\begin{tikzcd}
 & \mathcal{H}(X,Y) \arrow[dl] \arrow[dr] & \\
\int^{|X|}\mathcal{H}(X,Y)\mathrm{d}X \arrow[dr] & & \int^{|Y|}\mathcal{H}(X,Y)\mathrm{d}X \arrow[dl] \\
 & \int^{|Y|}\int^{|X|}\mathcal{H}(X,Y)\mathrm{d}X\mathrm{d}Y &
\end{tikzcd} \tag{7}$$

We can now describe the diamond (5) by invoking Proposition 16 on the function space construction Y^X.

▶ **Proposition 17.** *Let $\mathcal{H}\colon Bij \times Bij \to Set$ be given by $\mathcal{H}(X,Y) = Y^X$ with functorial action $\mathcal{H}(\alpha,\beta)(f) = \beta \circ f \circ \alpha^{-1}$. Then*

$$\int^{K} Y^X \mathrm{d}X \cong \mathcal{M}[K](Y)$$

is multisets with exactly K elements, and

$$\int^{N} Y^X \mathrm{d}Y \cong SP[N](X)$$

is set partitions with at most N blocks, i.e. $SP[N](X) = \{P \in SP(X) \mid |P| \leq N\}$. ⌟

Furthermore, if we define $MP[n,k]$ as multiset partitions of k with at most n blocks, we obtain the following improved categorical diamond.

$$\begin{tikzcd}
 & Y^X \arrow[dl, "acc"'] \arrow[dr, "mat"] & \\
\mathcal{M}\big[|X|\big](Y) \arrow[dr, "mc"'] & & SP\big[|Y|\big](X) \arrow[dl, "sc"] \\
 & MP\big[|Y|,|X|\big] &
\end{tikzcd} \tag{8}$$

All maps are coequalizers and natural in X, Y. This makes it clear that multisets and partitions arise from a fully symmetric situation, where we forget information along two independent axis: order (transposition $X \cong X$) and identity (substitution $Y \cong Y$).

10 Conclusion

In this paper we studied two orthogonal operations on lists from a fundamental perspective, namely: (1) accumulation of lists to multisets, which is stable under transposition, and (2) matching of lists to set partitions, which is stable under substitution. In subsequent work we wish to include the various distributions on these datatypes in the same perspective.

References

1 D. Aldous. Exchangeability and related topics. In P. Hennequin, editor, *École d'Été de Probabilités de Saint-Flour XIII – 1983*, number 1117 in Lect. Notes Math., pages 1–198. Springer, Berlin, 1985. doi:10.1007/BFb0099421.

2 M. Bona. *Handbook of Enumerative Combinatorics*. Discrete Mathematics and Its Applications. CRC Press, 2015.

3 T. Broderick, J. Pitman, and M. Jordan. Feature allocations, probability functions, and paintboxes. *Bayesian Analysis*, 8:801–836, 2013. doi:10.1214/13-BA823.

4 H. Crane. The ubiquitous Ewens sampling formula. *Statistical Science*, 31(1):1–19, 2016. doi:10.1214/15-STS529.

5 W. Ewens. The sampling theory of selectively neutral alleles. *Theoret. Population Biology*, 3:87–112, 1972. doi:10.1016/0040-5809(72)90035-4.

6 Z. Galil and G. Italiano. Data structures and algorithms for disjoint set union problems. *ACM Comput. Surv.*, 23:319–344, September 1991. doi:10.1145/116873.116878.

7 F. Hoppe. Pólya-like urns and the Ewens' sampling formula. *Journ. Math. Biology*, 20:91–94, 1984. doi:10.1007/BF00275863.

8 B. Jacobs. From multisets over distributions to distributions over multisets. In *Logic in Computer Science*. IEEE, Computer Science Press, 2021. doi:10.1109/lics52264.2021.9470678.

9 B. Jacobs. Multinomial and hypergeometric distributions in Markov categories. In A. Sokolova, editor, *Math. Found. of Programming Semantics*, number 351 in Elect. Proc. in Theor. Comp. Sci., pages 98–115, 2021. doi:10.4204/EPTCS.351.7.

10 B. Jacobs. Sufficient statistics and split idempotents in discrete probability theory. In *Math. Found. of Programming Semantics*, 2022.

11 A. Joyal. Une théorie combinatoire des séries formelles. *Advances in Mathematics*, 42(1):1–82, 1981. doi:10.1016/0001-8708(81)90052-9.

12 G. Kreweras. Sur les partitions non croisées d'un cycle. *Discrete Math.*, 1(4):333–350, 1972.

13 R. Stanley. Parking functions and noncrossing partitions. *Elect. Jour. of Combinatorics*, 2(4), 1997. doi:10.37236/1335.

14 R. Stanley and S. Fomin. *Enumerative Combinatorics*, volume 2 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 1999. doi:10.1017/CBO9780511609589.

15 S. Tavaré. The magical Ewens sampling formula. *Bull. London Math. Soc.*, 53:1563–1582, 2021. doi:10.1112/blms.12537.

16 E. Weisstein. Restricted growth string. MathWorld – A Wolfram Web Resource. http://mathworld.wolfram.com/RestrictedGrowthString.html, July 2022.